Department of Health and Human Services
National Institutes of Health
National Institute on Aging
Division of Behavioral and Social Research

# Harmonization Strategies for Behavioral, Social Science, and Genetic Research

Bethesda, Maryland
November 29–30, 2011

## WORKSHOP SUMMARY

**For Administrative Use**
February 5, 2012

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| Add Health | National Longitudinal Study of Adolescent Health |
| BSR | Division of Behavioral and Social Research |
| caBIG® | Cancer Biomedical Informatics Grid |
| CAMDEX | Cambridge Mental Disorders of the Elderly Examination |
| CESD | Center for Epidemiological Studies Depression Scale |
| CHARLS | China Health and Retirement Longitudinal Study |
| CTSA | Clinical and Translational Science Award |
| dbGaP | Database of Genotypes and Phenotypes |
| DBS | dried blood spot |
| DDI | Data Documentation Initiative |
| ELSA | English Longitudinal Study of Ageing |
| ELSI-BRASIL | Brazilian Longitudinal Study of Ageing |
| GWAS | genome-wide association studies |
| GxE | gene-environment interaction |
| GxG | gene-gene interaction |
| HHS | U.S. Department of Health and Human Services |
| HRS | Health and Retirement Study |
| IALSA | Integrative Analysis of Longitudinal Studies on Aging |
| ICPSR | Inter-University Consortium for Political and Social Research |
| ICs | NIH Institute and Centers |
| ISBER | International Society for Biological and Environmental Repositories |
| IT | information technology |
| KLoSA | Korean Longitudinal Study of Ageing |
| L.A.FANS | Los Angeles Family and Neighborhood Study |
| LASI | Longitudinal Aging Study in India |
| MHAS | Mexican Health and Aging Study |
| MIDUS | National Survey of Midlife Development in the United States |
| MxFLS | Mexican Family Life Survey |
| NACDA | National Archive of Computerized Data on Aging |
| NCDS | National Child Development Study |
| NHANES | National Health and Nutrition Examination Study |
| NHATS | National Health and Aging Trends Study |

| NHGRI | National Human Genome Research Institute |
| NHLBI | National Heart, Lung, and Blood Institute |
| NIA | National Institute on Aging |
| NICHD | Eunice Kennedy Shriver National Institute on Child Health and Human Development |
| NIDA | National Institute on Drug Abuse |
| NIH | National Institutes of Health |
| NLSY79 | National Longitudinal Survey, 1979 Cohort |
| NSHAP | National Social Life, Health, and Aging Project |
| | |
| $P^3G$ | Public Population Project in Genomics |
| PhenX | consensus measures for Phenotypes and eXposures |
| PROMIS | Patient-Reported Outcomes Measurement Information System |
| | |
| QTGEN | Qualitative Traits (QT) interval GENetics consortium |
| | |
| SEBAS | Social Environment and Biomarkers of Aging Study (in Taiwan) |
| SHARE | Survey of Health, Ageing, and Retirement in Europe |
| | |
| TILDA | The Irish Longitudinal Study of Ageing |
| | |
| UCLA | University of California, Los Angeles |
| USC | University of Southern California |
| | |
| WAIS | Wechsler Adult Intelligence Scale |
| WHAS | Women's Health and Aging Study |
| WLS | Wisconsin Longitudinal Study |
| | |
| XML | extensible markup language |

# Table of Contents

# National Institute on Aging
# Workshop on Harmonization Strategies for
# Behavioral, Social Science, and Genetic Research
# November 29–30, 2011

# EXECUTIVE SUMMARY

Behavioral, social, and economic research has benefited greatly from large, longitudinal cohort studies, many of which have been supported by the National Institute of Aging (NIA) Division of Behavioral and Social Research (BSR). Spanning a wide range of research disciplines, these studies have served as rich sources of information on the social and behavioral influences on aging and health, and many are starting to collect DNA to analyze possible genetic associations with these influences. However, the full potential of these studies in accelerating research can be realized only when they are pulled together. This aspiration can be facilitated by data and phenotype harmonization, which can facilitate cross-study comparative analysis and aid researchers in defining the level of phenotype granularity needed to study behavioral phenotypes in genetic studies. This will become especially important as the classification and recruitment of study populations shifts more toward a reliance on genetic knowledge or biological profiles, rather than particular phenotypes, and as researchers address issues of heterogeneity among groups previously defined by a common phenotype, disease, or condition.

On November 29–30, 2011, BSR convened a workshop to explore harmonization strategies for behavioral, social science, and genetic research. The workshop brought together harmonization experts, principal investigators on harmonization projects, and staff from BSR, the National Human Genome Research Institute, and the Eunice Kennedy Shriver National Institute on Child Health and Human Development. Workshop participants reviewed harmonization basics, existing harmonization efforts and issues, enabling tools and technologies, and the immediate needs of BSR, with a particular focus on phenotype harmonization and the informatics associated with cataloguing studies and data. Discussions from the workshop were intended to guide BSR as it defines the scope and priorities for building a unified harmonization strategy for promoting research and genetic studies within its portfolio.

Following welcoming remarks from Dr. Richard Suzman, BSR Director, and an outline of the workshop by Dr. Jennifer Harris, the first session focused on what harmonization means. Dr. Paul Burton, of the University of Leicester, discussed the importance of statistical power and effect sizes, and Dr. Isabel Fortier, of the McGill University Health Center and P$^3$G Consortium, described harmonization as a balance between the ability to share data across studies and the heterogeneity needed to meet specific study needs. Dr. Fortier also discussed prospective versus retrospective harmonization, and she emphasized defining the research question as a critical first step in any harmonization project. During the following discussion, workshop participants agreed that harmonization is only a tool and might not be appropriate for all research questions, and they emphasized the need for expertise to assess the inferential equivalence of variables across studies. Participants also called for NIH to support the building of a generic harmonization infrastructure that offers enough flexibility for small groups of researchers to achieve specific and precise harmonization.

The next session focused on the status of phenotype harmonization in the BSR portfolio. Dr. Eileen Crimmins, of the University of Southern California (USC), reported on the USC/University of California, Los Angeles (UCLA) "Light" Harmonization Meeting and provided a list of concepts that meeting participants had considered ripe for harmonization. She emphasized the importance of having a theoretical understanding of each concept or phenotype and of discussing concept dimensions, possible scales and approaches, and empirical evidence linking dimensions and measurements. Dr. David Weir, of the University of Michigan, described the Health and Retirement Study (HRS) family of studies and emphasized the need for enough ex ante harmonization to facilitate comparisons and replication of genetic associations with eliminating country-specific variations. Dr. Scott Hofer, of the University of Victoria, described construct-level comparisons and called for the development of phenotype maps and item libraries to help investigators understand how their measures map to common constructs. Drs. Chandra Reynolds, of University of California, Riverside, and Margaret Gatz, of USC, discussed efforts to harmonize twin studies from Minnesota, Sweden, and Denmark and noted the need for time, universal documentation standards, cross-site analyses, and other harmonization tools as challenges to harmonization efforts. Dr. Teresa Seeman, of UCLA, noted that biomarker collection is highly individualized but that there are opportunities for harmonization. The session closed with a discussion of whether NIA should support crosswalk studies and provide investigators with detailed information from existing large longitudinal studies.

The workshop then focused on existing harmonization efforts and enabling tools. In one session, Dr. James McNally, Director of the National Archive of Computerized Data on Aging (NACDA), described the archive as a basic foundation with tools to aid investigators as they harmonize phenotypes and data in a way specific to their needs. Drs. Jinkook Lee and Bas Weerman, both from the RAND Corporation, described the capabilities of the RAND Survey Meta Data Repository, which was designed based on the perspectives of study investigators. Dr. Erin Ramos, of the National Human Genome Research Institute, described PhenX, which was initially designed to facilitate standardization and prospective harmonization. She also acknowledged the importance of retrospective harmonization and described efforts to map PhenX measures to existing studies and resources. Dr. Fortier discussed the P$^3$G Harmonization Platform, which includes a catalog of 48 studies and a series of software tools to facilitate various steps in the harmonization process. Discussion emphasized the importance of researcher involvement in tool development, the benefits of a common data format and a common set of variables for complex domains, and the need to examine existing efforts for gaps and duplications. In a separate session, Dr. Burton discussed DataSHIELD, which facilitates individual-level data analysis across studies without data leaving its original site, and Dr. John (Jack) McArdle, of USC, discussed the use of structural equation and item response modeling in calibrating multiple longitudinal datasets and in accounting for time lags, missing data, and changes in measurement.

The final session involved a discussion of BSR needs, integration, and potential next steps. Although workshop participants suggested that NIA focus on investing in harmonization infrastructure, they also acknowledged current economic and fiscal constraints, as well as a need to educate researchers and study sections about the value of harmonization and calibration projects. Participants also discussed potential domains for harmonization, with the caveat that domain researchers should work with harmonization experts to avoid homogenization or reinventing the wheel.

**Executive Summary**

### *Emerging Themes*

- There is large agreement on the need for harmonization.
- Harmonization is not the same as homogenization; it is important for studies to maintain their unique focus and interests. Thus harmonization is intended to enhance original studies, but also to go beyond those studies in a systematic, planned way, using methods and tools that have already been developed. Harmonization should be viewed as an approach that can provide insight into critical issues about existing measures, both in ways to facilitate cross-study analyses and to identify measures that can capture critical domains.
- NIA can support development of infrastructure and a core set of tools, rather than separate silos. In light of current fiscal constraints, however, NIA is more likely to create support mechanisms to enable groups that already have coalesced, rather than create set-asides for infrastructure development.
- Despite the development of new tools and technology, there always will be a need for experts to analyze data and to inform harmonization efforts, tool development, archiving, and mapping, crosswalk, and calibration studies.
- There is a need for catalogs that are well maintained and easy to use, as well as for an assessment of existing studies, their content, and tools and variables available to plan new studies.
- It is important to consider both long- and short-term goals in unison when building a harmonization strategy. In the short term, it is clear that through the work of several investigators, there are areas that are ripe for harmonization. In the long term, it will be important to create a useful resource for genetic studies in the behavioral and social sciences in a holistic and dynamic way so that new phenotypic areas and/or new studies can be added.
- A harmonization strategy should include a way to systematically catalogue and archive how measures have been harmonized.
- A science of harmonization and associated methodology is being developed by various projects and initiatives. We should build upon those tools and methodologies for advancing harmonization according to the needs of behavioral and social research.
- Efforts should be made to identify and address gaps in the harmonization process.
- The harmonization strategy needs to be sustainable because it is labor and resource intensive; the agenda needs to go beyond the life of a particular grant.

### *Potential Topics or Phenotypes for Data Harmonization*

- Wealth
- Correlations of conscientiousness with health and wealth
- Diagnosis and subcategorization of highly prevalent chronic degenerative diseases such as Alzheimer's disease
- Well-being
- Time use
- Cognitive comparisons across developed and less developed countries
- Selfishness versus altruism
- Social stressors
- Areas highlighted at the USC/UCLA meeting:
  o Psychosocial measures

- Well-being
- Personality
- Depressive symptoms
- Stress
  - o Health outcomes
    - Heart disease
    - Cancer
    - Diabetes
    - Physical functioning
    - Cognitive functioning
  - o Behaviors
    - Drinking
    - Smoking
    - Risk-taking

## *Actions for BSR to Consider*

- Build upon the enormous work (including that not funded by BSR) that already has been conducted on a range of relevant activities (e.g., cataloging study holdings at NACDA, and RAND) to develop a suite of tools to facilitate harmonization (e.g., P3G harmonization platform), minimize duplicative effort on the part of investigators, and prioritize support for multiple complementary activities.
- Define research questions around which harmonization efforts could be organized, and encourage researchers from different disciplines to discuss and perhaps agree on the best way to organize variables and measures related to a domain.
- Organize workshops and conversations with other Institutes and Centers (ICs) to understand what harmonization efforts are under way at NIH and across HHS, as well as workshops where investigators can learn how to think about harmonization efforts and conduct longitudinal studies in the social and behavioral sciences. The experiences of the National Human Genome Research Institute (NHGRI), which has assessed ways to combine and harmonize genome-wide association studies (GWAS), can be used as a model for these conversations.
- Encourage collaborations among various groups that have conducted GWAS, now that new methods have been developed and chips are less expensive.
- Assemble study sections and educate them on the value of calibration studies and harmonize existing longitudinal studies that have genetic data.
- Encourage consensus on domain dimensions, or at least outline areas of disagreement, for example through domain profiles published in a journal such as the *Journal of Epidemiology*. Clinicians should be included in consensus development.
- Review what has been learned so far from harmonization efforts in the psychosocial sciences.
- Support hands-on workshops in which participants learn about available harmonization tools.

# National Institute on Aging
## Workshop on Harmonization Strategies for Behavioral, Social Science, and Genetic Research
## November 29–30, 2011

# WORKSHOP SUMMARY

## *Introduction*

Behavioral, social, and economic research has benefited greatly from large, longitudinal cohort studies. Many of these studies have been supported by the National Institute on Aging (NIA) Division of Behavioral and Social Research (BSR), but others are supported by other NIH Institutes and Centers (ICs), such as the National Human Genome Research Institute (NHGRI), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and the National Heart, Lung, and Blood Institute (NHLBI), as well as other organizations. Individually, these studies span a wide array of research disciplines and have served as rich sources of information regarding the social and behavioral influences on aging and health, for example the effects of adversity across the life course. Many studies are also beginning to collect DNA to facilitate the analysis of possible genetic associations with these influences. Pulled together, large, representative, longitudinal studies can accelerate genetic research within the behavioral and social sciences. However, the types of data collected, and the ways in which these data have been collected, often differ across studies.

As noted in opening remarks by Dr. Richard Suzman, BSR Director, and Dr. Jennifer Harris, BSR Consultant, harmonization of data and phenotypes is needed to facilitate cross-study comparative analyses and to aid in defining the level of phenotype granularity needed to study behavioral phenotypes in genetic studies. This will become increasingly important as the classification of study populations begins to rely more on genetic knowledge or biological profiles, rather than particular phenotypes, and as researchers address issues of heterogeneity among groups previously defined by phenotype, disease, or condition. Ultimately, harmonization can facilitate the realization of the potential these large studies offer and BSR's vision of a unique research resource that can maximize the use of existing and new data, thereby recouping additional returns on investments made by BSR and other agencies, as well as enable more efficient planning of studies and new research exploiting genomic tools in behavioral and social fields, attract new researchers to behavioral and social research, and stimulate greater integration of behavioral and social data into biomedical studies. However, it also was noted that harmonization does not equal homogenization and efforts should be taken not to over-homogenize.

Several harmonization efforts are under way, and strategies among these and future efforts will vary depending on the size, scope, and needs of the organization undertaking the effort. Yet all harmonization efforts must at least consider a common set of elements, including cataloguing, phenotype harmonization, informatics, ethical and legal issues, study descriptions, and data availability. An organization might not have to tackle all elements to enable its research, but it must understand them to develop an effective strategy and set priorities. In addition, it should be emphasized that harmonization is not synonymous with standardization and that the aim is not to

homogenize data or collection methods. Instead, harmonization strategies will differ by and depend on context.

On November 29–30, 2011, NIA BSR convened a workshop to explore and discuss harmonization strategies that will maximize the value of data within the behavioral and social sciences and accelerate research integrating these data with genetic and genomic inquiry. (See Appendix 1 for more extensive description of purpose and background.) Harmonization experts, principal investigators on harmonization projects, and staff from NIA, NHGRI, and NICHD reviewed harmonization basics, major efforts already completed or in process, and enabling technologies. Workshop participants also discussed the status of harmonization within the BSR portfolio, immediate Division needs, and potential next steps. Rather than focus on content areas, the workshop centered on the nuts and bolts of harmonization, particularly phenotype harmonization and the informatics associated with cataloguing studies and data. These discussions will guide the definition of the scope and priorities for building a unified harmonization strategy for promoting research and genetic studies within the DBSR portfolio.

## *What Does Harmonization Mean for Those Still Singing a Single Note?*

### Setting the Scene
*Paul Burton, PhD, University of Leicester*

Harmonization will facilitate the ability to analyze large, representative, longitudinal studies together, which can be especially important not only to compare what happens across different populations, but also to achieve the statistical power needed to identify and replicate small genetic effects. The statistical power achievable in a study depends on the nature of the effect, as determined by the actual effect size, measurement quality, the class of end point, and sample size. Statistical power also depends on the complexity of the problem. More power is achievable by studying direct effects, but many studies in the social and behavioral sciences focus on interactions. In addition, researchers must consider the number of unknowns, the nature of the causal pathway, and observable determinants versus latent variables. Allowing enough time for changes to occur, the amount of resources available to a study, and pragmatic restrictions are also factors in the level of statistical power achievable within a study.

Behavioral and social scientists have long known that any genetic effect they wish to observe is likely to be small. When Burton and others were designing the United Kingdom Biobank, for example, there was little evidence confirming genetic associations with complex disease, and odds ratios were less than 1.3. Even now with more evidence available, half of confirmed odds ratios from recent genome-wide association studies (GWAS) are less than 1.3, and a third are less than 1.2. In addition, as illustrated by the QTGEN study in 2009 and another study investigating loci associated with lung function, the use of common variants cannot escape small genetic effects, because many variants have an effect size that is less than a tenth of the standard deviation. Often these studies can observe genetic effects only because of their larger sample sizes, around 12,000 to 20,000 participants.

Burton and colleagues have developed a simulation-based approach to calculate realistic statistical power estimates for large studies. On the basis of a power calculation for a diabetes end point modeled on random hemoglobin A1C testing and building in genotyping errors,

environmental determinants, and frailty variants, they have found that a typical case-control study would need at least 2,000 cases to identify a genetic effect, with 10,000 cases being ideal. For investigations of gene-environment (GxE) or gene-gene (GxG) interactions, the sample size requirements are even larger, from 10,000 to 50,000 cases.

These sample size requirements, along with the possibility that effect sizes could be decreased or destroyed by confounders, raise the question of whether enough statistical power can be achieved at all. Yet this problem can be addressed through the use of randomization or Mendelian randomization; growing scientific knowledge enabling better study design; deep, high-quality phenotyping; and homogenous sampling. Statistical power also can be achieved by using disease-related or exposure-based sampling in a large cohort, rather than case-control studies, and by increasing the size of individual studies. Meta-analysis—not the traditional meta-analysis based on published results, but the sharing and pooling of raw data and samples and the promotion of harmonization between studies—also can aid in achieving adequate statistical power. Achieving adequate statistical power is feasible, so long as analyses are of the size stated, and individual studies can be any size, so long as they are well designed, networked, and harmonized.

## Discussion Points

- Many have suggested that sample sizes can be smaller with deeper, well-designed, homogenous phenotypes. However, there have been many genetic epidemiological studies where that has not proven to be true. Some have statistical power of 1 to 2 percent; thus even when phenotypes are improved, the statistical power is inadequate.

- A major aim for science is the understanding of causal pathways, but statistical power typically drops once investigators consider these pathways. Endophenotypes and the development of biomarkers reflecting elements of those pathways are crucial, but how well they are measured and how well they fit into the pathways under study will affect the level of statistical power.

- The ability to measure genetic variants has improved, for example through full sequencing as opposed to GWAS. However, these improvements do not negate the need for increased sample sizes.

- Some effects might be real but too small to worry about. In a time of economic constraints, behavioral and social researchers might have to agree on a threshold at which an effect is simply too small to pursue. However, that threshold will depend on the reasons for studying the effect, and in the push to understand causal pathways, a weak effect might suggest an important process. In addition, combining several small effects and testing them in new samples can begin to explain substantial observations. Careful thought is needed to understand what is being modeled and to avoid misinterpretation.

## Key Concepts and Practical Steps to Phenotype Harmonization
### Isabel Fortier, PhD, McGill University Health Center and P3G Consortium

Harmonization involves a balance between the ability to share data across cohorts and the maintenance of enough heterogeneity to meet a cohort's specific scientific needs. When faced

with studies that focus on the same broader question but ask it in different ways, researchers can try to identify the best variable to create among all the studies, design variables without concerns about heterogeneity, or find a balance between scientific validity and harmonization. Achieving that balance will depend on how each research question is defined, and it will require both a good catalog and an appropriate level of scientific expertise. Thus the harmonization program will vary across projects.

In an ideal world, harmonization would be achieved before data collection begins, for example by making questions, protocols, and measures the same across all cohorts. However, even with such stringent harmonization, a comparison among cohorts would reveal that questions were interpreted differently, meaning the measures would no longer be the same. This problem could be addressed through flexible or ex ante output harmonization, where cohorts have a common set of target variables but flexibility in specific question, protocols, and measures. Yet in this case, inferential equivalency must be assured. Prospective harmonization requires a consensus on compatible study designs and tools, and such consensus is challenging and can take many years.

In the real world, even if consensus is achieved for prospective harmonization, researchers must contend with retrospective harmonization using existing data. However, they must account for heterogeneity at the levels of the study, data collection, or source of information, and decisions about heterogeneity must be documented to facilitate future evaluations. In addition, successful retrospective harmonization requires access to study-specific data and related documentation, a respect for all ethical and legal requirements, proper recognition of intellectual property, ensured inferential equivalency, and proper data processing and integration. Study characteristics and database content must be catalogued, common variables must be identified to develop a core set of information, and the harmonization potential of participant studies must be evaluated.

To facilitate harmonization, emerging studies will need access to standard measures, guidelines, and standard operating procedures for data collection. However, such access is not always obvious for these studies. Emerging studies also need governance and generic consent models for access to data and use of it for broader purposes. They also need standard models for documenting data and samples, as well as efficient and flexible information technology (IT) tools. Likewise, investigators leading harmonization programs need well-organized data repositories that facilitate access. They also need access to comprehensive documentation on all aspects of a study that might influence its harmonization potential, as well as guidelines and IT resources. Harmonization should thus be considered a collaboration among studies and tool developers, with the understanding that the entire scientific community must pull together to foster this emerging field.

### *Discussion Points*

- Tools either exist or are under development to merge or facilitate the steps of harmonization.

- Ensuring inferential equivalency includes determining whether the relationship between variables of interest is similar across studies. This requires a validation of all variables and a focus on rigor and process.

- More models are using statistical methods to build in calibration, allowing researchers to test whether their questions or variables can be analyzed using pooled datasets.

## Discussion

Although harmonization can be useful in facilitating cross-study comparisons and achieving the statistical power needed to examine small genetic effects, it is expensive and difficult to do. In addition, the establishment of compatibility between studies is associated with the risk of losing information, yielding a weak analysis. Yet in some cases, such as the UK Biobank, the increase in sample size is worth the loss of information. Thus there is a tradeoff between compatibility and how much one can actually learn. In addition, harmonization can lead to multiple outcomes, depending on the studies chosen, and it can help researchers understand where mistakes could be made in combining studies or assuming equivalency. It should be noted that harmonization is a tool to answer research questions that cannot be answered in other ways. Thus, harmonization is not appropriate for every research question.

The ability to ensure inferential equivalency will depend on how variables are handled. For example, the establishment of comparability across Web-based surveys is hampered by the ability to manipulate survey responses with the layout of the questions. In addition, agreeing on common outcome measures for a variable can lead to rapid advances in the field, but only until new approaches are developed, and researchers should therefore ensure that the agreed-upon measure is not the only one they use. This has been particularly important in cases where subsequent studies have shown researchers that they have not measured what they thought they were measuring. Careful attention must be paid to the details surrounding each variable, particularly context, and a large amount of expertise is needed to assess equivalency.

In determining how to undertake harmonization, NIA must decide between establishing a model that allows harmonization regardless of the research question and developing silos of experts to develop collections of harmonized data. In other words, NIA must decide between developing software and facilitating scientific decisions. Several workshop participants indicated that current harmonization efforts often reinvent the wheel; infrastructure and software development is repeatedly supported, but scientific decisions must still be made. Workshop participants suggested that NIA or NIH support the development of a "grand package," a general harmonization model that would allow enough flexibility for smaller groups of studies to be harmonized in a more precise and specific way. Some participants also suggested that such a model include a phenotypic map, similar to what has been defined in the area of cognition.

### *What is the Status of Phenotype Harmonization in the BSR Portfolio?*

#### Harmonization Needs in Behavioral/Social Sciences: Insights from the USC/UCLA Meeting on Harmonization of Methods and Measures in Longitudinal Studies
*Eileen Crimmins, PhD, University of Southern California*

The University of Southern California (USC)/University of California, Los Angeles (UCLA) Light Harmonization Meeting brought together investigators from 15 large social surveys, which together include about 100,000 study participants. Before the meeting, investigators were asked to list all the measurements used in their studies and to provide information to enable comparisons among concepts and constructs, and at the meeting itself, meeting participants considered eight phenotypes. For each phenotype, an expert spoke about the basic meaning and dimensions of the concept, a variety of empirical approaches, and relationships across empirical

measures. Meeting participants then discussed the strength of correlations across measures and survey approaches.

These discussions revealed that some topics, such as sleep and time use, were barely covered in the 15 studies and needed further research to enable cross-study comparisons. Others, such as depressive symptoms and disability, were covered in all studies and were fairly easy to harmonize. Yet even among the concepts that were relatively harmonized, similarity of scales across studies was rare, and calibration was clear in some studies but not others. In addition, investigators were willing to harmonize but not necessarily able to integrate new measurements, and even if they were able, they had little incentive to do so. Questionnaires were too full, and studies were more focused on maintaining comparability over time. The genetic potential of harmonization was not considered.

The USC/UCLA meeting marked the beginning of a useful approach to harmonization, because every concept had a strong theoretical grounding. Meeting participants discussed all dimensions of a concept, considered some of the scales and approaches, and examined empirical evidence linking dimensions and measurements. The following concepts can be harmonized easily across NIA-supported studies, with the sample sizes needed to include genetic analyses:

| | |
|---|---|
| Psychosocial measures: | Well-being |
| | Personality |
| | Depressive symptoms |
| | Stress |
| | |
| Health outcomes: | Heart disease |
| | Cancer |
| | Diabetes |
| | Physical functioning |
| | Cognitive functioning |
| | |
| Behaviors: | Drinking |
| | Smoking |
| | Risk-taking |

Although harmonizing an entire set of concepts or variables is not warranted, investigators can make reasonable progress in harmonizing some outcomes that are measured well. Importantly, there is a set of variables within a theoretical model necessary to understand these outcomes.

### *Discussion Points*

- Concepts discussed at the meeting were too broad. Additional meetings might be needed to pursue the concepts in more detail. These meetings also might serve as an incentive to promote harmonization.
- Charts were developed to show potential commonalities in domains across studies. Investigators can refer to these charts as they consider whether they want to pursue harmonization within a concept.

- Although harmonization might be possible only for a subset of variables, how to analyze partially harmonized datasets is not clear. Harmonization might need to be stricter for outcome variables than for explanatory or causal variables, but there is no one rule. It could be that lack of coverage can be addressed by modeling, and the use of common items would thus reduce the need to measure everything for a variable. Experts must still perform the complex analyses.
- A map or taxonomy is needed of variables that are "ready for prime time," those that need more work, and those that do not lend themselves to harmonization at present. Calibration studies also are needed for measures to use in genetic studies, but whether to do these samples in or out of sample is not clear. NIA BSR should consider and perhaps release an announcement of what measures it would like to see.

## The HRS Family of Surveys
*David Weir, PhD, University of Michigan*

The Health and Retirement Study (HRS), a nationally representative panel study of persons aged older than 50 years, comprises biennial interviews; interviews with both members of a couple; linkages to administrative records; and multidisciplinary content on health, health services, labor force, economic status, family structure, and transfers. HRS emphasizes the rapid and public release of data. The HRS family of studies includes large, nationally representative studies from around the world. All of these studies participate in ex ante harmonization, in sample design as well as in content, and that harmonization is facilitated by encouragement and seed funding from NIA; by cooperation and collaboration among principal investigators, study directors, and study staff; by ongoing meetings and workshops; and by ongoing sharing of materials and personnel.

Although the studies must participate in harmonization to be considered members of the HRS family, the degree of resemblance between studies varies, as does the degree of resemblance among measurements. Studies most closely related to HRS include the English Longitudinal Study of Aging (ELSA), whose design is based on the HRS instrument; the Survey of Health, Ageing, and Retirement in Europe (SHARE) and The Irish Longitudinal Study of Ageing (TILDA), which share common core concepts; and the Korean Longitudinal Study of Ageing (KLoSA), which adheres closely to the HRS questionnaire. Others, such as the China Health and Retirement Longitudinal Study (CHARLS), the Mexican Health and Aging Study (MHAS), and the Japanese Study of Aging and Retirement, have diverged from the HRS questionnaire depending on their circumstances. Other studies, such as the Longitudinal Aging Study in India (LASI) and the Brazilian Longitudinal Study of Aging (ELSI-BRASIL), are still under development. The HRS family of studies is also distantly related to other BSR-supported longitudinal studies, such as the National Survey of Midlife Development in the United States (MIDUS) and the Wisconsin Longitudinal Study (WLS), and to other international studies such as the World Health Organization (WHO)-supported Study on Global Ageing and Adult Health (SAGE) and the Canadian Longitudinal Study on Aging (CLSA). However, few, if any of these studies have DNA available, so harmonization for genetic analyses will not be possible any time soon.

Within the health domain, the HRS family of studies is harmonized on a few chronic conditions, functional limitations and disability, and memory, and they may soon be harmonized on other cognitive measurements. They also are harmonized to a varying degree on depression, affect,

biomarkers, and genetics. However, the closeness of harmonization among these studies will likely be negatively correlated with the degree of heritability. The studies are harmonized on income and wealth, but these are downstream outcomes, and the studies are not harmonized on physical health.

There should be enough ex ante harmonization among studies to facilitate comparative research and the replication of genetic associations. However, there should not be so much harmonization that possible measurement improvements that could benefit all are eliminated or that country-specific variation is ignored. In the future, the HRS family of studies should advance ex poste harmonization as much as possible, sponsor workshops around cognition, and support research on the main elemental phenotypes linking genetics to complex behavioral outcomes and the best ways to measure them. Moreover, the HRS family should identify those measures that are of little value and can be eliminated to make way for others that should be added.

### *Discussion Points*

- TILDA has better physiology measurements than HRS and is better at collecting information on endophenotypes. CLSA is rich in health measures, but poor in economics.
- A bibliography of papers that cite more than one member of the HRS family has been constructed, and there has been some research on harmonization measures, but little attention has been paid to comparative measurement properties.
- SAGE now operates both through a grant and a contract, and the contract allows NIA more direct control, as well as an opportunity to suggest additional measures. SAGE will begin collecting DNA in the next round, using OraGene.
- It is not clear whether calibration studies were done for HRS, ELSA, and SHARE, but several papers have reported on cognition across these studies. This could be problematic in light of differences in form and language across studies, and it could become more difficult with other studies such as CHARLS. Calibration studies might prove inhibitory in countries with limited resources, but they still should be considered.

### Integrative Analysis of Longitudinal Studies on Aging (IALSA): Challenges and Needs for Quantitative Harmonization
### *Scott Hofer, PhD, University of Victoria*

The Integrative Analysis of Longitudinal Studies on Aging (IALSA) focuses on a coordinated analysis of individual datasets, with the goal of a high degree of comparability across studies at the concept level and an expectation that the pattern and magnitude of effects will be the same. Within this context, harmonization aims to obtain systematic answers to key questions and to provide evidence for generalizability. Harmonization occurs at the levels of research question, statistical models, and variables, and it allows for a synthesis of results to account for ways in which birth, cohort, country, culture, and issues of mortality and selection relate to outcomes and differences across studies. However, harmonization here does have its challenges, including how to proceed when measures differ. Some networks employ a lowest-common-denominator approach by whittling measures to their common definitions, whereas others might select studies with identical measures or compare standardized effects. IALSA has conducted construct-level comparisons, which in some ways can serve as a synthesis from other life course studies and help investigators learn more about how to use predictors from those studies.

Future harmonization efforts will require a single, universal platform that includes all metadata from longitudinal studies. Such a platform will not require a common data structure, because tools are available to read metadata in several formats. However, the platform will require input for potential data structures. In addition, because there is not sufficient overlap among construct indicators to allow harmonization of existing data, additional data should be collected in independent datasets where measures can be equated and calibrated across groups. Such a process can yield an item library, which would provide a phenotype map of how particular measures link with others, perhaps in multiple outcomes. The item-library approach can provide retrospective harmonization to compare results from past and current studies with those from future studies, evaluate both commonalities and differences among measures, and retain necessary study-specific heterogeneity by helping investigators determine where their items or scales map onto common constructs. Calibration and development of the item library will require a harmonization of quantitative measures, as well as a common multivariate item set with particular attention to "planned missingness" and the use of bilingual samples.

Construction of an item library faces some challenges. Some measures, such as those focused on memory or processing speed may not have items that can be analyzed using factor analysis or item response models and require test-level analysis (i.e., based on summary scores). Others, such as depressive symptoms and diagnostic checklists, are formative or mixed measures, and the starting point or context can influence the harmonized outcomes. Other challenges include the ability to maintain reliable and unique contributions of particular measures while maximizing commonalities for pooled data analyses, a dependence on the context of measurement, and the need for a well-defined theoretical framework, which can take several years of effort. Yet a focus on phenotype mapping and item libraries can increase understanding of how scales map to each other and thus have a larger scientific appeal than simply bringing biobanks together with longitudinal studies.

### *Discussion Point*

Several studies in the BSR portfolio focus on in-depth assessments of smaller study populations but include some of the scales used in larger longitudinal studies. Harnessing studies that have that type of overlap could prove valuable, especially as some surveys move toward biomarker measurement, but they must be representative and use systematic and reliable representations of variables.

### Twin Study Harmonization: Experience from "Gene-Environment Interplay of Social Contexts and Aging-Related Outcomes"
*Chandra A. Reynolds, PhD, University of California, Riverside; Margaret Gatz, PhD, University of Southern California*

The "Gene-Environment Interplay of Social Contexts and Aging-Related Outcomes" effort aims to harmonize twin studies representing more than 16,000 participants and 7,100 twin pairs from Minnesota, Denmark, and Sweden. To address the first aim of harmonizing social phenotypes and aging outcomes, investigators have agreed upon the structure of administrative variables, such as participation at each wave (at the individual and pair level), attrition, and mortality; identified common constructs and formed workgroups around themes such as depression or loneliness; and created measure spreadsheets with item descriptions. Data storage uses 7Zip

encryption software and makes data available only to investigators on a SharePoint site hosted by the Danish group. Dropbox (www.dropbox.com) is used to share organizational documents that are not related to the data itself. Per the project's data sharing plan, harmonized variables will be shared. Plans call for using $P^3G$ and the National Archive of Computerized Data on Aging (NACDA) as appropriate.

The harmonization effort has undertaken a depressive symptoms crosswalk as a case study for this workshop. Only two scales are used by the participating studies: the Center for Epidemiologic Studies Depression scale (CESD) in Minnesota and Sweden, and the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) in Denmark. Investigators established sufficient commonality of items across the two scales by noting what items and subscales correspond and by creating a linked sample (Mechanical Turk and Healthy Minds) where both the CESD and CAMDEX can be administered in counterbalanced order with unrelated material in between. All items were scored for depression, invalid responses were removed based on a check on scores on a filler vocabulary task, and item levels were assessed for comparability. On the basis of this evaluation, Mechanical Turk respondents were likely to be less educated and financially comfortable, and direct comparisons by age suggested that Mechanical Turk respondents were more depressed. The psychometric properties of the CESD and CAMDEX were robust by age and source (Mechanical Turk or Healthy Minds), with an overall correlation of 0.867 between measures. This crosswalk also suggested a strong underlying latent construct, to which many variables contribute, and that these measures might be driven more by affect items. The investigators will include all of the twin studies in conducting longitudinal twin analyses of this latent depression construct once crosswalk data analyses are complete.

The harmonization process is a time-intensive one, as illustrated by the time needed for biweekly workgroup meetings. The need for universal documentation standards, including standards for missing responses and standard names for variables at the metadata level, is another challenge, which could be addressed with the development of a data documentation toolbox. Other harmonization tools and cross-site analyses are also needed, but those that have been developed so far have not incorporated complex variables or are not yet live.

### *Discussion Point*

The work and decision making described here illustrate the importance of process documentation. The amount of effort underlying harmonization efforts should not be lost.

## Biomarker Harmonization
### *Teresa Seeman, PhD, University of California, Los Angeles*

Individuality is the current state of the art in biomarker or biological data collection, as study investigators decide what and how to measure based on their scientific questions. These decisions also are influenced by logistical and financial constraints. For example, decisions on whether to use venous collection methods or capillary methods such as dried blood spots (DBSs) or point-of-service meters can be influenced by cost concerns, geographic logistics such as state regulations on the amount of blood that can be drawn, and the availability and range of standard assays. Although DBS appears to be a simpler alternative, there are few written protocols for collection, handling, and processing, and there are no known laboratory standards.

The following table lists studies collecting venous blood, DBS, and DNA. LASI, the National Health and Aging Trends Study (NHATS), and the National Longitudinal Survey 1979 Cohort (NLSY79) are planning or considering DBS collection.

| Study | Venous Blood | DBS | DNA *Oragene †venous |
|---|---|---|---|
| Health and Retirement Study (HRS) | | X | X* |
| National Survey of Midlife Development in the United States (MIDUS) | X | | X*† |
| MacArthur Research Network on Successful Aging | X | | |
| National Social Life, Health, and Aging Project (NSHAP) | | X | X |
| National Longitudinal Study of Adolescent Health (Add Health) | | X | X |
| Women's Health and Aging Study (WHAS) | X | | |
| Wisconsin Longitudinal Study (WLS) | | | X |
| Los Angeles Family and Neighborhood Study (L.A.FANS) | | X | |
| Mexican Family Life Survey (MxFLS) | | X | |
| Costa Rican Longevity and Healthy Aging Study | X | | |
| Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan | X | | X† |
| Indonesian Family Life Survey (IFLS) | | X | |
| Chinese Health and Retirement Longitudinal Study (CHARLS) | X | | |
| Chinese Longitudinal Healthy Longevity Survey | | | X |
| English Longitudinal Study of Ageing (ELSA) | X | | |
| Whitehall Study | X | | |
| Newcastle 85+ Survey | X | | |
| National Health and Nutrition Examination Survey (NHANES) | X | | |
| Tsimane Health and Life History Project (Bolivia) | X | | |
| In-Depth Network – selected sites collect samples for specific projects | X | | |
| Survey of Health, Ageing and Retirement in Europe (SHARE) | | X | |
| Study on Global AGEing and Adult Health (SAGE) | | X | |

Coordination of biomarker data collection has centered mostly on the HRS family of studies, but cross-study comparisons have been hindered by differences in the protocols and types of sample collection within the United States, around the world, and even within DBS studies. Even when studies use the same protocol, they might use different materials or rely on different laboratories to read results. Thus calibrations and validations must be developed across protocols, laboratories, and even time periods. The USC/UCLA Center on Biodemography and Population Health is working on ways to facilitate the harmonization of biomarkers. The Center is also upgrading its website by adding a section to post the best protocols for a type of collection, incorporating videos demonstrating collection methods, and posting what is known about available laboratories and assays. The Center is conducting validation projects, for example on cross-laboratory DBS assays and point-of-service equipment, and conversations are under way for additional validations of DBS assays to facilitate international harmonization.

Although biomarker data collection is highly individualized at present, there are many opportunities for increased harmonization, because many studies are interested in similar concepts and resources are available to facilitate cross-study evaluations. However, harmonization of biomarker data will require centralized information, including a better catalog of what is available, how samples were collected and assays done, and who holds stored samples. Availability and accessibility of data and samples from different studies must also be catalogued.

### Discussion Points

- BSR has asked NIH-supported Clinical and Translational Science Award sites (CTSAs) about performing blood draws for its surveys. The CTSAs are considering it, but they are historically limited by their presence on university campuses and their requirement that participants come to them. Other groups are collecting blood and oral samples, for example for life insurance purposes, and others, such as UCLA, are working on establishing transports that will go into the communities.
- The International Society for Biological and Environmental Repositories (ISBER) has guidelines for biological sample collection. ISBER focuses on the life history of a sample, including laboratory collection, transport, and storage. Anyone working with biomeasures should consider becoming a member of ISBER, which has an active listserv, provides a learning environment, and is now discussing ways to obtain access to laboratory results.
- Issues with biomarker data collection also have been reviewed in a National Research Council publication on biosocial surveys (www.nap.edu).
- Both assays and cross-laboratory standards are improving for DBS. The USC/UCLA Center on Biodemography and Population Health is working with groups to fund additional assays, improve existing assays, and identify ways to calibrate across laboratories.
- Calibration is often a concern with biomarkers because studies use cut points. Percentiles can be useful across studies, regardless of calibration, but they might not work as well for specific points. Investigators and harmonization leaders should be aware of the context and limitations surrounding biomarker data collection.
- Discussions are under way regarding data and sample storage and, in some cases, ways to centralize repositories.

### Discussion

As discussed in several presentations in this session, harmonization efforts will require a catalog of what is available. One type of catalog could be a "phone book" or Web portal, where NIA simply provides summary information of who has what and how to contact them, and investigators are responsible for obtaining further information. A metadata catalog, for example a list of which studies use DBS collection, is essential and can be relatively easy to build. However, such a catalog will require constant interaction with investigators to ensure information is accurate, and it will be useful only to a limited degree. With another type of catalog, NIA would support the actual crosswalks and provide detailed information about what each study has measured and how. Preservation of all information is essential, as the science and changes in the ability to use data will drive how things should be catalogued.

Harmonization efforts to facilitate genetic studies in the behavioral and social sciences can learn from the experiences of the Alzheimer's disease research community. This community surveyed

everything it had and learned that 10,000 DNA samples could not be shared because appropriate consent had not been obtained. NIA and the Alzheimer's disease research community thus established a repository to store samples and a database that grows as needed, and it developed a materials agreement, a consent template, and uniform data-sharing plans. In 2005 NIA converted a dataset, to which investigators contributed with different measures and tests, into a universal dataset. Although investigators initially balked at having to conform to the universal dataset, NIA made funding contingent upon the contribution of data to this dataset. Now, data are available on approximately 30,000 individuals, and the data are uniform. However, these uniform data represent only a minimal dataset; investigators are free to add other data specific to their studies. The establishment of the universal dataset, and the collection of deep phenotypes and endophenotypes, has contributed to the discovery of 10 new genes associated with Alzheimer's disease.

Another model, at least for contributing data dictionaries, is the Database of Genotypes and Phenotypes (dbGaP), which requires investigators to submit their data dictionaries and allows other researchers to search investigator last names and study names. However, the dbGaP process is somewhat cumbersome, and such a hurdle could discourage investigators from using the data. The development or enhancement of a national resource will require sufficient investment to ensure that a database or catalog is user friendly.

### *How Do We Build on What Has Already Been Done?*

**The National Archive of Computerized Data on Aging (NACDA)**
*James McNally, PhD, University of Michigan*

NACDA, a component of the Inter-University Consortium for Political and Social Research (ICPSR) that has been supported by BSR since 1981, facilitates and supports gerontological research through thematic collection guidance and by collecting data on aging. It aims to become a seamless dataset and to conceptually link datasets so that investigators can use different pieces at different times. NACDA has approximately 20,000 registered users and 700 datasets related to issues of aging and the life course.

Harmonization, defined as any process that aims to improve comparability across databases, is a large part of what NACDA does and is considered part of the data life course. Although the Archive incorporates a vast amount of data, it still relies on experts to organize that data conceptually around issues of aging and to identify holes in its harmonized datasets. The archivists within NACDA are generalists, and they often interact with multidisciplinary groups. NACDA's harmonization approach is driven by the nature of its repository archival structure, but because different structures are appropriate for different sciences, it does not dictate that investigators use its approach. Instead, it provides a basic foundation, with datasets, variables, and tools that allow investigators to quickly identify, organize, and harmonize data in a way specific to their needs.

NACDA employs a standard processing pipeline, with closely monitored procedures, on every dataset regardless of size. This pipeline includes a standard distribution and secondary analysis process and a standardized set of products. Essential to the NACDA harmonization process is the Data Documentation Initiative (DDI) extensible markup language (XML), which provides

NACDA with the ability to control, manipulate, and translate data and provide uniform structure to its codebooks, regardless of the variable. Internal tools allow NACDA archivists to track data and metadata, and external tools include an active search page where users can search by filters including study, variable, subject, or geography. NACDA also provides a thesaurus for users who have an idea but do not know which data might be relevant. Search results provide the datasets or, if necessary, a link to the website where a dataset is available. An online analysis system allows users to specify and download customized datasets.

Because NACDA is a service organization, it emphasizes follow-up, monitoring, and user support. Archivists are interested in how the data are used, what problems might exist, and how NACDA can improve service delivery. NACDA therefore tracks user requests, generates reports for any interested investigator or funder, and provides information on the types of researchers using its datasets. In addition, its follow-up procedures aid NACDA itself in determining where best to promote its services. NACDA also offers the LEADS database, which harvests funding records from NIH, the National Science Foundation, and other funding organizations and allows NACDA to help investigators think about how to organize their data to facilitate archiving.

Specific projects include a harmonized dataset for the Collaborative Psychiatric Epidemiology Surveys. This dataset is useful not only for analysis, but also as a training tool on the actual process of harmonization. Other harmonization projects include baseline and follow-up data for Project Talent, HIV/AIDS patient records for Fenway Health, data for Americans Changing Lives, crosswalk and tailored analysis data for NHANES, and a biomarker registry for NACDA studies.

### *Discussion Points*

- NACDA prefers to treat longitudinal data as a seamless dataset, with one observation per person over time. It also prefers to provide tools for investigators to subset data themselves.
- NIA often points investigators toward NACDA as a potential resource, and it has several requests for applications (RFAs) that cite NACDA. Although NIA can require data sharing, it can only recommend NACDA as a potential resource. Yet an increasing number of review groups are paying attention to data sharing plans, particularly for international collaborations, and as a result investigators are communicating with resources and obtaining letters about data-sharing plans.
- Investigators often contact NACDA directly, and NACDA guides them on ways to facilitate sharing their data.
- Once NACDA archives data, the cost of maintaining it is minimal, but it might spend half its budget on a handful of studies. If needed, NACDA can write a grant or obtain internal funds to activate a dataset, or it will encourage investigators to write R03 or R21 grants. Data can be preserved and migrated to new platforms.
- All datasets at NACDA are publicly available, but users must complete a responsible use agreement and, in some cases, fulfill other requirements.
- NACDA does not yet link genotypes to phenotypes. HRS does so by providing dbGaP genotype data with an identification system and setting terms by which investigators can access the link.
- NACDA does not own data. Instead, it is a facilitator, allowing investigators to own their data while providing a way for them to share data at no added cost.

- NACDA also has a catalog of metadata from studies that are relevant to aging research but not part of the NACDA collections.

## RAND Survey Meta Data Repository
### JinkookLee, PhD, and Albert "Bas" Weerman, RAND Corporation

The RAND Survey Meta Data Repository was built based on the perspectives of principal investigators. Because many on the RAND team are or have been principal investigators, they are familiar with the HRS databases, and they worked with other investigators and advisory boards to clarify the repository structure, create common measures and a common dataset, and evaluate the harmonization potential of participating studies. Thus far, RAND has built metadata and concordance information on 11 studies from 25 countries; harmonized phenotype variables for cross-country, longitudinal study; and provided information and support for users, including where to obtain microdata, contextual information about country statistics from secondary sources, a Wiki system, and a help desk. Inputs to the RAND Survey Meta Data Repository come in different formats, but they are converted into the DDI format to facilitate compatibility with other harmonization efforts. Metadata from all waves of the HRS are in the DDI format.

On the Repository website, users can browse or search studies, and they can add relevant studies or items to a "shopping cart," where they can add variables and obtain links to relevant data. Like NACDA, the Repository allows users to search by topic, but the RAND team also has assigned topics to every variable or question contributed to the system. Thus, if a user searches for "depression," the repository lists all survey questions related to depression. The website also suggests other studies that might be relevant to a user's question, and it allows users to leave comments, for example if something is inaccurate.

The harmonization of phenotype variables for cross-country longitudinal study has followed the model of RAND HRS, which is convenient and frequently used. The harmonization potential of participating studies has been evaluated for several domains, including demographics, health, financial and housing wealth, family structure, and identifiers. Weights, working papers, and domain-specific user guides have been generated as well. Harmonized data files have been developed as a single data file with all longitudinal waves, in a format where one observation represents one respondent. For each variable, RAND has documented what the variable looks like, how questions have changed over time, and how the variable might compare to those used in HRS. Datasets are not stored on the RAND server; rather, RAND provides documentation of how metadata were harmonized and allows users to construct their own datasets. The repository uses RAND-HRS names for variables, allowing users to link to databases. If an item is substantially different from the RAND-HRS name, then the item is flagged so users can consider different ways to ask questions.

Through the Survey Meta Data Repository, RAND has developed documentation, within each domain, of how compatible and concordant surveys are, how programs were set, and where survey concepts are similar and different. Efforts thus far have created a minimal set of variables everyone can use, while allowing investigators the flexibility to add other variables specific to their interests. Since the repository was opened to the public in February 2011, the site has had more than 5,700 visits, and almost 100,000 pages have been reviewed. Future plans include updating the repository with newly available metadata indexed for cross-wave, cross-survey

concordance, expanding harmonized variables by adding new waves, surveys, and variables; and providing aggregate population- or subpopulation-level estimates of key harmonized variables. This will allow users to identify which country, year, and subpopulation they want to use, and obtain statistics as a dynamic interplay on the repository website, without having to download them. Although RAND does not touch restricted data, it aims to notify users when restricted data appear in a domain. RAND also plans to reach out through user workshops.

## Discussion Points

- The repository updates whenever client datasets update. However, most datasets update their process micro data, not their metadata. Users can see the version of the data RAND has when they search the repository.
- So far, because RAND has worked closely with principal investigators on data collection and analysis, it has not met with resistance from individual investigators, as they have understood from the beginning that this is a harmonization effort.

## PhenX
### Erin Ramos, PhD, MPH, National Human Genome Research Institute

PhenX (consensus measures for Phenotypes and eXposures) is a toolkit that provides almost 300 standard phenotype and environmental exposure measures across 21 domains. The key criterion for inclusion of a measure in PhenX is that it is of relatively low burden and well established with validation data. PhenX designers also have worked with the NIH Patient-Reported Outcomes Measurement Information System (PROMIS), but many PROMIS measures have not yet been validated. PhenX users can browse domains and see the keywords and synonyms assigned to a variable, the 15 measures selected for each research domain, the rationale for measure selection, and the roster for the working group charged with selecting measures for that domain. Users can add measures to a virtual basket and generate a report that provides protocols, rationales, and references, along with a data collection worksheet to help users incorporate measures into their studies. PhenX also offers an extensive search capability.

NHGRI initially designed the PhenX project with an aim toward standardization and prospective harmonization, with the rationale that standard measures would be needed to detect loci with small effect sizes, GxE, and GxG in large samples and to increase the potential for combined analysis. PhenX also was intended to provide standard measures that investigators could use for phenotypes outside their areas of expertise, thus positioning their studies for future collaborations and cross-study analyses, without the investigators having to conduct data harmonization processes of their own. Since the inception of PhenX, however, NHGRI also has acknowledged the need for and importance of retrospective harmonization and is therefore mapping PhenX variables to existing programs and resources, such as dbGaP, the Cancer Bioinformatics Grid (caBIG®), $P^3G$, and specific IC projects. More measures will be added as they become available. As is the case for other harmonization efforts, PhenX relies on experts to select measures and ensure they are comparable.

The PhenX cooperative agreement began in 2007, and the first domains went live in 2009. NHGRI has established a PhenX steering committee, identified liaisons to facilitate outreach to other NIH ICs, and engaged in public outreach to the scientific community. Through October

2011, PhenX has had close to 1 million page views and 220 visits per day. There are more than 650 registered users, half of whom have agreed to be contacted for feedback, and 146 countries have access to the toolkit. Users can share the toolkit with other potential users. An administrative supplement, PhenX Rising, has been launched to support investigators as they add PhenX measures to their ongoing studies, and NHGRI is working with Chinese investigators to translate all measures into Chinese. Validation studies are under way, and plans are in place for cross-study analyses to determine how well the toolkit is working.

### *Discussion Points*

- PhenX has relied on experts to do crosswalks and provide information, and NHGRI has worked to fill in the gaps.
- For each domain, the working group has selected and recommended measures after long and careful deliberation. There are some examples where the recommended measure is one observed in the clinic, with a notation of self-report measures investigators can use if they cannot collect data from the clinic.
- PhenX has an appendix with information about the measures that did not make it into the list of 15.
- Domains and variables relevant to NIA and BSR include socioeconomic status, MacArthur scales for wealth determination, and the social environment. NHGRI also is working with the National Institute on Drug Abuse (NIDA) to expand risk measures. Geographic measures tend to be domestic to the United States.
- The sensitivity and specificity of measures are not clear with respect to chronic degenerative diseases that are highly prevalent in essentially non-medical surveys. Nor is it clear what can be done at low cost and low effort to those measures in studies.
- The working group focused on the disease and trait domains included in genomic studies as a criterion for measure selection. However, it is not clear that many genomic studies have been done within the psychosocial and social environment domains.
- Although NHGRI has worked to map PhenX measures to existing variables, some PhenX measures, such as those for socioeconomic status, are not at the level behavioral and social scientists want to include in their studies. NHGRI can work with experts in these fields, as NIDA has done. In that case, NIDA provided supplemental funds, and NHGRI organized working groups to add sufficient detail for PhenX measures of substance abuse to be used by experts. The NHGRI/NIDA collaboration added new measures while establishing core measures that NIDA could encourage use of in funding announcements.
- There is a risk that PhenX users might use variables incorrectly because they do not know the theoretical underpinnings of a domain.
- Generating a mapping table might give NHGRI an idea on how well PhenX measures have mapped to other resources.[1]

---

[1] A compilation of PhenX measures in 13 selected NIA/BSR-funded studies (rev. November 2, 2011) is available by request from NIA/BSR staff.

## International Harmonization Platform (P³G)
### Isabel Fortier, PhD, McGill University Health Center and P3G Consortium

Although harmonization can be done by hand or with minimal materials, the development of a harmonization framework or software can facilitate the process by increasing transparency and quality; by making it faster, cheaper, and reproducible; and by allowing standardized documentation of outputs. The software also can make harmonization outputs accessible to other researchers, thereby facilitating further research. The International Harmonization Platform, a partnership between the Public Population Project in Genomics (P³G)consortium and several other groups, provides operational IT infrastructure that supports the management of harmonization projects. The platform includes a series of software applications that are open source and support each step of harmonization (see Appendix 3). Each application is individually functional, and the platform itself is flexible, allowing investigators to use only the applications they need.

- A comprehensive catalog provides general information, study background and objectives, information on how data were collected, and variable-related study questions for 48 studies. Similar information and keywords are also available for physical and cognitive measures.
- Once researchers have defined their research questions, the DataSchema software helps researchers identify and properly document core variables in their harmonization efforts. This software was developed in a process similar to that used for PhenX measures, through literature reviews and conversations with experts.
- Another software application aids in study selection by helping investigators define science-based rules to evaluate studies' harmonization potential with respect to the investigators' defined core variables. This software produces a table showing which studies can be included in an analysis. Investigators must define decision rules for equivalency.

Three pieces of software within the platform help investigators with data processing, integration, and analysis of a harmonized dataset. DataShaper allows users to enter their DataSchema algorithm for harmonization, Opal reads the algorithm and applies it to data, and Mica allows users to read harmonized data. The raw data remain at the study site. Mica also can be used to build a web portal for any harmonization project, facilitate web-based data collection, and work with other applications such as DataSHIELD. Each piece of software can be used separately, depending on an investigator's needs. For example, investigators who have local access limited to aggregated data can use Opal to read their harmonization algorithms but apply it to their local dataset.

### Discussion Points

- Many groups have developed search tools for their own work, but maintaining these tools requires a large amount of effort. A common platform will be useful for future harmonization efforts.
- DataShaper, Opal, and Mica are not yet DDI based, but discussions are under way to determine whether these applications will be made compatible with the DDI format.

## Discussion

As illustrated by the RAND Survey MetaData Repository, researcher involvement is critical in developing tools that will have value. RAND conducted a series of calibrations at the question, item, and variable levels in collaboration with researchers, but it also needed researchers to evaluate whether those calibrations would help in answering fundamental research questions. Likewise, care must be taken not to develop a fixed set of standards, because guidelines and harmonization processes will differ by discipline. NACDA, with its generic tools, can serve as the base of a pyramid, but groups will also create specialized tools, and these must be shared.

Some standardization is needed however. NIA-supported researchers are required to make their data available, although they do so in different ways. Agreement on a common format for tool development can facilitate the transition of data from a study dataset to an archive, and time and money can be saved if a principal investigator decides to use that format from the start. The ICPSR web page provides a link to a consortium of DDI developers and the international community.

Although a large amount of work has been done so far on harmonization, and although several tools have been developed, more work is needed to develop research on harmonization methods and to encourage further harmonization. A cluster of studies around a topic can be used to create a model to bring investigators into harmonization efforts. Tools should be developed to support smaller, sometimes specialized, groups that work on studies not linked to NACDA or other archives. Harmonization will benefit also from good metadata. The RAND Survey Meta Data Repository has so far focused its efforts on the HRS family of studies, but it is not yet clear how to link studies that do not map to HRS.

Although common software is needed for some domains, common language will be needed for others. Some domains of interest to BSR are highly complex and have a large number of components, and expert work is required to develop a common set of variables. The tools and software discussed during this session might not be needed for such an effort.

Inevitably, some researchers will reinvent the wheel when it comes to harmonization, but work should be done to avoid investing new research dollars in processes that have already been done. The research community might thus benefit from an online, searchable bibliography of existing harmonization efforts, including where the science is and how it is used. Investigators also can access NACDA, see what is there, and provide feedback on what is missing. Moreover, experts can review the steps of harmonization presented by Fortier and consider where gaps and duplications occur at each level.

ICPSR is a member of a large organization of repositories and has strong connections with British and European repositories. Thus the ability to strengthen data sharing and integration among these repositories exists. As experts consider where gaps and duplications exist within current harmonization efforts, they should work toward a holistic plan that can be linked with plans developed by other countries.

## *What Important Enabling Technologies and Approaches Do We Need?*

### DataSHIELD
#### *Paul Burton, PhD, University of Leicester*

In study-level meta-analysis, a summary statistic can be estimated from each study, and appropriately weighted means and standard errors can be calculated across all studies. This level of analysis is particularly useful if summary statistics are easy to obtain or cheap, fast, and convenient to construct, for example in GWAS. Study-level meta-analysis is relatively fast and easy, and it provides answers that are about the same as answers one would derive from individual-level meta-analysis. However, study-level meta-analysis also offers limited flexibility, and the increasing complexity of biomedical and behavioral science requires a capacity for exploratory analysis, for which flexibility is necessary.

Such flexibility can be gained through individual-level meta-analysis, where individual-level data from a series of studies are directly pooled into one large dataset, then analyzed as if that dataset were a single study. However, such analysis is often hampered by ethical, legal, and social restrictions on data sharing. For example consent documents for a study might state that study results will never be shared or that they will never be shared internationally. The need to obtain ethical, legal, and scientific permission to access the data can be difficult and time-consuming, thus discouraging or prohibiting investigators from performing individual-level data analyses.

DataSHIELD, which has been developed under BioShare and co-funded by the Birth Cohort Study in the United Kingdom, uses horizontal pooling to facilitate individual-level data analysis without data sharing. This tool links an analysis computer at a center with study data computers at external sites. The analysis computer sends commands to the data computers, the data computers return summary statistics, and a series of iterations occurs until parameters are updated and models converge. DataSHIELD results compare well with conventional analyses, and users receive results similar to those they might have received if they had done the analyses themselves. The datasets can be of any size, and all data storage remains at the data sites.

DataSHIELD has been implemented within Opal, with an R environment investigators can write to, for example to create new variables, and a second R environment that allows variables to be transferred from the first environment and returns analytical results. Three pilot projects are under development: one replicating analyses already done for international studies, another using DataSHIELD on a dataset in Scotland without using the Opal system, and a third implementing DataSHIELD in a study of childhood autism in Australia. Although the bulk of work in DataSHIELD so far has focused on horizontal pooling, DataSHIELD also is useful for study-level meta-analysis on vertically partitioned data.

Several safeguards are in place for DataSHIELD. All contributing studies must have ethical approval, all investigators and users must sign confidentiality agreements, and all study investigators must record information flows to facilitate troubleshooting. In addition, no new models are used until summary statistics are fully understood. DataSHIELD is useful for generalized linear models at present, but developers are also working on its potential to analyze random effects. Future plans include a software wrapper that can monitor and interpret incoming

and outgoing information flows and identify, record, and block any requests that might potentially disclose information.

### Discussion Points

- There is some concern that variables indicative of group characteristics, for example racial and ethnic composition, might allow users to identify data sources. However, these data are included in the larger, pooled data without leaving their storage site.
- Many scientific communities, for example the Alzheimer's disease community, have started to use cloud computing, which raises concerns about the possibility of data being identified. BioShare is using the cloud only to develop DataSHIELD on simulated datasets or unidentified datasets.

## Calibration Methods for the Harmonization of Quantitative Traits
### John McArdle, PhD, University of Southern California

McArdle and colleagues have explored the use of structural equation and item response models in calibrating data for harmonization. In one analysis,[2] they explored Cattell and Horn's theory of cognitive changes by assessing whether one or two factors (thinking versus knowing) were needed to assess changes in cognition over time. They compiled one sample of 41 studies encompassing approximately 13,000 Wechsler Adult Intelligence Scale (WAIS) protocols in normal aging people. They explored a reference variable design, grouping variables most associated with a factor, as well as a fractional block design in which variables were randomly chosen, and created power curves to look at the theories. In so doing, McArdle and colleagues found that the heterogeneity of studies, in this case heterogeneity in the measures chosen to make a conclusion, could be important. They also found that their models did not require all variables to be measured on all individuals under all conditions, and they suggested that structural equation models could encourage researchers to consider "planned incomplete" data analysis.

In a second analysis, McArdle and colleagues explored the principles of time lag and changing scales, focusing on WAIS data and data from the Berkeley Growth Study.[3] They found that a concept such as growth could still be plotted even if the measure changes, but that certain manipulations would be required. Combining datasets could accumulate statistical power, but the cognitive scales were not the same, the two datasets did not have enough overlap, and structural equation models depended partially on factors they could not test. McArdle and colleagues thus used item response models to map or translate scores from one scale to another, which allowed them to plot scores across time and ages.

---

[2]McArdle JJ. Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research* 1994;29:409-54.

[3]McArdle JJ and Woodcock JR. Expanding test-rest designs to include developmental time-lag components. *Psychological Methods* 1997;2:403-35; McArdle JJ et al. Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Developmental Psychology* 2002;38:115-42; McArdle et al. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods* 2009;14:126-49.

In a third analysis, McArdle and colleagues developed a framework to analyze data from Project Talent, which in 1960 aimed to collect high-quality achievement, aptitude, and interest data and examine its predictive ability over time. Project Talent initially measured 400,000 students across 1,200 high schools, but McArdle's framework used a tenth of this sample and chose the best representations of measures to be used for a common core of variables. Investigators were given a sample and asked to follow up with Project Talent participants. Data were divided into the central core, which all investigators had to collect, and data specific to each investigator. This process allowed for the correlation of any two items in the framework, demonstrating again that all information does not have to be collected for all participants.

Calibration involves harmonization, rather than standardization. Harmony is defined as agreement, accord, or a pleasing combination of elements in a whole, whereas standardization means conforming to or evaluating comparison with a standard. Yet although calibration will be essential in harmonization efforts, few understand the distinction between harmonization and standardization. Common factors are essential, because constructs of interest must be measured, no matter what else is used, and common items can provide a practical way to calibrate measures of a construct across studies. Common scales make it easy to determine whether a study can be harmonized with others, but some constructs might be lost. Likewise, common core sets of items and scales can be used to join constructs, but this works only to the degree at which constructs are correlated.

## Where Do We Go From Here?

### Discussion: Integration and Next Steps

NIA might support harmonization efforts best by networking specific elements into a coherent harmonization infrastructure. Using its experiences with the HRS family of studies as a model and starting point, NIA can create an infrastructure that keeps studies broadly compatible while conducting outreach to bring other investigators in, yet it also must enable more focus on specific behavioral and social science domains, research questions, and approaches to support deeper harmonization. In addition, NIA can encourage studies in domains, such as disability, where the evidence suggests that current methods do not provide realistic measures. Although NIA funding mechanisms are restricted to investigators in the United States, the resulting infrastructure is not necessarily limited to this country.

Crosswalk, mapping, and calibration studies could be highly useful to harmonization efforts and future research, as results from these studies would benefit everyone. Data collection for such studies could be accomplished easily and relatively inexpensively on the Internet and at least provide clues to which studies are worth pursuing further. This could be particularly useful for domains, such as conscientiousness or personality, where a wide constellation of measures vary at different stages of the life course. Calibration studies also could help investigators understand how measures are related and eliminate unnecessary measures from their surveys. Yet such studies appear to be incremental and might not fare well in a funding climate that demands grand statements about the potential impact of a study. The National Advisory Council on Aging has recently approved a program for secondary analysis and archiving, and the first request for applications has been released.

Although the broad research community worldwide is creating the ability to answer big questions, the demand for that capability is not apparent. This problem could arise from the time needed to develop a scientific discipline, or from a hesitation by researchers who do not know about these capabilities and thus assume these questions are unanswerable. In addition, investigators who might pose such questions could be penalized by study section reviewers who have not seen those types of applications before. Moreover, what produces health is still not understood. The medical community understands how to treat patients who have already developed a problem, but more research is needed into what determines health and how to maintain it. A better understanding of what constitutes health could lead to the larger questions that require harmonized data.

NIA has supported smaller studies that offer deeper phenotypes, biological specimens, and samples that could be genotyped, and leveraging these studies could improve understanding and perhaps calibration of the measures used in larger studies. However, investigators on some of these smaller studies have been the most resistant to data sharing and harmonization. Incentives are therefore needed to encourage investigators on smaller studies to participate in harmonization efforts. For example, investments in harmonization infrastructure might draw in investigators interested in a new resource, or the NIA Intramural Program could encourage collaboration with extramural investigators through joint meetings. Dr. Luigi Ferrucci, the new Scientific Director at NIA, has shown interest in this type of collaboration.

One workshop participant envisioned that NIA could provide investigators with a harmonization protocol or flow chart, similar to that presented by Fortier, so that investigators would know what tools and resources are available. NIA also could support a library that provides documentation on studies that have already been done. The International Harmonization Platform was designed with these goals in mind, and such a mechanism is in place among the HRS family of studies. NACDA has a generic cataloging system for many domains, and RAND has developed a strong system for addressing variables within certain domains. However, more is needed to build a catalog of methods that can apply for multiple domains beyond those covered by HRS. NACDA has the expertise for this kind of work and could take on this type of activity, freeing investigators to conduct their research. The RAND Survey Meta Data Repository also can extract metadata from non-HRS studies.

NIA has been hindered in supporting harmonization efforts by a lack of clarity regarding the amount of funding needed and where it is best applied. In addition, the return on investment, in terms of secondary applications, is not clear. A focus on the infrastructure, particularly the development of specific elements that could facilitate various steps of harmonization, could generate a large return on investment and be less costly to support. As suggested by Dr. Burton, such a focus could cost a total of $2 million over 5 years. He noted that he had received $1.5 million over 5 years to work with data from the 1958 Birth Cohort and that this work has yielded 450 publications since 2004. Lessons also could be learned from the Genetics Data Warehouse and dbGaP.

It is likely that in the current constrained funding climate, NIA will support already-assembled groups to move forward, as has happened in the areas of cognition, stress management, conscientiousness, and subjective well-being, rather than support the development of an infrastructure. In addition, BSR has proposed a Common Fund-supported activity encouraging

harmonization and, where necessary, calibration methods in behavioral, social science, and genetic research. This proposal has been combined with one from NHLBI on registries and with another proposal focused on moving beyond GWAS and identifying new targets for drug development. BSR is working to ensure that a behavioral phenotype component is maintained in the final proposal. NIA and NIH will need to ensure that the harmonization efforts they fund are compatible with efforts funded elsewhere, and they will have to consider how to avoid supporting duplicative efforts. NIA will need further input from workshop participants and others in the research community as it moves forward.

Workshop participants suggested several other steps for NIA to consider:

- Build upon the enormous work (including that not funded by BSR) that already has been conducted on a range of relevant activities (e.g., cataloging study holdings at NACDA, and RAND) to develop a suite of tools to facilitate harmonization (e.g., P3G harmonization platform), minimize duplicative effort on the part of investigators, and prioritize support for multiple complementary activities.
- Define research questions around which harmonization efforts could be organized, and encourage researchers from different disciplines to discuss and perhaps agree on the best way to organize variables and measures related to a domain.
- Organize workshops and conversations with other Institutes and Centers (ICs) to understand what harmonization efforts are under way at NIH and across HHS, as well as workshops where investigators can learn how to think about harmonization efforts and conduct longitudinal studies in the social and behavioral sciences. The experiences of the National Human Genome Research Institute (NHGRI), which has assessed ways to combine and harmonize genome-wide association studies (GWAS), can be used as a model for these conversations.
- Encourage collaborations among various groups that have conducted GWAS, now that new methods have been developed and chips are less expensive.
- Assemble study sections and educate them on the value of calibration studies and harmonize existing longitudinal studies that have genetic data.
- Encourage consensus on domain dimensions, or at least outline areas of disagreement, for example through domain profiles published in a journal such as the *Journal of Epidemiology*. Clinicians should be included in consensus development.
- Review what has been learned so far from harmonization efforts in the psychosocial sciences.
- Support hands-on workshops in which participants learn about available harmonization tools.

## Emerging Themes

- There is large agreement on the need for harmonization.
- Harmonization is not the same as homogenization; it is important for studies to maintain their unique focus and interests. Thus harmonization is intended not only to enhance original studies, but also to go beyond those studies in a systematic, planned way, using methods and tools that have already been developed. Harmonization should be viewed as an approach that can provide insight into critical issues about existing measures, both in ways to facilitate cross-study analyses and to identify the measures that can capture critical domains.

- NIA can support development of infrastructure and a core set of tools, rather than separate silos.
- Despite the development of new tools and technology, there will always be a need for experts to analyze data and to inform harmonization efforts, tool development, archiving, and mapping, crosswalk, and calibration studies.
- It is important to consider both long- and short-term goals in unison when building a harmonization strategy. In the short term, it is clear that through the work of several investigators, there are areas that are ripe for harmonization. In the long term, it will be important to create a useful resource for genetic studies in the behavioral and social sciences in a holistic and dynamic way so that new phenotypic areas and/or new studies can be added.
- A harmonization strategy should include a way to systematically catalogue and archive how measures have been harmonized.
- A science of harmonization and associated methodology should be developed.
- Efforts should be made to identify and address gaps in the harmonization process.
- The harmonization strategy needs to be sustainable because it is labor and resource intensive; the agenda needs to go beyond the life of a particular grant.

## BSR's Needs

The following needs were highlighted by Dr. Suzman and BSR staff:

- Development of long and short term strategies and prioritization of areas for harmonization.
- It is important to build upon what has already been done and reduce duplicative or overlapping efforts.
- Feedback from the research community on how best to support harmonization efforts, for example through competitive supplements or the establishment of special study sections.
- A formal way to identify research questions that are of interest to or demanded by consumers and could be answered by several datasets covering different cohorts.
- A list of harmonization studies NIA could support within the next 5 years.
- A way to harness the smaller NIA-supported studies that offer deeper phenotypes, biological specimens, and samples to genotype.
- Continued inclusion of longitudinal twin studies in harmonization efforts.
- Intensive modeling of phenotypes.
- A way to identify calibration data that already exist and when new calibration studies are needed, as well as a way for investigators in the NIA portfolio to learn about existing resources.

## Potential Topics or Phenotypes for Data Harmonization

The following topics or phenotypes were suggested by workshop participants, with an emphasis on a combined focus on domain and harmonization. A forum to allow domain researchers to link with harmonization experts was suggested.

- Wealth
- Correlations of conscientiousness with health and wealth

- Diagnosis and subcategorization of highly prevalent chronic degenerative diseases such as Alzheimer's disease
- Well-being
- Time use
- Cognitive comparisons across developed and less developed countries
- Selfishness versus altruism
- Social stressors
- Areas highlighted at the USC/UCLA meeting:
    o Psychosocial measures
        - Well-being
        - Personality
        - Depressive symptoms
        - Stress
    o Health outcomes
        - Heart disease
        - Cancer
        - Diabetes
        - Physical functioning
        - Cognitive functioning
        - Behaviors
        - Drinking
        - Smoking
        - Risk-taking

# APPENDIX 1: STATEMENT OF WORKSHOP PURPOSE AND BACKGROUND

## *Purpose*

The purpose of this workshop is to explore and discuss harmonization strategies that will maximize the value and use of data within the behavioral and social sciences and accelerate research integrating these data with genetic and genomic inquiry. Approaches and best solutions developed through activities funded by BSR (RAND Survey Meta Data Repository and The National Archives of Computerized Data on Aging) and other harmonization initiatives, such as PhenX and P$^3$G, will be presented and discussed.[4] The workshop will congregate a small group of harmonization experts and PIs. It will draw on case scenarios to focus discussions on approaches for retrospective and prospective phenotype harmonization, cataloguing studies and technologies that enable data sharing. Specific steps to achieve harmonization will be described and mapped to tools and resources that have been developed to help meet these harmonization steps. Discussions identify gaps, complementarities of efforts and further needs in order to establish a harmonization platform. The information presented and workshop discussions will provide valuable information for defining the scope, next steps and priorities for a unified harmonization strategy that will promote research within the BSR portfolio.

## *Background*

DBSR has invested substantially in large, longitudinal studies (i.e. *HRS, MIDUS, WLS, ELSA, including* twin research such as *The Swedish Twin Study on Adults, The Danish Twin Study, and VETSA).* These projects are among the world's richest sources of information about social and behavioral influences, aging and health; many of these projects collect biospecimens and it is becoming more and more common to collect DNA and analyze (1). Individually, these studies have contributed tremendously to the behavioral, social, and economic knowledge base on healthy aging; collectively, they hold enormous potential to accelerate genetic research within these fields. However, harnessing this potential and optimizing the scientific and financial return on the investments already made requires strategic harmonization efforts.

The ultimate goal is to develop a unique research resource that would enhance and promote research in BSS-G through: maximizing the use of existing and new data, enabling more efficient planning of BSS-G studies, attracting new researchers to BSS data, enabling novel research that exploits genomic tools and resources in conjunction with the BSS data, and stimulate greater integration of BSS data into biomedical studies.

Harmonization is pivotal to achieving these aims for several reasons. First, genetic investigation of behavior and research into gene-environment interactions require access to vast collections of genotypic data linked with social science data spanning demographic, behavioral, environmental, economic, and lifestyle factors. However, even the largest of most such collections is demonstrably too small (2-4) to provide the statistical power needed to detect effects reliably. Thus, from both scientific and economic perspectives the most effective and cost-efficient solution would be to pull together and harmonize data from the growing array of high-quality behavioral and social research projects that have (or plan to have) genotyped data or DNA.

Second, the 'phenotype to genotype' model will increasingly be replaced or complemented by a 'genotype to phenotype model' that relies on genetic knowledge and/or biological profiles to classify

---

[4] Public Population Project in Genomics, P$^3$G (http://www.p3g.org)

individuals into groups for study. This approach will be essential to help resolve issues of etiological heterogeneity. The optimal design and conduct of genetic research on complex traits will therefore hinge upon access to information about the availability of specific sets of data (genotypes, biomarkers and BSS measures) that are catalogued using a harmonized and searchable format. This would enable identification of highly targeted sub-samples while simultaneously providing needed information on available sample sizes.

## Harmonization Strategy

Critically, the concept of harmonization pertains to a range of harmonization components that are essential to achieve interoperability and data integration. Several of these harmonization elements are described in the recent NRC volume *Conducting Biosocial Surveys* (5). Examples include cataloguing, phenotype harmonization, evidence-based biospecimen handling, and a host of enabling technologies involving compatible informatics, secure data handling, consenting processes, and the development of meta-information systems to describe studies and data availability.

While not all of these harmonization elements must be tackled in order to enable the research, it is important to understand the full scope of harmonization in order to set priorities and develop a strategy that can be expanded upon in a useful way. One can imagine a scenario where phenotypes have been harmonized but data can't be shared because the IT systems for coding and secure exchange/use of the data or the ethical frameworks enabling data sharing are not in place. Furthermore, the scope of a specific harmonization strategy is context dependent, and may vary considerably depending on the overall purpose for harmonizing. For example, the strategy adopted by a small consortium of studies targeting a specific illness may have a much smaller scope than that adopted by an NIH institute setting up a long range plan for use of data and biosamples generated through multiple research portfolios. This workshop will provide an overview of harmonization and then focus on the most important issues for moving forward with the BSS-G agenda.

## A Unified Plan

There is a growing call within BSS-G for phenotype harmonization as witnessed by the increasing number of PI initiated harmonization activities [i.e. IALSA (Hofer), wellbeing (Smith & Steptoe), The HRS Family of Surveys, Several other BSS resources supported by BSR could also become integral components of an overall harmonization plan to serve BSS. These include BSR-funded data archives (i.e. NACDA, IPCSR, and the RAND Survey Meta Data Repository) and the NIA Population Studies Database. The proposed workshop will be invaluable for BSR to determine the best ways to build upon these interests and initiatives, avoid duplication of effort, and develop a unified harmonization plan to promote the BSR genetic research portfolio.

Jennifer R. Harris, PhD
Consultant to BSR, NIA

---

**APPENDIX 1: Statement of Workshop Purpose and Background**

**References**

1. Harris, J.R., Gruenewald, T., & Seeman, T., (2008). An Overview of Biomarker Research from Community and Population-Based Studies on Aging. (pp. 96-135). In M. Weinstein, J.W. Vaupal, & K.W. Wachter (Eds.) Biosocial Surveys, Washington, DC: National Academies Press.

2. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, & Elliott P. (2009). Size matters: just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol*. 2009 Feb; 38(1):263-73.

2. Wong MY, Day NE, Luan JA, Chan KP, Wareham NJ. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement*? Int J Epidemiol* 2003; 32(1):51-7.

4. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* 2009; 5(5).

5. Hauser RM, Pool R, & Cohen B (EDS). (2010). Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata. National Research Council, The National Academies Press, Washington, DC.

**APPENDIX 1: Statement of Workshop Purpose and Background**

# APPENDIX 2: PARTICIPANT ROSTER

**Paul Burton, PhD**
Professor of Genetic Epidemiology
Department of Health Sciences &
     Department of Genetics
University of Leicester
Email: pb51@le.ac.uk

**Eileen Crimmins, PhD**
AARP Chair
Davis School of Gerontology
University of Southern California
Email: crimmin@usc.edu

**Isabel Fortier, PhD**
Director
Research & Development for P$^3$G
     Consortium
McGill University Health Centre
Email: ifortier@p3g.org

**Margaret Gatz, PhD**
Professor of Psychology, Gerontology, and
     Preventive Medicine
University of Southern California
Email: gatz@usc.edu

**Robert Hauser, PhD**
Executive Director
Division of Behavioral and Social Sciences
     and Education
National Research Council
Email: Rhauser@nas.edu

**Taissa (Tess) Hauser**
Wisconsin Longitudinal Study
Email: hausert@ssc.wisc.edu

**Scott Hofer, PhD**
Professor & Research Chair
Adult Development & Aging
University of Victoria, Canada
Email: smhofer@uvic.ca

**Jinkook Lee, PhD**
Senior Economist
RAND Corporation
Email: jinkook@rand.org

**John McArdle, PhD**
Professor, Psychology & Gerontology
University of Southern California
Email: jmcardle@usc.edu

**James McNally, PhD**
Director
NACDA Program on Aging
Inter-University Consortium for Political &
     Social Research
Institute for Social Research
University of Michigan
Email: jmcnally@umich.edu

**Randall Jay Olsen, PhD**
Professor, Department of Economics
Director, Center for Human Resource
     Research
PI, National Longitudinal Surveys
Ohio State University
Email: olsen.6@osu.edu

**Chandra A. Reynolds, PhD**
Professor, Department of Psychology
University of California Riverside
Email: chandra.reynolds@ucr.edu

**Teresa Seeman, PhD**
Professor, Medicine & Epidemiology
David Geffen School of Medicine at UCLA
Email: tseeman@mednet.ucla.edu

**Albert "Bas" Weerman, MA**
Senior Information Systems Analyst
RAND Corporation
Email: Albert_Weerman@rand.org

**David Weir, PhD**
Research Professor and Director
Health and Retirement Study
University of Michigan
Email: dweir@isr.umich.edu

## NIA STAFF

**Farheen Akbar, MPH**
Research Program Analyst
Individual Behavioral Processes Branch
National Institute on Aging
Email: Farheen.Akbar@nih.gov

**Partha Bhattacharyya, PhD**
Health Program Specialist
Population & Social Processes Branch
National Institute on Aging
Email:BhattacharyyaP@nia.nih.gov

**John G. Haaga, PhD**
Deputy Director
Division of Behavioral & Social Research
National Institute on Aging
Email: HaagaJ@mail.nih.gov

**Jennifer Harris, PhD**
Consultant
Division of Behavioral & Social Research
National Institute on Aging
Email: HarrisJe@mail.nih.gov

**Tamara B. Harris, MD, MS**
Senior Investigator and Chief
Geriatric Epidemiology Section
Laboratory of Epidemiology, Demography,
    and Biometry
National Institute on Aging
Email: harrista@nia.nih.gov

**Chyren Hunter, PhD**
Deputy Director, Division of Extramural
    Activities
National Institute on Aging
Email: hunter@nia.nih.gov

**Jonathan W. King, PhD**
Program Director, Cognitive Aging
Individual Behavioral Processes Branch
National Institute on Aging
Email: kingjo@nia.nih.gov

**Marilyn Miller, PhD**
Program Director, Genetics of Alzheimer's
    Disease, Tau, and Hormone Research
National Institute on Aging
Email: millerm@nia.nih.gov

**Lisbeth Nielsen, PhD**
Program Director, Psychological
Development & Integrative Science
National Institute on Aging
Email: nielsenli@nia.nih.gov

**Georgeanne Patmios, MA**
Health Program Specialist
Division of Behavioral & Social Research
National Institute on Aging
Email: PatmiosG@mail.nih.gov

**John Phillips, PhD**
Labor Economist
Division of Behavioral & Social Research
National Institute on Aging
Email: john.phillips@nih.gov

**Winifred K. Rossi, MA**
Deputy Director, Division of Geriatrics and
    Clinical Gerontology
National Institute on Aging
Email: Winnie_rossi@nih.gov

**Erica Spotts, PhD**
Program Director
Division of Behavioral & Social Research
National Institute on Aging
Email: SpottsE@mail.nih.gov

**Richard Suzman, PhD**
Director, Division of Behavioral & Social
Research
National Institute on Aging
Email: SuzmanR@nia.nih.gov

**Barbara Torrey, MS**
Guest Researcher
Division of Behavioral & Social Research
National Institute on Aging
Email:Barbara.Torrey@nih.gov

## NICHD STAFF

**Regina Bures, PhD**
Program Official, Demographic &
    Behavioral Sciences Branch
National Institute of Child Health & Human
    Development
Email: regina.bures@nih.gov

## NHGRI STAFF

**Heather A. Junkins, MS**
Health Science Analyst
Office of Population Genomics
National Human Genome Research Institute
Email: junkinsh@mail.nih.gov

**Erin M. Ramos, PhD, MPH**
Epidemiologist
Office of Population Genomics
National Human Genome Research Institute
Email: ramoser@mail.nih.gov

## CONTRACTOR STAFF

**Rose Maria Li, MBA, PhD**
Managing Editor
Rose Li and Associates, Inc.
Email: rose@roseliassociates.com

**Frances McFarland-Horne, MA, PhD**
Senior Science Writer
Rose Li and Associates, Inc.
Email: fmhorne@roseliassociates.com

# APPENDIX 3: WORKSHOP AGENDA

## Harmonization Strategies for
## Behavioral, Social Science, and Genetic Research
### November 29–30, 2011

Gateway Building, 7201 Wisconsin Avenue, Bethesda, MD, Suite525C

| Nov 29 | Session | Speaker |
|---|---|---|
| **1.** | **WHAT ARE OUR GOALS FOR THE NEXT 26 HOURS?** | Jennifer Harris |
| 1:00 | Welcome | Richard Suzman, *BSR, NIA* |
| 1:15 | Purpose of the Workshop and Introductions | Jennifer Harris, *BSR, NIA* |
| **2.** | **WHAT DOES HARMONIZATION MEAN FOR THOSE STILL SINGING A SINGLE NOTE?** | Jennifer Harris, Chair |
| 1:30 | Setting the scene | Paul Burton, *University of Leicester* |
| 1:50 | Key Concepts and Practical Steps to Phenotype Harmonization | Isabel Fortier, *McGill University Health Centre and P³G Consortium* |
| 2:10 | Harmonization Considerations: What Is Unique to the Behavioral and Social Sciences? | Arie Kapteyn, *RAND Corporation (CANCELLED)* |
| 2:30 | *Discussion* | |
| 3:15 | Coffee Break | |
| **3.** | **WHAT IS THE STATUS OF PHENOTYPE HARMONIZATION IN THE BSR PORTFOLIO?** | Robert Hauser, Chair |
| 3:30 | Harmonization Needs in Behavioral/Social Sciences: Insights from the USC/UCLA Meeting on Harmonization of Methods and Measures in Longitudinal Studies (May 2011) | Eileen Crimmins, *University of Southern California, Davis* |
| 3:50 | The HRS Family of Surveys | David Weir, *University of Michigan* |
| 4:10 | Integrative Analysis of Longitudinal Studies on Aging (IALSA): Challenges and Needs for Quantitative Harmonization | Scott Hofer, *University of Victoria* |
| 4:30 | Twin Study Harmonization: Experience from "Gene-Environment Interplay of Social Contexts and Aging-Related Outcomes" | Chandra A. Reynolds, *University of California, Riverside* Margaret Gatz, *University of Southern California* |
| 4:50 | Biomarker Harmonization | Teresa Seeman, *University of California, Los Angeles* |
| 5:10 | *Discussion* | |
| 6:00 | ADJOURN FOR DAY | |
| 7:00 | Dinner at Shangri La Indian & Nepalese Cuisine* | |

*7345 A Wisconsin Ave, Bethesda, MD20814, USA
(301) 656-4444

# NIA WORKSHOP AGENDA (cont'd)

| Nov 30 | Session | Speaker |
|---|---|---|
| **4.** | **HOW DO WE BUILD ON WHAT HAS ALREADY BEEN DONE?** | Barbara Torrey, Chair |
| 8:00 | The National Archive of Computerized Data on Aging (NACDA) | James McNally, *Director, NACDA Program on Aging* |
| 8:30 | RAND Survey Meta Data Repository | Jinkook Lee and Bas Weerman, *RAND Corporation* |
| 9:00 | PhenX | Erin Ramos, *National Human Genome Research Institute* |
| 9:30 | International Harmonization Platform (P$^3$G) | Isabel Fortier, *McGill University Health Centre and P$^3$G Consortium* |
| 10:00 | *Discussion* | |
| 10:45 | Coffee Break | |
| **5.** | **WHAT IMPORTANT ENABLING TECHNOLOGIES AND APPROACHES DO WE NEED?** | Arie Kapteyn, Chair |
| 11:00 | DataSHIELD | Paul Burton, *University of Leicester* |
| 11:30 | Calibration Methods for the Harmonization of Quantitative Traits | John McArdle, *University of Southern California* |
| 12:00 | *Discussion* | |
| 12:30 | Lunch (brought in) | |
| **6.** | **WHERE DO WE GO FROM HERE?** | Paul Burton, Chair |
| 1:15 | BSR's Immediate Issues | Jon King, *BSR, NIA*<br>Erica Spotts, *BSR, NIA*<br>Lisbeth Nielsen, *BSR, NIA*<br>John Phillips, *BSR , NIA* |
| 2:00 | Discussion: Integration and Next Steps | |
| 2:45 | Conclusion and Wrap Up | Richard Suzman and Jennifer Harris |
| 3:00 | MEETING ADJOURNS | |

# APPENDIX 4: STEPS OF HARMONIZATION

**Define the research question(s)**

**Prospective Harmonization**

**Retrospective Harmonization**

Achieve consensus on compatible study designs, measures, and collection procedures

Catalogue study characteristics and database content

Identify common variables of interest

Ensure quality and consistency of common data collection

Evaluate harmonization potential of the participant studies

**Process study-specific data under a common data format and achieve quality control analysis**

**Integrate and, if relevant, transfer harmonized data**

**Achieve data analysis**