2013 Research Network Meeting

Harmonization of Cross-National Studies of Aging to the Health and Retirement Study

EXECUTIVE SUMMARY

National Institute on Aging

Behavioral and Social Research Program

Prepared by:

Jinkook Lee and Tania Gutsche, RAND

August 2013

A Research Network Meeting
Harmonization of Cross-National Studies of Aging to the HRS

This document summarizes the 2013 Research Network meeting for Harmonization of Cross- National Studies of Aging to the Health and Retirement Study (HRS). The meeting, held on July 17th and 18th in New Delhi, India, was supported by the Behavioral and Social Science Division, the National Institute on Aging (R24 AG037866). The meeting was organized by Principal Investigator, Jinkook Lee of RAND Corporation, and held in conjunction with the advisory meetings for the Longitudinal Aging Study in India (LASI) and the China Health, Aging, and Retirement Longitudinal Study (CHARLS).

Among the participants were the principal investigators of the HRS family of surveys from around the world, experts in aging research, and program officers at NIA. Appendix A lists all attendees. New study participants this year include the LASI co-investigators and team members (T.V. Sehker, Sanjay Mohanty, A.B. Dey, P. Hu, Jeremy Goldhaber-Fiebert, Ashok Posture, and Jennifer O'Brien), a new principal investigator of the Korean Longitudinal Study on Aging (Ki Seong Nahm) and his research team members (Yoon Sub Kwon and Tae Keun Eom), a co-investigator of the Brazilian Longitudinal Study of Health and Ageing (Olivera Cesar), a new principal investigator of the 1958 British Birth Cohort of the National Child Development Study (Alissa Goodman), and a co-investigator of the Global Aging Data Repository (Marco Angrisani).

The agenda (Appendix B) for the meeting included study updates and scientific innovations each survey has adopted or plans to introduce, the progress of the Global Aging Data Repository, results from ex ante and ex post harmonization initiatives on Biomarkers, Genetics, Cognition, Accelerometers, Subjective Well-being, and Social Isolation, and round-table discussion of the opportunities and next steps for harmonization.

I. STUDY UPDATES

The goals of the Research Network are to (1) support the development of new international studies with harmonized data, (2) maintain harmonization and data development among existing comparable studies; and (3) foster a cross-national research agenda.

Given these goals, the principal investigators (PIs) of HRS-family surveys sought to foster sharing and adoption of scientific innovations by bringing together both existing and emerging data producers and leading scientists. All existing HRS-family surveys were represented, including the Health and Retirement Study (HRS), the English Longitudinal Study of Ageing (ELSA), the Survey of Health, Ageing, and Retirement in Europe (SHARE), the Irish Longitudinal Study on Ageing (TILDA), the Mexican Health and Aging Study (MHAS), the Korean Longitudinal Study of Aging (KLoSA), the Japanese Study on Aging and Retirement (JSTAR), the Indonesian Family Life Survey (IFLS), the Chinese Health, Aging, and

Retirement Longitudinal Study (CHARLS), the Brazilian Longitudinal Study of Aging (ELSI), and the Longitudinal Aging Study in India (LASI). Two additional, related longitudinal studies, the WHO Study on global AGEing and adult health (SAGE) and the 1958 British Birth Cohort of the National Child Development Study, were also represented in this effort to foster further collaboration. Below we summarize reports from these representatives to the meeting.

Health and Retirement Study (HRS)

David Weir, the PI of HRS, gave a study update, noting remarkable growth in the number of peer-reviewed journal articles using HRS, demonstrating the study's scientific contribution. In 2010 and 2011, HRS added a mid-boomer cohort and expanded minority samples in both mid-boomer and early-boomer cohorts. Over time, as a result of efforts to increase scientific innovation, the length of HRS interviews has increased. To remedy this, the HRS team reviewed sections and items for potential cuts. In this review, international harmonization was one of its concerns, and recommended cuts will be posted to the HRS website for comments.

The HRS has undertaken two major innovations regarding biomarkers and genetics. HRS has collected dried blood specimen (DBS). Recently, Eileen Crimmins and Teresa Seeman conducted a validation project, comparing DBS assay results with venous blood sample. They found that DBS results are often much different from those obtained from matched whole blood, while correlations between assays from the DBS and venous blood varied by markers. Using the 2005 – 2008 NHANES, they computed weighted quantiles of each assay for the population age 50 or older. Using regression equations, they converted HRS DBS values to NHANES equivalents while preserving rank ordering in the DBS values. HRS also submitted a supplement to collect venous blood samples to measure markers of cardiovascular health, anemia, kidney function, liver function, vitamin D, immunosenescence, and inflammation.

HRS is at the forefront of genetics work by HRS-family surveys. HRS has genotyped 12,507 respondents who provided DNA samples and signed consent forms in 2006 and 2008. The genotyping was performed using the Illumina Human Omni-2.5 Quad beadchip, with coverage of 2.5 million single nucleotide polymorphisms (SNPs). The genotype data were deposited in the NIH GWAS repository (dbGaP), which include imputation of about 22 million DNA variants from the 1000Genomes Project. Currently, about 57 approved dbGaP projects downloaded the HRS genetic data. HRS plans to add data from 3,500 respondents in 2010 and 3,000 respondents in 2012.

HRS has established several linkages to administrative data. Dr. Weir highlighted three major linkages: Medicare, Veteran's Administration, and Census Business Register. Establishing such linkages requires many negotiations but can yield rich details otherwise not possible. He also shared plans for the next wave. He invited all interested study teams to the interview training session, which has proven to be a good way to learn interview techniques and training tools. HRS provided "train-the-trainer" support to SHARE, LASI, and JSTAR and is willing to provide such support to new surveys.

English Longitudinal Study of Ageing (ELSA)

James Banks, a co-PI of ELSA, shared Wave 6 (2012/13) outcomes, news on development of Wave 7 (2014/15), and a few analytical developments. Wave 6 fieldwork is almost complete with a few nurse visits outstanding. Wave 6 conducted 10,319 CAPI interviews and achieved a household response rate of 79.4% (84.2% for non-refreshment sample). The nurse visits had a response rate of 78%, the core

self-completion questionnaire had a response rate of 93%, and the sexual activity self-completion questionnaire had a response rate of 81%. Noteworthy new contents for Wave 6 include time-use, affect, well-being, social care, and adaptive number series of cognitive tests.

While the funding for Wave 7 is still pending, the ELSA team is currently working on developing new contents for it. It planned a pilot test for August, 2013, a rehearsal in November, 2013, and fieldwork starting in April 2014. Main new elements for Wave 7 include hearing assessment, a modified social-care module, and a test for mild cognitive impairment. The cognitive test includes backward counting, serial 7s, naming objects, and animal naming. ELSA will rotate literacy and numeracy tests implemented in Wave 7. The addition of an adaptive number series test has raised concerns about the length of the cognitive module; researchers will further consider which tests should be rotated for it.

With funding from the Economic and Social Research Council (ESRC) and NIA, ELSA researchers genotyped DNA from the Wave 2 sample and is genotyping the Wave 4 sample. ELSA researchers used the same GWAS chip as HRS and plan to release the first data from approximately 8,000 respondents in late 2013 (N~8,000). ELSA researchers secured additional funding from ESRC and the Medical Research Council (MRC) for the FRAIL project to combine biological data from multiple levels, including proteome and metabolome together with genome.

Survey of Health, Ageing, and Retirement in Europe (SHARE)

Axel Borsch-Supan, the coordinator of SHARE, noted the sample was expanded in many countries in Wave 4 (2010/11), resulting in 150,000 interviews from 19 countries. Noteworthy new content of Wave 4 are social network questions, based on a name generator. The survey asks respondents, “From time to time, most people discuss things that are important to them with others […]. Who are the people with whom you discussed things that were important to you?” SHARE allowed respondents to list as many as seven names and then asked about role relations, network composition, living distance, contact frequency, support exchange, and satisfaction with network.

Wave 5 fieldwork is in progress with three new countries, Croatia, Luxembourg, and Catalonia, but not Hungary, which dropped the survey. A key innovation in Wave 5 is biomarkers: SHARE-Germany collected dried blood spots and plans to assay cholesterol, HbA1c, and CRP.

The Irish Longitudinal Study on Ageing (TILDA),

James Smith, the Chair of TILDA advisory board, presented the study updates on behalf of TILDA team. Wave 2 data collection was complete in December of 2012 and had a retention rate of 89.4%. The fieldwork took place from April to December subsequent to pilot testing in February and interview training in March and April. Only physical measures (i.e., grip strength and timed up and go) were collected in Wave 2 with about 97%, while other health assessments are collected in every other wave.

For Wave 3, which will begin in February, 2014, TILDA plans to collect health assessments both at a health center and in homes. TILDA anticipates collecting nearly 5,500 health assessments , using only one health center in Dublin (rather than two as for Wave 1). New health-assessment features include a food-frequency questionnaire, the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), National Adult Reading Test (NART, a measure of IQ), iPhone ECG to assess arrhythmia, a sound-induced flash illusion test to measure multisensory perception, Near Infrared Spectroscopy (NIRS) to measure cerebral perfusion, ultrasound of calf and stomach to measure sarcopenia, chair

stands, and accelerometry for a sub-sample.

Wave 3 also added new economic content, including level of physical activity at work, expenditure, numeracy, financial literacy, union membership and coverage, childhood economic status, and marriage bar and labor supply history. New social questions for Wave 3 include family transfers, social connectedness adopted from SHARE, coping inventory for stressful situation, and a modified version of the HRS resilience module.

Mexican Health and Aging Study (MHAS)

Dr. Smith also made a presentation on behalf of the MHAS team. The MHAS team collected Wave 3 data in 2012, re-interviewing survivors of 2003 Wave 2 and interviewing relatives of the deceased. Wave 3 MHAS added a sample to yield a representation of the Mexican population aged 50 and over in 2012. Given the nine years that had elapsed since Wave 2, the MHAS team achieved a remarkable 88.2% retention rate. Wave 3 adopted a CAPI mode rather than the paper-and-pencil mode used for Wave 2.

Wave 3 collected biomarker data for a sub-sample (N=2,089). From intravenous blood, researchers measured CRP, total and HDL cholesterol, thyroid, and vitamin D prevalence. Using point-of-service devices (Hemocue and Ac1Now), interviewers measured Hb and HbA1c. Other new content includes psychosocial questions. MHAS team is currently cleaning Wave 3 data and preparing to release them to the public in late 2013.

The MHAS team is currently designing and planning 2014 follow-up interviews as well as links to other data sources. County (municipio) level data they seek to link are those population census 2010, vital statistics (mortality by cause), Directory of health facilities from the Ministry of Health, Seguro Popular (new public insurance for previously uninsured) affiliation by year from 2004 to 2012, and social security (IMSS) records.

Korean Longitudinal Study of Aging (KLoSA)

A new KLoSA PI, Ki Seong Nahm, provided a brief overview of the structural change in the study team and an update on the study. The Korea Employment Information Service (KEIS) took over release of Wave 2 data; for Wave 3 and subsequent waves, KEIS has charge of both data production and release. Dr. Nahm became a new PI for the project for Wave 4 in 2012 and currently plans to lead Wave 5 in 2014. The retention rate for Wave 4 interview was 76.2% of respondents who participated in Wave 1.

The KLoSA has administered off-year surveys with different contents. In 2007, the off-year survey, using a modified version of the ELSA's life-history interview, collected job-history data. In 2009 and 2011, the off-year survey collected information about retirement and opinions about public policy. In 2013, the off-year survey collected detailed information about labor supply and activities after retiring from main job.

KLoSA team released an English version of data, codebook, and user guide for Wave 3 and is currently preparing these for Wave 4. They also seek to refresh the sample with respondents born in 1962 and 1963 for Wave 5 and to add a cohort to make the sample representative of the population 45 or older for Wave 6 in 2016.

Japanese Study on Aging and Retirement (JSTAR)

Satoshi Shimizutani, the PI of JSTAR, reported on progress and challenges for the survey. The JSTAR project started in 2005. Wave 1 in 2007 collected data from 8.000 individuals in 5 municipalities (Adachi, Kanazawa, Sendai, Shirakawa, Takikawa) and had a response rate of 60%. It added two municipalities (Naha, Tosu) for Wave 2 in 2009, and three more municipalities (Chofu, Hiroshima, Tondabayashi) for Wave 3 in 2011/12. It has conducted exit interviews as needed since 2009.

JSTAR collected interview data using CAPI and a self-administered questionnaire. It includes a nutrition-intake survey, the first such among HRS-family surveys. JSTAR used two-step stratified sampling, first choosing municipalities that allow linkages to official medical/LTC/health check data, then choosing individuals 50 to 74 years old, randomly selected from household registration rolls.

JSTAR researchers planned to implement Wave 4 in 2013, but their funding was sufficient to include only 7 municipalities, forcing them to raise additional funding. The JSTAR team also recognized a need to create a "Survey Research Center" to make JSTAR operations more stable and to recruit a "refresher sample" for Wave 5. Expanding the survey to other municipalities is a continuing goal.

Indonesian Family Life Survey (IFLS)

John Strauss, PI of IFLS, updated the status of the study. IFLS is an on-going longitudinal survey conducted by RAND and Survey Meter. In addition to individual and household information, IFLS collects detailed information on the communities where IFLS households are located and from the facilities that serve residents there. Wave 1, fielded in 1993 – 94, interviewed 16,300 individuals in 7,200 households. Since then three follow-up waves were completed with very high retention rates. Wave 5 will be fielded in 2014 – 15.

IFLS plans to introduce the following new modules for Wave 5: additional cognition modules, including number series from HRS; personality module, including BFI 15 questions; subjective well-being questions on positive and negative affect; HbA1c assays from dried blood; and balance and timed walk performance tests.

China Health, Aging, and Retirement Longitudinal Study (CHARLS)

Yaohui Zhao, PI of CHARLS, reported the study progress. The CHARLS team is currently preparing the first follow-up study, recruiting and training interviewers and developing tracking protocols. They have developed a longitudinal design for their questionnaire, deciding preloads, and determining changes and follow-up questions. Their exit interview is adapted from the HRS exit interview and will use a "verbal autopsy," discussed further below. For subjective well-being measures, the baseline CHARLS included a single-item general life satisfaction question. The CHARLS team plans to add a 12-item experienced subjective well-being measure to the longitudinal follow-up wave. For waves 3 and 4, CHARLS plans to add an adaptive cognition test, personality measures, time-use measures, an additional well-being measure using a day-reconstruction method, and an accelerometer.

CHARLS interviewers are distinct from those for other HRS-family surveys. The CHARLS team recruited interviewers among students from Peking University and other universities. Altogether, the CHARLS has 443 interviewers, whose mean age is 19, and 55% of whom are female. Peking University offers a semester course on "socioeconomic surveys: theory, methodology, and practice" at its main campus:

the course consists of large lectures with small practice sessions and includes intensive 12-day interviewer training after the end of the semester. Peking University also offers a course on "analysis of socioeconomic surveys."

The CHARLS field team comprises 46 roaming teams, each covering about three counties, each of which include 9 – 10 villages. Each team has about 10 members. Advance team members verify the accuracy of baseline sampling and conduct a community and policy survey, while the main team conducts individual interviews. Interviewers must receive a team leader's approval to abandon efforts to gain cooperation from a non-responding household head.

In tracking study participants, the advance team first canvasses baseline households to identify movers. The main team then locates contact information of movers. The team leader approves the transfer of the movers to headquarters, and then the headquarters review each case of a mover and transfer the mover to the team in the region where the move now resides. All teams stay in the field for the whole survey period so, ideally, they will identify all movers and re-interview them. Additional trackers later search for movers not identified and tracked.

At headquarters, a nine-person supervisor team manages five to six roaming teams. This supervisor team consists of CHARLS staff and outstanding student interviewers from Wave 1. They check on progress, boost morale, resolve field difficulties, and approve final interview status. CHARLS headquarters also has a 20-person quality-control team and help desk. The help desk approves proxy requests and addresses IT and questionnaire issues.

The quality-control team checks the first two households of every interviewer by listening to recordings and calling back if no recording is found. The quality-control team also records key questions at important skipping points and effort questions (e.g., subjective, mental health, vignettes) and identifies problem cases based on short interviews. The quality-control team also randomly checks about 10% of the interview.

The CHARLS team boasts a quick turn-around in public data release, releasing 2008 pilot data just five months after data collection and the national baseline survey data 11 months after data collection. The team will release data from the 2012 first follow-up interview of the pilot sample in September 2013. Their dissemination efforts include exhibits at international conferences such as PAA, IAGG, and IUSSP, domestic conferences, and Peking University courses. As a result of their efforts, the number of users and publications are rising rapidly.

Among the innovations the CHARLS team has introduced is verbal autopsy, a method used to ascertain the cause of death based on an interview with next of kin or other caregivers. Verbal autopsy uses a standardized questionnaire to elicit information on medical history, symptoms, signs, and circumstances preceding death. It provides a reasonable direct estimation of cause structure of mortality at a community or population level, but may not be an accurate method for attributing cause of death at individual level.

The CHARLS team is collaborating with the WHO's SAGE-China team in adopting 2012 WHO verbal autopsy. This has a separate questionnaire for each of three age groups (less than 4 weeks, 4 weeks to 14 years, and 15 years or older). The listed causes of death are those based on (1) frequency of a specific cause of death and its importance for global mortality levels; (2) causes that public-health interventions can address; and (3) feasibility of ascertaining a particular cause of death being through a

limited number of questions that are amenable to software analysis. The China Center for Disease Control conducted a pre-test of the Chinese verbal autopsy questionnaire for 14 deaths, comparing the results with death certificates. Verbal autopsy yielded eight matching cases and discrepancies for the other six. There was no evidence to suggest that these discrepancies were related to translation. The verbal autopsy questionnaire requires 20 to 30 minutes to administer.

For Wave 1 of the CHARLS, venous blood specimen was collected in collaboration with the China Center for Disease Control (CDC) and the Capital Medical University (CMU) in Beijing. The CDC completed the blood-cell count test in the field, and the CMS conducted the pre-test for cystatin C assay in mid-2012. The following bioassays were completed on all stored venous blood samples (about 12,000) by July, 2013: total, LDL, and HDL cholesterol, triglyceride, CRP, HbA1c, glucose, BUN, creatinine, and uric acid. The CHARLS plans an additional bioassay for 2013 and 2014 on a random sample of 6,000 venous blood samples for cystatin C. For 2015, the survey will collect whole blood, with various options being considered in an implementation plan.

The WHO Study on Global AGEing and Adult Health (SAGE)

Somnath Chatterji, PI of SAGE, presented the study's progress and future directions. SAGE stored dried blood spots from 45,000 respondents and has been developing the following assays: Hb, HbA1c, hs CRP, EBV, HIV, HDL and total cholesterol, triglycerides, and IL6. SAGE plans to develop assays for telomerase, TREC, and CMV and also plans to collect salivary samples for DNA.

SAGE completed Wave 1 data collection at all national sites, including China, India, Russia, Mexico, South Africa, and Ghana, as well as eight INDEPTH sites (India, South Africa, Ghana, Kenya, Tanzania, Bangladesh, Viet Nam, and Indonesia). SAGE is developing a longitudinal follow-up for Wave 2, planning waves 2 for 2013 and wave 3 for 2015.

Regarding future direction, SAGE aims to improve understanding of future fatal and non-fatal health outcomes and determinants. It continues to develop methods to improve measurement, including those for health states, self-reported morbidity, risk factors, and effect of interventions. The SAGE team plans to link survey data to demographic surveillance data from INDEPTH. They also plan to add measures of personality and life histories based on other surveys, diagnostic measures of dementia, and harmonized measurements of cognition and the costs of dementia.

Brazilian Longitudinal Study of Aging (ELSI)

Fernanda Lima-Costa, the PI of ELSI, reported progress in developing ELSI, the Brazilian Longitudinal Study of Health and Ageing. ELSI will be a population-based cohort study of 10,000 adults aged 50 or older and representative of the nation. The ELSI team is planning biennial face to face interviews as well as nurse visits and blood tests and will allow proxy interviews as necessary.

The ELSI team is currently developing an instrument with a household and an individual questionnaire. The household questionnaire will have household roster, family wealth, family monthly expenditure, and income modules. The individual questionnaire will have questions on demographics, cognition, neighborhood context, status for the elderly, life-course history, occupational history, current work condition, retirement and pension, individual wealth, intergenerational transfers, health behaviors, women's health, physical health, physical function, oral health, use of health services and medications, sleeping, depressive symptoms, and psychosocial items.

The ELSI biomarker protocol includes: height, weight, waist circumference, blood pressure, pulse rate, grip strength, balance tests, gait speed, chair rising, saliva sample for DNA extraction, total and HDL cholesterol, glycosylated hemoglobin, C-reactive protein, and Vitamin D. Dr. Limia-Costa noted the distance to be used for gait speed is an open question given space limitations in poor areas.

The ELSI team planned to pretest the questionnaire in August and September of 2013 and to organize an international consultant meeting in October or November of 2013. It has already secured from the Brazilian Ministry of Health (DECIT) for the baseline study, to be conducted in 2014 and 2015.

1958 British Birth Cohort of the National Child Development Study

Finally, Alissa Goodman, the PI of the 1958 British Birth Cohort, introduced the National Child Development Study. This is a long-running panel study that has followed 17,415 children born in a single week in UK in 1958 at ages 0, 7, 11, 16, 23, 33, 42, 46, and 50. The 1958 Birth Cohort Study is currently in the field to interview respondents now 55 years old, and plans a major follow-up at age 60. This is a unique resource on ageing and presents an opportunity for a "cross-walk" with other aging studies, particularly with ELSA, about early life experiences and later life social, economic, and health behaviors and outcomes.

II. HARMONIZATION INITIATIVES

Richard Suzman, Director for the Behavioral and Social Research Program at the National Institute on Aging (*NIA*), discussed **NIA program priorities for harmonization**. Several research network participants also presented the progress and findings on several harmonization projects, including the Global Aging Data Portal, Biomarker validation, Genetics, Cognition, Accelerometers, Subjective Well-being, and Social Isolation.

NIA Perspectives

Dr. Suzman emphasized the need for harmonization, not homogenization. Homogenized studies may have inadequate or insufficient DNA data for scientific advances. He recognized that there is always tension between maintaining past measure consistency that is usable for cross-national and other comparisons and generating new approaches and measures. Continuous improvement is necessary to improve measurement quality, and multiple ways to measure identical or similar concepts or constructs have proven useful. He noted that a core set of measures (around 10 – 20%) that may be compared with new measures or approaches could be useful.

Suzman noted that the scientific values of surveys depend on the importance of the questions they can answer, the number of users and researchers, the number of publications in top journals, and the extent to which they may form a synthetic cohort for genetic analysis.

Francis Collins, the NIH Director, has called for a prospective cohort of more than 500,000 to discover

genetic-environment interactions. For identifying main genetic effects, sample sizes of 2,000 to 10,000 are needed; for main lifestyle effects, sample sizes must be 2,000 to 20,000. For identifying gene and lifestyle and gene-gene interactions, a sample size must be at least 10,000, usually needs to be 20,000, and ought to be at least 50,000 for a comprehensive platform. Given the requirement of large sample sizes for discovery of genetic effects, a need for synthetic and harmonized cohorts has emerged, increasing the value of comparative and cross-national studies.

Also, there is a need for behavioral synthetic cohorts. According to Schroeder et al. (2007)[1] in NEJM, behavioral patterns account for 40 percent of premature mortality. Other contributors to premature mortality are: genetic predisposition – 30%; social circumstances – 15%; environmental exposure – 5%; and health care – 10%. Many GWAS-discovered risk variants affect genes that exert their influence through behaviors (e.g., FTO) or have behavioral phenotypes (e.g., APOE).

Dr. Suzman noted that comparative research has played an important role in understanding behavioral determinants of health. Chris Murray is developing a new measure for behaviors affecting health and disability effected life years. Other key work includes the Banks, Marmot, Smith paper in JAMA (2006)[2] on Anglo-U.S. comparisons, as well as Avendano et al.'s paper in AJPH (2009)[3] on the European Union.

NIA funds many national cohorts: HRS and international HRS-like cohorts, the Panel Study on Income Dynamics (PSID), Midlife in the United States (MIDUS), National Health & Aging Trends (NHATS), National Social Life, Health & Aging Project (NSHAP), Project Talent, many Twin and Cognition cohorts, and U.S. regional cohorts such as the Wisconsin Longitudinal Study and the Northern California, KP/UCSF cohort. Some of these studies are broad and general. Others are narrow and examine outcomes in-depth, with questionnaires that spend several hours on cognition, personality or cardiac functions. Ideally, deep studies should produce brief and inexpensive measures that capture the essence of the outcome – even binary measures, if need be for use in the field.

Behavioral and Social Research Division, NIA has taken several actions for harmonization. It has held competitions for networks and harmonization studies, workshops on harmonization and topical areas, such as measuring lifetime stress, subjective well-being, conscientiousness, and economic phenotypes for genetic analyses.

Two NIA –funded networks to harmonize cross-national research and data are: network on measurement of biological risk (USC/UCLA) and Harmonization of Cross-national Studies of Aging to the Health and Retirement Study (RAND). The Survey Meta Data Repository is a resource developed for harmonization. It provides a concordance/search engine for HRS-family survey data to support cross-national research. It also allows searches for comparable questions across surveys. Finally, it includes a set of identically defined variables constructed for cross-country analysis (see http://mmicdata.rand.org/megametadata).

---

[1] Schroeder, S.A. (2007) We can do better – Improving the health of the American people, New England Journal of Medicine, 357: 1221- 1228.

[2] Banks, J., Marmot, M., Oldfield, Z. & Smith, J. (2006) Disease and disadvantage in the United States and in England, Journal of American Medical Association, 295: 2037-2045.

[3] Avendano, M., Glymour, M., Banks, J., & Mackenbach, J.P. (2009) Health disadvantage in US adults aged 50 to 74 years: a comparison of the health of rich and poor Americans with that of Europeans, American Journal of Public Health, 99: 540 – 548.

Several NIA workshops have brought together behavioral scientists, social scientists, and geneticists together. Reports summarizing these workshops are available at http://www.nia.hih.gov/research/dbsr/featured-reports/. For example, participants at the USC/UCLA R24 network harmonization meeting concluded that topics such as sleep and time use need further research to enable cross study comparisons. Yet, they noted, current surveys cover well topics such as depressive symptoms and disability, with items fairly easy to harmonize. One outstanding challenge is that many questionnaires are too long and studies are more focused on maintaining comparability over time. NIA also organized a workshop to explore harmonization strategies to maximize the value and use of data within the behavioral and social sciences and to accelerate research integrating these data with genetic and genomic inquiry.

Dr. Suzman noted the importance of both ex ante and ex poste harmonization. To facilitate comparative research and the replication of genetic associations, sufficient ex ante harmonization among studies is required. HRS-family surveys should also perform ex post harmonization, such as that by the Survey Meta Data Reporitory, as much as possible. Several considerations are needed to determine how much ex ante harmonization is desirable. Above all, it is important to ensure comparative research and replications are possible. As noted earlier, researchers need variants of measures to improve them. Researchers must also account for country-specific variations.

Several phenotypes and measures may be suitable for data harmonization. These include health outcomes (e.g., cardiovascular diseases, depression); wealth; cognitive functioning; health behaviors; time use; physical activity; subjective well-being; time discount function, risk taking; personality measures, including conscientiousness; stress, including lifetime stress; childhood experiences; and loneliness and social isolation. Other suitable ideas and suggestions for harmonization may evolve over time. For example, researchers may harmonize a birth cohort such as the 1958 cohort to a study starting later in life such as ELSA to model initial conditions. Adding common modules for calibration or competing measures would also yield many insights. A research network might work to improve measurement. Indeed, there is a new RFA for research network, which elicited an application on stress over the life course. From an in-depth assessment, a short and simple measure may be developed for use in population surveys.

Harmonization Initiatives

Several meeting participants discussed recent harmonization initiatives for HRS-family surveys. These included Global Aging Data Repository, biomarker validation studies, genetics initiatives, new cognition measures, subjective well-being, accelerometers, and social isolation/network.

Global Aging Data Repository

Jinkook Lee, Bas Weerman, and Marco Angrisani updated the progress and innovations of the **Global Aging Data Repository**. The repository facilitates cross-country analyses using the HRS-family surveys. It currently contains: (1) metadata from and searchable concordance information on eleven surveys from 25 countries and (2) harmonized data files for six surveys, presented in user-friendly, individual-level longitudinal files merged with key household and spouse-level variables and a set of identically defined variables for cross-country analysis. The harmonized data files are available for HRS, ELSA, KLoSA, JSTAR, CHARLS, and LASI.

They demonstrated the newest innovation of this project, presenting harmonized variable estimates for population and sub-population levels in interactive tables and graphs. With this new feature, users can specify the aggregate statistical information they want to see by selecting (1) harmonized variables of interest, (2) population or sub-population of interest (e.g., Swedish women age 50 or older), (3) survey year(s), and (4) preferred format of display (table or graph). Without downloading micro data, users can easily compare descriptive statistics of all harmonized variables on the site. These tables and graphs illustrate cross-country differences at a point in time as well as longitudinal trends within countries over time. Such information can motivate researchers to further examine causes of cross-country differences or similarities and thereby facilitate scientific advances.

Biomarker Validation Studies

Perry Hu presented lessons learned from **Biomarker** validation studies, especially those from his experience with assaying dried blood specimen (DBS) in three Asian countries for CHARLS, IFLS, and LASI. The DBS assays for these studies are: C-reactive protein (CRP) and hemoglobin (Hb) for CHARLS, CRP for IFLS, and CRP, Hb, Epstein-Barr Virus (EBV) antibody titer, and Glycosylated homglobin (HbA1c) for LASI.

He outlined three main steps for bioassays: training lab personal, securing equipment, and obtaining test reagents and supplies. Three studies had varying training protocol. CHARLS had a five-day training session at Thomas McDade's laboratory at Northwestern University. IFLS and LASI had a prior training as a part of the SAGE training protocol, with lab personnel re-trained for five days at their own lab (although for some staff it was the first training because of staff turnover). Dr. Hu concluded that there is no "magic number" of training days, but prior experience as well as number of assays are important to consider in determining an ideal number of training days. The level of technical support available during the pre-test is key to bioassay success, and is even more important when there is a significant delay from time of training to implementation of bioassays. Details of equipment, reagents, and supplies are also important. Although labor costs are lower in developing countries, other costs, such as reagents, may be higher. There are many challenges in ensuring the availability and shipping of reagents and supplies.

Dr. Hu emphasized the importance of two additional steps for successful bioassays: assay validation using a validation sample and on-going quality control for the study sample. The goal of pre-tests for DBS-based assays is to evaluate the technical skills and transfer of knowledge capabilities. Evaluating the lab environment is equally important, as temperature and humidity can affect results. It is important to evaluate the reliability and validity of the assay results, using both study and validation samples.

For CHARLS and IFLS, validation samples were obtained by recruiting volunteers, collecting two DBS cards with two blood spots per card. These cards were then sent to a U.S. lab as well as an international lab. Specifically, serum samples were sent to University of Vermont or University of Washington for serum-based assays and Northwestern University or University of Washington for DBS-based assays.

For LASI, DBS cards with five blood spots per card were created from venous blood specimens, which were collected from volunteers recruited through the USC/UCLA Biodemography Center. Serum samples were sent to UCLA Clinical laboratory, and the DBS cards were sent to University of Washington and NARI lab for assays. A temperature monitor was included in shipments to NARI to maintain low temperatures, though this was expensive ($1000 for 50 cards).

From these exercises, we learned that a large quantity of dried blood spots is needed to track change over time. Ideally, studies would have validation samples from the same individual over time. Retesting of specimens is also possible for some assays. For validation, IFLS was able to include a validation sample for every 1,500 samples. LASI had many more samples, so five validation samples could be included in every fourth plate. Due to the large sample, LASI researchers were able to track samples from same individuals over microplates.

Altogether, the three most critical elements for DBS studies are

1. Conduct well designed pretest and monitoring
2. Good pretest and quality control
3. Keep in touch with lab

Currently, there are no widely-accepted criteria to define acceptable assay quality and or variability between labs. For harmonizing biomarkers, there are many elements to consider. There are different protocols in sample collection, shipment and processing; different assay methodologies; and laboratories that operate for limited periods of time.

Initiatives in Genetics

David Weir discussed the current state of **Genetics** initiatives for HRS-family surveys. Behavioral scientists are increasingly interested in genetics, as genes affect both behavior and biology, and through biological mechanisms behavior and environments affect health. The integration of genetic data and theory into science is a big transformation waiting to happen. The HRS family of studies is a relatively big consortium in the world of social science, and therefore in position to lead the integration of genetics and social science.

HRS began collecting DNA in 2006 and storing the information without funding or definite plans. The GWAS model was then identified as the best approach for HRS, as it has broad coverage of genomes, permitting exploratory work in a wide range of health and behavioral phenotypes, especially longitudinally defined phenotypes. Simple GWAS genetic discovery now requires many studies in combination for statistical power and credibility. Yet simple discovery is not the only thing one can do with DNA.

In the past, twin studies or other family studies were the norm for genetic work. Twin studies indicate how much of various traits are heritable, but when you look at GWAS studies there were significantly fewer heritable traits. This is so-called “missing” heritability. From a dataset, researchers can create measures of relatedness between pairs of individuals based on how many SNPs they have in common. Researchers can then look at how phenotype variation co-varies with relatedness and estimate heritability of traits.

How genes and environments may interact is quite complicated. Some genetic variation may only be relevant in certain environmental conditions, and some social factors may influence how genes work. DNA starts in a fertilized egg, and the same DNA produces all different cells. Epigenetics recognize that only some genes get expressed, while others are mute. Expression is the production of transcriptome (RNA) which then shapes proteins. RNA is now addressable by chips and sequencing.

Weir argued for building larger databases of DNA matched to similarly-measured behavioral

phenotypes, and work on models of social science that integrate genetic information to provide exogenous instruments and control for heterogeneity. HRS, CHARLS, ELSI, and SAGE are committed to doing genetic work.

New Cognition Measures

James Banks, David Weir, and John Strauss presented preliminary analyses of new **cognition** measures, an adaptive test of number series, adopted by HRS, ELSA, IFLS, and MHAS. Number series is a measure of fluid intelligence, measuring how well one reasons. It appealed to the HRS team, because numbers are easier to harmonize than words or images. The HRS team implemented the number series in 2004, 2006, and 2008 experimental modules.

One aspect to work through when adding cognition modules is methodology. McArdle and Woodcock developed a block-adaptive approach using normed unpublished items from Woodcock-Johnson test. They started with a pool of 15 items, with batch one as 3 items taken from the middle, and subsequent routing depending on how well respondents answered the initial three items. To develop adaptive testing, "pre-normed" items (i.e., prior data on the relative difficulties of items) are required. From 45 available questions, each person gets about 5 to 6 questions using adaptive testing strategy, from which researchers can assign scoring based on the psychometrics behind the items. Adaptive testing creates efficiency by identifying individual ability with relatively few items and thereby improving respondents' experience as well.

Analyses of HRS experimental modules find that short adaptive sequences from block-adaptive tests have high weight on first items asked. That is, seeding by education is efficient, but the final result is anchored by the first question asked.

Both HRS and ELSA use identical module questions but different training questions. HRS felt examples in 2010 were not difficult enough to give respondents a sense of the questions so added another training question. ELSA trained respondents with the harder example but as a result many could not get through all the examples. The HRS takes longer as it gives two examples. There are more ineligible respondents in the HRS and proxies are ineligible to be tested.

The interview model for ELSA was completely face-to-face, while HRS conducted half its face-to-face and half by telephone. Interviewers gave spouses two different lists so that they wouldn't overhear questions they would later be asked. HRS had a cognition module in 2010 and 2012, but there were issues making longitudinal comparisons between them. The interviewer is heavily trained and instructed to allow respondents to use pen and paper to make notes, but allow only certain amount of time. Even if respondents provide wrong answers to practice questions, the interviewer can still move them on to the test.

Overall, number series adaptive testing seems to be performing well in both surveys. Calibration distribution provides a metric for scoring. Banks and Weir presented unweighted distribution of sample means, patterns across demographics, and the relationship with other cognitive tests. They noted the age pattern on the testing shows a predicted decline. Female scoring was lower, with a large decline by age. Controlling for education, memory and numeracy are closely associated with the number-series score.

Instead of using the McArdle and Woodcock's score based on psychometric properties, Banks and Weir created their own scoring scheme that is not bounded by the first question in sequence and therefore more forgiving of mistakes made earlier in sequence. This scoring scheme resulted in six levels of fluid intelligence, explaining 90% of the variation in the score. This scheme showed good correlation and co-variation patterns. For example, respondents with the best memory are more likely to have higher scores for fluid intelligence. However, the coding of the questions remains though; and it is not clear what calibration should be when you are thinking across countries.

Strauss discussed the IFLS/MHAS New Cognition Tests. These resulted from comments by Willis on cognition measures in China at the Harmonization meeting 2011. Wong, the MHAS PI, and Strauss as IFLS PI after that session discussed whether the questions used in the HRS would work in countries where SES is much lower. One of the worries they had in Mexico or Indonesia or China was large fraction of respondents may skip out before the test. They collaborated with McArdle and Dick Woodcock who designed adaptive number series tests. They did simple pilots in Indonesia and Mexico, but not in China due to costs. Paid partially with funding from the NIA Biodemography center, the small pilot in Indonesia cost $11,000. That in Mexico cost about $14,000 and was paid by additional funding from the University of Michigan. Both, Strauss would later note, showed that a small amount of money can have a large impact. Following a PAA discussion this year, Hans Peter Kohler also noted plans to use the New Cognition Tests in rural Malawi.

In piloting this cognitive test on number series, the first question researchers considered was whether they could even use numbers, given low levels of numeracy. The pilot tests therefore used both numbers and pictures. They found that in Indonesia, 90% of respondents "strongly preferred" numbers over the pictures. Pictures add complexity rather than ease given that respondents still have to count items. Questions with pictures also took three to four minutes, on average, longer than those with numbers, though the number of correct answers for each type of test did not differ.

Following these initial pilot tests, the IFLS and the MHAS teams did formal pretests in Indonesia and Mexico. McArdle and Woodcock designed 15 questions. In Indonesia they tested 390 (in February 2013) and approximately 300 in Mexico more recently. In Mexico, almost everyone answered the first question, on completing the sequence 1 2 3 _, correctly, while in Indonesia 15% do not.

Strauss presented some descriptive statistics on these tests. They showed that even 14% of Indonesian respondents with no education still get some questions correct, but also that the tests are successful in discriminating among abilities.

Among Mexican respondents, nearly all answered the easiest question correctly, and none received a zero score. Distribution of scores varied greatly between Mexico and Indonesia, leading Smith to note the importance of scoring.

Meeting participants discussed whether SHARE and SAGE will consider similar questions and, if so, what questions would they replace. Banks noted ELSA accommodated such questions by removing the animal naming test. Others noted it is not necessary to do all the cognitive test in each wave, and that some cognitive tests might, for example, be rotated.

Subjective Well-Being

Kaptyen discussed a harmonization initiative on **subjective well-being**. Subjective well-being can be defined in different ways through evaluative, experience, and eudemonic measures. Evaluative measures have been used in hundreds of surveys, including HRS, ELSA, and SHARE. Experience measures are newer and less well known. They are more costly to collect (e.g., Day Reconstruction Method), as ideally one wants to cover many episodes in a day and then elicit many affect categories. It is really time use measurement plus elicitation of mood. Eudemonic measures have been less popular but are used in major surveys such as the ONS survey (British Statistical Office).

Kaptyen and his colleagues compared well-being measures adopted by HRS, ELSA, SHARE, and ONS surveys. They found various measures were used, and that even surveys using the same questions often used different scales, raising the need for calibration. Kaptyen and colleagues conducted an experiment to investigate the relations among four sets of evaluative, three sets of experienced measures, and one eudemonic measure, using the American Life Panel (ALP). The ALP is the first population survey to collect jointly all these different types of measures , yielding new insights on issues such as the relation among the three types of measures, the number of underlying dimensions, the importance of response scale, test-retest reliability, and external validity.

Kapteyn and colleagues administered all evaluative well-being measures in two waves of ALP. In each wave, respondents answer one experienced well-being measure. The order of the evaluative well-being questions was randomized. Kapteyn presented the results, including the administration time for each measure, test-retest reliability, and the results of factor analysis. The experiment found that all evaluative measures load on the same factor. It found that the eudemonic measure does not appear to represent a separate factor but rather mainly loads on the common evaluative factor. Yet it also found that the positive and negative experienced affect measures load on different factors, thus confirming that positive and negative affect are not simply opposite poles on the same scale. The researchers found that that negative affect can be represented by one or two factors, depending on the scale used.

The relation of evaluative and experienced measures with demographics is markedly different. Evaluative well-being increases monotonically and almost linearly with income, but there is no such relation between experienced well-being and income. Evaluative well-being has a U-shaped relation with age but experienced well-being does not. Health and labor market status, which have clear and significant effects on evaluative well-being, do not appear to have a consistent influence on experienced well-being; what influence they do have is related to the kind of response scale used in eliciting items.

In general, the relation between experienced measures and demographics appears to be much weaker than that between evaluative measures and demographics. The work of Kapteyn and colleagues is the first systematic study of the relation between different subjective well-being concepts. It found that response scales matter, and in particular that a 2-point scale appears to be too crude.

Smith suggested Kapteyn create a translating slide with recommended content to be used by other studies. Kapteyn recommended the ELSA's experienced well-being measure with a 7 point scale and the SHARE's single-item evaluative well-being measure for such a translating slide. Discussants also noted that it may be useful to link evidence Steptoe found. For devising policy implications, Suzman noted that looking at suffering and misery is more important than joy.

Accelerometers

Kapteyn presented results of another harmonization experiment, an **accelerometer** being done in three

countries: England, Holland, and the United States. Accelerometers, first developed by UK Geneactiv, measure acceleration, skin temperature, and day light. Dutch data, collected through an Internet panel in the Netherlands like the ALP, are now available. These include 850 observations collected in a 13-week period among 70 to 90 panel members weekly who wore the device for 8 days, using 300 accelerometers. The experiment also asked for self-reported physical activities for one weekday and one weekend, as well as global questions about physical activities and six anchoring vignettes describing different people doing different levels of physical activities and asking respondents to rate the level. Respondents require a few days to get used to wearing the device.

Preliminary analyses of Dutch data indicate systematic difference between reported and actual physical activities. While Kapteyn reported only mean levels of activity at this early stage, accelerometer data can be summarized in a much more sophisticated manner. Similarly, vignettes seem to pick up some of the differences from self-reports and actual physical activities occurred but need to be explored further to determine how much of the difference vignettes can capture.

Social Isolation

The last harmonization initiative discussed was **social isolation**. Axel Borsch-Supan discussed a new module of social network introduced in SHARE. SHARE asks all respondents to name up to 7 personse with whom they most often discuss important things. This name generator, which takes approximately ten minutes to complete, can determine the social network for each person. Follow up questions include those on consumption items, accommodation, well-being, and social exclusion. This social network module was included in Wave 4 and Wave 6 and will be repeated every other wave.

Borsch-Supan presented some results from social-network questions as well. Europe is divided in many respects, with less wealth and health in the South and East reflected in in number of friends there as well. There are more diverse social networks in the North and West, but family networks are strong in the South and East. Borsch-Supan also recognized social networks are an important buffer, protecting against depression and cognitive decline, and that active and healthy aging is associated with social cohesion on country level. He further noted that time in retirement is negatively related to the number of non-family social network members, especially for early retirement. Having few non-family social network members leads to lower memory scores, over and above job-related effects.

Smith then presented some findings on social network and supports in England and United States. This work began from the recognition that Americans have worse health than the English, and examined cross-country differences in social networks as a potential explanation for cross-country differences in health, morbidity, disease, mortality. Both HRS and ELSA collect information about marital status and partnership, ties with children, other relatives, friends, voluntary and civic organization, and religious organizations. They examine various characteristics of networks, including presence or absence, frequency of contact, size of close ties, positive quantity of social support, conflict and negative interactions, and yield a summary index.

Smith and colleagues, using a summary indes, found that social networks are more extensive in England than in the United States. Nevertheless, there is no widespread consensus on what are the 'right' measures for social networks especially for international comparisons. The measurement of social network and social integration is not sufficiently mature and requires more research. Future research also must account for reporting style differences between countries.

III. OPPORTUNITIES FOR INTERVENTIONS

Hidehiko Ichimura and David Bloom led roundtable discussion on opportunities for interventions.

Ichimura discussed how longitudinal population surveys can be used to examine **the effects of natural disasters** and other events. Specifically, he discussed a special survey in Sendai as part of JSTAR to capture the extent of damage caused by the Great East Japan Earthquake in 2011. As JSTAR had baseline survey data prior to the earthquake, the JSTAR team developed a set of questions for pre- and post- earthquake comparisons. They administered these not only in Sendai but also in Kanazawa, another JSTAR city comparable to Sendai in many respects but not directly affected by the earthquake.

The JSTAR team examined how the earthquake affected "taste" for risk aversion and discounting, ; how families coped with event, i.e. family transfers, donations, savings; and who was most vulnerable to health and economic disruptions of the earthquake. For this purpose, the team administered a seven-page self-completion questionnaire in September 2011, six months after the earthquake. The response rate was 50.5%. The questionnaire included questions on family loss and injuries; how respondents spent the first five months after the earthquake; housing damages and car damage; job loss or job related damage; help received from family members or friends; help received from the government; whether they obtained bank loans; and their personal mental condition.

A risk-attitude question was tied to the uncertainty of receiving a raise: "if you were given a chance to take either job, which one do you prefer: a. a job with 90% chance of increasing the income by 50% and 10% chance of increasing the income by 5%; b. a job with 100% chance of increasing the income by 10%." Preliminary analysis suggests there was no significant change in risk aversion for both men and women.

Natural disasters can lead to losses in the sample, particularly should such events displace respondents and make tracking and follow-up more difficult. Crisis events can also require special attention. For example, two-year intervals for the HRS made it difficult for researchers to capture effects of events such as the late 2008 financial crisis in the United States. Kapteyn tried to overcome such time intervals by administering more frequent interviews using the ALP panel, enabling researchers to follow respondents over shorter intervals.

Regarding specific interventions, Bloom presented his vision for a purposeful survey-initiated intervention: a **Sight Intervention** study in LASI. Bloom recognized that there are a large number of Indians who have sight problems which may be remediable. The WHO reports 50-60 million Indians with impaired vision; *Times of India* reports the order of magnitude is even higher. Much remains unknown about prevalence of sight problems for individuals and communities. There is a shortage of eye doctors and treatment centers, especially in rural areas where two-thirds of the population lives. Eye disease is a major health problem in India and will worse as aging continues.

For individuals, poor vision affects well-being and quality of life. It also affects families, given the need for caregivers to look after and assist those with poor vision. LASI could perhaps generate estimates of the prevalence of eye problems. Eye diseases such as cataracts, glaucoma, and macular degeneration may have effects beyond those of poor vision.

The LASI team therefore plans to conduct traditional screening to determine refractive error and give respondents needing them prescriptions for corrective lenses. In a follow-up study, it will record whether eye tests are taken with corrective lenses as well as the weather and light of the test area. Field personnel will have a whole new suite of devices for diagnostics as well as an eye doctor who will train personnel in rural areas and ultimately dispense any needed corrective lenses.

Potential issues with such an intervention include maintaining protection of human subjects and the integrity of the sample's composition. One idea to help address these issues is to use a sub-set of pilot sample from a northern state and a southern state. These subsamples would help measure the cost of the intervention and its benefits so as to calculate a return on investment.

A similar study in rural China (Ganzou province) gave eyeglasses to school children. The study sent in eye doctors, screened children, gave them glasses if necessary, and followed them over time. Smith asked if other HRS-family studies have conducted an intervention study. SHARE has given different money incentives for participation to increase response rate. Ichimura suggested looking at similar respondents rather than randomizing.

There are dangers in interventions. For example, JSTAR respondents are told how their eating habits deviate from the norm, which may change their behavior. In the Yucatan, one of the poorer states in Mexico, a randomized social security program gave approximately $75 to persons in their 70s in some towns, while those in other towns were not given anything. Kapteyn and colleagues designed the Yucatan intervention and have been studying the results. This new initiative presents a promising new area of research but calls for careful considerations to prevent potential spoiling of the sample as well as behavioral change.

IV. PRIORITIES FOR CROSS-COUNTRY COMPARISONS AND NEXT STEPS FOR HARMONIZATION

David Wise led the roundtable discussion on priorities for cross-country comparisons. Wise opened the session by illustrating an example of possible comparisons. He recommended comparing health-care systems as public-pension systems are compared. Health care systems can be characterized by components such as age eligibility, with differences related to those in health. OECD compared the U.S. health-care system to those in five European countries about ten years ago, and such comparison needs to be updated.

Another example of comparisons is work at older ages. Researchers might collect information about what aspects of employment, such as part-time work, old workerss would find appealing. Regarding the demand for older workers, researchers might collect external data such as that on regional shocks or survey employers (particularly senior executives who make hiring decisions). Such data would facilitate analysis of how employment and production shocks affect labor markets, particularly for older workers.

Wise called for more comparable data on more countries for better harmonization. Time spans on retrospective questions significantly influence the comparability and therefore potential for cross-country analysis. Therefore, adjustments "since last wave" which refers to two year time period for most of the HRS type surveys or "last year" would be important. He noted that many researchers do not use the first year of HRS and AHEAD data but focus their analysis on the waves after harmonization.

John Phillips led the discussion on the next steps for the harmonization network. He highlighted the value of communication among network members to preserve and improve harmonization among surveys over time. He noted that the number of countries in the HRS-family of surveys is growing and is expected to continue doing so. Given the research network meets only once every two years, he recognized a need for other mechanisms, such as regional or thematic meetings, to supplement harmonization efforts.

The cognition experiments discussed by Weir, Banks, and Strauss illustrate what harmonization meetings can accomplish, as do the verbal autopsy in the CHARLS and SAGE-China. Even the existence of the LASI survey demonstrates the success of these harmonization surveys. Future meetings should highlight innovations or results from this organization.

Such meetings can be of tremendous value for personnel of new HRS-family surveys, although it may not be necessary for all members to be involved. Future researchers may benefit from documentation of knowledge that has been accumulated through these meetings. Regional meetings might also prove helpful to researchers in such areas.

Future pilot work might be supported by R24 applications. Network grants are small but the network could perhaps apply for a conference grant or achieve cost efficiency by organizing together with a study advisory meeting. Phillips encouraged the involvement of more psychologists.

David Bloom noted that recently the *Journal of the Economics of Aging* was developed and includes a number of meeting participants on its editorial board. Bloom noted the possibility of having a special issue connected to this group meeting.

The next Harmonization meeting is planned at RAND in Santa Monica in 2014, after the Rand Summer Institute. Axel Borsch-Supan raised an alternative in conjunction with SHARE, specifically having papers comparing two or three surveys at a November 2014 user conference in Berlin.

APPENDIX A
List of Attendees

Name	Affiliation
Angrisani, Marco	University of Southern California/RAND
Arokiasamy, P.	LASI, International Institute of Population Studies
Banks, James	ELSA, University of Manchester
Bloom, David	LASI, Harvard University
Borsch-Supan, Axel	SHARE, University of Munich
Chatterji, Somnath	SAGE, WHO
Cohen, Isabella	Harvard University
de Olivera, Cesar	ELSI, University College London
Goldhaber-Fiebert, Jeremy	LASI, Stanford University
Goodman, Alissa	NCDS, Institute of Education, University of London
Gutsche, Tania	University of Southern California /RAND
Hu, Perry	UCLA Medical School
Ichiumura, Hide	JSTAR, University of Tokyo
Kapteyn, Arie	University of Southern California /RAND
Lee, Jinkook	LASI, RAND
Lima-Costa, Fernanda	ELSI, Federal University of Minas, Brazil
Nahm Ki Sung	KLoSA, Korea Employment Information Services
O'Brien, Jennifer	LASI, Harvard
Phillips, John	NIA
Posture, Ashock	LASI, International Institute of Population Studies
Shankar, Aparna	University College London
T.V. Sekher	LASI, International Institute of Population Studies
Shimizutani, Satoshi	JSTAR, Research Institute of Economy, Trade & Industry
Smith, Jim	RAND
Strauss, John	IFLS and CHARLS, University of Southern California /RAND
Suzman, Richard	NIA
Tae Keun Eom	KLoSA , Korea Employment Information Services
Weerman, Bas	University of Southern California /RAND
Weir,David	HRS, University of Michigan
Wise, David A.	Harvard/NBER
Yoon Sub Kwon	KLoSA, Korea Employment Information Services
Zhao, Yaohui	CHARLS, Peking University

Appendix B

2013 Harmonization Research Network Meeting

Wednesday, July 17, 2013

Taj Palace Hotel, New Delhi, India

1pm-1.15pm	Welcome Jinkook Lee and Jim Smith	
1.15pm-2.45pm	Study Updates	
	HRS	David Weir
	ELSA	James Banks
	SHARE	Axel Boersch-Supan
	TILDA	Jim Smith
	MHAS	Jim Smith
	KLoSA	Ki-Sung Nam
	JSTAR	Satoshi Shimizutani
	IFLS	John Strauss
	CHARLS	Yaohui Zhao
	SAGE	Somnath Chatterji
	ELSI	Fernanda Lima Costa
	NCDS (1958 Cohort)	Alissa Goodman
2.45pm-3.45pm	Global Aging Data Repository	Jinkook Lee, Bas Weerman, and Marco Angrisani
3.45pm-4.30pm	Biomarkers	Perry Hu
4.30pm-5.15pm	Genetics	David Weir

2013 Harmonization Research Network Meeting

Thursday, July 18, 2013

9am - 9:30am	NIA program priorities for harmonization	Richard Suzman
9:30am-10:30am	Cognition HRS/ELSA New Cognition tests IFLS/MHAS New Cognition tests	 David Weir and James Banks John Strauss
10:30am-10:45am	Break	
10:45am-11:30am	Accelerometers and Subjective Well-being	Arie Kapteyn
11:30am-12:15pm	Social Isolation	Axel Boersch-Supan and Jim Smith
12:30pm-1:30pm	Lunch	
1:45pm-2:45pm	Roundtable discussion on opportunities for interventions Natural disasters and other events Purposeful survey-initiated interventions	 David Bloom Hidehiko Ichimura
2:45pm-3:30pm	Roundtable discussion on priorities for cross-country comparison	David Wise
3:30pm-3:45pm	Break	
3:45pm-4:45pm	Roundtable discussion on next steps The path ahead: next harmonization meeting, regional and thematic meetings Research outlets: a special issue? New entrants Other…	John Phillips