Ageing & Society **31**, 2011, 1–19. © Cambridge University Press 2011
doi:10.1017/S0144686X1000070X

The harmonisation of longitudinal data: a case study using data from cohort studies in The Netherlands and the United Kingdom

PETER A. BATH*, DORLY DEEG† and JAN POPPELAARS†

ABSTRACT

This paper presents a case study of the challenges and requirements associated with harmonising data from two independently-conceived datasets from The Netherlands and the United Kingdom: the Longitudinal Aging Study Amsterdam (LASA) and the Nottingham Longitudinal Study of Activity and Ageing (NLSAA). The objectives were to create equivalent samples and variables, and to identify the methodological differences that affect the comparability of the samples. Data are available from the two studies' 1992–93 surveys for respondents born during 1908–20, and the common data set had 1,768 records and enabled the creation of 26 harmonised variables in the following domains: demographic composition and personal finances, physical health, mental health and loneliness, contacts with health services, physical activity, religious attendance and pet ownership. The ways in which the methodological differences between the two studies and their different selective attrition might lead to sample differences were carefully considered. It was concluded that the challenges of conducting cross-national comparative research using independent datasets include differences in sampling, study design, measurement instruments, response rates and selective attrition. To reach conclusions from any comparative study about substantive socio-cultural differences, these challenges must first be identified and addressed.

KEY WORDS – data harmonisation, older people, cross-national comparison, sample attrition.

Introduction

Demographic changes across Europe have resulted in an increase in both the absolute and relative number of older people (Walker 2005), and

* Department of Information Studies, University of Sheffield, Sheffield, UK.
† Department of Psychiatry, Vrije Universiteit, Amsterdam, The Netherlands.

stimulated increased research into the factors associated with the health and wellbeing of older people, into the determinants of increased longevity, and into ways of maintaining healthy and disability-free lives. This information is important for the development and planning of services for older people. For many years, research on health and wellbeing in older people has concentrated on analysing data from studies in single countries or from different countries separately. Comparisons between studies and across countries have been undertaken through reviews of the published literature, which allow the formation of a 'cumulative knowledge base' on specific issues (Curran and Husong 2009: 81). Such an approach enables findings from individual studies to be confirmed or refuted in other settings, and provides evidence of country (or study) differences but have a fundamental limitation: usually it is unclear whether the observed differences arise from: (a) methodological differences between the studies, (b) a defect or error in the comparative method, or (c) actual population differences.

One way to develop a better understanding of older people in multiple countries is to undertake studies with consistent designs and methods. This eliminates the first listed cause of any differences and greatly reduces the likelihood of the second, so leaving any observed differences attributable to actual differences between the populations or to random variation. Setting up identical studies in two countries is costly and difficult, however, partly because regional and national funding bodies are unlikely to support research in another country (Casado-Díaz, Kaiser and Warnes 2004). Many studies of older people have addressed specific issues in single countries, and commonly aspects of their design and emphases reflect local cultural and institutional arrangements or preoccupations (not least concerning health-care delivery). Such studies are rarely comparable with studies of similar issues in other countries. An alternative approach is to use data from existing longitudinal studies of older people (Minicuci *et al.* 2003) and to develop cross-national data sets by harmonising the variables. While this approach makes use of the available data, careful attention has to be paid to differences in sampling, design and measurement instruments (Hofer and Piccinin 2009). The process of integrative data analysis (Curran and Hussong 2009), in which one data set (formed from pooling two or more separate samples) is used for statistical analysis, is an emerging method within the social sciences (Curran 2009), and provides new opportunities for analysing data on older people. A specific problem with using *longitudinal* studies in this way is the loss of participants and attrition bias between the baseline and follow-up surveys through mortality or for other reasons.

Aims and data sources

The overall aim of this study was to develop harmonised data from two independent cohort studies of older people in The Netherlands and the United Kingdom: the Longitudinal Aging Study Amsterdam (LASA) (Deeg, Knipscheer and van Tilburg 1993) and the Nottingham Longitudinal Study of Activity and Ageing (NLSAA) (Morgan 1998). More specifically, the objectives were to:

- Identify equivalent samples of older people from the LASA and NLSAA data sets.
- Harmonise variables with comparable content from the two studies.
- Describe any methodological differences between the two studies and discuss the challenges and requirements for harmonising data from two independently-conceived longitudinal datasets.
- Develop recommendations for data harmonisation for future cross-national research.

The LASA data

The methodology of LASA is described in detail elsewhere and only a brief account is provided (Deeg, Knipscheer and van Tilburg 1993). LASA has a nationally-representative sample of people aged 55–85 years (*i.e.* born between 1908 and 1937), with over-sampling of men and the oldest age groups to ensure sufficient numbers at the follow-up. The sample was recruited from the 3,805 respondents for the NESTOR study in 1992 of Living Arrangements and Social Networks of Older Adults (LSN), which had a response rate of 62.3 per cent (Knipscheer *et al.* 1995). About 10 months after the LSN interview, the participants were approached for the first LASA cycle in 1992–93 (Deeg, Knipscheer and van Tilburg 1993). By the start of the LASA baseline study, there were 3,679 surviving LSN participants. Of these, 3,107 took part in the interviews and tests, yielding a response rate of 84.5 per cent; the 15.5 per cent non-response included 3.6 per cent ineligibility through frailty, 1.1 per cent not contacted after eight or more attempts, and 10.7 per cent refusals. Non-response was associated with higher age but not with gender (Deeg *et al.* 2002). Although only a few of the LASA variables had been collected in the precursor LSN study (age, gender, marital status and self-rated health), tests showed a significant association between the LSN measure of health rating relative to peers in 1992 and the profile of the follow-up interviewees in 1992–93 ($p = 0.003$): people who rated their health as a little worse than that of their peers in LSN (1992) were more likely not to participate in the 1992–93

LASA interview than people who rated their health as much better than their peers in LSN (1992) (odds ratio (OR) = 2.15; 95 % confidence interval 1.25–3.71; $p = 0.006$). The baseline inquiry was a face-to-face interview, after which the interviewer left a self-completion and return questionnaire. Among those interviewed, 74.1 per cent returned completed questionnaires, with a slight over-representation of the younger respondents (Deeg *et al.* 2002). The questions from LASA used in this study are described below and reproduced in Table 1.

The NLSAA data

The methodology of NLSAA is described in detail elsewhere (Morgan 1998) and only a brief account is provided here. Three areas of Greater Nottingham were used to generate a study population similar to the average national pattern for England and Wales. All community-dwelling people aged 65 or more years in the survey areas were identified. From the resulting 8,409 older people, a random sample of 1,299 non-institutionalised individuals were invited to participate, of whom 1,042 agreed (406 men and 636 women), giving an 80 per cent response rate. There was over-sampling of the oldest ages to allow sufficient numbers for follow-up surveys.

The baseline survey was conducted between May and September 1985, and the follow-up surveys in 1989 and 1993. People who had participated in 1985 and who were still alive and resident locally were contacted and invited to participate in the follow-ups (Morgan 1998). The main reasons for attrition from the sample were death, refusal, emigration and lost trace. In 1989, of the 781 people remaining in the sample, 690 were re-interviewed (88.3 % response). In 1993, of the 540 people remaining, 426 were contacted successfully and 410 interviews satisfactorily completed (75.9 % response) (Morgan 1998). The third wave of interviews began in May 1993 and completed by the end of the year. The questions from the interview schedule used in this study are described below and reproduced in Table 1.

Data harmonisation procedures

To obtain equivalent and unbiased samples of older people from the two studies, the sampling, design and measurement instruments for each were reviewed and similar sub-samples and variables selected. New variables in each dataset were created and the data for the selected samples were merged into a single combined dataset. First, it was important to specify the two sampling frames. To reiterate, LASA was a nationally-representative survey conducted during 1992–93 among 3,107 respondents

TABLE 1. *Questions and response categories in LASA and NLSAA and new variables created following data harmonisation*

	LASA		NLSAA		Harmonised	
D	Question	Categories or scale	Question	Categories or scale	Variable	Categories
Demography	Age on day of interview	Continuous variable	Age on day of interview	Continuous variable	Age group	<80 years/80+
	Sex	Male/female	Gender	Male/female	Gender	Male/female
	Are you unmarried, married, divorced or widowed?	Never married/ married/divorced/ widowed	Marital status	Married/single/ widowed/separated or divorced	Marital status	Single/married/ divorced/ widowed
Finance	Are you in paid work or do one or several hours per week or short-term temporary work?	No/yes	Are you in full-time or part-time employment or voluntary work?	No/full-time employment/part-time employment/full-time voluntary/part-time voluntary	Paid job at present?	No/yes
	Do you receive income from a pension?	No/yes	Do you receive an old age/retirement pension?	No/yes	Receiving pension	No/yes
	Are you satisfied with your income level?	Dissatisfied/a little dissatisfied/not dissatisfied or satisfied/ a little satisfied/satisfied	Do you feel satisfied or dissatisfied with your present financial position?	Completely satisfied/ fairly satisfied/fairly dissatisfied/completely dissatisfied	Expressed satisfaction with income/or present financial position?	No/yes
Health	Do you have osteoarthritis? Do you have rheumatoid arthritis?	No/yes	Would you say you suffer from arthritis or rheumatism?[1]	No/yes	Has rheumatism or arthritis	No/yes
	Do you have heart disease or have you had a myocardial infarction?	No/yes	Would you say you suffer from heart trouble?[2]	No/yes	Has heart disease?	No/yes
	Do you unintentionally lose urine sometimes?	No/yes	Would you say you suffer from leakage of urine?[3]	No/yes	Has incontinence?	No/yes
Perceived H	How in general is your health?	Excellent/good/fair/ good and bad/poor	How would you rate your present health?	Excellent/good/ average/fair/poor	Self-rated health	Excellent/ good/good and bad/fair, average
	How is your health compared to your age peers?	Much better/little better/ DK, just as good/ a little worse/much worse	Compared with men or women of your own age, do you think you are:	Much more healthy/ more healthy/ as healthy/less healthy/ much less healthy	Perceived health relative to peers	Note 4

TABLE 1. (*Cont.*)

		LASA		NLSAA		Harmonised	
D		Question	Categories or scale	Question	Categories or scale	Variable	Categories
MH		Cognitive impairment	MMSE scale	Cognitive impairment	CAPE scale	Cognitive impairment	Standardised scale
		Anxiety	HADS-A scale	Anxiety	Anxiety sub-scale	Anxiety	Standardised scale
		Depression	CES-D scale	Depression	Depression sub-scale	Depression	Standardised scale
L		During the past week have you felt lonely?	Rarely or never/some of the time/often/ mostly or always	How often do you feel lonely?	Often/sometimes/ seldom/never	Loneliness	Rarely/never/ at times/seldom/ often/at times/ mostly, always/ never
Health services contacts		In the past half year have you have had contact with a Family doctor?	No/yes	When did you last see general practitioner (family doctor)?	Last week/Last month/ within last 6 months/ more than 6 months ago	Seen family doctor in last 6 months?	No/yes
		Medical specialist?	No/yes	When did you last see a hospital doctor?	Last week/last month/ in last 6 m./>6 m. ago	Seen hospital doctor in last 6 months?	No/yes
		District nurse?	No/yes	When did you last see community (district) nurse?	Last week/last month/ in last 6 m./>6 m. ago	Seen district nurse in last 6 months?	No/yes
		Health visitor?	No/yes	When did you last see a health visitor?	Last week/last month/ in last 6 m./>6 m. ago	Seen health visitor in last 6 months?	No/yes
Physical activity		Do you go for walks (for shopping, daily trips, like visiting someone)?[5]	No/yes	Did you walk outdoors yesterday. If so, was the walk typical/usual?[6]	Yesterday/2 days ago/ 3 days ago/4–7 days ago/in last month/ longer ago	Walking outside in the last two weeks (LASA)/ month (NLSAA)?	No/yes
		Do you at times cycle? (*e.g.* for shopping, daily activities, like visiting someone)	No/yes	How much time do you spend cycling (leisurely, level and purposeful, fast, varying gradients)	No/yes	Goes cycling?	No/yes

	Do you do household tasks at times? (light, *e.g.* dusts, irons, cooks; heavy, window cleans, wash or scrub floor)	No/yes	Household tasks: Light, *e.g.* dusts, tidies; Moderate, *e.g.* cleans windows, hoover; Heavy, *e.g.* polish, scrub floors	Minutes per week	Does any household activities?	No/yes
	Do you do gardening? Have you worked in the garden in the last two weeks?	No/yes	Light, *e.g.* weeding, pruning; Moderate, *e.g.* raking, hoeing; Heavy, *e.g.* digs, mows lawn	Minutes per week	Does gardening?	No/yes
	Have you played sports in the last two weeks?	No/yes	Time on hobbies, recreations involving moderate physical activity?	Minutes per week	Sport/leisure participation?	No/yes
R	Involved in church or religious organisation?	No/yes	Do you attend religious services or meetings?	No/yes	Attends church or religious service	No/yes
P	Do you have pets?	No/yes	Do you have a pet?	No/yes	Do you have a pet?	No/yes

Notes: LASA: Longitudinal Ageing Study Amsterdam. NLSAA: Nottingham Longitudinal Study of Activity and Ageing. D: Domain. H: Health. L: Loneliness. MH: Mental health. P: Pet owner. R: Religious activity. DK: Don't know. Some questions have been abbreviated from the administered versions. 1. Any part of the body including any persistent joint pain. 2. Including angina, rheumatic heart disease, palpitations, heart attack, poor valve operation. 3. All degrees of incontinence from occasional leakage to total incontinence. 4. Five categories derived from similar categories in LASA and NLSAA: much better, much more healthy/a little better, more healthy/don't know, just as good, about as healthy/a little worse, less healthy/much worse, much less healthy. 5. Have you been for a walk in last two weeks? 6. If no, when was the last time the amount of walking you did outdoors was typical/usual? Total time spent walking/shopping.

between the ages of 55 and 85 years. The response rate was 62 per cent, which is relatively high for a survey in The Netherlands. The sample was drawn from the population registries of 11 municipalities in three culturally-distinct areas in the west, north-east and south of the country. Turning to the NLSAA sample, it was developed first by using electoral ward statistics from the 1981 population census to identify three areas of Greater Nottingham that in aggregate had a study population with a similar profile to that of England and Wales in terms of age, gender, socio-economic class composition, ethnicity and the number of elderly people living alone. Then, using Nottinghamshire Family Practitioner Committee patient registration lists, which specified age and gender, all patients aged 65 or more years living in the community (*i.e.* excluding those living in residential or nursing homes) in the designated study area were identified.

Second, to minimise age and period effects, only the participants in both studies who were born during the same years and who were interviewed at similar times were included. All of the LASA respondents were born between 1908 and 1937, and the NLSAA respondents were born anytime up to, and including, 1920. The pooled analysis sample included those born between 1908 and 1920 who were interviewed in both studies' follow-up surveys during 1992–93 (LASA) or 1993 (NLSAA). Finally, as NLSAA did not include persons living in long-term care institutions, institutionalised participants were excluded from the LASA sample.

The measures and scales

The next step was to develop a common set of socio-demographic, financial, behavioural, social, psychological and physical health status variables in a new database. The exact wording of the relevant variables in LASA and NLSAA were examined. Both English translations of the LASA questions, and where appropriate the original English wording of pre-existing scales, were examined by the first author to determine whether the variables and categories had the same face value and to assess their comparability. The possible comparator variables were then discussed with the second author and a consensus reached. To create the harmonised variables, a standard procedure of 'transform and recode' was applied to one or both of the original study measures. Existing codes for categories were merged and re-labelled in each study depending on the precise wording and the ordering of the categories. The study-specific categories are presented in Table 1 together with the harmonised variable names and categories. The study-specific scales for cognitive impairment, anxiety and depression were standardised, as described below, to create harmonised mental health measures.

Selective attrition

Analyses of the NLSAA data were conducted to test for the effects of selective attrition on the pooled analysis samples. Chi-squared tests and logistic regression analyses were applied to the NLSAA sample to test the null hypothesis that there was no association between variables measured at baseline sample and participation in the 1993 interviews among those born during 1908–1920.

Results

Data harmonisation

The harmonised data file had 1,768 records and 47 harmonised variables for socio-demographic attributes (age, gender, marital status, living arrangements), personal finances (currently in paid job, receiving pension, satisfaction with income), physical health (presence of heart disease, diabetes, rheumatism or arthritis, incontinence, occurrence of cerebrovascular accident), self-rated health, mental health (cognitive impairment, anxiety, depression), contact with health and social care services (family doctor, hospital doctor, district nurse, home help care), physical activity (household activities, walking, cycling, gardening, sports or leisure participation), and social activity (church or religious service attendance, pet ownership and loneliness).

The socio-demographic and personal finances variables

The participants in LASA and NLSAA were asked their gender and exact date of birth (day, month, year) from which it was possible to calculate an exact age at interview. Although respondents in both LASA and NLSAA were asked to state their marital status, the precise wording was not available for NLSAA and the response categories differed slightly. LASA respondents were asked if they had never been married, whereas NLSAA respondents were asked if they were single. To create a harmonised variable, it was assumed that these response categories had the same meaning; that is, that LASA respondents who said that they had 'never married' were equivalent to NLSAA respondents who answered 'single'. In addition, a LASA response category was 'divorced', whereas NLSAA used 'separated or divorced'. In the harmonised variable, these categories were considered equivalent (although it is possible that LASA respondents who were separated answered 'married' rather than 'divorced'). The four categories in the harmonised variable were therefore 'single/married/divorced/widowed'.

The questions and response categories on paid work (LASA) or employment (NLSAA) were slightly different: both studies asked whether the declared employment status was 'at this moment' (LASA) or 'currently' (NLSAA), LASA used a dichotomous response (no/yes) while NLSAA used several categories for full-time or part-time employment and voluntary work. The harmonised variable was necessarily a simple dichotomy for being in paid work (no/yes). For NLSAA, a response of in full-time or part-time employment was taken as equivalent to 'yes', and a response of in full-time and part-time voluntary work as 'no'.

The questions on receipt of a pension were similar in LASA and NLSAA and both studies used 'no/yes' response categories, which was adopted for the new variable. In LASA, people were asked whether they were satisfied with their income, with five response categories including a neutral category (not dissatisfied or satisfied). In contrast, the equivalent question in NLSAA asked whether people felt 'satisfied' or 'dissatisfied' with their present financial position, with four response categories and no neutral response. To harmonise these variables, the new variable was whether the person expressed satisfaction with their income or present financial position and 'no/yes' responses were used. Among the LASA respondents, those who said that they were 'dissatisfied', 'a little dissatisfied' or 'not dissatisfied or satisfied' were categorised as 'no', and those who said that they were 'a little satisfied' or 'satisfied' were categorised as 'yes'. Among the NLSAA respondents, those who said that they were 'fairly dissatisfied' or 'completely dissatisfied' with their income or present financial position were categorised as 'no', and those who said that they were 'fairly satisfied' or 'completely satisfied' were categorised as 'yes'.

The health-related variables

Several similar variables relating to the health of the respondents were identified in LASA and NLSAA, including whether arthritis, heart diseases and incontinence were reported. In LASA, participants were asked whether they had rheumatoid arthritis or osteoarthritis, and if so, whether it was in the knees, hips or hands, whereas in NLSAA, the respondents were asked whether they suffered from arthritis or rheumatism in any part of the body (including any persistent joint pain). These questions all used 'no/yes' responses, so a harmonised variable (has rheumatism or arthritis) was created. The LASA respondents were asked whether they had heart disease or had had a myocardial infarction (no/yes), whereas the NLSAA respondents were asked whether they had heart disease with several examples provided (no/yes), so a harmonised variable (has heart disease? no/yes) was created. Both the LASA and NLSAA respondents were asked

whether they were incontinent (no/yes), and although the precise wording of the questions was slightly different, both studies sought information on the frequency of the problem ('sometimes' in LASA and from 'occasional' to 'total' in NLSAA). These variables were harmonised into a single variable (has incontinence; yes/no).

Perceived health measures

The LASA and NLSAA respondents were asked two similar questions about how they rated their health and how they rated it relative to their peers. Although these 'self-rated health' questions were worded similarly and three of the response categories were identical and in the same order (excellent, good, –, –, poor), the response category 'fair' was third in the LASA sequence and fourth in NLSAA. The fourth category among the LASA responses was 'sometimes good, sometimes bad', and the third category for NLSAA was 'average'. The harmonised variable had four response categories: the three shared categories were retained and the differing third and fourth categories were merged into 'sometimes good or sometimes bad/fair/average'. Turning to the relative health variables, although the words used in the LASA and NLSAA questions were slightly different, the overall meaning was the same. There were five response categories in both studies but they were phrased differently (indicating 'better/worse' in LASA and 'more/less healthy' in NLSAA), and the middle LASA response category included 'don't know', which was not available to the NLSAA respondents. Nonetheless, as the ordered categories in the LASA and NLSAA questions were considered sufficiently similar, the harmonised variable was given the five response categories (for the phrasing *see* Table 1).

Mental health measures

Cognitive impairment, anxiety and depression were assessed by both studies but different measures and scales were used. Cognitive impairment was assessed in LASA using the Mini Mental State Examination (MMSE) (30-point scale) (Folstein, Folstein and McHugh 1975) and in NLSAA using the Information/Orientation sub-scale of the Clifton Assessment Procedures for the Elderly (CAPE) (12-point scale) (Pattie and Gilleard 1979). To standardise these scales, the MMSE scores were divided by 30 and the CAPE scores divided by 12. For anxiety, LASA used the anxiety sub-scale of the Hospital Anxiety and Depression scale (HADS-A) (21-point scale) (Zigmund and Snaith 1983), and NLSAA used the anxiety sub-scale of the Symptoms of Anxiety and Depression (SAD) scale (21-point scale) (Bedford, Foulds and Sheffield 1976). Depression was

assessed in LASA using the 60-point Center for Epidemiologic Studies Depression Scale (CES-D) scale (Radloff 1977), and in NLSAA using the 21-point depression sub-scale of the SAD (Bedford, Foulds and Sheffield 1976). To standardise these scales, the CES-D scores were divided by 60 and the SAD depression scores divided by 21.

Loneliness measures

Questions on loneliness were asked in both LASA and NLSAA, although the exact questions and the context in which they were asked differed. The LASA loneliness question was an item of the CES-D, whereas NLSAA's question was an element of the Life Satisfaction scale (Morgan *et al.* 1987). The LASA question asked about the frequency of feeling lonely during the last week, whereas the NLSAA question asked how often the person felt lonely. The response categories were also quite different, and were presented in the opposite orders in the two questionnaires ('rarely/never' was the first response category in LASA; 'often' was the first category in NLSAA). To harmonise the variables, it was therefore necessary to regard each set of responses as a four-point ordered scale, with the first response in LASA being equivalent to the final response category in NLSAA.

Contacts with health-care services

LASA and NLSAA asked about contacts with each of the following health-care services: family doctor or general practitioners; medical specialists or hospital doctors; district or community nurses; and health visitors. The LASA respondents were asked if they had had contact with these services during the previous six months (no/yes), whereas the NLSAA respondents were asked when they had last had contact with the services (with four response categories: within the last week/last month/last six months/more than six months ago). The harmonised variable had to be simplified to a dichotomy, whether the person had received or had contact with the specified service during the previous six months (no/yes), with the first three NLSAA response being conflated to 'yes'.

Physical activity measures

Variables relating to participants' physical activity were derived from analogous questions in LASA and NLSAA that had subtle but important differences, about the types of activity covered, the regularity or frequency of activity, and the reference period. To take walking, for example, the LASA respondents were asked about walking as for shopping and daily

activities but not for a tour or recreation during the two weeks before the interview (whether they at times went out for a walk; whether they had been out for a walk in the past two weeks, how many times they had been out for walk in that period, and how long they had been out each time). In contrast, the NLSAA respondents were asked about the last day on which the amount of walking they had done was 'typical or usual', and the total time they had spent walking or shopping (excluding leisure walking, *e.g.* hiking) that day. Although the specification of walking as a purposeful activity or for shopping was the same in the two studies, the reference periods differed (the previous two weeks in LASA, but the amount either yesterday or on the most recent typical day within the last month in NLSAA). Even if the LASA figure is divided by 14 to give minutes per day, the statistic is not comparable with the NLSAA figure because it represented activity on a 'typical' day, whereas LASA collected the aggregate duration over two weeks. The harmonised variable had to be a simple dichotomy, whether the person went out walking (no/yes). LASA respondents who were bed-ridden or wheelchair-bound, who said that they did not go out for walks, or who had not been for a walk during the last two weeks were coded 'no', and those who had been for a walk during the previous two weeks were coded 'yes'. NLSAA respondents who had spent no time walking on the last typical day were coded 'no', and those who had spent some time walking were coded 'yes'.

The same procedure was applied to the variables about other activities, namely indoor household tasks, cycling, gardening, and sports or recreational pursuits requiring at least a moderate degree of physical activity. The collected information on the frequency, regularity and time spent on the activities was not comparable in the two studies, only whether a respondent undertook the activity at all, for which dichotomies were created (no/yes). For indoor household activities, the LASA respondents were asked separately whether they undertook *light* (*e.g.* dusting, ironing, cleaning) or *heavy* (*e.g.* window cleaning, scrubbing the floor) household activities; whereas the NLSAA respondents were asked separately whether they undertook *light* (*e.g.* dusting, tidying up, ironing), *moderate* (*e.g.* cleaning windows, mopping) or *heavy* tasks (*e.g.* polishing furniture, scrubbing floors). Although the specified activities were very similar, the gradations of the required effort were incompatible, so it was believed most appropriate to conflate the grades and create a variable for whether or not household tasks were performed. Similarly, the LASA respondents were asked separately about gardening and digging the garden, whereas the NLSAA respondents were asked about *light*, *moderate* and *heavy* gardening tasks. The harmonised variable covered all gardening tasks.

Participation in religious organisations and pet ownership

The LASA respondents were asked whether they were members of or involved in organisations, and those who did were asked whether they visited a church or organisation with a religious or life-contemplation goal. The NLSAA respondents were asked whether they attended religious services, gatherings or meetings and offered three response categories: 'never' (excepting annual mass, weddings or funerals), 'sometimes', and 'often'). The new variable was whether the participant attended a religious service or organisation (no/yes). LASA and NLSAA asked almost identical questions about whether the respondent owned a pet and both used the binary 'no/yes' response categories, so the harmonised variable replicated this form.

Analyses of attrition in NLSAA

We turn to the testing of the null hypotheses that there were no associations between the baseline characteristics of the 1985 NLSAA sample and who was interviewed at the follow-up in 1993. Using the variables selected for data harmonisation, we first undertook a series of chi-squared tests to examine the association between the equivalent variables from 1985 and whether those still alive in 1993 participated in the NLSAA follow-up survey or not. There was a significant association between participation in the 1993 interviews among survivors and self-rated health in 1985 ($p = 0.009$), and with whether they did any gardening in 1985 ($p = 0.006$), but no association between the other 1985 variables and participation among the survivors in the 1993 interviews. We tested these results further using separate logistic regression models to determine how the 1985 attributes predicted whether the 1993 survivor participated in the interview in 1993 or not (Table 2). In the NLSAA, people with poor self-rated health in 1985 were more likely not to participate in the 1993 interview compared with people with excellent self-rated health. People who did not do any gardening in 1985 were more likely not to participate in the 1993 interview than people who gardened in 1985.

Discussion

This paper has described how harmonised data were developed from two independent cohort studies of nationally-representative samples of older people in The Netherlands and the United Kingdom, and discussed the challenges of this approach for comparing older people in different countries. It builds upon an extensive literature of studies that have

TABLE 2. *Predictors of attrition among older people who participated in the NLSAA in 1985 and who were alive in 1993*

1985 variable (reference category)	Category	OR	95% CI	*p*
Self-rated health (excellent)	Good	1.16	0.57–2.35	0.687
	Average/fair	1.52	0.72–3.20	0.275
	Poor	7.87	2.19–28.24	0.002
Does gardening? (Yes)	No	1.99	1.24–3.19	0.004

Notes: Results of separate logistic regression analyses of which 1985 variables and categories associated with non-response in 1993 (dependent variable). NLSAA: Nottingham Longitudinal Study of Activity and Ageing. OR: odds ratio. CI: confidence interval.

undertaken comparative social research (*e.g.* Fleishman and Shmueli 1994; Minicuci *et al.* 2003; Nikula *et al.* 2003; Shanas *et al.* 1968). A central issue is the extent to which the results of such a comparison are generalisable to the wider populations of Dutch and British older people (Deeg 2002). Any observed differences could be non-substantive (*i.e.* a result of methodological differences between the two studies, or attrition within them, or problems in the data harmonisation method) or substantive, indicating real differences in the health and wellbeing of the two populations of older people. Various factors could contribute to non-substantive differences and highlight challenges in undertaking cross-national comparisons using this approach.

First, the *sampling* for follow-up interviews may have different impacts on different studies. The (purposeful) over-sampling of people in the older age group in LASA and NLSAA, and of men in LASA, may have resulted in higher observed frequencies in specific categories, particularly if there were age, cohort or gender-related differences for particular variables. This can be overcome by weighting or controlling for particular groups in subsequent analyses. Second, different intervals from baseline to follow-up interview may affect later response rates. The LASA and NLSAA follow-up studies were ten months and four years after the original survey, respectively. Selective, and differential, mortality and non-mortality-related attrition may result in follow-up samples being biased, and therefore not representative of the wider population. Although mortality is non-random, it occurs naturally in both the overall population and the study sample (Deeg 2002). Therefore, for there is no reason to suggest otherwise, this is unlikely to have led to bias in either study's sample. When considering non-mortality-related attrition, refusal, failure to re-establish contact and the inclusion/exclusion of institutionalised participants may lead to sample bias in individual studies, particularly as the rate of institutionalisation depends on a country's health- and social-care

policies. Further analyses of the NLSAA respondents suggested that there was limited attrition within the sample. Examining the effects of non-mortality-related attrition helps at least understand, if not discount, this as a possible source of bias in follow-up surveys.

Third, differences in the *phrasing of questions and response categories* in the survey instruments used in separate studies and data harmonisation may create apparent differences. Respondents in LASA were asked about loneliness over the last week, whereas respondents in NLSAA were asked about the frequency of loneliness (Table 1): people may respond to these questions in different ways, particularly in relation to sensitive or emotional issues, or negative feelings. Differences in response categories may also affect participants' responses to certain questions, *e.g.* two equivalent categories for how respondents rated their health relative to peers, were 'much better' (LASA) *versus* 'much more healthy' (NLSAA). Differences in the context of questions and response categories in different studies might have affected the participants' responses. Participants in LASA were asked whether they had heart disease or had had a myocardial infarction whereas participants in NLSAA were asked whether they had heart trouble, and the examples provided were angina, rheumatic heart disease, palpitations, heart attack, and poor valve operation.

The use of *different instruments* may affect levels of response in different studies. The specific domains of cognitive impairment, anxiety and depression were measured using different scales, *e.g.* the MMSE and CAPE scales were used to measure cognitive impairment in LASA and NLSAA, respectively. The use of different scales to measure the same concept can be a source of error (Shanas *et al.* 1968), and it is possible that differences in the wording of specific scale items may affect the reported levels in cross-national studies. Despite the development of taxonomies and classification systems (*e.g.* the International Classification of Diseases (ICD) during the last few decades, numerous instruments and tools are used in different studies for measuring socio-demographic variables (*e.g.* education, occupation) and different diseases (including physical disabilities and psychiatric conditions), which hinders comparative research.

Fourth, the *timeframe* for questions and response categories may affect participants' responses, *e.g.* for the use of health- and social-care services. Participants in LASA were asked, 'Have you seen your doctor in the last six months?' (no/yes), whereas NLSAA participants were asked, 'When did you last see your doctor?' (last week, last month, within last six months, more than six months ago): recall bias may affect responses to having seen a doctor during the last six months or not. Differences in the organisation, funding and delivery of services by older people in different

countries could also create real differences in reported use. Similarly, differences in the expectations of families to provide support and care for older people may also affect the utilisation of services, and also how older people report their use of professional care services.

Fifth, the *context* in which otherwise similar questions are asked may affect participants' responses. The question on loneliness in LASA was asked as part of the CES-D scale (*i.e.* one of several questions relating to depression), and in NLSAA this was asked as part of the Life Satisfaction scale. The question immediately preceding the loneliness question may have affected the participant's response, *e.g.* in LASA it was whether they felt that during the last week they had talked less than usual, whereas in NLSAA it was how satisfied they felt with their life today.

Sixth, differences in the exact wording and meaning of the questions and response categories in the two studies arise partly from the *different languages* being used: Dutch by LASA and English by NLSAA. The translation of questions from English into Dutch in LASA, *e.g.* the MMSE and CES-D scales, may have changed the meaning or nuance and affected participants' responses. Similarly, translating originally Dutch questions into English for reporting purposes may have changed the meaning and affected the authors' understanding of the concepts being measured. Superficially equivalent words and phrases in, *e.g.* the adjective for 'excellent' or 'good' in relation to one's own health, may have a subtle difference in meaning in different languages (Shanas 1968). Additionally, the authors' first languages are Dutch (DD, JP) and English (PB), and using English as the language of communication may have resulted in differences in understanding during discussions on data harmonisation (Jackendoff 2009). Cross-country data harmonisation therefore needs to consider whether language differences between individual studies and among researchers affect observed responses.

Conclusions

Careful consideration of the methodological challenges faced when combining data from different cohort studies of older people using different methodologies, particularly when the studies are from different countries, should minimise bias in harmonised data sets and permit valid comparisons. Any subsequently observed differences between the samples should then indicate substantive or real differences between the populations of older people, *e.g.* cross-cultural differences and/or random variation in the populations, rather than artefactual differences arising from methodological differences. We are confident that the harmonised data we

developed from the two nationally-representative samples can now be used for comparative purposes. Additionally, we make recommendations for future comparative research.

First, we recommend that when designing comparative analyses from extant studies, the overall sampling and design are carefully considered to avoid the harmonised samples being non-representative of the populations of older people in each country. Second, the selective attrition between baseline and follow-up surveys in longitudinal studies should be examined. The effects of any differences in the study timeframes should also be considered. Third, the original measurement instruments should be examined carefully for differences in wording of questions and response categories. Fourth, the context in which questions in the studies are asked should be considered. Finally, international gerontology organisations could make recommendations for standard tools, *e.g.* for measuring health, wellbeing, and levels of activity, to be used in cohort studies of older people. We hope that providing this rationale for our approach and these recommendations will help others in undertaking cross-national comparisons of health and wellbeing among older people.

Acknowledgements

This particular study was supported by the British Council UK–Netherlands Partnership Programme in Science. LASA is facilitated primarily by The Netherlands Ministry of Health, Welfare, and Sports, and by the ~~Vrije Universiteit~~. We thank the reviewers and the editor, Tony Warnes, for suggestions that helped us improve the paper.

References

Bedford, A., Foulds, G. A. and Sheffield, B. F. 1976. A new personal disturbance scale: DSSI/SAD. *British Journal of Social and Clinical Psychology*, **15**, 4, 387–94.

Casado-Díaz, M. A., Kaiser, C. and Warnes, A. M. 2004. Northern European retired residents in eight southern European areas: characteristics, motivations and adjustment. *Ageing & Society*, **24**, 3, 353–81.

Curran, P. J. 2009. The seemingly quixotic pursuit of a cumulative psychological science: introduction to a special issue. *Psychological Methods*, **14**, 2, 77–80.

Curran, P. J. and Husong, A. M. 2009. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological Methods*, **14**, 2, 81–100.

Deeg, D. J. H. 2002. Attrition in longitudinal population studies: does it affect the generalizability of the findings? An introduction to the series. *Journal of Clinical Epidemiology*, **55**, 3, 213–5.

Deeg, D. J. H., Knipscheer, C. P. M. and van Tilburg, W. (eds) 1993. *Autonomy and Well-being in the Aging Population: Concepts and Design of the Longitudinal Aging Study Amsterdam*. Netherlands Institute of Gerontology, Bunnik, The Netherlands.

Deeg, D. J. H., van Tilburg, T. G., Smit, J. H. and de Leeuw, E. D. 2002. Attrition in the Longitudinal Aging Study Amsterdam: the effect of differential inclusion in side studies. *Journal of Clinical Epidemiology*, **55**, 4, 319–28.

Fleishman, R. and Shmueli, A. 1984. Patterns of informal social support of the elderly: an international comparison. *The Gerontologist*, **24**, 3, 303–12.

Folstein, M., Folstein, S. and McHugh, P. R. 1975. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 3, 189–98.

Hofer, S. M. and Piccinin, A. M. 2009. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, **14**, 2, 150–64.

Jackendoff, R. 2009. *Language, Consciousness, Culture: Essays on Mental Structure*. MIT Press, Cambridge, Massachusetts.

Knipscheer, C. P. M., de Jong Gierveld, J., van Tilburg, T. G. and Dykstra, P. A. (eds) 1995. *Living Arrangements and Social Networks of Older Adults*. VU University Press, Amsterdam.

Minicuci, N., Noale, M., Bardage, C., Blumstein, T., Deeg, D. J. H., Gindin, J., Jylha, M., Nikula, S., Otero, A., Pedersen, N. L., Pluijm, S. M. F., Zunzunegui, M. V. and Maggi, S. 2003. Cross-national determinants of quality of life from six longitudinal studies on aging: the CLESA Project. *Aging Clinical and Experimental Research*, **15**, 3, 187–202.

Morgan, K. 1998. The Nottingham Longitudinal Study of Activity and Ageing: a methodological overview. *Age and Ageing*, **27**, Supplement 3, 5–11.

Morgan, K., Dallosso, H. M., Arie, T., Byrne, E. J., Jones, R. and Waite, J. 1987. Mental health and psychological well-being among the old and very old living at home. *British Journal of Psychiatry*, **150**, 6, 801–7.

Nikula, S., Jylha, M., Bardage, C., Deeg, D. J. H., Gindin, J., Minicuci, N., Pluijm, S. M. F. and Rodriguez-Laso, A. 2003. Are IADLs comparable across countries? Sociodemographic associates of harmonized IADL measures. *Aging Clinical and Experimental Research*, **15**, 6, 451–9.

Pattie, A. H. and Gilleard, C. J. 1979. *Manual of the Clifton Assessment Procedures for the Elderly* (*CAPE*). Hodder and Stoughton, Sevenoaks, UK.

Radloff, L. S. 1977. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 3, 385–401.

Shanas, E. 1968. Health and incapacity in later life. In Shanas, E., Townsend, P., Wedderburn, D., Friis, H., Milhoj, P. and Stehouwer, J. (eds), *Old People in Three Industrial Societies*. Routledge and Kegan Paul, London, 18–48.

Shanas, E., Townsend, P., Wedderburn, D., Friis, H., Milhoj, P. and Stehouwer, J. 1968. *Old People in Three Industrial Societies*. Routledge and Kegan Paul, London.

Walker, A. (ed.) 2005. *Growing Older in Europe*. Open University Press, Maidenhead, UK.

Zigmund, A. S. and Snaith, R. P. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* **67**, 6, 361–70.

Accepted 20 July 2010

Address for correspondence:
Peter Bath, ~~Department of Information Studies~~,
University of Sheffield, Regent Court,
211 Portobello Street, Sheffield S1 4DP, UK

E-mail: p.a.bath@sheffield.ac.uk