11.03.12

Description of the Process of Creating the Harmonized Database

Presented to the Israeli Gerontological Data Center

Introduction

This paper describes the process of integrating five large-scale Israeli surveys of older adults into a harmonized database. The present report begins with an outline of the literature on previous harmonization efforts. Next, the present work is described, as well as the five different databases that were included in the harmonized file. The harmonization process required several adjustments to some of the inconsistencies across the surveys. These will be described in the following sections. We conclude with remarks regarding limitations and possibilities for future research based on the harmonized data.

In the past 30 years, data harmonization has become an important tool in research that requires comparisons between different data sets. In the harmonization process, researchers use data from different sources (different countries, different cohorts, different surveys, etc.), compare variables from these data sets, and provide a unified form for these variables. In particular, the goal of data harmonization is to create data that conceptually measure the same variables in the same units.

Harmonization is essential for learning whether a certain trend or behavior is specific to culture, cohort or study (Burkhauser & Lillard, 2005; Minicuci, et al., 2003).

In some cases, data can be easily harmonized across studies. For instance, given information on the month and year an individual was born, it is straightforward in both concept and practice to define age in common units of years. For variables
other than age and sex, the task of data harmonization may be more complex. The harmonization of variables such as “marital status” typically requires that a researcher merge multiple categories in one data set to the number of categories available in the data set that has the fewest categories defined. A most complex example of harmonization often involves the variable of income. The challenge of data harmonization of income is to measure income received over a similar time period (a calendar year, for instance). Moreover, harmonization efforts need to account for widely different tax schemes present in different countries (Burkhauser & Lillard, 2005; Minicuci, et al., 2003). In this review, different types of harmonization, and various examples of harmonization are covered.

Most harmonization projects conduct cross-national comparative research in order to investigate trends of human behavior that are beyond a particular society, culture, or population group. The need to use data from different populations and compare them yielded two types of research efforts. The first is ex-ante harmonization, where entirely harmonized questionnaires are developed and introduced into new surveys, or otherwise a new questionnaire is adjusted to suit ones in existing surveys. The second is ex-post harmonization, where results are reported ways adjusted to already available surveys. The former method has proven to be difficult to implement in practice (Burkhauser & Lillard, 2005). Therefore, most of the harmonization efforts mentioned in this review refer to cross-national ex-post harmonization projects.

The Comparison of Longitudinal European Study on Aging (CLESA) project is one example of this type of harmonization. The CLESA project encompasses cross-national data on health, physical and cognitive functioning, sociodemographic information, and social situations among older persons from six longitudinal studies.
which were conducted in six countries: Finland, Italy, the Netherlands, Spain, Sweden, and Israel. Subjects were older people aged 65-89 (N = 11557) living in the community or in institutions. The goal was to create a Common Data Base (CDB) and to investigate the distribution of some selected common variables among the six countries of the CLESA project (the CDB contained 111 final variables). Drawing on the CDB, common risk factors for mortality and functional decline and the determinants of quality of life and health of older individuals were identified (Minicuci, et al., 2003).

An important part of the CLESA group’s work was the harmonization of Activities of Daily Living (ADL) and Instrumental ADL (IADL) functioning variables. They examined the comparability of IADLs across countries and the association of IADLs with age, gender, and socioeconomic status. Comparison of data from the six CLESA countries yielded three common IADL items: preparing meals, shopping, and doing housework. All six data sets included a variety of variables measuring IADLs. The questions in the different countries were not identical, but addressed the same functional activity, more or less. The response categories were harmonized by dichotomizing the responses into two categories: able and unable. The socioeconomic indicator was created by harmonizing data about longest lifetime occupation. After the harmonization process, associations of IADLs with age, gender and socioeconomic status were analyzed by logistic regression models, and the scale properties were examined. The results indicated that associations between IADLs and background variables were similar across countries, despite cultural differences and differences in original variables. Moreover, the reliability of the three items scale was satisfactory. Hence, IADLs were comparable across countries (Nikula, et al., 2003).
The CLESA group produced various research projects based on the comparison of the six longitudinal studies. One study focused on the measure of Self-Rated Health (SRH), which may have different implications in various social and cultural settings. The study examined whether there are cross-national differences in the associations between status characteristics, health conditions, ADL, and SRH, using logistic regression analyses (Bardage, 2005).

An earlier, successful attempt at harmonizing existing data sets is the Luxembourg Income Study (LIS). The LIS project began in 1983. It contains information on household income and labor market outcomes from 29 countries in Europe, North America, the Far East and Australia from 1968 till the present. The main goals of the project were: harmonizing cross-national data, testing the feasibility of creating a unified database of surveys from different countries, and promoting comparative research (Burkhauser & Lillard, 2005; Smeeding, 2004).

The LIS data has been a major asset in the growing literature on cross-national comparisons of poverty and inequality. It provides a variety of research options, such as a general comparison of economic measures between countries. Allegrezza, Heinrich, and Jesuit (2004) analyzed poverty rates and income inequality for Luxembourg between the mid-1980s and 2000. They found that levels of poverty and income inequality in Luxembourg are among the lowest in the world and that the ‘depth’ of poverty is comparatively low.

Another option is to conduct comparisons between population subgroups across countries. Researchers have used the LIS data to analyze women's economic status, or economic gender inequality, across industrialized countries. Gornick (2004) synthesized the conclusions of over 60 LIS studies about gender and the labor market.
His review produced the following outcomes: (1) women are left behind men in labor market status in nearly every LIS country and time period; (2) motherhood has a negative effect on women's employment rates and earnings nearly everywhere, while for men parenthood has little effect (or a positive effect) on these factors; (3) single mothers everywhere face a heightened risk of low income and/or poverty; (4) gender inequality factors vary across countries; and (5) variation in policies explains a substantial share of the observed variation in outcomes (Gornick, 2004).

Researchers have also used the LIS for comparing research tools and scales. Buhmann, Rainwater, Schmaus, and Smeeding (1988) reviewed the availability of equivalence scales and examined the ramifications of choosing a certain measure of income inequality and poverty on results. Their examination showed that the choice of equivalence scale is likely to affect levels of measured poverty and inequality systematically, and in turn the rankings of countries or population subgroups within countries. Therefore, a careful consideration of the implications derived from cross-national comparisons of poverty and/or inequality is warranted.

All the above research examples were possible due to a thorough harmonization process, which could successfully provide comparable data. Next, we describe the development of the present harmonized database, and the original samples that were unified in this process.

Participants

The current project is comprised of data from five existing large-scale samples of the Israeli elderly population:

(1) **CALAS (Cross-Sectional and Longitudinal Aging Study)**. This database contains a nationally representative sample of Israelis aged 75 or older ($M = 83.52, SD$
= 5.42), interviewed during 1989-1992. The current, first-wave database includes 1369 self-respondents dwelling either in the community or in institutions. The CALAS conducted a multidimensional assessment of socio-demographic, physical, functional, cognitive, psychological, and social characteristics. More details may be obtained in a variety of publications based on CALAS data (e.g., Cohen-Mansfield, Shmotkin, & Hazan, 2010; Shmotkin, Blumstein, & Modan, 2003a, 2003b).

(2) IMAS (the Israel Multi-disciplinary Aging Study): This database contains a nationally representative sample of Israelis aged 65 or older ($M = 77.14$, $SD = 7.78$), interviewed during 1999-2001. The total database includes 721 self-respondents dwelling in the community. Using nearly the same survey questionnaire of the CALAS, the IMAS conducted a multidimensional assessment of socio-demographic, physical, functional, cognitive, psychological, and social characteristics (for more details, see Shmotkin, Lerner-Geva, Cohen-Mansfield, Blumstein, Eyal, Shorek, Kave, & Hazan, 2010; Shrira, Palgi, Ben-Ezra, & Shmotkin, 2011).

(3) SHARE-Israel (Survey of Health, Ageing, and Retirement in Israel). This database is the Israeli component of the large-scale European survey. It contains a nationally representative sample of Israelis aged 50 or older ($M = 63.39$, $SD = 10.12$) and their spouses regardless of age, interviewed during 2005-2006. The current, first-wave Israeli database includes 2598 community dwellers in 1771 households. The SHARE conducted a multidimensional assessment of socio-demographic, physical, functional, cognitive, psychological, and social characteristics. More details may be obtained in a variety of publications based on SHARE-Israel data (e.g., Litwin, 2009; Litwin & Sapir, 2008; Shmotkin & Litwin, 2008).
(4) **LSWBS (The Life-Span Well-Being Survey).** This database contains a convenience sample of Israelis throughout the adult life span (age 20 and older), interviewed during 1985-2003 (\(M = 49.82, SD = 20.50\)). The total database includes 7221 community and institution dwellers. Participants completed socio-demographic and psychological indices (more details may be obtained in studies that used earlier, more restricted versions of this database, e.g., Shmotkin, 1998; Shmotkin & Lomranz, 1998).

(5) **BSE (Biographical Survey of Elderly).** This database contains a convenience sample of Israelis aged 60 and older, interviewed during 2001-2003 (\(M = 74.08, SD = 8.04\)). The total database includes 776 community and institution dwellers. Participants answered to semi-structured questions about their life story, and completed socio-demographic and psychological indices (for more details, see Shmotkin, Berkovich, & Cohen, 2006; Shrir & Shmotkin, 2008).

The surveys were approved for ethical treatment of human participants by the institutional review boards of the Chaim Sheba Medical Center (CALAS and IMAS), the Hebrew University of Jerusalem (SHARE), and Tel Aviv University (LSWBS and BSE), respectively.

**Measures**

The five databases differed in the central variables that were measured. While the BSE and LSWBS databases were oriented towards assessing well-being in later life, the IMAS, SHARE, and CALAS were based on variables that measure functional domains of physical and mental health. The different surveys did share some similar or identical indices. The following section will describe the common and unique
variables of each database (see Table 1 for a complete legend of the original variables in each survey and the final variables in the harmonized database).

*Background characteristics* included age, gender, geographic origin (born in Israel, in Europe/United States, or in the former Soviet Union), marital status, and number of children. Participants were classified into one of seven education levels according to the International Standard Classification of Educational Degrees (United Nations Educational, Scientific and Cultural Organization, 1997).

*Life satisfaction* was measured in all surveys using a single item measure. Participants were asked to rate how satisfied they were with their lives in general on a scale ranging from 1 (very dissatisfied) to 4 (very satisfied).

*Depressive symptoms* were assessed in all databases except the BSE, by the Center for Epidemiological Studies - Depression scale (CESD) (Radloff, 1977). The scale items span negative affect, lack of well-being, psychosomatic reactions, and interpersonal distress. Each item specified a depressive symptom (e.g., “I felt sad”). Participants were asked to rate the frequency they had experienced each item in the past week on a scale ranging from 0 (*almost none of the time*) to 3 (*almost all of the time*). Four items were phrased positively (e.g., “I was happy”) and reverse coded. The respondent’s overall score was the mean rating of the items.

*Health* was measured by a basic subjective health rating in all databases. Participants were asked to rate their self-perceived health either on a 4-point scale (BSE, IMAS, and CALAS) or a 5-point scale (SHARE, LSWBS).
The SHARE, IMAS, and CALAS databases included other, more comprehensive indices of physical and mental functioning:

*Physical health and functioning*

1. Comorbidity: participants indicated whether they had ever suffered from any problem out of a checklist of 26 common medical conditions. The illnesses consisted of heart disease (e.g., myocardial infarction, coronary thrombosis, or any other heart problem including congestive heart failure), high blood pressure, high cholesterol, stroke or cerebral vascular disease, diabetes or high blood sugar, chronic lung disease such as chronic bronchitis or emphysema, asthma, arthritis (including osteoarthritis or rheumatism), osteoporosis, cancer or malignant tumor (including leukemia or lymphoma, but excluding minor skin cancers), stomach or duodenal ulcer or peptic ulcer, Parkinson disease, cataracts, and hip fracture or femoral fracture.

2. Medication use: respondents were asked to count and display the distinct medications they were currently taking. Number of medications was determined by summing listed medications that the participants reported taking at least once a week.

3. Contact with doctors: participants indicated the number of times they had visited a physician in the past year.

4. Health behaviors were assessed with five indicators: (a) Vigorous activities and (b) moderate-energy activities were assessed by summing participants’ rating of how often they engaged in vigorous physical activity (e.g., sport, heavy housework), and in activities that required a low or moderate level of energy (e.g., gardening, cleaning the car), respectively. Both items included a scale ranging from 1 (hardly ever or never) to 4 (more than once a week). (c) Smoking behavior was assessed by asking the participants if they smoked at the present time. The item was scored as 0
(no) or 1 (yes). (d) Alcohol consumption was assessed by asking the participants to rate how often they drank alcoholic beverages in the past 6 months on a scale ranging from 1 (not at all in the last 6 months) to 7 (almost every day). (e) Weight was assessed by first calculating the body mass index (BMI). BMI scores were further divided to below 18.5 (1 = underweight), 18.5–24.99 (2 = normal), and 25 and above (3 = overweight).

5. Physical functioning was assessed using four different indicators: (a) Activities of daily life were assessed by summing everyday activities (e.g., dressing up, showering; adapted from Katz, Downs, Cash, & Grotz, 1970) that participants reported to have been limited in their performance (possible range = 0–6). (b) Instrumental activities of daily life were assessed by summing activities (e.g., shopping, working around the house or garden; adapted from Lawton & Brody, 1969) that participants reported to have been limited in their performance (possible range = 0–7). (c) Activities with movement were assessed by summing activities (e.g., walking 100 m, climbing stairs; adapted from Nagi, 1976) that participants reported to have been limited in their performance (possible range = 0–10). (d) Subjective disability was assessed by asking participants to rate whether they had been limited because of a health problem in activities that people usually do on a scale ranging from 1 (not limited) to 3 (severely limited).

Mental health and functioning

1. Cognitive functioning was tested using seven indicators: (a) Time orientation was the sum of accurate responses that participants gave in request to name the current year, month, day of the month, and day of the week (possible range = 0–4). (b) Verbal learning was the sum of words that participants spontaneously recalled from a 10-
word list immediately after it was read to them (possible range = 0–10). (c) Verbal recall was the sum of words spontaneously recalled by participants from the list 5 min after the verbal learning task (possible range = 0–10). (d) Word fluency was the sum of correct names of animals that participants could think about within a 1-min trial. (e) Arithmetic was the sum of correct answers that participants gave to four arithmetic questions (possible range = 0–4). (f) Subjective reading and (g) writing were measured with two items asking the participants to rate their reading and writing skills, respectively, on a scale ranging from 1 (poor) to 5 (excellent). The final cognitive functioning score was the mean of the ratings, with a higher score indicating higher cognitive functioning.

2. Social exchanges were assessed with four indicators: (a) Conflicts inside family and (b) conflicts outside family were assessed by five items and one item, respectively. The first five items asked the participants whether they had experienced conflict with parents, parents-in-law, partner/spouse, children, and other family members. The final item inquired about conflicts with friends, coworkers, or acquaintances. (c) Feeling rewarded was assessed by asking the participants to rate three items asking (i) whether they were always satisfied with the balance between what they had given to their partner and what they had received in return, (ii) whether they had always received adequate appreciation for providing help in their family, and (iii) whether they had always been satisfied with the rewards for their efforts in their major activities (e.g., job, looking after home).

3. Cumulative life-event distress was assessed using two indicators: (a) Self- and (b) other-oriented adversity were assessed with the Traumatic Events Inventory, which includes 17 difficult life events (Shmotkin & Litwin, 2009). Self-oriented adversity
was the number of confirmed events in which the primary infliction was upon the self (e.g., “was the victim of violence or abuse”; possible range = 0–8), and other oriented adversity was the number of confirmed events in which the primary infliction was upon another person (e.g., “experienced the death of a spouse”; possible range = 0–9).

If confirming the experience of an event, the participants were further asked to specify their age when the event had first taken place and to rate the effect of the event on their life on a scale ranging from 1 (little) to 3 (great).

The BSE and LSWBS databases included specific measures of well-being:

1. The Self Anchoring Scale (Cantril, 1965), which measures an evaluation of one’s life as a cognitive element of SWB. It consists of a hierarchy of 11 levels, where the top score (10) represents the best possible life for the respondent and the lowest (0) represents the worst. The component of this scale that refers to the present time was also used in the IMAS database.

2. The Affect Balance Scale (Bradburn, 1969), which measures affective SWB by assessing its positive and negative components separately. Each are measured by five items that that tap recent emotional experiences, referring to recent occurrences of feelings such as "pleased" and "proud" (Positive Affect, or PA) or "depressed" and "bored" (Negative Affect, or NA). Respondents were asked to note how often these feelings occurred in the past week, using a 4-step scale with scores of 1 (never) to 4 (often). The mean ratings of each participant for the positive affect and negative affect items make up the scores for each measure.

3. The Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), examines satisfaction with life and is therefore a cognitive measure of SWB. It
contains five items referring to general judgments of one's life (e.g., "The conditions of my life are excellent") and rated by respondents on a 7-step scale ranging from 1 (strongly disagree) to 7 (strongly agree). The score was the items' mean rating.

Finally, there were some measures that were specific to each database, chosen based on the unique characteristics of each group:

The LSWBS database included a measure of hostile-world scenario (HWS). Based on the HWS scale developed by Shmotkin (2008), this is a 72-item measure of one’s beliefs and reactions to hostile-world representations (e.g., accidents, war, aging, illness, death). It is made up of the two dimensions: negative engagement and positive engagement with such representations. The participants’ use of each of the six declarative functions was also measured by the Declarative Functions Questionnaire (Shmotkin, 1998). This 51-item questionnaire examines one’s tendency to rely on different functions while reporting one’s SWB, with separate indexes for each function (self expression, self presentation, self deception, self reinforcement, self simulation, and defensive pessimism). The Big Five personality dimensions (neuroticism, extroversion, openness to experience, conscientiousness, and agreeableness) were measured using the 44-item Big Five Inventory (BFI; John, Donahue, & Kentle, 1991).

The BSE database included measures of participants’ outstanding life chapters, termed anchor periods. Interviewees were asked to rate two such positive meaningful periods (the happiest and most miserable) and two negative anchor periods (most miserable and most difficult). Participants also rated their happiness and suffering in each of the given anchor periods, resulting in differential SWB types (Shmotkin, 2005).
Data Harmonization Procedure

The harmonization of the data sets into a unified database allowed us to create an exceptionally large sample of Israeli elderly participants \(N = 14679\) with great variability on several indices central to functioning, well-being, and on background characteristics. Tables 2-7 present descriptive statistics, intercorrelations, and alpha coefficients of the main study variables in the original databases and in the final, harmonized database.

The goal of our harmonization efforts was to produce data measures that are consistent across surveys, out of existing databases that measured the same conceptual variables but sometimes differed in the original measurement tools (Burkhauser, & Lillard, 2005). We based the harmonization process on methods previously reported in the literature (see Bardage et al., 2005; Minicuci et al., 2004 for some examples of harmonization of disability and health measures).

The first step of the harmonization process was to search for parallel indices measuring similar concepts across the databases. To illustrate, similar scales of subjective (self-rated) health status were administered in all five databases. In addition, the Center for Epidemiological Studies - Depression scale (CES-D; Radloff, 1977), a widely used measure of depression, was administered in its full form in three databases (the CALAS, IMAS and LSWBS) and in short form in the SHARE. As another example, life evaluation was rated in certain variations by a single item in the CALAS, IMAS, and SHARE, whereas it was measured by the Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) in the LSWBS and BSE. In the present study, the standard we relied on for creating a common variable was that
information pertaining to the specific measure in question had to be included in at least three databases. Following the initial search, more than 150 items were drawn for harmonization.

The next step was to harmonize each variable across the relevant databases so that measuring units or result categories were identical. For several variables such as age or gender this step was straightforward and easily carried out, involving little or no change to the original variable (in one database, BSE, age had to be calculated by subtracting year of birth from the year the interview was taken).

In other instances, particular measures were taken according to certain limitations imposed by differences in variable formations across databases. These measures included creating new variables that matched research goals, such as creating the Holocaust survivor variable needed for selecting the comparison groups. Often, comparable measures were available in only some of the databases (e.g., a mini-mental test for cognitive status is available only in CALAS, IMAS, and SHARE; a measurement of subjective well-being is available in SHARE, LSWBS, and BSE). The following section will delineate the main harmonization considerations and principles.

**Main harmonization principles.** Selecting the variables that were common to all databases involved examining the similarities and differences in the wording of the questions and the response categories of each item. If both were identical or similar enough, the item was chosen and phrased in a manner that retained the common essence measured by the variable. In most cases, transformations were needed to match variable categories between databases.

For example, the harmonization of the variable “marital status” required merging multiple categories in one database to the number of categories available in
the database that had the fewest categories defined. A few variables required
dichotomizing the response categories in several databases, thus saving the variable
from being entirely discarded while losing some of the variance. For example, all 5
databases included subjective income variables with a scale range reaching up to 5
categories in one database. However, since a single database included a dichotomous
scale only, the variables in all 5 databases were recoded into a "sufficient/insufficient"
scale. Several variables required the transformation of interval or ratio scaled variables
into categorical scales (numerical or ordinal). For example, Level of Education was
measured in categories in some databases while in others it was measured in total
number of years of studying, requiring the recoding of number of years of studying
into specific categories.

In many cases the harmonization process required the creation of algorithms
and application of statistical techniques to verify the degree of correlations and the
internal validity of certain measures. For example, 3 databases included variables
regarding specific income sources. Using specific algorithms for each database we
summed the total number of income sources thus creating an ordinal objective income
scale of 6 levels.

The final unified database resulting from the harmonization process, contained
a number of 103 variables that covered the following content areas: sociodemographic
characteristics (age, gender, education, marital status, place of origin, religion,
religious level, age at immigration, number of living children, number of income
sources, etc.), Holocaust (survivorship, types of Holocaust experiences, parental
holocaust fate, etc.), health habits (smoking, alcohol consumption, physical activity,
etc.), health status (diseases, medications, CESD depression scale, doctor visits,
weight, height, etc.), physical functioning (ADL, IADL, NAGI scales), and subjective wellbeing (Diener scale, SWLS scale, Bradburn scale).

Convergent and divergent validity of the newly adapted measures were tested by specific databases that conduct multiple assessments in the same domains (e.g., the SHARE uses several different indicators of depression and also includes a large section of cognitive assessment; the LSWBS includes a 5-measure battery of subjective well-being).

**Summary and future directions**

The present work reported a description of the process of harmonization of 5 different databases of elder Israeli populations. While several variables were similar and, after modifications, appear across all surveys (life satisfaction, general background variables, Holocaust experiences), some were unique to certain databases (functional health measures in the IMAS, CALAS, and SHARE, and SWB measures in the BSE and LSWBS databases, for instance). Thus, the harmonized and final database is also a modular one, which will enable future cross-tabulations of many variables. One notable example for future studies on these data involves the effects of trauma in later life, as this database includes 1859 Holocaust survivors, not counting the comparison groups. This large-scale collection of samples of older adults can provide researchers with unprecedented information that pertains to late-life functioning and well-being in the face of traumatic experiences.
References


