Psychological Methods
2009, Vol. 14, No. 2, 150–164

1082-989X/09/$12.00 DOI: 10.1037/a0015566

Integrative Data Analysis Through Coordination of Measurement and Analysis Protocol Across Independent Longitudinal Studies

Scott M. Hofer and Andrea M. Piccinin
Oregon State University

Replication of research findings across independent longitudinal studies is essential for a cumulative and innovative developmental science. Meta-analysis of longitudinal studies is often limited by the amount of published information on particular research questions, the complexity of longitudinal designs and the sophistication of analyses, and practical limits on full reporting of results. In many cases, cross-study differences in sample composition and measurements impede or lessen the utility of pooled data analysis. A collaborative, coordinated analysis approach can provide a broad foundation for cumulating scientific knowledge by facilitating efficient analysis of multiple studies in ways that maximize comparability of results and permit evaluation of study differences. The goal of such an approach is to maximize opportunities for replication and extension of findings across longitudinal studies through open access to analysis scripts and output for published results, permitting modification, evaluation, and extension of alternative statistical models and application to additional data sets. Drawing on the cognitive aging literature as an example, the authors articulate some of the challenges of meta-analytic and pooled-data approaches and introduce a coordinated analysis approach as an important avenue for maximizing the comparability, replication, and extension of results from longitudinal studies.

Keywords: longitudinal, integrative data analysis, meta-analysis, data pooling, longitudinal studies

Scientific progress in understanding developmental and aging processes is optimally based on the evaluation and extension of theoretical and empirical findings from within-person data. It is well understood that cross-sectional designs rely on untenable assumptions and are fundamentally limited for understanding individual-level change processes (Hofer, Flaherty, & Hoffman, 2006; Hofer & Sliwinski, 2001; Kraemer, Yesavage, Taylor, & Kupfer, 2000; Molenaar, Huizenga, & Nesselroade, 2003; Wohlwill, 1973). Longitudinal designs provide the best basis for describing patterns of change and for understanding the interdependency among developmental and aging-related processes and influences of risk and protective factors across the life span.

Integrative Research on Longitudinal Studies of Development and Aging

Remarkable national and international efforts have produced numerous longitudinal studies of developmental and aging-related processes. Although longitudinal information is time and effort intensive to collect, it is required to address central questions in developmental research relating to intraindividual change and variation and, particularly important for research on aging, for inference to defined populations conditional on attrition and mortality. Given the profound investment of time, energy, and funding that these studies require, it is not uncommon for them to be multidisciplinary in nature. Existing longitudinal studies, therefore, represent an enormous wealth of information on within-person changes in a variety of domains, including cognition, health, personality, affect, lifestyle, and well-being. These studies have already provided important information and permit further opportunities for describing and explaining developmental and aging-related changes and cross-process dynamics, as well as for identifying influential factors associated with early and late life outcomes.

Scott M. Hofer and Andrea M. Piccinin, Department of Human Development and Family Sciences, Oregon State University.

This article and the Integrative Analysis of Longitudinal Studies of Aging (IALSA) research network were supported by Grant 1R01AG026453 from the National Institute on Aging, National Institutes of Health. We acknowledge the contributions of Duane Alwin, Daniel Bontempo, Lesa Hoffman, Mike Martin, Martin Sliwinski, Avron Spiro III, and the collaborating IALSA members for their efforts in the development of the network.

Correspondence concerning this article should be addressed to Scott M. Hofer, who is now at the Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, British Columbia V8W 3P5, Canada. E-mail: smhofer@uvic.ca

Relative to research reports from cross-sectional age-comparative studies, accumulation of knowledge and development of theory from a within-person perspective has progressed slowly. Given the requirements of data collection in longitudinal research, long intervals often pass between within-person findings and opportunities for replication. Aggravating this slow process are the differences across studies in measures, samples, design characteristics, and statistical analyses, which limit direct comparison of study results (Freese, 2007; Tooth, Ware, Bain, Purdie, & Dobson, 2005). In particular, variation in statistical analysis and evaluation of particular models with restricted reporting of results make direct comparison of findings difficult. The diversity of research interests relative to the number of longitudinal studies has also led to somewhat unique analyses and specific statistical models, which have not yet been evaluated in other relevant data sets. Consequently, there is currently little basis for evaluating results from longitudinal studies of aging within a meta-analytic framework. Nevertheless, one of the clearest next steps in the developmental aging field is the evaluation, confirmation, and extension of theoretical and empirical findings in available within-person data.

Numerous calls have been made for increased interdisciplinary, international, and collaborative efforts as a means to focus developmental research on within-person processes (Bachrach & Abeles, 2004; Butz & Torrey, 2006; National Research Council, 2000, 2001a, 2001b). The use of existing data on within-person change (and between-person differences in within-person change) is one powerful way to evaluate and extend current theory and hypotheses that have been developed primarily from a cross-sectional, between-person comparison perspective.

Replication in the Context of Longitudinal Studies

Replication of research findings across independent longitudinal studies is essential for a cumulative and innovative developmental science. Researchers use extant scientific evidence to structure, justify, and extend research and to develop theory, and they may often base decisions on one or a few reports. Replication of results from longitudinal studies is necessary to protect against Type I errors and uncritical acceptance of empirical findings and to clarify the sensitivity of results to measurement, design, and statistical model decisions.

Research findings and conclusions often vary across independent studies. Certainly, no one study can measure and control for all extraneous influences, particularly when results may be influenced by differences in birth cohort or culture. However, in many cases, differences in the statistical analysis and presentation of results make comparisons across studies ambiguous. In general, this between-study variability points to the need for skepticism regarding a single instance of a result and to the importance of multiple replications in the evaluation of scientific findings. Replication is essential for scientific progress—replication once is good, but replication multiple times is better, because results are usually not as straightforward as they might first appear (e.g., Hendrick, 1990; Lindsay & Ehrenberg, 1993; Lykken, 1968; C. L. Park, 2004; Rosenbaum, 2001; Wilkinson & Task Force on Statistical Inference, 1999).

Lykken (1968) described different types of replication. Literal replication involves the exact duplication of sampling procedure, conditions, measurement, and analysis methods. Operational replication involves duplication of the minimal essential conditions, such as sampling, measurements, or experimental conditions. In the longitudinal study context, this can also apply to the use of similar statistical models and analysis procedures across studies. Constructive replication, most pertinent to long-term longitudinal studies, provides a broad test of validity of methods and approaches in that research findings should generally hold across studies that implement different samples, measures, and designs. Except in relatively rare instances, longitudinal observational studies differ from one another in many ways and provide few opportunities for exact or literal replication (except within certain countries or multiple-cohort designs). Measurement differences can be magnified in cross-cultural or cross-national data, where variation is inevitably introduced because of differences in language, administration, and item relevance (i.e., culture). These differences, however, can be a strength for constructive replication opportunities in the longitudinal context, permitting evaluation of the generalizability of research findings across independent samples, measures, and designs.

A number of analysis strategies permit evaluation of the replicability and generalizability of results. At one end of this continuum is sequential independent replication. This is science as it is typically performed, where the published results of a particular study are replicated in an independent data set. For observational studies in particular, there can be a broad range of how similar the sample, context, measurement, design, and statistical analysis are to the original study, and it is important to take these into consideration when comparing results across studies (e.g., Van Dijk, Van Gerven, Van Boxtel, Van der Elst, & Jolles, 2008).

The next level involves meta-analysis (e.g., Cooper & Hedges, 1994; Sutton & Higgins, 2008) of the existing literature, which combines standardized effects from a set of published findings to estimate the general effect and to understand why studies differ in their results. Meta-analysis relies on assumptions regarding the comparability of research results across studies but permits assessment of study-level characteristics affecting the pattern of results.

A third level includes methods for combining individual-level data sets within a simultaneous analysis, which is known as data pooling (i.e., integrative data analysis; Curran & Hussong, 2009), pooled data meta-analysis (also known as individual participant meta-analysis; Cooper & Patall, 2009), and mega-analysis (McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). These methods permit evaluation of both study-level and individual-level effects (Smith, Williamson & Marson, 2005a, 2005b; Stewart & Parmar, 1993; Thompson & Sharp, 1999) and have been used very effectively in a variety of substantive areas and types of data.

Finally, there are methods that permit questions that go beyond what can be learned from any particular data set. Generalized evidence synthesis (Ades & Sutton, 2006; Spiegelhalter, Abrams, & Myles, 2004; Spiegelhalter & Best, 2003) and data fusion provide a means of combining data from multiple sources for the analysis of models that cannot be evaluated in any single data source.

An alternative, and the primary focus of this article, is coordinated analysis with replication, the collaborative analysis of multiple independent data sets in ways that optimize comparison of results across studies. The aim of this approach is to maximize the data value from each study while making results as comparable as possible by coordinating measurement and statistical analysis protocol across studies. This does not preclude the evaluation of alternative models and the extension of models in particular data sets but focuses on maximizing opportunities for direct comparison of results. Results from such coordinated analyses can potentially be summarized by a multilevel meta-analysis for the evaluation of differences across studies related to sample composition and other study characteristics.

Comparing results across longitudinal studies can present a number of challenges. In a cross-national context, a key issue is the comparability of outcomes and covariates based on different measurement instruments that may differ in language, difficulty, number of items, and range of measurement. The difficulty in making direct comparisons of effects of or on these measures is that there is no natural metric on which to scale these effects. This is further compounded by differences in sample composition, including differences in birth cohort, culture, and social system. In the next section, we briefly describe some of the many potential differences across samples, measures, and designs that can have an effect on cross-study comparison. We then discuss the current potential for meta-analysis or pooling specific to within person analysis of longitudinal data on aging, introduce a research model for the coordinated analysis of such data, and summarize the benefits of coordinated analysis of longitudinal studies on aging.

Sources of Heterogeneity Within a Cross-National Longitudinal Study Context

Differences across long-term longitudinal studies can be seen as an impediment to the direct cross-study comparison that is essential for gauging the generalizability of results. Replication of findings from longitudinal studies is often not straightforward and requires special treatment given the variety of complex design and analysis approaches as well as differences across studies in terms of samples (e.g., birth cohort, culture), time (e.g., differing assessment intervals, retest effects), and measures (e.g., reliability, sensitivity, language).

However, the variety of samples, measurements, contexts, and research designs, particularly in the area of longitudinal aging research, is also an advantage for replication of research findings, referred to as generalized causal inference (Shadish, Cook, & Campbell, 2001). Understanding the generalizability of results requires that research be replicated across a representative range of samples and contexts to which the findings would be expected to generalize. A thorough treatment of any particular research question, therefore, might require a range of strategies to detect the sensitivity of a finding to the conditions under which it is found, including the use of different indicators of the same construct, different populations, and different research designs. In the sections that follow, we outline some of the important differences across longitudinal studies that represent both challenges and opportunities for identifying and understanding systematic developmental and aging-related processes. It is important to explicitly address these often ignored differences, both in cross-study analysis and in a general review of previous findings.

Sample Characteristics

Population representativeness, birth cohort, socioeconomic, racial–ethnic, educational, and cross-national differences are important to consider when interpreting and comparing scientific findings on developmental, aging, and health processes.

Population Representativeness

Population representativeness is, of course, critical for making inferences to defined populations. However, participation in longitudinal studies (initial and ongoing) is demanding, and population inference—even in studies that are among the most rigorous in terms of initial sample representativeness—is limited by selectivity at the first occasion of measurement and by subsequent attrition and mortality selection (e.g., Hofer & Hoffman, 2007). Although longitudinal studies differ in modes of population representativeness, this does not necessarily limit what can be learned about basic psychological processes, particularly if results are generalizable (i.e., systematic) across studies differing in sampling characteristics. Inclusion of variables in the statistical analysis that account for population heterogeneity (i.e., stratification, composition) may, in some cases, serve to adjust for differences across samples and permit a stronger basis for comparison.

Birth Cohort

In studying contemporary cohorts of older adults, researchers must also be sensitive to matters of historical location. Cohorts born early in the 20th century have experienced dramatic and rapid changes in their lifetimes and have had significant experience with war, in particular. These experiences may be critical but largely hidden variables that lie beneath much scientific knowledge about aging. In addition, there is evidence for differing effects of mortality selection across birth cohorts (e.g., Janssen, Peeters, Mackenbach, & Kunst, 2005).

Several major longitudinal studies obtained multiple sequential cohort samples to permit comparisons across birth cohorts, cross-sectionally and longitudinally within studies (e.g., Schaie, 1965). Others, such as the Gothenburg H-70 study, have focused on a single cohort. Most longitudinal studies, however, are comprised of samples heterogeneous in terms of age/cohort at the first occasion. Comparison of results across longitudinal studies usually involves comparison of populations differing in average birth cohort, having experienced unique historical contexts and changes, such as educational experiences and health care. Such comparisons are important for understanding human development broadly and can be expected to remain important for comparison with future studies for understanding broad contextual differences and historical shifts that affect developmental and aging outcomes.

Nationality/Culture

Numerous longitudinal studies are available from Australia, North America, and Europe and are increasing in number elsewhere in the world. Differences in social welfare policies and programs as well as other macrosocial influences, even within Western societies, may have significant effects on developmental, aging, and health-related outcomes.

Socioeconomic Status

Education, occupational status, and income, the most widely measured dimensions of socioeconomic status, are often moderately correlated but not interchangeable, so cross-study work should be based on the same dimension. In addition, the meaning of these variables can differ considerably across time and place. Educational attainment, in years or credentials, varies a great deal across birth cohorts, with significantly more widespread completion of secondary and postsecondary education in recent decades. Data on occupational position (highest achieved, longest held, final, or current) and status are often precoded into broad categories, which may be difficult to reconcile across studies. Although direct use of raw individual or household income values would likely be problematic, it may be possible to generate societal-level measures of income inequality, which has been linked to a range of health-related outcomes.

These dimensions of socioeconomic status are important to consider for explaining results within and across longitudinal studies. In the area of cognitive aging, for example, education is sometimes used as a proxy for *cognitive reserve*, with research focused on whether higher levels of schooling act as a protective factor in cognitive aging and dementia by retarding the rate of change in cognitive decline and, therefore, acting to buffer the processes of normal aging (e.g., see Anstey & Christensen, 2000; Dufouil, Alpérovitch, & Tzourio, 2003; Stern et al., 1994). The results of this body of research are mixed: Some studies have shown no interaction of schooling and rates of change, whereas others have found such a buffering effect. The education variable, however, as a study-level characteristic and an individual differences variable, clearly requires careful treatment in cross-study comparison, because there are marked country and birth cohort differences in educational attainment (Piccinin et al., 2006).

Race/Ethnicity

In societies such as the United States, socioeconomic and racial/ethnic comparisons must be jointly understood given their interactive and interdependent natures (e.g., Anderson, Bulatao, & Cohen, 2004; Manly, 2008). Whitfield and Morgan (2008) emphasized the use of culturally appropriate models and suggested that prior to making such comparisons, it would be wise to understand within-group processes because it may be more informative to study the relevant constellation of mechanisms within each group, rather than assuming that the same factors apply in both.

Selection/Attrition/Mortality

Numerous studies have demonstrated the relatively strong link among age-related outcomes, participant nonresponse, and survival. The mortality selection dynamic cannot be understood by single-occasion sampling of different age groups in which population mortality has already occurred to different degrees and possibly for different reasons. Unlike cross-sectional designs, longitudinal data provide the opportunity to directly address both attrition and mortality selection. This is essential for understanding aging-related changes in psychological and health outcomes (e.g., Harel, Hofer, Hoffman, Pedersen, & Johansson, 2007; Hofer & Hoffman, 2007; Kurland, Johnson, & Diehr, 2007). Comparisons across studies should be sensitive to these selection issues, because differences in interval length between assessments as well as initial sample characteristics, such as age, socioeconomic status, and health, all contribute to the prevalence and impact of missing information on longitudinal results. As existing longitudinal studies mature, modeling these differences becomes more feasible.

Measurement Characteristics

Longitudinal studies, by definition, require repeated assessment of individuals. Particularly for longitudinal studies of aging, samples are often followed over many years and are sometimes criticized for providing only limited knowledge as judged by the current state of biological and psychological measurement. Given ongoing developments in measurement and biological evaluation, current studies and any future longitudinal studies will eventually be considered dated. However, inference regarding within-person change cannot otherwise be obtained, and these trade-offs must be acknowledged and embraced as a fundamental feature of developmental science based on long-term within-person assessments (e.g., Duncan & Kalton, 1987).

Constructs/Measurements

A major step in comparing results across studies involves identifying comparable variables. A wide variety of construct operationalizations can be found in a single nation (e.g., Weiner, Hanley, Clark, & Van Nostrand, 1990), and this variety is magnified in cross-national (or cross-cultural) and longitudinal research. Language and item relevance or meaning are more likely to differ across nations. Longitudinal study investigators must find a balance between maintaining continuity of measurement and keeping up with the latest standards. Inevitably, investigators initiating research in different historical periods tend to implement measures current to that period. Regardless of whether different studies use different variables to identify particular constructs, most studies permit comparison of constructs at the primary factor level, and in some cases, sufficient overlap of items or measures across studies permits factor analysis and the testing of invariance within a pooled data analysis (e.g., Bontempo & Hofer, 2007).

Change in Measurement Over Different Life Periods

It is also often necessary to use different items or measures at different points in the life span to capture relevant aspects of a concept. For example, different intelligence tests are appropriate for children and adults; the meaning of frequent crying in measures of psychopathology changes through childhood, adolescence, and adulthood; work-related questions may be less or not at all relevant following retirement. Curran and Hussong (2009; Curran et al., 2008) and McArdle et al. (2009) made use of item response theory (IRT) methods to address changing items or overlapping sets of items that permit models of change in a common construct over time.

Design and Analysis Characteristics

Assessment Interval

The sampling of time must be carefully considered in the design, analysis, and interpretation of results from longitudinal studies. Typical longitudinal studies have intervals of 6 months to several years between assessments. Different sampling intervals (e.g., minutes, hours, days, weeks, years) will provide windows on different aspects of change processes, producing results that require different interpretations for both within-person and between-person processes (Boker & Nesselroade, 2002; Martin & Hofer, 2004). The measurement interval is also critical for the prediction of outcome variables and the evaluation of leading indicators (Gollob & Reichardt, 1987, 1991). The frequency of assessments may be seen as a nuisance by busy participants or may help to maintain continuity and encourage continued participation. Less frequent intervals decrease the opportunity to identify influences producing change and to understand reasons for participant attrition. For example, in a variety of longitudinal studies with relatively similar per year attrition rates (1%–12%), the actual per interval attrition rates ranged from 3% to 54%. Although the per year attrition rates were more comparable across studies, the impact of the attrition in terms of number of data points available for analysis per participant was more variable and provided less opportunity to model attrition and change processes occurring between assessments.

Retest Effects

Another concern related to measurement intervals is the impact of retest effects, particularly over the first few assessments. Measurements taken closer in time may reasonably be expected to show greater performance increases (or smaller decreases) due to any number of related influences, including warm-up effects, initial anxiety, and test-specific learning of content and strategies. These gains are also likely to be differential, related to ability, age, or task difficulty, and to interact in unknown ways with age-related changes. For example, retest gains in older adults may manifest as a slower decline, rather than greater involvement.

Despite many attempts to separate the effects of age and retest, these effects are tightly confounded in any typical longitudinal study. Decomposition is not possible without strong and untenable assumptions (Thorvaldsson, Hofer, Berg, & Johansson, 2006; Thorvaldsson, Hofer, Hassing, & Johansson, 2008). Measurement burst designs (Nesselroade, 1991) offer a promising alternative for future longitudinal data collection, but analysts of currently available studies and data must continue to wrestle with this issue.

Alternative Models of Time

In addition to differences in the way time has been sampled across studies, there are differences across research reports in the way the time metric has been modeled. Earliest reports used repeated measures ANOVAs or *t* tests, which focused solely on fixed effects. The advent of mixed models opened an entirely new perspective on the analysis of longitudinal data. Early growth models of this type were restricted to using balanced data and hence a time in study metric, but as software has advanced, age-based and then process-based models have emerged. For example, time is often better treated more flexibly and directly in terms of evolving time-dependent processes other than chronological age, such as disease progression (e.g., time before or since diagnosis of dementia; Sliwinski, Hofer, & Hall, 2003; Sliwinski, Hofer, Hall, Bushke, & Lipton, 2003; Sliwinski & Mogle, 2008), measured physiological changes, mortality or years of life remaining (see Thorvaldsson, Hofer, Hassing, & Johansson, 2008), or events such as retirement or widowhood (Alwin, Hofer, & McCammon, 2006), to understand the effects of stress and psychosocial interactions. These more recent models provide a useful perspective for describing and explaining average change and individual variation in change relative to common, possibly causal, processes. They do not, however, easily translate from one metric to another, so reports and conclusions cannot be directly compared across metrics.

Replication of findings from complex analyses is challenging, particularly in the case of longitudinal studies that vary widely in terms of samples, measures, and designs. Indeed, there are often theoretical and empirical reasons (e.g., differences in birth cohort) for differences, and these must be carefully considered when synthesizing research findings.

Feasibility of Comparative Models in the Context of Longitudinal Studies on Cognitive Aging

In the literature on cognitive function in older adulthood, there is currently only a limited basis for synthesizing research findings from within-person designs. In the context of maximizing opportunities for the synthesis of research based on longitudinal studies of aging, we discuss the current potential for meta-analysis of available longitudinal results and pooled data analysis of longitudinal data and then introduce a coordinated analysis approach.

Current Potential for Meta-Analysis of Longitudinal Studies of Cognitive Aging

Meta-analysis was developed to provide a means to statistically evaluate the similarity of results across studies. When the degree of similarity of methods across studies is adequate to justify more stringent comparative strategies, meta-analytic methods are used to summarize findings and to identify and address questions about potential sources of heterogeneity across research findings (e.g., Higgins & Thompson, 2002). Meta-analysis is a powerful tool. It has, however, mainly been used and is most practical with experimental, clinical trial, or intervention data and restricted variable sets that have exact or similar outcome measures and minimal variation in study design.

As discussed earlier, meta-analysis of existing reports from observational longitudinal studies is more challenging and is currently limited in at least two ways. The first limitation stems from the paucity of published information about particular intraindividual focused research questions. For example, in the area of aging-related change in cognitive functioning, replications or comparisons across studies are relatively rare and usually do not permit a strong basis for comparison of major findings. The second factor limiting direct comparison and replication of results is variability across studies in terms of participant sampling and available variables, and this is further complicated by noncomparable statistical models and often idiosyncratic and limited reporting of statistical results (e.g., Freese, 2007; Tooth et al., 2005). For example, H. L. Park, O'Connell, and Thomson (2003), intending to conduct a meta-analysis of cognitive decline in community-based prospective cohort studies with low attrition, whittled 5,990 abstracts down to 19 articles. They then concluded that heterogeneity due to population, country, measure, follow-up (intervals and number), and attrition differences required that they reduce their goal to a narrative review. Coincidentally, with a different research question, Anstey, von Sanden, Salim, and O'Kearney (2007) also found 19 publications with "measures compatible with at least one other article" (p. 5). They proceeded with meta-analyses of the relative risk of four possible outcomes (Alzheimer's disease, vascular dementia, any dementia, and yearly change on MMSE) for smokers and nonsmokers or former smokers, and they based their analyses on subsets of three or four studies at a time with corresponding measures. The limited number of comparable studies in any one category meant that they were unable to investigate sources of heterogeneity. It was also necessary for them to obtain smoking data from the authors of the publications because seven of the studies reported smoking results only incidentally, with the relevant information unavailable in the published articles. A search of the current literature reveals few meta-analyses of longitudinal questions.

Meta-analysis of longitudinal results is further complicated by the variety of decisions made in the design and analysis, as described earlier, and by the conditional nature of the results to these decisions. Additional factors include whether the set of covariates deemed necessary in an analysis is available in all or most studies. In many cases, particular variables are unavailable or have been measured in dissimilar ways. Correspondingly, the tasks of harmonizing the measurement and implementing the meta-analysis on comparable outcomes become more difficult. Certainly, meta-analysis can be performed on diverse measures, but this basis for comparison relies on assumptions regarding measurement equivalence at a broad construct level, the nonequivalence of metrics of outcomes and predictors, and related issues regarding postanalysis standardization decisions (e.g., Becker & Wu, 2007).

Current Potential for Pooled Data Analysis of Longitudinal Studies of Cognitive Aging

There has been long-standing interest in collaboration and pooling of longitudinal study data (e.g., Riegel & Angleitner, 1975; Rose, 1976). Pooled analyses can be implemented to address individual rather than study-level effects or to address questions about subgroups of individuals too small to be studied with adequate power in a single data set. Pooled raw-data analyses, as opposed to pooling of summaries, are required to address questions related to heterogeneity due to both study-level (e.g., design features or inclusion criteria) and individual-level (e.g., education level or age) effects (Stewart & Parmar, 1993). Pooled meta-analyses have been shown superior in terms of determining individual-level effects (e.g., Smith et al., 2005a, 2005b) but to date have been implemented in only relatively restricted circumstances. Observational studies, in particular, differ in sampling and design characteristics that are related to essential questions of internal and external validity; sources of such biases must be accounted for in the model to evaluate their influence on results and in explaining heterogeneity between studies (e.g., Turner, Spiegelhalter, Smith, & Thompson, 2009). When data are identical or sufficiently comparable across studies, pooled analysis of raw data across studies permits the analysis of influences associated with rare events (i.e., evaluation of apolipoprotein E subtypes on cognitive functioning), provides increased power for the detection of associations and interactions, provides more reliable estimates of population-level change, and permits a basis for evaluation of hypotheses regarding sources of mixed findings (e.g., differences in educational attainment) across studies.

Pooled data analysis is a powerful method that can proceed when measurements are identical or can be equated: by fiat, through cocalibration with IRT models, or with latent variable approaches based on item- or scale-level data across studies (see Cooper & Pattell, 2009; McArdle et al., 2009). Unfortunately, for a majority of longitudinal studies of normal cognitive aging, opportunities for pooled data analysis with these methods are limited or may require untenable assumptions. The potential for pooling depends very much on the feasibility of pooling variables that are not operationally defined in the same way. Although it might be possible to use standardized variables (e.g., *T* scores) or proportion correct, this would require assuming that the measurement properties of the variables were relatively comparable and linear—that gains or losses operated in the same way across different measures. For the most part, these have not been determined or evaluated for any of the measures used across studies, so it is hard to predict the impact on a pooled analysis. Another potential inroad here is supplemental data collection in independent samples to permit cocalibration and pooled data analysis (see Curran et al., 2008; McArdle et al., 2009).

A single pooled or mega-analysis may not always provide the best answer to a particular research question, however. A variety of issues should be considered prior to such an undertaking. In the field of cognitive aging, for example, the age, birth cohort, and education ranges of the samples may differ. In the longitudinal context, the interoccasion intervals and number of occasions may differ. Combining data from studies with nonoverlapping age ranges (e.g., 55–70 years vs. 80+ years) can result in study-level differences in outcomes that are confounded with age differences. Extrapolating beyond the data in particular studies requires a too heavy reliance on the assumption that the same processes/associations hold across a wider range than that for which one has evidence in any particular study.

Potential for a Coordinated Replication/Meta-Analysis Approach

One approach for using existing data, without relying solely on publicly available data, is to collaborate on the coordinated analysis of data and synthesis of research findings. A major strength of collaborative, coordinated research, as opposed to use of multiple archived data sets, is that the investigators associated with each study are major partners in the analysis and synthesis of particular research questions, bringing essential substantive expertise related to particular study characteristics. This serves to realize the full potential for maximizing each study's data value while permitting rigorous comparison. Collaborative approaches can accelerate results from longitudinal studies and provide a basis for direct comparison of results across studies, such as meta-analysis.

In many cases, a collaborative, coordinated research approach is optimal for the evaluation and report of both parallel and alternative models on the same data as well as models incorporating individual- and study-level characteristics to account for disparities across studies differing in birth cohort and nationality. A major goal of a coordinated analysis approach is the maximization of opportunities for reproducible research (e.g., Gentleman & Lang, 2007; King, 2007) through open access to analysis scripts and output for published results, permitting quick modification and evaluation of alternative models related to published articles and application of similar models and variable operationalization to other studies. We believe that direct and immediate comparison and contrast of results across independent studies, based on the open availability of analysis protocol, scripts, and results, results in the most solid accumulation of knowledge and is the most powerful way to build developmental science (Piccinin & Hofer, 2008).

Several large-scale collaborations are already in existence, for example, the National Alzheimer's Coordinating Center, the Collaborative Alcohol-Related Longitudinal Project (Fillmore et al., 1988, 1991), the Asia Pacific Cohort Studies Collaborative (APCSC) Group (1999). Examples of smaller scale parallel analyses are also available (e.g., Duncan et al., 2007; Nguyen & Zonderman, 2006). Major benefits of collaborations and parallel analyses can include accelerated accumulation of scientific knowledge, earlier understanding of the stability and generalizability of the findings, and greater statistical power for the study of infrequent events. As Wulf (1993) and the Society of Collaboratories have indicated, an efficient and effective network requires good use of communication and computation technologies, in addition to good personal relations among the investigators. The APCSC, for example, coordinates most correspondence through regular e-mails, but also issues a quarterly newsletter and minutes of the Executive Committee meetings, arranges teleconferences as needed, and maintains a password-protected link on their Web site that gives all collaborators access to APCSC documents.

Such parallel analyses can be conducted independently or can be conducted in a more centralized way by a designated group, and there are advantages to both approaches (Piccinin & Hofer, 2008). Centralized analyses (e.g., Fillmore et al., 1988, 1991; Johnstone et al., 1991; Minicuci et al., 2003) can be more efficient and uniform and can facilitate careful scrutiny of sampling and measurement differences across studies. Coordinated independent analyses (e.g., Duncan et al., 2007; Piccinin et al., 2006; Thorvaldsson et al.'s, 2008, replication of Sliwinski et al., 2006) may provide an evaluation of generalizability more akin to what might typically appear in the literature. As in many situations, a combination of both approaches may be most productive.

Although the cognitive aging literature does not currently contain the information necessary to conduct meta-analyses of within-person questions, it will be possible to take advantage of such methods to evaluate the consistency of findings produced in planned parallel analyses. As in Fillmore et al.'s (1988, 1991) alcohol work, the APCSC's medical research, and Duncan et al.'s (2007) work on school readiness, parallel analysis provides the multiple study data that are necessary to estimate average effect sizes, identify statistically significant heterogeneity in effect size across studies, and evaluate the impact of specific cross-study differences on these inconsistencies.

We are developing a collaborative system for coordinated analysis, evaluation, and communication of results from independent longitudinal studies of aging. Working from the conservative assumption that cross-study sampling, design, and measurement differences often preclude pooling or require more extensive measurement or harmonization work than is feasible or useful, our approach is to primarily make use of parallel independent analyses, using pooled data analysis where applicable. This general approach to understanding key substantive questions makes use of alternative models on the same data as well as meta-analysis incorporating individual- and study-level characteristics to account for disparities across studies differing in birth cohort and nationality. The outcome of this direct and immediate comparison and contrast of results across independent studies, based on open availability of analysis protocol, scripts, and results, is the accumulation of knowledge regarding aging-related processes based on replicated evidence.

A Coordinated Research Model for Integrative Data Analysis

Given the key issue of cross-study comparison, attention to comparability of measurements and statistical models is a critical aspect of a coordinated approach. The evaluation of alternative models on the same data to permit direct comparison of results across models (within and across studies) also aids in the determination of why results might differ. Longitudinal research is challenging, and coordinating analysis across studies is more so given the diversity of study designs, samples, and variables. These challenges are not insurmountable, however, and there is great promise for new collaborations that integrate recent theoretical perspectives for within-person change, developments in statistical analysis of within-person data, and the remarkable number of completed and ongoing longitudinal studies.

A coordinated research model is essentially a system for collaboration. The aims are to enhance communication and collaboration among national and international investigators, to facilitate reproducible research, to archive the analysis and measurement alignment process, to provide a stronger basis for cumulative science based on optimal comparison and replication of results across longitudinal studies, and to permit quick entry into completed analyses, replication in additional studies, and extension of statistical models and substantive hypotheses of within-person change. In the next section, we describe a general research model suitable for analysis of existing data.

Integrative Analysis of Longitudinal Studies on Aging (IALSA): An International Collaborative Research Network

The IALSA research network is a collaborative research infrastructure for coordinated interdisciplinary, cross-national research aimed at the integrative understanding of within-person aging-related changes in health and cognition. The IALSA network currently comprises over 30 longitudinal studies on aging, which span eight countries and have a combined sample size of over 70,000 individuals. These studies include a mix of representative, volunteer, and special population samples (Piccinin & Hofer, 2008). Within the network, data have been collected on individuals from birth to over age 100 (mainly adulthood), with birth cohorts ranging from 1880 to 1980 and with historical periods from 1946 to the present. Between-occasion intervals range from 6 months to 17 years (the majority are 1–5 years), with between 2 and 32 (mainly 3–5) measurement occasions spanning 4 to 48 years of within-person assessment.

IALSA is an open and extensible international network of people, data, and methods collaborating in the analysis and synthesis of existing longitudinal data. Other study investigators may request or be invited to participate on the basis of their expertise and/or the relevance of their data with respect to particular questions of interest.

Overview of Infrastructure

Central to a continued program for coordinated analysis and replication is the establishment of a research network involving key investigators of major longitudinal studies on aging and investigators with experience in longitudinal design and statistical analysis. This vital infrastructure for collaboration facilitates the identification and solution of critical issues in aging research, provides central administration for project management as well as analysis and synthesis of results, and emphasizes broad dissemination of analytical and substantive knowledge to gerontological researchers.

There are numerous ways in which to enhance communication and involvement across research teams. Face-to-face meetings can provide a forum for analysis, dissemination, and discussion of results for current projects and for the development of new projects. Annual research meetings comprised of all network members and project-focused meetings at conferences or other venues facilitate research and encourage further developments. Web-based conferencing provides another form of day-to-day communication among investigators across research sites, augmenting regular teleconferences. Seminar series also provide a structured forum for interaction, training, and communication among the investigators across projects and research sites.

Web site

To support multiple concurrent interactions among investigators across wide geographic distances and time zones, a secure Web site is used for data management (where applicable), progress reports, preliminary results, and statistical analysis scripts, which are available to all investigators. The Web site is used to manage permissions, authorship agreements, and access to data sets that are public and to those with data-sharing agreements. Protocol, annotated statistical analysis scripts (e.g., SAS, SPSS, Stata, Mplus), and the results of such analyses are readily accessible to all investigators, facilitate direct and simultaneous comparison of results across studies, and archive the research process for future use and extension. Although communication technology of this sort is not always critical to the success of a collaborative system, it is clearly facilitative.

Searchable Study-Variable Meta-Database

Identifying studies with sufficient measures for evaluation of specific hypotheses is made possible by access to a searchable database listing the measures used by each study. We have developed such a meta-database that can eventually be linked to study protocol and exact details regarding the particular measurements used. Although relatively few studies have identical measures, especially in the multivariate context, there is great commonality at the primary and secondary factor construct level, and we have made it possible to search at any level of construct across studies.

Data Sharing and Authorship Agreements

All data remain property of the respective longitudinal study primary investigators. Use by others is permitted in the context of a range of general as well as specific data-sharing agreements.

Overview of Research Process

Major strengths of the research process are the coordinated analysis according to protocol, the harmonization of measurement coding and analysis, and the direct comparison of results across studies with the opportunity for immediate evaluation of differences when found and for additional analyses to reconcile such differences.

The research process for the coordinated analysis of longitudinal studies on aging is shown schematically in Figure 1. First, the process begins with a proposed research issue that delineates the problem, briefly cites relevant research, and details preliminary protocol for analysis and structure of results. Second, the searchable database is used to identify studies with targeted variables and characteristics that permit the analysis to be performed. Investigators on these studies are alerted to the proposal and invited to collaborate on developing the protocol in terms of available variables (coding differences) and plans for analysis. Third, preliminary analyses begin with finalizing a protocol for aligning or harmonizing variables, studies, and individual-level covariates and for reporting results. Fourth, analyses are then per-

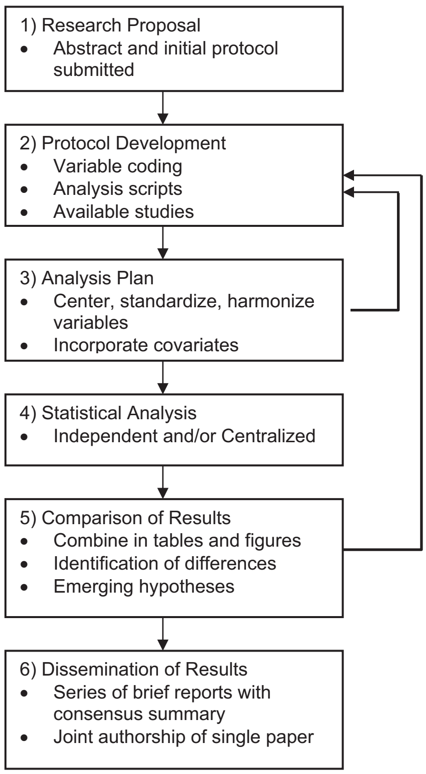

Figure 1. Coordinated research process.

formed independently by each group of researchers and reported in common format. Fifth, results are combined in tables and figures to identify differences and permit the discussion of (a priori or post hoc) alternative models and follow-up analyses; meta-analysis is performed. Sixth, the process is completed by submission for publication of each study's findings and a summary article describing the cross-study comparison and meta-analysis of results.

Research Proposal

Research questions can be proposed by any member of the network. Proposals should include adequate detail for other investigators to decide the appropriateness of their data and their level of interest in participating: a brief background and rationale, a list of dependent and independent variables, and a suggested analytical approach. In some cases, the initiator may already have a completed manuscript or published article to replicate. Using the study-variable meta-database, the proposing investigator identifies the most appropriate potential collaborators and invites them to participate. Project priorities and timelines are determined by the participating investigators.

Protocol Development

Collaborative interactions among research teams lead to more specific decisions regarding the aligning of measurement operations (e.g., data-coding procedures) and to the development of an analysis protocol potentially comprising alternative sets of analyses. The initial steps in variable coding are based on information in the study-variable database. Script development relies primarily on the data set on which the script template is based. To the extent that details of the sample characteristics are known, decisions about coding of the variables and centering of covariates, such as age and education, can be determined at this point, but some of these decisions will have to be modified on the basis of initial analysis of all of the data sets.

For cross-study analysis and comparison, we consider three levels of linkage: broad construct, narrow construct, and identical indicator. Across most studies, broad conceptual replication at the construct level (e.g., comparing different measures of verbal ability across studies) is possible in almost all domains. In many of the studies, replication on more similar variables, for example, comparing memory for different word lists across studies, is possible. On a smaller subset of studies, opportunities are available for direct comparison of identical measures and, in some cases, pooled data analysis.

Extension of the Statistical Analysis Plan

Collaborative interactions across research teams further refine decisions about the aligning of measurement operations (e.g., data-coding procedures) and the analysis protocols comprising alternative sets of analyses.

Measurement operationalization. An important aspect of this step is the evaluation and optimization of available measures for cross-study analyses. Depending on the specific application, in consultation with primary investigators from the affiliated studies, a variety of strategies can be used to maximize comparability of estimates from the affiliated studies and to allow straightforward evaluation of individual-level effects in meta-analytic models. Given the challenges for direct cocalibration of measurements across many longitudinal observational studies, we focus on preanalysis and postanalysis approaches for comparing results on a common metric. Preanalysis approaches range from (a) deciding on a common centering or reference point, standardizing to a common metric (e.g., T scores based on between-person differences at Time 1 or on a reference group with particular characteristics, such as age range and education level), use of proportion correct/endorsed items on instruments of different lengths, or the use of international diagnostic standards to (b) more involved methods, such as the common denominator methods described by Minicuci et al. (2003; Zunzunegui et al., 2006), where commonalities are identified and algorithms or scoring criteria developed to (c) psychometric methods, such as factorial invariance and IRT methods. Almost all of the affiliated studies have collected item-level data that could, in principle, be used as the basis for analysis.

For background variables (i.e., sociodemographic) used in most analyses, an aligning process involving all affiliated studies is being implemented, initially gauging study differences in age, sex, and education. Additional measures of socioeconomic status will be added to this process to the extent possible, though these tend to be measured in a greater variety of ways and are more difficult to reconcile across countries and generations. Although this entails a certain amount of work at the outset, it ensures from the start that the characteristics of all the studies are taken into account in the planning of appropriate comparisons. It also facilitates the inclusion of new or external studies into a comparative framework. Measurement operationalization involving variable coding, centering, and possibly standardization of particular outcomes necessarily involves only those studies with relevant data.

In most cases, analyses are best performed on the raw data from each study. Results across studies can be readily compared on the basis of general conclusions and pattern and statistical significance of results. This basis for comparison is sufficient for scientific progress and is necessary when the congruence across measures indicating a similar construct is low. Summary statistics can be transformed postanalysis to a common metric to permit comparison of effect size within a meta-analytic framework.

Development of analysis scripts. To facilitate implementation of the analyses and to ensure similar processing of the data from the different studies, the initial template developed for the lead study is distributed and modified by each collaborating team as appropriate for their own data. During this process, issues that arise with respect to the appropriateness of the planned operationalization of the included variables can be relayed to the group, which may decide that a change or an extension to the protocol is warranted. For example, the initial protocol for an early project proposed following the lead of other population comparison studies (e.g., Huisman et al., 2004), categorizing education into low, middle, and high following the conventions described by the International Standard Classification of Education. These categories correspond to the International Standard Classification of Education's 0–2 (preprimary, primary, and lower secondary education), 3 (upper secondary education), and 4–6 (postsecondary education). However, this coding resulted in sparsely populated cells across generations, so years of education were used instead. Although this does not solve the issue of comparing samples with different underlying characteristics, it does permit similar operationalization of education across analyses. When evaluating findings from a set of such studies, it is important to consider their location in the underlying matrix of sampling characteristics.

Statistical Analysis

Analyses are performed independently by the research team for each study or can be analyzed by a statistical core, which ensures the availability of resources for implementation of the agreed-on models. This step of the process is facilitated by the interactive Web site, which provides access to protocol and statistical analysis scripts (e.g., SAS, SPSS, Stata, Mplus; with documentation) and allows researchers to upload the results of such analyses.

Comparison of Results

In many cases, parameters are obtained from models that are based on different variables, different measurement intervals, and different population and sampling characteristics. We can compare results in terms of general patterns of effects, such as direction and magnitude. This is the most basic level, providing evidence for cross-study validation of particular research findings. Meta-analysis can take into account sociodemographic and other sample characteristics and so control for study-level characteristics and evaluation of moderation.

Our approach to maximizing the comparability of results from the different studies includes two main efforts: aligning measurement and analysis operations and identifying stratification or other methods for dealing with country or sampling differences across studies. To reduce the impact of constraints and data loss through common denominator problems, researchers are also encouraged to conduct more extensive analyses on the core research questions in each study, making use of more elaborated versions of the key variables and adding relevant variables that might be unique to their own project. In this way, both maximally comparable and maximally rich methods can be applied to each research question to make full use of individual study data.

The situation may also arise in which a particular study is ideal for addressing some research question and is also a good match on most of the variables for a particular project, but it is missing a covariate (e.g., total cholesterol) or has no variance on a particular variable (e.g., the Normative Aging Study sample is men only, H-70 is a single age sample). One solution to including this study along with the others is to rerun the analyses of interest in the other studies, leaving out the problematic variable so that comparisons can be made on the same subset for each study. Clearly, this additional work would be warranted only in certain situations, and if a good number of studies with all relevant data were already providing results, it might not be the best choice.

Dissemination of Results

This coordinated research process leads to publication for both independent and jointly authored research and maintains attention to appropriate allocation of authorship credit. The major publication model is one of independent analysis and write up as a series of brief reports, with a jointly authored introduction and capstone paper of cross-study research synthesis and discussion of overall research findings. The secondary model is one of joint authorship of a single paper making use of multiple data sets with authorship determined at initiation and reconsidered at completion.

Summary: Benefits of the Coordinated Analysis Approach

Replication is the hallmark of a successful science. A collaborative, coordinated analysis framework can provide a broad foundation for cumulating scientific knowledge by facilitating efficient examination of multiple studies in ways that maximize comparability of results. The goal of such a framework is to maximize opportunities for reproducible research (e.g., Gentleman & Lang, 2007) through open access to analysis scripts and output for published results, permitting modification and evaluation of alternative models related to published articles and application of similar models and variable harmonization to other studies. A collaborative network impacts future science through reevaluation of existing data and planning for future data collections.

When research findings do not agree, researchers are left with uncertainty regarding the sources of the differences. Replicating findings across longitudinal studies of developmental and aging-related processes is challenging because of the different measures, designs, and statistical analysis performed. Cooperative networks—in addition to their central focus of cross-study and cross-national comparison of research findings—provide new opportunities for addressing sources of difference. Strengths of a collaborative research network include the consideration of alternative approaches and statistical models to evaluate key hypotheses and the evaluation of the sensitivity of results to alternative hypotheses and models. Such efforts can make the most of currently available data and provide an opportunity to move beyond current barriers to progress.

The availability of samples from different birth cohorts is invaluable for comparison of both current and future studies in order to understand the historical and cultural differences across generations. Indeed, cross-national comparisons and tests of hypotheses across birth cohorts defined by changes in historical socioeconomic status, education, and societal health outcomes are a major potential outcome of an international research network. Planning of future studies would be facilitated by open access to a searchable database for identifying studies with particular constructs or measures that would be available for evaluation of particular research questions. An organized summary of available data also provides a basis for informed decisions regarding optimal or essential test batteries that future studies might use to permit comparison to existing longitudinal studies.

Typically, science proceeds sequentially, with replication of results often taking years in the case of longitudinal studies. A key component of a collaborative, coordinated analysis approach is the immediate replication of research findings achieved through cooperative parallel analysis of independent studies and simultaneous publication. The opportunity for the evaluation and report of alternative models on the same data and the immediate follow-up of alternative hypotheses and accounting for disparities by individual- and study-level characteristics increases knowledge rapidly. Major benefits of collaboration with parallel analyses include accelerated accumulation of scientific knowledge, earlier understanding of the stability and generalizability of the findings, and greater statistical power for the study of infrequent events. Differences in language, culture, history, demographics, design, and measurements across longitudinal studies are important for establishing evidence of the generalizability of developmental and aging-related processes and must be considered in understanding cross-study differences. It is important that current and future studies permit analytical opportunities for quantitative comparison across samples differing in birth cohort and country given the historical shifts and cultural differences that may have an effect on late life processes and outcomes. These differences across studies, although presenting challenges for researchers to take into account in a cumulative science, may best be resolved through a collaborative research process.

References

Ades, A. E., & Sutton, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169,* 5–35.

Alwin, D. F., Hofer, S. M., & McCammon, R. (2006). Modeling the effects of time: Integrating demographic and developmental perspectives. In R. H. Binstock & L. K. George (Eds.), *Handbook of the aging and the social sciences* (6th ed., pp. 20–38). San Diego, CA: Academic Press.

Anderson, N. B., Bulatao, R. A., & Cohen, B. (Eds.). (2004). *Critical perspectives on racial and ethnic differences in health in late life.* Washington, DC: National Research Council.

Anstey, K. J., & Christensen, H. (2000). Education, activity, health, blood pressure and Apolipoprotein E as predictors of cognitive change in old age: A review. *Gerontology, 46,* 163–177.

Anstey, K. J., von Sanden, C., Salim, A., & O'Kearney, R. (2007). Smoking as a risk factor for dementia and cognitive decline: A meta-analysis of prospective studies. *American Journal of Epidemiology, 166,* 367–378.

Asia Pacific Cohort Studies Collaborative Group. (1999). Determinants of cardiovascular disease in the Asian Pacific region: Protocol for a collaborative overview of cohort studies. *Cardiovascular Disease Prevention, 2,* 281–289.

Bachrach, C. A., & Abeles, R. P. (2004). Social science and health research: Growth at the National Institutes of Health. *American Journal of Public Health, 94,* 22–28.

Becker, B. J., & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science, 22,* 414–429.

Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research, 37,* 127–160.

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 153–175). Oxford, England: Oxford University Press.

Butz, W. P., & Torrey, B. B. (2006, June 30). Some frontiers in social science. *Science, 312,* 1898–1900.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Sage.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14,* 81–100.

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple prospective studies: The role of item response theory in integrative analysis. *Developmental Psychology, 44,* 365–380.

Dufouil, C., Alpérovitch, A., & Tzourio, C. (2003). Influence of education on the relationship between white matter lesions and cognition. *Neurology, 60,* 831–836.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Duncan, G. J., & Kalton, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review, 55,* 97–117.

Fillmore, K. M., Grant, M., Hartka, E., Johnstone, B. M., Sawyer, S., Spieflman, R., & Temple, M. T. (1988). Collaborative longitudinal research on alcohol problems. *British Journal of Addiction, 83,* 441–444.

Fillmore, K. M., Hartka, E., Johnstone, B. M., Leino, E. V., Motoyoshi, M. M., & Temple, M. T. (1991). Preliminary results from a meta-analysis of drinking behavior in multiple longitudinal studies. *British Journal of Addiction, 86,* 1203–1210.

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology? *Sociological Methods & Research, 36,* 153–172.

Gentleman, R., & Lang, T. (2007). Statistical analyses and reproducible research. *Journal of Computational & Graphical Statistics, 16,* 1–23.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58,* 80–92.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 243–259). Washington, DC: American Psychological Association.

Harel, O., Hofer, S. M., Hoffman, L. R., Pedersen, N., & Johansson, B. (2007). Population inference with mortality and attrition in longitudinal studies on aging: A two-stage multiple imputation method. *Experimental Aging Research, 33,* 187–203.

Hendrick, C. (1990). Replications, strict replications, and conceptual replications: Are they important? *Journal of Social Behavior and Personality, 5,* 45–48.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Hofer, S. M., Flaherty, B. P., & Hoffman, L. (2006). Cross-sectional analysis of time-dependent data: Problems of mean-induced association in age-heterogeneous samples and an alternative method based on sequential narrow age-cohorts. *Multivariate Behavioral Research, 41,* 165–187.

Hofer, S. M., & Hoffman, L. (2007). Statistical analysis with incomplete data: A developmental perspective. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 13–32). Mahwah, NJ: Erlbaum.

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology, 47,* 341–352.

Huisman, M., Kunst, A. E., Adersen, O., Bopp, M., Borgan, J.-K., Correll, C., et al. (2004). Socioeconomic inequalities in mortality among elderly people in 11 European populations. *Journal of Epidemiology and Community Health, 58,* 468–475.

Janssen, F., Peeters, A., Mackenbach, J. P., & Kunst, A. E. (2005). Relation between trends in late middle age mortality and trends in old age mortality—Is there evidence for mortality selection? *Journal of Epidemiology and Community Health, 59,* 775–781.

Johnstone, B. M., Leino, E. V., Motoyoshi, M. M., Temple, M. T., Fillmore, K. M., & Hartka, E. (1991). An integrated approach to meta-analysis in alcohol studies. *British Journal of Addiction, 86,* 1211–1220.

King, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research, 36,* 173–199.

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry, 157,* 163–171.

Kurland, B., Johnson, L. L., & Diehr, P. (2007). *Longitudinal data with follow-up truncated by death: Finding a match between analysis method and research* (UW Biostatistics Working Paper Series). Seattle: University of Washington.

Lindsay, R. M., & Ehrenberg, A. S. C. (1993). The design of replicated studies. *American Statistician, 47,* 217–228.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–159.

Manly, J. J. (2008). Race, culture, education, and cognitive test performance among older adults. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook on cognitive aging: Interdisciplinary perspectives* (pp. 398–417). Thousand Oaks, CA: Sage.

Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology, 50,* 7–11.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14,* 126–149.

Minicuci, N., Noale, M., Bardage, C., Blumstein, T., Deeg, D. J., Gindin, J., et al. (2003). Cross-national determinants of quality of life from six longitudinal studies on aging: The CLESA project. *Aging and Clinical Experimental Research, 15,* 187–202.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with life-span psychology* (pp. 339–360). Dordrecht, the Netherlands: Kluwer.

National Research Council. (2000). *The aging mind: Opportunities for cognitive research.* Washington, DC: National Academy Press.

National Research Council. (2001a). *New horizons in health: An integrative approach.* Washington, DC: National Academy Press.

National Research Council. (2001b). *Preparing for an aging world: The case for cross-national research.* Washington, DC: National Academy Press.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palermo (Eds.), *Visuals of aesthetics, the environment, and development: The legacy of Joachim F. Wohwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nguyen, H., & Zonderman, A. (2006). Relationship between age and aspects of depression: Consistency and reliability across two longitudinal studies. *Psychology and Aging, 21,* 119–126.

Park, C. L. (2004). What is the value of replicating other studies? *Research Evaluation, 13,* 189–195.

Park, H. L., O'Connell, J. E., & Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *International Journal of Geriatric Psychiatry, 18,* 1121–1134.

Piccinin, A. M., & Hofer, S. M. (2008). Integrative analysis of longitudinal studies on aging: Collaborative research networks, meta-analysis, and optimizing future studies. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook on cognitive aging: Interdisciplinary perspectives* (pp. 446–476). Thousand Oaks, CA: Sage.

Piccinin, A. M., Hofer, S. M., Anstey, K. J., Deary, I. J., Deeg, D. J. H., Johansson, B., et al. (2006, November). *Cross-national IALSA coordinated analysis of age, sex, and education effects on change in MMSE scores.* Paper presented at the annual Gerontological Society of America conference, Dallas, Texas.

Riegel, K. F., & Angleitner, A. (1975). The pooling of longitudinal studies of aging. *International Journal of Aging and Human Development, 6,* 57–66.

Rose, C. L. (Ed.). (1976). *Collaboration among longitudinal aging studies, 1972–1975* (Publication No. 8, Research Report Series). Boston: Veterans Administration Outpatient Clinic.

Rosenbaum, P. R. (2001). Replicating effects and biases. *American Statistician, 55,* 223–227.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64,* 92–107.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Sliwinski, M. J., Hofer, S. M., & Hall, C. (2003). Correlated and coupled cognitive change in older adults with and without clinical dementia. *Psychology and Aging, 18,* 672–683.

Sliwinski, M. J., Hofer, S. M., Hall, C., Bushke, H., & Lipton, R. B. (2003). Modeling memory decline in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging, 18,* 658–671.

Sliwinski, M. J., & Mogle, J. (2008). Time-based and process-based approaches to analysis of longitudinal data. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook on cognitive aging: Interdisciplinary perspectives* (pp. 477–491). Thousand Oaks, CA: Sage.

Sliwinski, M. J., Stawski, R. S., Hall, C. B., Katz, M., Verghese, J., & Lipton, R. (2006). Distinguishing preterminal and terminal cognitive decline. *European Psychologist, 11,* 172–181.

Smith, C. T., Williamson, P. R., & Marson, A. G. (2005a). Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine, 24,* 1307–1319.

Smith, C. T., Williamson, P. R., & Marson, A. G. (2005b). An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *Journal of Evaluation in Clinical Practice, 11,* 468–478.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation.* New York: Wiley.

Spiegelhalter, D. J., & Best, N. G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine, 22,* 3687–3709.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Journal of the American Medical Association, 271,* 1004–1010.

Stewart, L. A., & Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference? *Lancet, 341,* 418–422.

Sutton, A. J., & Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27,* 625–650.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18,* 2693–2708.

Thorvaldsson, V., Hofer, S. M., Berg, S., & Johansson, B. (2006). Effects of repeated testing in a longitudinal age-homogeneous study of cognitive aging. *Journal of Gerontology: Psychological Sciences, 61*B, P348–P354.

Thorvaldsson, V., Hofer, S. M., Berg, S., Skoog, I., Sacuiu, S., & Johansson, B. (2008). Onset of terminal decline in cognitive abilities in non-demented individuals. Onset of terminal decline in cognitive abilities in individuals without dementia. *Neurology, 71,* 882–887.

Thorvaldsson, V., Hofer, S. M., Hassing, L., & Johansson, B. (2008). Cognitive change as conditional on age heterogeneity in onset of mortality-related processes and repeated testing effects. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook on cognitive aging: Interdisciplinary perspectives* (pp. 284–297). Thousand Oaks, CA: Sage.

Tooth, L., Ware, R., Bain, C., Purdie, D. M., & Dobson, A. (2005). Quality of reporting of observational longitudinal research. *American Journal of Epidemiology, 161,* 280–288.

Turner, R. M., Spiegelhalter, D. J., Smith, G. C. S., & Thompson, S. G. (2009). Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172,* 23–47.

Van Dijk, K. R. A., Van Gerven, P. W. M., Van Boxtel, M. P. J., Van der Elst, W., & Jolles, J. (2008). No protective effects of education during normal cognitive aging: Results from the 6-year follow-up of the Maastricht Aging Study. *Psychology and Aging, 23,* 119–130.

Weiner, J. M., Hanley, R. J., Clark, R., & Van Nostrand, J. F. (1990). Measuring the activities of daily living: Comparisons across national surveys. *Journal of Gerontology: Social Sciences, 45,* S229–S237.

Whitfield, K., & Morgan, A. A. (2008). Minority populations and cognitive aging. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook on cognitive aging: Interdisciplinary perspectives* (pp. 384–397). Thousand Oaks, CA: Sage.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.

Wulf, W. A. (1993, August 13). The collaboratory opportunity. *Science, 261,* 854–855.

Zunzunegui, M. V., Rodriguez-Laso, A., Otero, A., Pluijm, S. M. F., Nikula, S., Blumstein, T., et al. (2006). Disability and social ties: Comparative findings of the CLESA study. *European Journal of Ageing, 2,* 40–47.

Received August 6, 2008
Revision received January 22, 2009
Accepted January 29, 2009 ■