2011 Harmonization of Cross-National Studies of Aging Meeting

EXECUTIVE SUMMARY

National Institute on Aging

Behavioral and Social Research Program

Prepared by:

Jinkook Lee and Tania Gutsche, RAND

November 2011

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

EXECUTIVE SUMMARY

This document summarizes the first Research Network meeting for Harmonization of Cross-National Studies of Aging to the Health and Retirement Study (HRS). The meeting was held on August 2 and 3, 2011, in Beijing, China, and 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. Among the participants were the principal investigators of aging surveys from around the world, experts in aging research, and program officers at NIA. Appendix A lists all attendees.

I. PURPOSE

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 first of planned biennial meetings of 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 HRS, the Mexican Health and Aging Study (MHAS), the English Longitudinal Study of Ageing (ELSA), Survey of Health, Ageing, and Retirement in Europe (SHARE), the Korean Longitudinal Study of Aging (KLoSA), the Japanese Study on Aging and Retirement (JSTAR), the Indonesian Family Life Survey (IFLS), the Irish Longitudinal Study on Ageing (TILDA), the Chinese Health, Aging, and Retirement Longitudinal Study (CHARLS), and the Longitudinal Aging Study in India (LASI). PIs from two emerging surveys in Brazil and Thailand also attended, presenting their plans and seeking advice.

The agenda (Appendix B) for the first PI meeting included scientific innovations each survey has adopted or plans to introduce, cross-country comparability in selected key domains, demonstration of the RAND Survey Meta Data Repository, and round-table discussion of the challenges faced by various surveys.

II. SIGNIFICANCE

HRS has achieved remarkable scientific success, as demonstrated by its impressive number of registered users (9,937) and publications (18,276) through March 2009. Such scientific success, along with rising concerns over global aging, has generated substantial interest elsewhere in similar data. The result has been a number of surveys designed to be comparable with HRS.

HRS-family surveys provide remarkable opportunities for cross-national comparisons because they were designed with harmonization as a goal (Lee, 2010)[1]. Common features include nationally-representative samples of individuals near retirement age or older, approximately biennial interviews, and multi-disciplinary core domains, including health, retirement, economic status, and family and social network. A more detailed description of each survey can be found in Appendix C.

The scientific opportunities that cross-national, comparative studies offer in evaluating policy on health and retirement outcomes are widely recognized. Yet there are few empirical studies exploiting such opportunities. This is partly due to the complexity of multidisciplinary and longitudinal data as well as to the difficulty associated in learning multiple surveys and their contexts (e.g., policies, institutions). Our proposed research network aims to address this challenge by encouraging communication among data producers for multiple countries, promoting the development and dissemination of harmonized information and tools developed for cross-country analyses, and reaching out to current and potential users.

The PIs of existing HRS-family surveys have been closely collaborating and seeking to include PIs of emerging surveys as they develop. This early-stage collaboration has helped enable the harmonization of research, sampling designs, and key concepts in all HRS surveys. For example, this collaboration led the PIs to recognize unnecessary idiosyncratic differences in questions and subsequently to modify question wording to enhance cross-country comparability. Continued collaboration among PIs of existing surveys is also important given the innovation that can continually occur for panel studies. All longitudinal studies face a challenge in innovating without sacrificing stability of measurements. Nevertheless, existing surveys can make adjustments and modifications should scientific merits warrant.

[1] Lee, J. (2010). "Data Sets for Pension and Health: Data Collection and Sharing for Policy Design," *International Social Security Review*, 63, (3-4), 197 – 222

III. SETTING THE STAGE

Dr. Richard Suzman, Director for the Behavioral and Social Research Program at the National Institute on Aging (*NIA*), gave a presentation on the importance of comparative analysis in HRS-family surveys. He noted that some of the most important findings have come from comparative studies. For example, an NBER project led by David Wise and Jonathan Gruber yielded a seminal paper on the importance of public-pension incentives in determining retirement ages in developed countries (Gruber & Wise, 1998)[2]. More than 70 percent of the world's aging population is now included HRS-family surveys. Suzman he predicts that large-scale harmonized surveys will become increasingly more important given that harmonization is the foundation of comparative analyses.

The Denver Summit Communiqué acknowledged the critical importance of comparable data sets for international research:

> We discussed the idea of "active aging" -- the desire and ability of many older people to continue work or other socially productive activities well into their later years, and agreed that old stereotypes of seniors as dependent should be abandoned. We considered new evidence suggesting that disability rates among seniors have declined in some countries while recognizing the wide variation in the health of older people. We discussed how our nations can promote active aging of our older citizens with due regard to their individual choices and circumstances, including removing disincentives to labor force participation and lowering barriers to flexible and part- time employment that exist in some countries. In addition, we discussed the transition from work to retirement, life-long learning and ways to encourage volunteerism and to support family care- giving. …. We agreed that it is important to learn from one another how our policies and programs can promote active aging and advance structural reforms to preserve and strengthen our pension, health and long-term care systems. Our governments will work together, within the OECD and with other international organizations, to promote active aging through information exchanges and cross-national research. We encourage collaborative biomedical and behavioral research to improve active life expectancy and reduce disability, and have directed our officials to identify gaps in our knowledge and explore developing comparable data in our nations to improve our capacity to address the challenges of population aging into the 21st Century. (http://www.g8.utoronto.ca/summit/1997denver/g8final.htm).

[2] Gruber, J., & Wise, D. (1998). "Social Security and retirement: An international comparison," *the American Economic Review*, 88 (2), *Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association*, 158 – 163.

Suzman further discussed the value of an international research agenda, pointing out that countries are adapting with policy changes and therefore providing an opportunity to analyze natural experiments in several NIA Behavioral and Social Research Areas of Emphasis, including(1) increasing health expectancy, (2) health, work, and retirement, (3) interventions and behavioral change, (4) genetics, behavior, and the social environment, and (5) the burden of illness and the efficiency of health systems. Macroeconomic challenges require microeconomic responses, and demographic and institutional variation makes detection of causality easier. Cross-national research requires cross-national comparable data, but it is important to balance comparability and experimentation/individuality.

Suzman noted several current activities at the forefront of aging research, including: (1) better collection of information on early-childhood experience, (2) HRS and genetics, (3) economic phenotypes for genetic analyses, (4) personality measures, especially conscientiousness, (5) subjective well-being and time use, and (6) harmonization network and meetings.

He further discussed the value of adding genetics to longitudinal studies on aging. One use of genetic analysis is for instrumental-variable analysis to unravel causality. By incorporating genetic information, researchers can reduce unobserved heterogeneity and make vectors in behavioral models more exact. Genetic analysis will also improve understanding of population stratification. Longitudinal data will provide variation in incidence, course and outcome in subgroups identified by "high-risk genes" or "high-risk experiences." By adding genetics to longitudinal studies, researchers can further refine the molecular basis of behavioral phenotypes. Phenotypes of particular interest in middle and later life are those related to economic behaviors (e.g., impulsivity, risk taking, time preference), social behaviors (e.g., caregiving, marital quality, social connectedness), and health behaviors (e.g., physical activity, management of disease, compliance with medical regime). Harmonization of these phenotypes across studies is particularly important.

For genetic epidemiology, sample size is an important consideration. The required sample size varies by study design, effect sizes, bio-clinical complexity, and still other characteristics. For binary outcome of case control studies, the minimum required sample size would be 2,000 – 5,000 to estimate genetic main effects, 2,000 – 20,000 to estimate life-style main effects, and 10,000 – 20,000 to estimate gene-lifestyle interactions (50,000 for a comprehensive platform). Suzman suggested that combining studies can help genetic analysis requiring large samples are needed. Harmonization by Nancy Pedersen of studies of twins provides one example of how combining samples can aid analysis of genetic effects. In response to NIA RFA AG10-006 Effects of Gene-Social Environment Interplay on Health and Behavior in Later Life, a consortium of seven longitudinal twin studies were combined to explore the basis for the association of social factors and aging outcomes (N=16,000). The participating studies include SATSA (Swedish Adoption/Twin Study of Aging), OctoTwin (Origins of Variance in the Old-Old), GENDER (unlike-

sex Swedish twin pairs), TOSS (Twin and Offspring Study in Sweden), LSADT (Longitudinal Study of Aging Danish Twins), MADT (Middle-Aged Danish Twins), and MTSADA (Minnesota Twin Study of Adult Development and Aging). The targeted phenotypes included those related to early-life adversity, early-life socioeconomic status, loneliness, social support, physical health, functional health, and depression.

Recent work suggests personality, especially conscientiousness, as a promising area of research. Conscientiousness refers to a hierarchically-organized family of traits that describe individual differences in the propensity to be self-controlled, responsible to others, hardworking, orderly, and to follow rules (Roberts, 2011).[3] Specifically, conscientiousness items from "big five inventory" include: is a reliable worker; tends to be disorganized (reverse-scored); tends to be lazy (reverse-scored); perseveres until the task is finished; does a thorough job; does things efficiently; makes plans and follows through with them; is easily distracted (reverse-scored); or can be somewhat careless (reverse-scored). NIA is promoting research on conscientiousness.

The NIA mission to improve health and wellbeing received a boost from the May 25, 2011 endorsement by President Obama and Prime Minister Cameron of "collaboration between world-class longitudinal studies in the U.S. and UK, with the potential to transform our understanding of issues such as childhood obesity, cancer, aging, and emotional well-being."[4] The leaders also welcomed the involvement of the Economic and Social Research Council in partnership with the National Institute on Aging at the National Institutes of Health in developing a U.S. National Research Council Panel on Measuring Subjective Wellbeing. This effort, Suzman said, may generate new insights for social and economic policies. Suzman noted a need for an experimental measurement of well-being, particularly the negative aspects of well-being (or misery), insights on which may be more valuable for policymakers.
Other presentations and discussions on day one included those by PIs on innovations and future plans in their surveys of aging, beginning with those from developed countries followed by those from developing countries and new initiatives; on the RAND Survey Meta Data Repository for harmonized data and contextual statistics; and on Biomarkers and Genetics.

[3] Conscientiousness and Healthy Aging Workshop, Division of Behavioral and Social Research, National Institute on Ageing, http://www.nia.nih.gov/ResearchInformation/ExtramuralPrograms/BehavioralAndSocialResearch/Psychology.htm

[4] http://www.whitehouse.gov/the-press-office/2011/05/25/joint-fact-sheet-us-uk-higher-education-science-and-innovation-collabora

IV. STUDY INNOVATIONS

David Weir, Health and Retirement Study (HRS)

The HRS, begun in 1992, has built over time with the addition of new cohorts. HRS researchers, Weir said, are now working to expand the minority sample. HRS had an oversample of blacks and Hispanics, but a smaller sampling rate in "baby-boom" cohorts has resulted in smaller absolute minority sample sizes. Using supplemental funding from NIA and the Social Security Administration (SSA), the HRS sought to double the number of minority respondents in baby-boom cohorts in the 2010 wave.

Over time, HRS has added data collected by other modes, including self-completion mail-in questionnaire and Internet survey. The pace of HRS research has accelerated; every third day a paper is published using the HRS. To date, 1,100 peer-reviewed journal publications and more than 200 dissertations have used the HRS.

The HRS has broadened its content over time. In 2006, the HRS introduced an "Enhanced Face-to-Face interview" that included physical measures, biomarkers, and a psychosocial self-administered questionnaire. These interviews use an alternating half-sample design.

Weir warned that high response rates can mask selection bias. He recommended continual checking of distributions against those in major national surveys and official statistics. Two keys to success in longitudinal studies of aging are to (1) bring respondents back after missing a wave and (2) use proxy respondents for the cognitively impaired. The strategy of bringing respondents back even if missing a wave has had some success. In 2002, 12 percent of those interviewed had missed at least one wave since 1992. Kapteyn et al. (2006) has shown this practice reduces attrition bias[5].

Administrative linkages provide another opportunity to enrich the survey data. HRS data has been linked with individual data in Social Security earnings and benefit records, Medicare records, the National Death Index, and the Social Security Death Master File. HRS data are currently being matched with individual records of the Department of Veterans Affairs. Through employer matching, the HRS is also linked to pension plans, and its records are currently being matched to employer records maintained by the Census Bureau.

[5] Kapteyn, Arie , Pierre-Carl Michaud, James P. Smith and Arthur van Soest. (2006). Effects of Attrition and Non-Response in the Health and Retirement Study. Santa Monica, CA: RAND Corporation. http://www.rand.org/pubs/working_papers/WR407.

The HRS provides a unique and valuable means to study dementia given its (1) national sample, (2) large, representative sample (race/ethnicity, education, wealth), and longitudinal data allowing (3) analysis of the complex interactions among genetic, medical, behavioral, and social factors affecting the risk for cognitive decline, and (4) analysis of the multiple outcomes of dementia (health, family, economics, public programs).

The ADAMS is an add-on study to the HRS supplemented in 2002 – 03. It used leading in-home diagnostic assessments. The sample size for the original assessment was 856 and reassessed in 2004 – 10. Key findings from the ADAMS indicated that prevalence of dementia is lower than popular estimates, while that of mild cognitive impairment (MCI) and cognitive impairment with no dementia (CIND) is higher. Analysis of the ADAMS has also found that some, but not all, CIND is likely to be early-stage Alzheimer's disease. Based on the estimates from the ADAMS, trends and outcomes of impairment (e.g., costs, institutionalization, or mortality) are then projected for the HRS sample.

James Nazroo, English Longitudinal Study of Ageing (ELSA)

Key research areas of the ELSA, launched in 2002, include: (1) health trajectories, disability, and healthy life expectancy; (2) the determinants of economic position and resources in later life; (3) the relationship between economic position and health; (4) the timing of retirement and post retirement labor-market activity; (5) the nature of social networks, social support and social participation, and (6) household and family structure, and the transfer of resources. Given the need for research on linked dynamic relationships in all these research areas, there is a need for longitudinal, multidimensional and multidisciplinary, and credibly detailed in each of these dimensions.

Similar to the HRS, the ELSA uses multiple modes for data collection. It includes a core interview, self-completion questionnaire, links to administrative records, and add-on studies. The ELSA also introduced nurse visits to collect health data in Wave 2 and life-history interviews in Wave 3. Its currently linked administrative records include those on mortality, cancer registration, hospital episodes, national insurance records, benefits, and tax credits.

Over time, the ELSA expanded its contents. At first, it drew heavily from the HRS, later introducing innovations. ELSA has since added content on (1) DNA and molecular biomarkers, including inflammatory markers, cortisol, lipids, blood sugar, HbA1c, and anemia; (2) performance tests; (3) measures of blood pressure, lung function, and anthropometry; (4) questions on symptoms and functioning (sleep, pain, vision, etc.); (5) psychosocial measures of well-being, quality of life, effort-reward imbalance, demand and control, subjective social status, and altruism; (6) social, civic, and cultural participation; (7) quality of social networks and social support.

For Waves 5 and 6, the ELSA continues to bring innovations in:

(1) religiosity and spirituality, including data on religious affiliation, frequency of attendance at a service and of prayer or meditation, participation in organized religion, and importance of religion (in general, on a daily basis, and to provide meaning and purpose in life);
(2) discrimination and abuse, including HRS measures on respect, service, harassment (e.g., for gender, race, age), and New Zealand Longitudinal Study of Ageing measures on vulnerability to abuse
(3) time-use and well-being measures developed from the HRS); the proposed measures ask about previous day and intend to measure both hedonic and eudemonic dimensions. Emotional experiences are asked for various activities carried out on previous day
(4) risk and time preferences and regret, with a sub-sample of 1,000 respondents age 50 to 74 years of age being asked to play in which they can win real money, allowing researchers to observe which respondents take risks or postpone rewards, regret, etc.
(5) cortisol in hair; hair analysis is initially developed for diagnosis, drug screening, etc., but more recently used in large population studies to understand epidemiology and aetiology. It requires a specific amount of hair, 2 cm long, cut from close to the scalp. Hair sample is easy to collect and stable, if properly packaged and frozen. To assess homeostasis, a sample insensitive to short term fluctuation is needed. Hair sample offers an "integrated" measure reflecting chronic exposure over two months so it can be used for longitudinal analysis to identify change in/acute exposure, but very costly and time-consuming. It needs to separate exogenous contaminants from endogenous concentration. Cortisol does not capture diurnal variation, but may be less relevant for older people and is demonstrated to be related to factors indicative of stress (employment grade, gender, and race/ethnicity. Pilot studies suggested a high success rate in collection of hair sample: less than 1% refusal and sample adequate for 85% of cases. Quality of analysis appears unrelated to physical factors, such as color of hair, frequency of washing, and hair treatment.
(6) Sexual experiences and functioning , for which almost no UK data exists, despite their effects on well-being, morbidity, and mortality. Changes in sexual function with age, across generations, and over time are expected. Development work was completed using the European Male Ageing Study (EMAS), also the National Survey of Sexual Attitudes and Lifestyles (NATSAL) and National Social Life, Health and Aging Project (NSHAP). Self-completion questionnaire has been developed, covering attitudes, sexuality, libido, frequency, functioning, medication, worries and satisfaction.

The ELSA also plans to adopt HRS measures on positive and negative affect, personality, and fluid intelligence. The ELSA is also developing new measures for health and social care arrangements

particularly designed for UK institutions, polypharmacy and adherence to medication, using an accelerometer to measure physical activities, and a diary module of diet and nutrition. Researchers have some concern about measures for diet and nutrition, consumption, and time use, given the high time commitments required to capture adequately detailed information.

Axel Boersch-Supan, Survey of Health, Ageing and Retirement in Europe (SHARE)

The SHARE started in 2004 with 11 countries: the Netherlands, Germany, Austria, Denmark, Belgium, France, Switzerland, Spain, Italy, Greece, and Sweden. It added four countries for Waves 2 and 3: the Czech Republic, Poland, Ireland, and Israel. It added another five countries for Wave 4: Estonia, Luxembourg, Hungary, Slovenia, and Portugal. Wave 4 will include 80,000 respondents 50 years of age or older. Wave 4 fieldwork is in progress. There has been some trouble in collecting data in Greece, although the SHARE has secured matching funds from the Greek government. The SHARE has become an international organization with its own legal structure, ERIC.

Among innovations of the SHARE have been life histories and development of an archive of historical context variables. The SHARE has established administrative record linkage in Germany, linking employment histories to earnings relevant for public pensions and health insurance, earnings points, and rehabilitation activities. It is seeking such linkages in Sweden, Denmark, and France.

The SHARE also started biomarker collection in Germany in cooperation with German National Cancer Center, the University of Washington Lab, and the USC-UCLA Biodemogrphy Center led by Eileen Crimmins and Teresa Seeman. The biomarker protocol includes: anthropometrics, collection of dried blood specimen, a puff test, grip strength measurement, and blood pressure measurement.

The SHARE has lower response rates than other HRS-family surveys, a result of wide-spread European concerns regarding data confidentiality. To increase response rates, the SHARE carried out an “Incentive and Nonresponse Experiment” in Germany for 2,000 participants in a 4,000-member refreshment sample. The experiment included participations of as much as 40 euros as well as additional training of interviewers. It resulted in many complaints, seven law suits, and bad press for its use of marketing methods.

Wave 4 fieldwork is still in progress, but appears to be yielding lower response rates. Boersch-Supan expressed concern over continuing decreases, about 5 percent per wave, in response rates.

Hide Ichimura, Japanese Study on Aging and Retirement (JSTAR)

The JSTAR began pilot work in 2005, collected its baseline wave in 2007, and is now in its third wave. The survey has the same structure as the SHARE but a different sampling design. Its first-wave sample included five municipalities: Adachi, Kanazawa, Sendai, Shirakawa, and Takikawa. It added two more municipalities, Naha and Tosu, for its second wave, and three more, Chofu, Hiroshima, and a city in Osaka for its third wave. The JSTAR selected municipalities in part for their willingness to provide administrative records. Within each municipality, the JSTAR obtains a random sample stratified by location. The JSTAR chose this sampling plan in order to raise the response rate and obtain administrative data. The JSTAR obtained endorsements from sampled municipalities. It notes this endorsement both in a letter to respondents as well as in official monthly newsletters of each municipality. The JSTAR team also works with officials of sampled municipalities to answer respondents' questions about the survey.

The JSTAR links its results with official records of health and care usage, including health-examination records (anthropometrics, vision, hearing, blood pressure, urinalysis, stool test, X-ray, and blood tests, including neutral lipid, HDL, LDL, AST, ALT, gamma-GTP, BS, HbA1c), among respondents 65 and older.

Ichimura noted that many JSTAR research questions relate to how individuals make decisions and react to different incentives in their environment. For this purpose, the municipality-based sample offers the advantage of having many individuals who face uniform socio-economic environment which can be identified without asking individuals about their environment. Municipality-based sampling allows researchers to separate the effect of different incentives for individuals from those of heterogeneous socio-economic environments in which individuals are placed in a flexible way without resorting to a regression-based fixed-effect model.

One of the JSTAR municipalities, Sendai, was recently hit by an earthquake and tsunami. The JSTAR team planned follow-up research there on (1) change of cohabitation to predict the need to future temporary housing; (2) change in various health measures; (3) change in social networks; (4) within-family transfers and public transfers; and (5) stability of risk aversion and discount factor measures.

Among the innovations of the JSTAR has been collection of food-intake data using a brief diet-history questionnaire (BDHQ). The BDHQ is based on Sasaki's Diet History Questionnaire, originally developed

from Willet's dietary questionnaire. The BDHQ takes about 15 to 20 minutes to complete and was validated by Kobayashi et al (2011).[6]

Other innovations of the JSTAR include (1) identifying potential care givers to parents; (2) measuring risk aversion by using lottery choices for income increases of varying certainty; (3) measuring hyperbolic discount factor in Wave 2; and (4) measuring time use separately for weekdays and weekends in a leave-behind questionnaire. Suzman encouraged all HRS-family surveys to include time-use questionnaires, noting the JSTAR instrument could be adopted for that purpose.

Rose Anne Kenny, The Irish Longitudinal Study on Ageing (TILDA)

Ireland's population is projected to age rapidly: 11 percent were 65 or older in 2011, a proportion that is expected to increase to 14 percent in 2021 and to 19 percent by 2031. In response, the Irish government and philanthropic sources have funded the TILDA to provide information on aging in Ireland.

The TILDA conducted two pilot studies between 2006 and 2009, and a baseline study in 2010, which included an at-home interview, health assessment at a Health center, and a self-completion questionnaire. Archiving of Wave 1 data was to be completed by December 2011, and biennial follow-up waves are planned.

The TILDA used a clustered random-sample design, drawing 640 clusters of addresses from a geographic directory. The largest populations in Ireland are in Cork and Dublin, which were therefore chosen for the location of the two TILDA health centers. The sample drew 40 addresses per cluster, resulting in a total sample of N=25,600 total households. TILDA sent an advance letter by post to 60 percent of households and a hand-delivered letter to 40 percent. Response rates have been quite high by European standards: 62% for the core interview, 84% for the self-completion questionnaire, and 86% for the health assessment at Health centers. TILDA obtained successful interviews from 8,507 respondents in 6,282 age-eligible households (i.e., household with at least one individual at age 50 or older). Early research findings have focused on comparisons of health assessments conducted at home and in health centers.[7]

---

[6] Kobayashi et al. (2011). Comparison of relative validity of food group intakes estimated by comprehensive and brief-type self-administered diet history questionnaires against 16 dietary records in Japanese adults, *Public Health Nutrition*, 14 (7): 1200 – 1211.

[7] Kearney, P.M. et al. (2011). Comparison of centre and home-based health assessments: early experience from the Irish Longitudinal Study on Ageing (TILDA), *Age and Ageing*, 40: 85 – 90.

One of the first innovations of TILDA, Kenny noted, has been in the study of the prevalence, diagnosis, and treatment of atrial fibrillation. Using data from Wave 1 of TILDA, bioengineers have developed a new algorithm to distinguish Atrial Fibrillation from Sinus Rhythm. This has prompted policymakers to develop public-awareness campaigns to raise awareness of atrial fibrillation and stroke prevention.

Chonggak Shin, Korean Longitudinal Study of Aging (KLoSA)

Since its pilot work in 2005, the KLoSA has completed three waves of data collection. Wave 1 and 2 data are publicly available from the Korean Labor Institute. The Korean Employment Information Services (KEIS), a new home institution for the KLoSA, will soon release the Wave 3 data collected in 2010. KLoSA has allowed proxy interviews but has not had a refreshment sample. Wave 4 data collection is planned for the fall of 2012.

Among the innovations of the KLoSA is the use of a calendar method enabling researchers to investigate the impacts of the financial crisis in 1998 on older Koreans' careers, wealth, and health. Shin, who was recently appointed as a new PI for the project, plans to institute a panel of experts to obtain advice and support.

John Strauss, Indonesia Family Life Survey (IFLS)

Since being launched in 1993, the IFLS has collected 4 waves of data. The IFLS is best known for its excellent and innovative methods of tracking. The IFLS, conducted in 13 provinces, has sought to follow all household members since 1997 except if they move from the country. Strauss emphasized the importance of tracking over long distances so as to minimize selection bias.

Yaohui Zhao, China Health and Retirement Longitudinal Study (CHARLS)

The CHARLS research team includes scholars from all over China and abroad. The operations team full-time employees and interviewers from universities. The study is funded by the National Institute on Aging, the World Bank, the Chinese National Science Foundation, and the Peking University.

The CHARLS conducted a pilot study in 2008 in two states, Gansu and Zhejiang, interviewing 2,685 individuals and achieving a response rate of 85%. The CHARLS has had two user conferences for its more than 1,300 users, one-third of who are from outside China. Zhao noted that among the innovations of the CHARLS are the use of GPS and photographing of buildings in data collection.

Jinkook Lee, Longitudinal Aging Study in India (LASI)

The LASI pilot study is the first multi-disciplinary survey to use the Computer-Assisted Personal Interview and to collect a rich set of biomarkers and performance measures in India. Other innovations of the pilot studies include: (1) collection of GPS data; (2) detailed surveys of physical environment, including access to water, sanitation, and indoor air quality; (3) perceived neighborhood characteristics; (4) self-reported incidence of infectious diseases; (5) health-care utilization measures differentiating types of health care providers; (6) new methods in collecting data on social connectedness; and (7) an experimental module on expectations for illiterate population.

Building on lessons and experiences of the pilot study, the LASI team developed a fieldwork-monitoring protocol for quality control, refined the instrument, and recognized the extent of self-report biases associated with diagnosed diseases and therefore expanding biomarker measures. The team has put together research proposals to raise funding for the baseline study, seeking National Institutes of Health (NIH) support to collect data from 18,000 representative individuals 45 or older and their spouses irrespective of age, in the 15 largest states (which account for 90% of the national population). The LASI team is also seeking funding inside India to collect a supplemental sample of about 12,000 respondents from remaining states as well as booster sample for larger states.

For the baseline study, the team has considered (1) including additional measures of wellbeing, time-use, consciousness, mobility, and additional cognitive tests suitable for illiterate population; (2) developing more assays, such as those for micronutrient deficiency (e.g., iron, Vitamin C), and Apo lipoproteins A and B; (3) collecting data on nutritional intakes in an experimental module; and (4) collecting information about local policies on ration cards and prices. The team was also considering modification of the protocol for the lung-function test from a spirometer to a peak-flow meter with greater ease of use; assaying cytomegalovirus (CMV) instead of Epstein-Barr virus (EBV) antibodies; and dropping the vision test.

The team is also planning to develop a community survey, including (1) mapping distances to health-care providers, parks, other facilities, and polluted areas; (2) surveying local government (panchayati raj) regarding policies and community histories of major events (e.g., disease epidemics) by use of a calendar method; and (3) collecting information about environmental health through direct assessment of water quality as well as indoor and outdoor air quality.

Dararatt Anantanasuwong, Health, Aging and Retirement in Thailand (HART)

The Health, Aging, and Retirement in Thailand (HART) survey has completed its pilot study with funding from the National Research Council of Thailand (NRCT) and matching funds from the National Institute on Development Administration. The pilot HART survey was conducted in 2009, interviewing 1,500 respondents 45 or older (preferably household head, or, if not available, spouse of the household head or any age eligible) from 1,500 households, drawn from Bangkok and vicinity (Nonthaburi, Pathumthani, and Samuthprakarn) and the northeast province of Khon Khaen (one of the poorest). The pilot data have been stored in a server at the School of Applied Statistics and are available in Thai-language Excel files in Thai. A HART homepage is under construction.

In April 2011, the HART team received funding from the National Higher Education Commission (NHEC) of the Thai Ministry of Education to conduct a pilot Wave 2. Fieldwork began in July and was to be complete by September of 2011. The second pilot wave also included exit interviews with proxies of persons from the first wave who had died. From the baseline pilot, the HART team recognized some of the study limitations, including the effects of a lack on the team of health or medical experts as well as of a full-time researcher and programmer, and restriction of government research funding for procuring computer hardware or software or offering incentives to respondents.

The HART team planned to conduct a national study in 2013 with a baseline sample of 5,600 households. The team has submitted a proposal for this study to the NRCT and was awaiting a decision. The team was also developing a proposal to establish a Longitudinal Study Center with working space and a full-time staff. Once established, the center would seek formal collaboration with other domestic and international research institutions and provide training for longitudinal panel survey techniques and analyses.

Rebecca Wong, Mexican Health and Aging Study (MHAS)

The MHAS, which started in 2001, is modeled on the HRS. After completion of two waves, the study stalled and will collect a third wave in 2012. Contents for Wave 3 will include biomarkers, anthropometric measurements, respondent vignettes, psychosocial characteristics, food intake, time use, and linkages to administrative records. The original sample was drawn from 32 states, making sample design and field work expensive but follow-up work easier. The MHAS had a high response rate, 90 percent, and a follow-up rate of 93 percent. The MHAS gave a token gift of reading glasses to respondents as an incentive. Earlier waves indicated one in five households had a child working in the United States.

Fernanda Lima-Costa, Estudo Longitudinal de Saúde e bem-estar dos Idosos (ELSI)

Brazil, with 190 million inhabitants, is experiencing one of the most rapid aging processes in the world. According to the Brazilian Institute for Geography and Statistics, 10% of the population in 2008 was 60 or older. Most Brazilians lived in urban areas (83%); and more than one in four (28%) were illiterate.

The general principles in developing the ELSI are (1) enabling international comparisons; (2) drawing a nationally representative sample of 15,000 persons 50 or older; (3) taking an interdisciplinary approach to economy, health-services utilization, and physical and mental health questions; (4) including the themes that are strategic for the Brazilian government, such as public pensions and the Unified Public health system, in reducing social inequalities and improving wellbeing and health conditions; and (5) investigating separate and joint effects of environmental and genetic factors on health outcomes for a highly mixed population of African, European, and Native American ancestry.

The Ministry of Health declared the ELSI as a research priority and intends to sponsor most of the study. The ELSI team plans to develop the instrument and conduct pretests in 2013.

Other Issues

There are plans to field an HRS-family survey in Argentina. The Statistical Bureau of Argentina and the University of Cordoba have committed to field a national the survey with technical assistance from University of Wisconsin and the University of Texas Medical Branch. The project leaders were not able to attend this meeting, but will attend future ones.

Participants also discussed procedures for working with restricted data in the surveys. NIA and SSA have a data safeguard for dissemination and restricted data procedures for surveys in early stages of data dissemination, and the HRS model is a good model for NIA and SSA safeguards. Other protocols include cover use of data terminals at universities and fingerprint analysis allowing researchers to run but not print analyses.

Participants also discussed questions regarding policy for data records leaving the country of origin, de-identified or not. It was recommended that studies have linked data in the country, and extract summary variables. Most users do not need microdata, but rather derived values for earnings, present values, and other variables. Keeping the data in the home country constitutes an Enclave model of data archiving. Some participants recommended new data collectors start with the most liberal European model of data usage before restricting it for the research community. All agreed that users should get the minimum they need for their research and that what happens in one country effects all countries.

V. The RAND Survey Meta Data Repository

The RAND Survey Meta Data Repository seeks to facilitate cross-country studies on aging by providing information and tools for HRS-family surveys around the world. The repository team has improved the usability of the site and created a set of derived variables, identically defined across four surveys, HRS, ELSA, SHARE, and KLoSA. The team has also developed a video demonstration on using the website (www.metadata.rand.org). It has sought support from all data producers by linking to their study sites. The project has been well received, and the team has been encouraged to expand the scope of the site. The team also discussed with meeting attendees plans to improve the usefulness of the site. Among suggestions participants made were to include response rates and retention rate information in the repository; add more surveys on aging, such as that on Costa Rica; and to list in the repository papers using HRS-family survey data.

VI. DOMAIN-SPECIFIC HARMONIZATION

Jinkook Lee and her collaborators have prepared a set of domain-specific user guides, examining comparability of surveys for key domains of income, wealth, financial transfers, work, retirement, chronic medical conditions, cognition, and expectations. Domain experts reviewed the documentations and provided feedback at the meeting, both on how to enhance harmonization and how to advance cross-country research. The domain-specific user guides are forthcoming as a series of RAND Working Papers[8].

[8] Hu, P. & Lee, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Chronic Medical Conditions, WR-861/1, Santa Monica, CA: RAND Corporation
Zissimopoulos, J., Lee, J., & Carroll, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Financial Transfer, WR-861/2, Santa Monica, CA: RAND Corporation
Delavande, A., Lee, J., & Yoong, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Expectations, WR-861/3, Santa Monica, CA: RAND Corporation
Zamarro, G., & Lee, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Employment and Retirement Measures, WR-862/4, Santa Monica, CA: RAND Corporation
Angrisani, M., & Lee, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Income Measures, WR-861/5, Santa Monica, CA: RAND Corporation
Angrisani, M., & Lee, J. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Wealth Measures, WR-861/6, Santa Monica, CA: RAND Corporation
Shih, R.A., Lee, J., & Das, L. (2011). Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Cognition, WR-861/7 , Santa Monica, CA: RAND Corporation

Biomarkers and Genetics

David Weir discussed blood-collection practices for three groups of surveys: (1) ELSA, TILDA, and CHARLS, which collect whole-blood samples (as ELSI plans to do); (2) HRS, IFLS, German SHARE, and LASI, which collect dried blood specimens (DBS); and (3) MHAS, KLoSA, SHARE (outside Germany), JSTAR, and HART, which do not collect any blood. Limited blood collection can limit the number of possible assays. The accuracy of assay also needs to be validated, often by examining the correlation with whole blood on matched subjects and the predictive value of assay for outcome of interest (e.g., mortality). One valuable resource for biomarkers is the UCLA-USC NIA P30 Aging Center Network on Measures of Biological Risk led by Eileen Crimmins and Teresa Seeman.

HRS-family surveys collect a standard battery of physical measures, including height, weight, waist and hip circumference, blood pressure, grip strength, lung function (puff test or other), gait speed (some studies also include chair stands), and balance. TILDA has the most comprehensive and innovative physical performance measures. HRS, ELSA, German SHARE, IFLS, CHARLS, and LASI have standard measures, as ELSI plans to adopt. SHARE (outside Germany), JSTAR, KLoSA, and MHAS have a subset of these measures. Weir encouraged researchers to use physical performance measures and to have a protocol for ascertaining mortality. He also recommended that all studies work to get DNA collection and consent.

Genetic analysis of HRS-family surveys has two steps, collecting DNA and analyzing genomes. For extracting DNA, whole-blood collection has the most promise, while saliva collection is a very competitive alternative. Oragene collection kits for saliva are not expensive (approximately $20/kit). One possible funding strategy is to propose saliva collection as a supplement to a study grant. The cost of high-quality genotyping is also rapidly decreasing, and telomere length can be measured. The HRS started collecting saliva in 2006. This year, DNA for 13,000 respondents genotyped using the Illumina 2.5 million SNP chip, will be deposited in dbGap. Imputations were currently under way based on the 1,000 genomes project, with another 7,000 cases to be added over the next two years. Telomere length measurements on 6,000 samples will also be completed this year. The harmonization of phenotype measures across the HRS network makes the studies collectively a rich potential source of greater statistical power.

Rose Anne Kenny introduced an extensive set of biomarker measures that TILDA has adopted. In addition to the standard measures adopted by most HRS-family surveys, TILDA includes heart-rate variability measured with the Medilog Darwin AR12; visual acuity measured with the Logmar chart; contrast sensitivity measured using the Stereo Optical Corporation's Functional Visual analyzer; retinal

photograph measured using the NIDEX, Non-Mydriatic Auto-Fundus Camera; macular pigment optical density measured using the Macular Metrics Densitometer; and bone density measured using the Achilles Insight Heel Ultrasound. TILDA also includes an extensive array of cognitive tests, measuring (1) global cognition, using cognitive assessment (MOCA) and mini mental state examination (MMSE); (2) attention, using Sustained Attention Response Time; (3) visual memory, using CAMDEX Picture Memory Test, on acquisition, free recall, recognition; (4) speed of processing, using choice reaction time test; and (5) executive function, using Timed Colour Trails 1 & 2 of Visual reasoning test of CAMDEX. In measuring gait speed, TILDA used a GAITRite sensored mat, not only timing normal walk, but also walk with manual task and walk with cognitive task. These measures had been incorporated under the primary research interests in pathophysiology, disease prevalence, and norms in clinical tests.

For measuring biomarkers in harmonized studies, TILDA researchers recommend (1) measure blood pressure sitting AND standing (repeat one minute after standing) and ask if respondent became dizzy; (2) standardize TUG (timed up and go); (3) experiment with different diagnostic methods; (4) add protocol notes to metadata site so experiments can be duplicated and data can be better understood; and (5) consider joint work, such as a joint grant with HRS and TILDA on macular degeneration and how it may suggest Alzheimer's Diseases. Other participant concerns included areas validity checks for laboratories, variance in biomarkers, and future plans for genetic analysis.

Economic Resources

James Smith and Yan Shen discussed economic resources. In measuring economic resources, it is important to be comprehensive both across and within sources. Specifically, in measuring income, it is important (1) to collect data on the main components of income, such as work, assets, transfers, pensions, and government transfers, not just an aggregate income; (2) to be able to separate respondents, especially partners, by their main source of income; (3) to pick the best reference period (e.g., past calendar year); (4) to determine whom to ask questions (e.g., each respondent, financial respondent); (5) to decide whether to ask pre- or post-tax income. Angrisani and Lee (2011) discuss income questions, measurement units (e.g., individual, couple, or household), and handling of taxation across surveys.

In measuring wealth, it is essential to collect comprehensive measures of both household and pension wealth. It is also important to identify asset changes that reflect asset price and savings or dis-savings. In collecting wealth data, researchers should consider (1) who knows about wealth; (2) whether asset ownership is joint or separate; (3) how to handle housing; and (4) how to handle business and farm ownership. Agrisani and Lee (2011) also provide information on wealth measures across surveys,

including who answered questions on wealth, units of measurement, and how housing, business, and farm ownership are handled.

Consumption is particularly important in developing countries. The IFLS, CHARLS, and LASI have collected data on consumption. The time taken to measure consumption varies by survey by methods used to measure it and their levels of disaggregation. Past research involves old household budget surveys, and greatly compressed categories in World Bank version. The World Bank is experimenting with ways to measure consumption. HRS has surveyed consumption by mail since 2002. Consumption diaries have been used, but consumption in SHARE diaries did not match known results. Consumption surveys must ask different items in different countries, but researchers need to carefully consider the best periods (e.g., weekly, monthly, yearly) for minimizing recall bias.

Meeting participants noted the importance of monitoring differences in consumption. Food is unequally distributed within the household, and clothing can be individually attributed sometimes. Food intake may indicate individual levels of consumption, but outcomes such as weight and height might also indicate individual food intake or nutrition consumption.

EuroMark, a European tax calculation system, may eventually offer some insights on international comparisons of income. Questions regarding appropriate financial units for analysis, e.g., whether to assess income for couples or households are likely to persist, especially in developing countries.

Family and Transfers

Kathleen McGarry and Xian Lei discussed family and transfers. McGarry emphasized the importance of family across fields and topics. Family can affect a variety of outcomes, such as emotional, financial, and physical well-being, as well as economic, demographic, and health behaviors. Family is also key to understanding models of behaviors, as in life cycle model, family as insurer, and altruism/exchange models of family transfers. Survey questions about family are included in sections on economic resources, retirement, health, and wellbeing.

Studying changes in the definition of family, with increases in stepfamilies, divorce, and cohabitations, resulting in more kin with weaker ties, is also relevant to studying the effects of family. The role of family has changed with increases in women's labor force participation. Public policies need recognize changes in family and their effects for Social Security, pension, long-term care, and health-care programs. The role of families may be changing most rapidly in countries where rapid economic development has created large differences in lifetime resources across generations and resulted in greater geographic mobility as well as changes in the direction of transfers and remittances. How

individuals fare as a result of these changes will depend in large part on how families respond to them and how their roles are reshaped by them.

McGarry also highlighted the value of cross-country studies in studying the effects of families. There are many cross-country differences in public policies and programs, demographics, and cultural norms. Using cross-country studies allows researchers to exploit such variation to test models of behavior and predict outcomes of policy changes, including those for public provision of long-term care, laws regarding distributions of estates, college tuition, and generosity of pension programs. Demographic and cultural differences may influence such issues as who provides care within a family as well as investment in and cash transfers to children. Changes in mobility for the young also vary by country: in some countries, more children live further from parents; in others, children stay longer at a parent's home. Differences in life expectancy will influence differences by country in grandparents' care for grandchildren, with differences in contributing to differences in burden of care.

Family data in the HRS and its sister surveys are extremely rich but have not been exploited fully, partly due to difficulty in using them. RAND has recently created a user-friendly version of RAND HRS family data, linking family data across waves, imputing missing values, and carrying forward variables for longitudinal data.

The user guide on transfer data also provides useful information for family researchers, but its scope is limited to financial transfers. McGarry called for further attention to other types of transfers, such as those related to caregiving. Transfer mechanisms may differ by country or subpopulation given differences in social norms in providing care and financial assistance. Differences in public programs may lead to substitution in types of assistance, and differences in type of care might vary across income distributions.

There are also differences by survey in capturing transfers to co-resident kin. It is difficult to measure in-kind transfers and who helps whom, and it is also difficult to value the benefit of returns to scale in consumption. To enhance comparability for cross-country studies, surveys should scale transfers to a common timeframe (e.g., last month, 2 years). Choosing an appropriate length for such a timeframe is important: if the time frame is too short, then prevalence of transfers would be under-estimated due to their infrequency. Their amounts may also be overestimated if the measures are scaled to a longer period. Finding the right periodicity is particularly important if there are substantial transfers for "once in a lifetime" events like birth, college graduation, marriage, and new home, and if receipt of transfers varies over time.

Regular and infrequent transfers may also have different purposes, with one, for example, being a source of support and the other serving as insurance against negative shocks. Wording for financial transfer questions also varies, although it is not clear how this may influence responses. Several surveys use a minimum threshold for recording transfers; such censoring may result in lower estimates for prevalence measures.

Finally, McGarry noted that with data on both spouses, researchers can analyze measures of emotional well-being, correlation in health behaviors and outcomes, and correlation in cognition. HRS-family surveys collect little data on siblings. Interviews of children might be useful for measuring genetic and phenotype differences and similarities.

Lei goals and contents of the user guide for financial transfers. Data on financial transfers between respondents and their family and friends are available in most HRS-family of surveys. ELSA does not measure financial transfers but only the amount of transfers received. The concept of financial help is similar in most of the surveys with the exception of JSTAR. Some issues researchers should consider in cross-country studies of financial transfers are (1) the narrow scope and definition transfers as defined in JSTAR, focusing on food and living expenses, which may not be comparable to that in other surveys; (2) inconsistent censoring and need to consider purchasing-power parity; (3) whether to include or exclude co-residents in transfers; (4) inconsistent periodicity across surveys; and (5) whether amounts can be summed to an aggregate measure.

Work and Retirement

David Wise and Hidehiko Ichimura led the discussion on harmonization of work and retirement measures. Wise called for attention to more fundamental issues of what we want to know about work and retirement. Once researchers know what they want to know, he said, they can determine what questions they need to ask in each country (e.g., hours worked per week, number of paid vacation days per year). In making cross-country comparisons from collected survey data, researchers need to judge whether they can calculate comparable answers across countries. At the same time, researchers may want to know different things about work and retirement in different countries, meaning they should consider “do the questions and answers tell what we want to know within the country?”

Wise also emphasized the importance of question framing. There has been considerable work on question framing and how it affects answers. For example, questions on losses may elicit different substantive answers than those on gains. Question order also matters, as does form. For example, for employment status, questions asking respondents to choose a single most appropriate status will elicit different answers than those asking a respondent to choose all answers that apply.

Ichimura compared the work and retirement modules of the HRS and the SHARE. Both surveys provide information on how long a respondent worked, how much the respondent earned for the work, and whether a respondent is classified as retired. He found the data to be mostly comparable except on a few measures. He noted for some occupations such as research hours worked may not be well defined but comprise a mixture of investment and on-the-job training as well as hobby and work hours. Researchers may find it useful to distinguish hours one is required to spend at work from those spent for self-improvement for work. Without an explicit guideline for respondents, answers may be left to interpretation. Whether some occupations, such as farming, are considered jobs can vary by countries. So may the definition of retirement, with the "retired" in some countries being defined as those not working or looking for work.

Meeting participants noted several best practices in the coding of occupation and industry sectors. European surveys use International Labour Organization (ILO) classifications. All agreed industrial and occupational coding is a difficult task requiring a great deal of time and effort, and that different coders over different waves may produce different results. Zhao noted that a number of Chinese jobs that might be captured in the CHARLS do not exist in ILO coding—leading her to use Chinese Census coding of jobs for the survey. Strauss said that IFLS researchers get descriptions of the jobs then compares strings of code, using Indonesian words, to code jobs by ILO classifications.

Health

James Nazroo and Jinkook Lee led the discussion in health. Nazroo noted the great benefit in cross-country studies on health, including on questions such as (1) changes in life expectancy, healthy life expectancy, and dependency; (2) impact of health on economic, social, and civic participation; and on consumption and service use; (3) conditions and causal pathways for aetiology; (4) drivers of population variation and inequalities; and (5) underlying biological processes that might affect economic and social outcomes. All these questions can benefit from international research, including comparative analyses across different institutional, demographic, and cultural contexts.

Nazroo identified a set of measures that all surveys should measure, including (1) self-reports of diagnosis, symptoms (depression, CVD, etc.), global assessments of health, functioning (ADLs, sleep, vision, hearing, balance, falls, incontinence), health behaviors (smoking, alcohol, diet, exercise), and health service use, including specific interventions and quality of care; (2) direct assessments of functioning (performance tests – physical and cognitive), measures of systems (hypertension, respiratory function, vision), anthropometry, and physical activity; (3) DNA and biomarkers from blood, saliva, urine, hair; and (4) administrative data on health records and mortality. He noted the value of

direct assessments and biomarkers, calling for further data collection efforts to diagnose conditions (hypertension, diabetes, and raised cholesterol) using directly assessed blood pressure, glycated hemoglobin, and blood cholesterol, to allow for the identification of undiagnosed disease and poorly controlled diagnosed disease.

Nazroo then recognized constraints for harmonization. Questions must to reflect national and regional institutions (for example, who diagnosed the condition) as well as national and regional context (for example, infectious disease, nutrition, sanitation). Researchers should be cautious even when measures appear harmonized. Apparent differences may be the result of the impact of different institutions on detection rates. In understanding difference, researchers need to be culturally informed on interpretation of questions and responses. Even performance tests may be influenced by culture. Mode effects may exist, as may variations across (and within) laboratories.

Nazroo called for a more comprehensive review of health conditions, given as Hu and Lee's working paper focuses only on chronic medical conditions. Nazroo also noted remarkable comparability in chronic medical conditions, but marked variation in coverage of symptoms, functioning, and treatment.

Lee discussed important population health issues in developing countries. She identified in particular the issues of under-nutrition (of food generally and micronutrients specifically), infectious diseases (such as tuberculosis and malaria) jaundice, and injuries associated with accidents and adverse weather.

Lee noted that an epidemiological transition is in progress in many developing countries, with a resulting increased burden of chronic non-communicable diseases. She noted several important contributors to Years Lived with Disability (YLD), including eye diseases, hearing loss, dementia, musculoskeletal diseases, and heart diseases, as listed in the Global Burden of Disease report, as well as dementia, cerebrovascular diseases, musculoskeletal diseases, neuropsychiatric diseases and eye diseases, as listed by the 10/66 Dementia group.

Health environment is another important issue in developing countries. Unsafe water, sanitation, hygiene, indoor air pollution from solid fuel use, and outdoor air pollution are some of the key health environment issues that contribute to environmental burden.

Finally, Lee noted the issue of health-care service. Health-care providers in in developing countries, differ from those in developed countries. In particular, unlicensed practices and traditional care are more prevalent in developing countries, raising the need to document them. Health-care access is another important issue. There have been several approaches in the HRS-family surveys measure distance to health-care facilities and health-insurance coverage.

Cognition and Expectation

Robert Willis and Kenneth Langa discussed cognition and expectations. Willis noted the importance of measuring cognition and the challenges of doing so in larger surveys. In the United States, older adults are faced with an increasing scope of cognitively-demanding decisions, such as those for savings and wealth management, health and long-term care, and retirement. For example, increased longevity and advances in medical technology present older Americans with complicated health decisions. Increased choices due to growth of 401(k), the decline of defined benefit pensions, and the advent of Medicare D, also force older individuals to make cognitively-demanding decisions.

The goal of the original cognition measures in HRS was to track cognitive aging and risk of dementia. This later led to the ADAMS study. In 2010, the HRS added new measures of "higher-order" reasoning to better understand (1) how people deal with complexity of financial and health decisions; (2) how they will respond to policy changes; and (3) determinants of inequalities in health and wealth.

Willis noted that the theory of fluid and crystalized intelligence (Gf/Gc) suggests that cognitive ability is multi-dimensional, comprising a fluid-intelligence component for thinking and reasoning, and a crystallized-intelligence component for accumulating knowledge. Fluid intelligence grows rapidly during childhood and adolescence but declines thereafter, while crystalized intelligence continues to grow through most of life. Willis noted parallels between the Gf/Gc theory and the theory of human capital. Fluid intelligence is an "exhaustible resource" used in combination with knowledge, effort, and inputs from teachers, schools, and on-the-job training to produce durable crystalized intelligence or human capital over a life cycle. Cognitive abilities, both fluid and crystalized intelligence, are key resources in determining labor-market earnings, inputs into children's human capital, decisions about saving, wealth, and retirement, and decisions about health and health care.

The HRS introduced several new cognitive measures in 2010, based on findings from collaborative separate but related research projects, including a psychology project (CogUSA) led by Jack McArdle and an economics project (CogEcon) led by Willis. Criteria for choosing measures of fluid intelligence included ability to (1) administer in a survey with multiple modes, including telephone and personal interviews; (2) administer in a short period of time; (3)measure a full range of abilities; and (4) be related to economic and health decisions and outcomes. From the standard 47 items of the Woodcock-Johnson number series test (NS47), McArdle developed an "adaptive test" that has no more than 6 items, takes about 3 – 4 minutes (original 47 items take 20 – 30 minutes to administer), and can work in personal, telephone, and Internet survey modes. The adaptive number series test is based on item

response theory and matches difficulty of question to ability of individual, beginning, for example, with an easier item for a poorly educated person and a harder item for a better education one.

Langa offered a medical perspective on the importance of "brain health" for nearly all aspects of daily function and well-being with wide-ranging consequences for individuals, their families, and their communities. Brain health is affected by the complex interactions of biological, social, environmental, and economic variables. Cross-nation variation in these may help to identify key strategies (medical, social, public policy) for keeping the aging brain healthy and happy. The impact of cognition on health, well-being, and daily function makes it a high priority for cross-national aging studies.

Langa noted that about 25 percent of older adults require medical decision-making in the last days of life, but are too cognitively impaired to make those decisions (Silverira et al., 2010)[9]. This underscores the importance of patient-family-physician discussions regarding goals of care for older adults. End-of-life care is a high-priority policy topic in the United States given its high costs and concern over "futile" care.

According to the NIH Toolbox, cognition is defined as "the process of knowing and, more precisely, the process of being aware, knowing, thinking, learning and judging" (http://www.nihtoolbox.org/WebPart%20Pages/Cognition.aspx). Key sub-domains of cognition include:

- Executive function, the capacity to plan, organize, and monitor the execution of behaviors that are strategically directed toward a goal.
- Episodic memory, cognitive processes involved in the acquisition, storage and retrieval of new information.
- Working memory, a limited-capacity storage buffer that becomes overloaded when the amount of information exceeds capacity.
- Processing speed, defined as either the amount of time it takes to process a set amount of information, or conversely, the amount of information that can be processed within a certain unit of time, measuring mental efficiency.
- Language, a set of mental processes that serve to translate thought into symbols (words, gestures) that can be shared among individuals for purposes of communication.
- Attention, the allocation of one's limited capacities to deal with an abundance of environmental stimulation and the foundation for all other types of mental processes.
- Reading, the cognitive process of deriving meaning from written or printed text.

[9] Silverira, M.J., Kim, S.Y.H., & Langa, K.M. (2010). "Advance Directives and Outcomes of Surrogate Decision Making before Death, New England Journal of Medicine, 362: 1211 – 1218.

Langa noted the importance of measuring both threshold dementia and early cognitive decline. Dementia threshold refers to the point at which an individual can no longer independently manage ADLs and IADLs. Its impact on individuals, families, and society are most significant. Cross-national studies might benefit from analysis of the cultural differences in defining "independence." Interventions for prevention of cognitive decline may only be effective very early in its course, and therefore the current scientific focus at NIA and Alzheimer's Association is on early diagnosis of cognitive decline. Measures for dementia thresholds include tests of verbal memory (immediate and delayed recall) and orientation administered to respondents and proxies, including IQCODE, AD8, Blessed Dementia Rating Scale, and an assessment of ADL/IADL limitations. Potential measures to assess early cognitive decline include number series, speed of processing, reaction time, animal fluency, prospective memory, and trail-making tests.

Langa noted there are relatively good opportunities exist for cross-national studies of moderate to severe impairment (dementia) using memory and orientation items. Identifying early and mild levels of cognitive decline or impairment is a measurement challenge, but new measures (e.g. number series, processing speed), especially when used longitudinally, show promise. Identifying early cognitive changes using surveys fits well with current basic science and clinical epidemiology priorities. Key challenges in cross-country measurement include identifying reliable "culture-free" measures that are brief and therefore able to fit in comprehensive surveys. Measures for low-literacy or illiterate individuals also require further attention. Langa emphasized the importance of methods for recruiting proxies and nursing home residents.

Meeting participants discussed the need to have newer surveys adopt cognition measures. Currently, 50 to 60 percent of the NIA budget focuses on Alzheimer's disease and dementia. Susann Rohwedder and Robert Willis's (2010)[10] finding of the impact of work on cognition is important to watch. If it is replicated, further analyses would need to examine early retirement and cognitively demanding hobbies and estimates on what fraction of the population's cognitive decline can be influenced by retirement. Our goal is to understand the optimum cognitive environment. To do so, it is useful to collect information on of antihypertensive and calcium blockers, because they may influence cognition and Alzheimer's disease.

Willis also reviewed the user guides on expectation. He noted that probability beliefs are central in all forward-looking economic models (which maximize expected utility subject to uncertain constraints). He categorized three types of probabilities that can affect decision making: (1) events without personal control or personal information (e.g., stock returns, housing prices, government policies); (2) events with personal information (e.g., survival chances, health events); (3) events with personal control (e.g.,

[10] Rohwedder, S., & Willis, R. (2010). "Mental retirement," *Journal of Economic Perspectives*, 24: 119 – 138.

retirement, preventative health care, savings, portfolio choice). Although economists have long assumed that people have probabilistic beliefs that influence decisions, they have seldom attempted until Manski (1990, 2004)[11] made the case for numerical probabilities. Delavande et al. (2011)[12] showed subjective probabilities can be measured in developing countries and are useful in analysis.

Surveys often ask respondents about the probabilities they perceive for survival, bequests, retirement, and stock returns. Willis noted the importance of asking questions with varying thresholds and suggested directly asking about summary statistics of distribution (e.g., expected age at retirement) and noise in measures (e.g., what does 50-50 mean? equal probability or uncertainty?). He recommended modeling survey response processes to relate theoretical concepts to survey answers. In mental models, one might ask how people form expectations (e.g., on what information do they base beliefs; how do they process this information; why do answers vary across respondents) and how expectations affect decisions.

In harmonizing expectation questions, researchers should consider which queried probabilities would be useful. Subjective survival questions measure outcomes common to people in all countries and should be asked in the same way in all countries, although one might use stones in India, and percent chance in the US. In many cases, however, expectations measures most useful for research may be important in one country but not another. Measuring expectations offers a window into the future that may be very different than the current situation. The expectations, in turn, affect decisions on a wide variety of issues (e.g., what are implications for saving and retirement if people do not expect to live with children?).

Well-being

Richard Suzman and Arie Kapteyn led the discussion on well-being. The mission of NIA is to improve the health and well-being of older Americans through research. To that, it is important to understand the predictors of healthy aging and what would improve the lives of older people. The Division of Behavioral and Social Research (BSR) supports research to "produce a scientific knowledge base for maximizing active life and health expectancy." This knowledge base is required for informed and effective public policy, professional practice, and everyday life. The BSR division also encourages the practical application of behavioral and social research.

---

[11] Manski, C. (1990). "The use of intentions data to predict behavior: A best case analysis," *Journal of the American Statistical Association*, 85: 934 – 940.
Manski, C. (2004). "Measuring expectations," *Econometrica*, 72: 1329 – 1376.

[12] Delavande, A., Giné, X., & McKenzie, D. (2011). "Measuring subjective expectations in developing countries: A critical review and new evidence," *Journal of Development Economics*, 94: 151 – 163.

Suzman recognized the need for better metrics on quality of life. While Gross Domestic Product and National Income Accounting statistics were among the greatest contributions of economics of last century, helping guide economic policy, they have major limitations in their failure to consider non-market goods, externalities, consumer surpluses, and inequality. Subjective well-being is an important indicator for quality of life. Suzman said that understanding the end of life may yield unique insights for understanding subjective well-being. Psychology and economics also have distinct but integrated goals of measuring a good life or utility including

- for psychology, a desire to understand the components of a good life, how people experience and evaluate their lives, the interplay between experiences and evaluations, and the impacts on health and well-being
- for economics, a desire to measure utility (what people value) and progress (improvements in life quality) to inform individual decision making and public policy

The Commission on the Measurement of Economic Performance and Social Progress, created in 2008 by French President Sarkozy, seeks to develop a new method of economic calculation of social progress (supplements to GDP) that will include indicators of well-being. An offshoot of this activity has been the study of broad measurements of well-being. NIA has funded a proposal to develop a measurement handbook, and the National Academy of Sciences (NAS) convened a "Workshop on the Role of Wellbeing Measures in Public Policy" in November 2010 (http://www.nia.nih.gov/NR/rdonlyres/276F9943-8F65-47A2-B8E9-A06AC6D259F4/0/2010Nov_Workshop_WB_Measures.pdf). A formal NAS panel will report on measures for use by statistical agencies in about 12 months.

Suzman identified two types of subjective well-being measures: experienced wellbeing and evaluative wellbeing. Experienced wellbeing measures include those on good and bad emotions, joy and pain, or other daily experiences. Evaluative wellbeing measures consider retrospectively what respondents like or dislike. Specific means to measure well-being include:

- Experienced well-being: PANAS (hedonic), experience sampling, day reconstruction method
- Evaluative well-being: satisfaction with life, Ryff scales (eudemonic)

These measures must be taken separately, as they have distinct social, demographic, and personality correlates. They also appear to have distinct physiological correlates, so they are not expandable.

Suzman discussed details, such as sample questions and their predictive ability, for several types of specific measures.

- Satisfaction with life measure captures overall evaluation of life and is viewed as a cognitive component of well-being. Some researchers have focused on global assessments of overall life

satisfaction, such as the standard measure of Diener and colleagues, while others have examined domain-specific satisfaction with work, income, social relationships, and neighborhood. A measure might ask respondents to indicate, on a seven-point scale, with such items as, "in most ways my life is close to my ideal."

- Ryff well-being scales (eudemonic) are based on theory of well-being that distinguishes between hedonic and eudemonic (Aristotelian) conceptions of happiness. Psychological constructs of well-being measured with these scales include autonomy, positive relations with others, environmental mastery, personal growth, purpose in life, and self-acceptance. A measure might ask respondents to assess, on a five-point scale, their confidence in their own opinions, even those contradicting general consensus.
- Day Reconstruction Method (DRM)[13] asks respondents to reconstruct the previous day by completing a structured self-administered questionnaire. The respondent first produces a short diary on a sequence of episodes in the previous day. This method draws on insights from cognitive research with event-history calendars and facilitates retrieval of autobiographical memory through multiple pathways. Its episodic reinstatement format attenuated biases commonly observed in retrospective reports. Respondents are then encouraged to describe key features of each episode, including (1) when the episode began and ended, (2) what they were doing, (3) where they were, (4) whom they were interacting with, and (5) how they felt on multiple affect dimensions. There is some question over the proper time period to assess for DRM. For example, does assessment of a single day make sense for a survey that may be administered only once annually or even less frequently? Because there is day-to-day variability in affective well-being, it is not clear how stable such measures are. Researchers concerned about longer periods of time may wish to ask about those. Pain studies conducted by Stone and colleagues indicate a single end-of-day measure functions as well as one-week recall.

Suzman also discussed findings from the Gallup World Poll of 132 countries in 2006. Deaton (2008)[14] examined data for an 11-point life-satisfaction ladder and found that satisfaction varies with age and with the objective country characteristics, particularly the levels and rates of change in per capita income and life expectancy. Deaton also found that subjective well-being was sensitive to question context in the Gallup survey, and that such questions should be buffered by others that will not affect them. DRM measures can help overcome question-context problems by focusing on specific episodes.

[13] Kahneman, D., Krueger, A.B., Schkade, D.A., Schwartz, N. & Stone, A.A. (2004). "A survey method for characterizing daily life experience: the day reconstruction method," *Science*, 306: 1776 – 1781, doi:10.1126/science.1103572

[14] Deaton, A. (2008). "Income, health, and well-being around the world: Evidence from the Gallup World Poll," *Journal of Economic Perspectives*, 22: 53 – 72, doi:10.1257/jep.22.2.53

A pending RFA on subjective wellbeing: advances in measurement applications to aging (R01) seeks capitalize on recent advances in measuring both experienced (e.g., reports of momentary positive and rewarding, or negative and distressing states) and evaluative (e.g., cognitive judgments of overall life satisfaction or dissatisfaction) well-being. There is global interest in well-being measurement as a critical index of the success or failure of economic, social and health policies. New daily approaches are on horizon, including Jacqui Smith's current NIA Challenge grant, work by the Princeton Center for Research on Affect and Well-being, and Arthur Stone's work on a short form of DRM. A hybrid between Gallup and DRM approaches may yield new insights. In addition to appropriate scales and DRM timelines, researchers may wish to address methodological questions such as sensitivity of measures to big societal changes, whether measures should focus on positive or negative well-being, and measures that are most relevant to public policy. Suzman noted this field is evolving quickly and moving toward "what makes life worthwhile." Nevertheless, measures of subjective well-being will always be imperfect, and raise questions of what contribution, if any, they make beyond GDP and other existing measures, and whether such measures are credible enough for government statistical agencies to monitor.

Kapteyn presented some prior findings on cross-country studies on subjective well-being and discussed methodological issues for cross-country comparisons and subjective well-being measures adopted in HRS-family surveys. He noted that the correlation between percent of persons satisfied with life and GDP is 0.383 (Englehart et al., 2004)[15] and that, according to the OECD, the correlation between percent satisfied with life and the level of poverty is -0.583, with this correlation being higher, at -0.685, for persons 65 or older.

Kapteyn noted the need to correct for response bias by using vignettes in conducting cross-country studies. Preliminary findings based on Gallup data, confirm that differences in life satisfaction across countries are much smaller than raw data suggest.

Kapteyn summarized various subjective well-being measures adopted by HRS-family surveys. All HRS-family surveys have some form of a depression scale (CESD) and global well-being scale (often the five-item Diener scale), but only three surveys—ELSA wave 6, HRS in Consumption and Activities Mail Survey (CAMS), and JSTAR—collect data on time use. Kapteyn, Smith, and Van Soest (2009)[16] found that global life satisfaction appears to be well described by satisfaction in four domains, income, relations, job, and health. Researchers may therefore wish to collect satisfaction measures for these four domains.

[15] Englehart, M. et al. (2004). Human beliefs and values: A cross-cultural sourcebook based on the 1999 – 2002 values surveys, Siglo XXI Editores, Mexico.

[16] Kapteyn, A., Smith, J.P., & Van Soest, A. (2009). Comparing life satisfaction, WR-623-1, RAND Corporation: Santa Monica, CA.

Kapteyn called for further research to learn more about measurement properties and relations. Doing this would require collecting different measures in the same surveys, using varying modes and differing countries, to study scale differences. Such measurement calibration is not unique to subjective well-being measures, and needs to be addressed for cross-country studies.

Discussion on Working in Developing Countries and Starting a New Study

The meeting ended with a round-table discussion on challenges of working in developing countries and starting a new study. Among the challenges of working in developing countries are measuring labor-force participation and nutrition. Life-cycle profiles of earnings and income in developing countries are more difficult to measure due to a widespread work in agriculture and informal economic sectors. Time-use studies can help capture labor-force participation in informal labor market.

For nutrition, allocation within household varies, but it is hard to get individual levels. There may be discrimination against women, or other contributors to differences. Researchers may wish to examine nutrition through health outcomes (e.g., anthropometric measures) rather than nutritional intakes. JSTAR is the only survey to measure and track individual-level nutrient intake, and other surveys may wish to adopt its approach.

Cognition may be difficult to assess in developing countries. Illiteracy among respondents mean many cognitive tests (e.g., counting backward) employed in developed countries are not suitable for developing countries.

For starting a new study, government support may be helpful if not required. In Latin America, government agencies must be involved, but this is not the case for other countries. Government involvement can have mixed effects. Some meeting participants suggested that the more the government is kept from the sample and the substance, the better the research will be. Nevertheless, there are huge research benefits to gaining access to government administrative data. Getting the right government involvement can be important. For example, the HRS is conducted through corporative agreement with an advisory committee rather than a federal contract. This structure allows for numerous scientific freedoms. If the HRS were run through a government contract, researchers would have had difficulty setting the agenda for it.

Incentives can help achieve high response rates, but, in implementing them, researchers must be sensitive about acceptable practices within a country. Local scholars can help identify question sensitivities, as well as those questions most likely to be relevant in each country.

VII. Conclusions

John Phillips of NIA led concluding discussions. Meeting participants agreed the sessions were extremely useful, and they wished to continue the momentum and relationships established. Future action items included:

1. Allowing more junior researchers to attend future meetings, perhaps by requesting supplemental grants to each survey allowing three persons rather than just a project leader to attend
2. Holding additional regional meetings or thematic meetings focused on a single topic area (e.g., cognition)
3. Holding other small-scale meetings, e.g., a small-group meeting at a more established study site or a small-scale meeting in conjunction with professional conferences such as the annual meeting of the Population Association of America or the RAND Summer Institute
4. Hosting a research conference on comparative analysis
5. Launching experiments on particular topics, such as time use or performance measures

Finally, researchers were encouraged to identify where identical or varying measures would be most appropriate.

Appendix A: Attendees

Dararatt Anantanasuwong
Professor, School of Development Economics, National Institute of Development Administration, Bangkok, Thailand

Axel Boersch-Supan
Director, Munich Center for the Economics of Aging (MEA) at the Max-Planck-Institute for Social Law and Social Policy Amalienstrasse, Munich, Germany

Tania Gutsche
Deputy Director, Labor & Population, RAND Corporation

Hide Ichimura
Professor, Graduate School of Economics, University of Tokyo

Arie Kapteyn
Director, Labor & Population, RAND Corporation

Rose Anne Kenney
Professor, Mercer's Institute for Successful Ageing, St James's Hospital and Neurosciences Institute, Trinity College, Dublin, Ireland

Kenneth Langa
Professor, Division of General Medicine, University of Michigan, Veterans Affairs Ann Arbor Center for Clinical Management Research

Jinkook Lee
Senior Economist, Labor & Population, RAND Corporation

Xiaoyan Lei
Assistant Professor, National School of Development, Peking University

Jeremy Lupoli
Programmer, RAND Corporation

Fernanda Lima-Costa
Professor, Epidemiology at Oswaldo Cruz foundation and Federal University of Minas, Brazil

Kathleen McGarry
Professor, Department of Economics, University of California, Los Angeles

Cesar De Oliverira
Senior Research Associate, Epidemiology & Public Health, University College London

Mariana Lopez Ortega
Instituto de Geriatria, Institutos Nacionales de Salud, Mexico

Drystan Phillips
Programmer, RAND Corporation

John Phillips
Division of Behavioral and Social Research, National Institute on Aging

James Nazroo
Professor, Sociology, University of Manchester, UK

Yan Shen
Professor, National School of Development, Peking University

Chonggak Shin
Economist, Korea Employment Information Services

James P. Smith
Distinguished Chair in Labor Markets and Demographic Studies, RAND Corporation

John Strauss
Professor, Department of Economics, University of Southern California

Richard Suzman
Director, Division of Behavioral and Social Research Program, the National Institute on Aging

Bas Weerman
Senior Information System Analyst, RAND Corporation

David Weir
Research Professor and Director, Health and Retirement Study, University of Michigan

Robert Willis
Professor of Economics, Research Professor in Survey Research Center and the Population Studies Center of the Institute for Social Research, University of Michigan

David A. Wise
Professor, John F. Kennedy School of Government, Director, Program on Economics of Aging, National Bureau of Economic Research

Rebeca Wong
P. & S. Kempner Distinguished Professor in Health Disparities, University of Texas Medical Branch, Sealy Center on Aging

Yaohui Zhao

Professor, National School of Development,

Peking University