Published in final edited form as:

Psychol Methods. 2009 June ; 14(2): 150–164. 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

Department of Human Development and Family Sciences, Oregon State University

Abstract

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 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, we 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 will optimally be 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 (Molenaar, Huizenga, & Nesselroade, 2003; Hofer, Flaherty, & Hoffman, 2006; Hofer & Sliwinski, 2001; Kraemer, Yesavage, Taylor, & Kupfer, 2000; 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 lifespan.

Corresponding author: Scott M. Hofer, Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, BC, V8W 3P5, Canada.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/journals/met.

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. While 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.

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 until the opportunity for replication of within-person findings. Aggravating this slow process are the differences across studies in measures, samples, design characteristics and statistical analysis 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 and Torrey, 2006; National Research Council, 2000; National Research Council, 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. We use extant scientific evidence to structure, justify and extend research, and to develop theory, and may often base decisions on one or 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, replication

multiple times, better, because results are usually not as straightforward as they might first appear (e.g., Hendrick, 1990; Lindsay & Ehrenberg, 1993; Lykken, 1968; Park, 2004; Rosenbaum, 2001; Wilkinson and 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). For example, measurement differences can be magnified in cross-cultural or cross-national data where variation is inevitably introduced due to 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 result of a study is evaluated across independent studies. 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 in order 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, known as data pooling (i.e., integrative data analysis (Curran & Hussong, this volume), pooled data meta-analysis, (aka individual patient meta-analysis; Cooper & Patall, this volume), and mega-analysis (McArdle et al., this volume), which permit evaluation of both study-level and individual-level effects (Smith, Williamson & Marxon, 2005a, 2005b; Stewart & Parmar, 1993, Thompson & Sharp, 1999). These methods have been used very effectively in a variety of substantive areas and types of data.

Beyond this are methods that permit questions that go beyond what can be learned from any particular data set. Generalized evidence synthesis (Ades & Sutton, 2006; Spiegelhalter & Best, 2003; Spiegelhalter, Abrams, & Myles, 2004) 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 paper, 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 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 comparison 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. We briefly describe some of the many potential differences across samples, measures, and designs that can have an effect on cross-study comparison in the section below. 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 in order 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 below, 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 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 inference 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). While 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, we 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 in order to permit comparisons across birth cohorts, cross-sectionally and longitudinally within studies (e.g., Schaie, 1965). Others, such as the Gothenburg H70 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 will usually involve 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 macro-social 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 SES, 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 post-secondary education in recent decades. Data on occupational position (highest achieved, longest held, final, or current) and status are often pre-coded into broad categories, which may be difficult to reconcile across studies. While 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 SES 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, Alperovitch, & Tzourio, 2003; Stern et al., 1994). The results of this body of research are mixed, with some studies showing no interaction of schooling and rates of change, while others finding 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, as there are marked country and birth cohort differences in educational attainment (Piccinin et al., 2006).

Race/Ethnicity—In societies such as the U.S., socioeconomic and racial-ethnic comparisons must be jointly understood given their interactive and interdependent nature (e.g., Anderson, Bulatao, & Cohen, 2004; Manly, 2008). Whitfield & Morgan (2008) emphasizes the use of culturally appropriate models and suggests that prior to making such comparisons we would be wise to understand within-group processes as it may be more informative to study the relevant constellation of mechanisms within each group, rather than assume that the same factors apply in both.

Selection/Attrition/Mortality—Numerous studies have demonstrated the relatively strong link between 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 et al., 2007; Hofer & Hoffman, 2007; Kurland, Johnson, & Diehr, 2007). Comparisons across studies should be sensitive to these selection issues, as differences in interval length between assessments as well as initial sample characteristics such as age, SES and health will all contribute to the prevalence and impact of missing information on longitudinal results. As existing longitudinal studies mature, it will become more feasible to model these differences.

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 study will eventually be “dated”. However, inference regarding within-person change cannot otherwise be obtained and these tradeoffs 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. The measures can differ at a number of levels, and even within a single nation large operational differences can be found (e.g., Weiner, Hanley, Clark & Van Nostrand, 1990). When considering cross-cultural or cross-national data sets these differences can be magnified: regardless of whether the same measure has been used, differences are inevitably introduced due to language, administration, and item relevance. A balance must be found between optimal similarity of administration, similarity of meaning, and significance of meaning—avoiding unreasonable loss of information or lack of depth.

It can be difficult to gauge differences across studies with samples from different birth cohorts or different countries, in large part because the measurements themselves differ. Certainly, measures used 30 or 40 years ago may not be the ones used in more recently initiated studies. 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 permit factor analysis and test of invariance within a pooled data analysis (e.g., Bontempo & Hofer, 2007; Cooper & Pattall, this volume).

Change in Measurement over Different Life Periods—It is also often necessary to use different items or measures at different point in the lifespan in order 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 (this issue; Curran, Hussong, Cai, Huang, Chassin, Sher & Zucker, 2007) and McArdle, Grimm, Hamagami, Bowles, & Meredith (this issue) make use of 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 temporal sampling frame of individual change and variation must be carefully considered in both the design and analysis of longitudinal studies. It is best to assume that different sampling intervals (compared within and across studies) produce results that will require different interpretations for both within-person and between-person processes (Boker & Nesselroade, 2002; Martin & Hofer, 2004). For example, within-person correlation will indicate potentially different processes across temporal sampling of relatively short intervals (minutes, hours, days, or weeks) and certainly in contrast to correlated change across multiple years, as is the case for many of the longitudinal studies on aging. Consideration of measurement interval is similarly critical for the prediction of outcome variables and for establishing evidence regarding leading versus lagging indicators (Gollob & Reichardt, 1987; 1991). A related issue is that of interval censoring and the resolution by which different time-varying events have been measured and can be modeled and compared across studies.

Retest Effects—The selection of intervals between measurements is also critical for separating effects of repeated testing (i.e., learning) from those of development/aging over longer periods of time. Estimates of longitudinal change may be attenuated due to the gains occurring as a result of repeated testing, potentially persisting over long intervals. Complicating matters is the potential for improvement to occur differentially, related to ability level, age, or task difficulty, and which may be due to any number of related influences, including warm-up effects, initial anxiety, and test-specific learning, such as learning content and strategies for improving performance. Differential retest gains such as these confound the identification of differential age-related changes (e.g., in older adults, retest may not be manifest as an increase in performance, but as an attenuated decrease in performance). In most studies, retest effects are perfectly confounded with within-person changes (i.e., temporal spacing for test exposure and change between occasions are identical or highly correlated) and do not permit decomposition of effects at an individual level (Thorvaldsson, Hofer, Berg, & Johansson, 2006; Thorvaldsson, Hofer, Hassing, & Johansson, 2008).

Alternative Models of Time—In addition to sampling time within individuals, there are numerous ways to conceptualize and model change over time and the choice of temporal metric is critical for the interpretation and understanding of change processes and for cross-study comparison of results. Typically, change models are based on chronological age, or on time-in-study with chronological age included as a covariate, making level and rate of change conditional on age. Age-based and time-based models are equivalent in single or narrow age-cohort samples, but in age-heterogeneous samples the use of age-based models may not be appropriate without explicit test of convergence of between-person age differences and within-person age-changes. However, 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/since diagnosis of dementia; Sliwinski, Hofer, & Hall, 2003a; 2003b; 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, & MacCammon, 2006) to understand the effects of stress and psychosocial interactions. Such models provide a useful perspective for describing and

explaining average change and individual variation in change relative to common, possibly causal, processes.

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 has been developed in order to provide a means to evaluate statistically the similarity of results across studies. When the degree of similarity of method across studies is adequate to justify more stringent comparative strategies, meta-analytic methods are used to summarize findings and to identify and address questions regarding 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 which have exact or similar outcome measures and minimal variation in study design.

As discussed above, meta-analysis of existing reports from observational longitudinal studies is more challenging and currently limited in at least two ways. The first limitation stems from the paucity of published information regarding particular intra-individual 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 the 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, Park, O'Connell and Thompson (2003), intending to conduct a meta-analysis of cognitive decline in community-based prospective cohort studies with low attrition, whittled 5990 abstracts down to 19 papers and then concluded that heterogeneity due to population, country, measure, follow-up (intervals and number) and attrition differences required they reduce their goal to a narrative review. Coincidentally, with a different research question, Anstey, von Sanden, Salim and O'Kearney (2007) also find 19 publications with "measures compatible with at least one other article". They proceed 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 non- or former smokers, based 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 the smoking data from the authors because seven of the studies only reported smoking results incidentally with the relevant information unavailable in the published manuscripts. 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 above, and the conditional nature of the results to these decisions. Additional factors include whether the set of covariates deemed necessary in an analysis are 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 post-analysis 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, Williamson & Marson, 2005a; 2005b), but have to date 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, et al., 2009). When data are identical or sufficiently comparable across studies, pooled analysis of raw data across studies permit the analysis of influences associated with rare events (i.e., evaluation of apoE subtypes on cognitive functioning), provide increased power for the detection of associations and interactions, provide more reliable estimates of population-level change, and permit 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 in cases where measurements are identical or can be equated: by fiat, through co-calibration using IRT models, or with latent variable approaches based on item or scale-level data across studies (see Cooper & Pattell, this volume; McArdle et al., this volume). Unfortunately, for a majority of longitudinal studies of normal cognitive aging, opportunities for pooled data analysis using 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. Since, for the most part, these have not been determined or evaluated for any of the measures used across studies, it is hard to predict the impact on a pooled analysis. Another potential inroad here is supplemental data collection, in independent samples, to permit co-calibration and pooled data analysis (see Curran et al., 2008; McArdle et al., this volume).

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 inter-occasion intervals and number of occasions may differ. Combining data from studies with non-overlapping age ranges

(e.g., 55–70 versus 80+) can result in study level differences in outcomes that are confounded with age differences. Extrapolating beyond the data in particular studies requires 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 publically 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 papers 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, will result in the most solid accumulation of knowledge and is the most powerful way to build developmental science (Piccinin & Hofer, 2008).

Several large-scale collaboratories (Wulf, 1993) are already in existence (e.g., the National Alzheimer's Coordinating Center (NACC), the Collaborative Alcohol-Related Longitudinal Project (Fillmore et al., 1988; 1991), the Asia Pacific Cohort Studies Collaborative (APCSC) Group, 1999); and 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 and the Society of Collaboratories (SOC) indicate, 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 emails, but also issues a quarterly newsletter and minutes of the Executive Committee meetings, arranges teleconferences on an as needed basis, and maintains a password protected link on their website 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. For example, Thorvaldsson, Hofer, Berg, Skoog, Sacuiu, and Johansson (2008) used data from the Gothenberg H-70 study to explicitly replicate the terminal decline findings of Sliwinski et al., (2006) in Einstein Aging Study, finding consistent results following the same analysis protocol. In contrast, core staff from the Collaborative Alcohol-Related Longitudinal Project conducted parallel analyses of primary data from relevant subsets of the 39 affiliated studies and combined the results using meta-analysis (Fillmore et al., 1988; 1991). Their reports have presented results from combined analyses. Similarly, the Asia Pacific Cohort Studies Collaborative Group (1999) has reported

pooled analyses while appropriately considering study differences. Duncan and colleagues (2007) examined the effect of school readiness on later school reading and math achievement in six observational longitudinal studies using identical statistical models and culminating in a meta-analysis.

There are advantages to centralized analysis as well as to independent approaches to parallel analysis. While centralized analyses facilitate careful scrutiny of sampling and measurement differences across studies, coordinated independent analyses may better protect against capitalizing on chance and overmanipulation of data. As in many situations, a combination of both approaches may be most productive. Centralized analysis, as in the Collaborative Alcohol project (Johnstone et al., 1991) or the European CLESA Project (Minicuci, et al., 2003) allows the clearest view of the individual study differences, as a single set or group of eyes becomes familiar with the sampling and other idiosyncrasies of each dataset. This facilitates the identification of specific differences that might be due to sampling, etc., leading naturally to a priori tests of hypotheses regarding the source of divergent results. Independent sequential replication, as in the case of the Thorvaldson et al. (2008) replication of Sliwinski et al. (2006), may result in a more powerful replication but may be more limited in terms of testing hypotheses regarding differences across research findings.

While the normal cognitive aging literature does not currently contain the information necessary to conduct meta-analyses of the 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's alcohol work, the APCSC's medical research, and Duncan et al. (2007; 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 will often preclude pooling or will 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 understand 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 are critical aspects 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) will also aid in the determination of why results might differ. Longitudinal research is challenging, and coordinating analysis across studies 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; facilitate reproducible research; archive the analysis and measurement alignment process; provide a stronger basis for cumulative science based on optimal comparison and replication of results across longitudinal studies; and 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 is currently comprised of over 30 longitudinal studies on aging, spanning eight countries, with a combined sample size of over 70,000 individuals. These studies represent 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 100 (mainly adulthood), with birth cohorts ranging from 1880 to 1980, and historical periods from 1946 to the present. Between-occasion intervals range from 6 months to 17 years (the majority 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 based on 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 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.

Website—In order to support multiple concurrent interactions between investigators across wide geographic distances and time-zones, a secure website is used for data management (where applicable), progress reports, preliminary results, and statistical analysis scripts which are available to all investigators. The website is used to manage permissions, authorship agreements, and data access for data sets that are public and for those with data sharing agreements. Protocol, annotated statistical analysis scripts (e.g., SAS, SPSS, Stata, Mplus; with

documentation) 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. While communication technology of this sort is not always critical to the success of a collaborative system, it is clearly facilitative.

Searchable Study-Variable Meta-Data Base—Identifying studies with sufficient measures for evaluation of specific hypotheses is made possible by access to a searchable data base listing the measures used by each study. We have developed such a meta-data base that can eventually be linked to study protocol and exact details regarding the particular measurements used. While 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 PIs. 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 opportunity for immediate evaluation of differences, when found, and additional analyses to reconcile such differences.

The research process for the coordinated analysis of longitudinal studies on aging is shown schematically in Figure 1. 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 (1). 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 (2). Preliminary analyses begin with finalizing a protocol for aligning or harmonizing variables, studies and individual-level covariates, and for reporting results (3). Analyses are then performed independently by each group of researchers and reported in common format (4). 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 (5). The process is completed by submission for publication of each study's findings and a summary paper describing the cross-study comparison and meta-analysis of results.

1. 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 or published manuscript to replicate. Using the study-variable meta-data base, the proposing investigator identifies the most appropriate potential collaborators and invites them to participate. Project priorities and timelines are determined by the participating investigators.

2. 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 comprised, potentially, of alternative sets of analyses. The initial steps in variable coding will be based on information in the study-variable database. Script development will rely primarily on the dataset upon which

the script template is based. To the extent that details of the sample characteristics are known, decisions regarding 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 based on initial analysis of all of the datasets.

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.

3. Extension of the Statistical Analysis Plan—Collaborative interactions across research teams will further refine decisions regarding the aligning of measurement operations (e.g., data coding procedures) and the analysis protocols comprised of 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 PIs from the affiliated studies, a variety of strategies can be employed 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 co-calibration of measurements across many longitudinal observational studies, we focus on pre-analysis and post-analysis approaches for comparing results on a common metric. Pre-analysis 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 T1 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 Minicucci 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 SES measures 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. While this entails a certain amount of work at the outset, it will ensure from the start that the characteristics of all the studies are taken into account in the planning of appropriate comparisons. It will also facilitate the inclusion of new or external studies into a comparative framework. Measurement operationalization involving variable coding, centering, and possibly standardization of particular outcomes will necessarily involve 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 based on 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 post-analysis 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, who 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 (ISCED). These categories correspond to ISCED 0–2 (pre-primary, primary, and lower secondary education); 3 (upper secondary education), and 4–6 (post-secondary education). However, this coding resulted in sparsely populated cells across generations, so years of education were used instead. While 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 will be important to consider their location in the underlying matrix of sampling characteristics.

4. 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-upon models. This step of the process is facilitated by the interactive website which provides access to protocol and statistical analysis scripts (e.g., SAS, SPSS, Stata, Mplus; with documentation) and for upload of the results of such analyses.

5. Comparison of Results—In many cases, parameters will be 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 in terms of direction and magnitude across studies. 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.

Maximizing Individual Study Data: 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, each study is also encouraged to conduct more extensive analyses on the core research questions, 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.

The situation may also arise that a particular study may be ideal for addressing some research question, and also a good match on most of the variables for a particular project, but is missing a covariate (e.g., total cholesterol) or has no variance on a particular variable (e.g., the NAS sample is men only, H-70 is a single age sample). One solution to including this study along with the others is to re-run 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.

6. 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 papers and application of similar models and variable harmonization to other studies. A collaborative network will impact future science through reevaluation of existing data and planning for future data collections.

When research findings do not agree, we 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 comparison and test of hypotheses across birth cohorts defined by changes in historical SES, 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 data base 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 will increase 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, demographic, 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, while presenting challenges for us to take into account in a cumulative science, may best be resolved through a collaborative research process.

Acknowledgments

This manuscript and the Integrative Analysis of Longitudinal Studies of Aging (IALSA) research network were supported by a grant from the National Institute on Aging, National Institutes of Health (1R01AG026453). We would like to acknowledge the contributions of 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.

References

- Ades AE, Sutton AJ. Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *Journal of the Royal Statistical Society A* 2006;169:5–35.
- Alwin, DF.; Hofer, SM.; McCammon, R. Modeling the effects of time: Integrating demographic and developmental perspectives. In: Binstock, RH.; George, LK., editors. *Handbook of the aging and the social sciences*. 6. San Diego: Academic Press; 2006. p. 20-38.
- Anderson, NB.; Bulatao, RA.; Cohen, B., editors. *Critical perspectives on racial and ethnic differences in health in late life*. Washington, DC: National Research Council. Washington.; 2004.
- Anstey KJ, Christensen H. Education, activity, health, blood pressure and Apolipoprotein E as predictors of cognitive change in old age: A review. *Gerontology* 2000;46:163–177. [PubMed: 10754375]
- Anstey KJ, von Sanden C, Salim A, O’Kearney R. Smoking as a risk factor for dementia and cognitive decline: A meta-analysis of prospective studies. *American Journal of Epidemiology* 2007;166(4):367–78. [PubMed: 17573335]
- Asia Pacific Cohort Studies Collaborative Group. Determinants of cardiovascular disease in the Asian Pacific region: Protocol for a collaborative overview of cohort studies. *Cardiovascular Disease Prevention* 1999;2:281–289.
- Bachrach CA, Abeles RP. Social science and health research: Growth at the National Institutes of Health. *American Journal of Public Health* 2004;94:22–28. [PubMed: 14713689]
- Becker BJ, Wu MJ. The synthesis of regression slopes in meta-analysis. *Statistical Science* 2007;22:414–429.
- Boker SM, Nesselrode JR. A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research* 2002;37:127–160.
- Bontempo, DE.; Hofer, SM. Assessing factorial invariance in cross-sectional and longitudinal studies. In: Ong, AD.; van Dulmen, M., editors. *Handbook of methods in positive psychology*. Oxford University Press; 2007. p. 153-175.
- Butz WP, Torrey BB. Some frontiers in social science. *Science* 2006;312:1898–1900. [PubMed: 16809524]
- Cooper, H.; Hedges, LV. *The Handbook of Research Synthesis*. New York: Russell Sage; 1994.
- Curran PJ, Hussong AM, Cai L, Huang W, Chassin L, Sher KJ, Zucker RA. Pooling data from multiple prospective studies: The role of item response theory in integrative analysis. *Developmental Psychology* 2008;44:365–380. [PubMed: 18331129]
- Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*. 2008 this issue.
- Dufouil C, Alperovitch A, Tzourio C. Influence of education on the relationship between white matter lesions and cognition. *Neurology* 2003;60:831–836. [PubMed: 12629242]
- Duncan GJ, Kalton G. Issues of design and analysis of surveys across time. *International Statistical Review* 1987;55:97–117.
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, Pagani L, Feinstein L, Engel M, Brooks-Gunn J, Sexton H, Duckworth K, Japel C. School readiness and later achievement. *Developmental Psychology* 2007;43:1428–1446. [PubMed: 18020822]

- Fillmore KM, Grant M, Hartka E, Johnstone BM, Sawyer S, Spieflman R, Temple MT. Collaborative longitudinal research on alcohol problems. *British Journal of Addiction* 1988;83:441–444. [PubMed: 3395726]
- Fillmore KM, Hartka E, Johnstone BM, Leino EV, Motoyoshi MM, Temple MT. Preliminary results from a meta-analysis of drinking behavior in multiple longitudinal studies. *British Journal of Addiction* 1991;86:1203–1210. [PubMed: 1836407]
- Freese J. Replication standards for quantitative social science: Why not sociology? *Sociological Methods & Research* 2007;36:153–172.
- Gentleman R, Lange T. Statistical analyses and reproducible research. *Journal of Computational & Graphical Statistics* 2007;16:1–23.
- Gollob HF, Reichardt CS. Taking account of time lags in causal models. *Child Development* 1987;58:80–92. [PubMed: 3816351]
- Gollob, HF.; Reichardt, CS. Interpreting and estimating indirect effects assuming time lags really matter. In: Collins, LM.; Horn, JL., editors. *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC, US: American Psychological Association; 1991. p. 243–259.
- Harel O, Hofer SM, Hoffman LR, Pedersen N, Johansson B. Population inference with mortality and attrition in longitudinal studies on aging: A two-stage multiple imputation method. *Experimental Aging Research* 2007;33:187–203. [PubMed: 17364907]
- Hendrick, C. Replications, strict replications, and conceptual replications: Are they important?. In: Neuliep, JW., editor. *Handbook of Replication Research in the Behavioural and Social Sciences*. 1990. p. 45–48.
- Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;21:1539–1558. [PubMed: 12111919]
- Hofer SM, Flaherty BP, Hoffman L. 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* 2006;41:165–187.
- Hofer, SM.; Hoffman, L. Statistical analysis with incomplete data: A developmental perspective. In: Little, TD.; Bovaird, JA.; Card, NA., editors. *Modeling ecological and contextual effects in longitudinal studies of human development*. Mahwah, NJ: LEA; 2007. p. 13–32.
- Hofer SM, Sliwinski MJ. Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology* 2001;47:341–352. [PubMed: 11721149]
- Huisman M, Kunst AE, Adersen O, Bopp M, Borgan JK, Correll C, Costa G, Deboosere P, Desplanques G, Donkin A, Gadeyne S, Minder C, Regidor E, Spadea T, Valkonen T, Mackenbach JP. Socioeconomic inequalities in mortality among elderly people in 11 European populations. *Journal of Epidemiology and Community Health* 2004;58:468–475. [PubMed: 15143114]
- Janssen F, Peeters A, Mackenbach JP, Kunst AE. NEDCOM. 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* 2005;59:775–781. [PubMed: 16100316]
- Johnstone BM, Leino EV, Motoyoshi MM, Temple MT, Fillmore KM, Hartka E. An integrated approach to meta-analysis in alcohol studies. *British Journal of Addiction* 1991;86:1211–1220. [PubMed: 1751844]
- King G. An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research* 2007;36:173–199.
- Kraemer HC, Yesavage JA, Taylor JL, Kupfer D. How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry* 2000;157:163–171. [PubMed: 10671382]
- Kurland, B.; Johnson, LL.; Diehr, P. UW Biostatistics Working Paper Series. University of Washington; 2007. Longitudinal data with follow-up truncated by death: Finding a match between analysis method and research.
- Lindsay RM, Ehrenberg ASC. The design of replicated studies. *The American Statistician* 1993;47:217–228.
- Lykken DT. Statistical significance in psychological research. *Psychological Bulletin* 1968;70:151–159. [PubMed: 5681305]

- Manly, JJ. Race, culture, education, and cognitive test performance among older adults. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks: Sage Publications; 2008. p. 398-417.
- Martin M, Hofer SM. Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology* 2004;50:7-11. [PubMed: 14654720]
- McArdle JJ, Grimm K, Hamagami F, Bowles R, Meredith W. Modeling life-span growth curves of cognition using longitudinal data with changing scales of measurement. *Psychological Methods*. this issue.
- Minicuci N, Noale M, Bardage C, Blumstein T, Deeg DJ, Gindin J, Jylha M, Nikula S, Otero A, Pedersen NL, Pluijm SM, Zunzunegui MV, Maggi S. CLESA Working Group. Cross-national determinants of quality of life from six longitudinal studies on aging: the CLESA project. *Aging and Clinical Experimental Research* 2003;15:187-202.
- Molenaar, PCM.; Huizenga, HM.; Nesselrode, JR. The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of Developmental Systems Theory. In: Staudinger, UM.; Lindenberger, U., editors. *Understanding human development: Dialogues with life-span psychology*. Dordrecht; Kluwer: 2003. p. 339-360.
- National Research Council. The aging mind: Opportunities for cognitive research. Committee on Future Directions for Cognitive Research and Aging. In: Stern, Paul C.; Carstensen, Laura L., editors. *Commission on Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.; 2000.
- National Research Council. New horizons in health: An integrative approach. In: Singer, BH.; Ryff, CD., editors. *Committee on Future Directions for Behavioral and Social Sciences Research at the National Institutes of Health*. Washington, DC: National Academy Press; 2001.
- National Research Council. Panel on a Research Agenda and New Data for an Aging World, Committee on Population and Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press; 2001. *Preparing for an aging world: The case for cross-national research*.
- Nguyen H, Zonderman A. Relationship between age and aspects of depression: consistency and reliability across two longitudinal studies. *Psychology and Aging* 2006;21:119-126. [PubMed: 16594797]
- Park CL. What is the value of replicating other studies? *Research Evaluation* 2004;13:189-195.
- Park HL, O'Connell JE, Thomson RG. A systematic review of cognitive decline in the general elderly population. *International Journal of Geriatric Psychiatry* 2003;18:1121-1134. [PubMed: 14677145]
- Piccinin, AM.; Hofer, SM. Integrative analysis of longitudinal studies on aging: Collaborative research networks, meta-analysis, and optimizing future studies. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks: Sage Publications; 2008. p. 446-476.
- Piccinin, AM.; Hofer, SM.; Anstey, KJ.; Deary, IJ.; Deeg, DJH.; Johansson, B.; Mackinnon, AJ.; Spiro, A.; Thorvaldsson, V. Cross-national IALSA coordinated analysis of age, sex, and education effects on change in MMSE scores. In: Hofer, SM.; Piccinin, AM., editors. *Integrative Analysis of Longitudinal Studies on Aging: Accounting for Health in Aging-Related Processes*; Paper symposium conducted at the annual Gerontological Society of America Conference; Dallas, TX. 2006 Nov.
- Riegel KF, Angleitner A. The pooling of longitudinal studies of aging. *International Journal of Aging and Human Development* 1975;6:57-66. [PubMed: 1150336]
- Rose, CL. Collaboration among longitudinal aging studies, 1972-1975. Veterans Administration Outpatient Clinic; Boston, MA: Jun. 1976 Publication No. 8, research Report Series
- Rosenbaum PR. Replicating effects and biases. *American Statistician* 2001;55:223-227.
- Schaie KW. A general model for the study of developmental problems. *Psychological Bulletin* 1965;64:92-107. [PubMed: 14320080]
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2001.
- Sliwinski MJ, Hofer SM, Hall C. Correlated and coupled cognitive change in older adults with and without clinical dementia. *Psychology and Aging* 2003a;18:672-683. [PubMed: 14692856]

- Sliwinski MJ, Hofer SM, Hall C, Bushke H, Lipton RB. Modeling memory decline in older adults: The importance of preclinical dementia, attrition and chronological age. *Psychology and Aging* 2003b; 18:658–671. [PubMed: 14692855]
- Sliwinski, MJ.; Mogle, J. Time-based and process-based approaches to analysis of longitudinal data. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks: Sage Publications; 2008. p. 477–491.
- Sliwinski MJ, Stawski RS, Hall CB, Katz M, Verghese J, Lipton R. Distinguishing preterminal and terminal cognitive decline. *European Psychologist* 2006;11:172–181.
- Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* 2005a;24:307–1319.
- Smith CT, Williamson PR, Marson AG. 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. *J Eval Clin Pract* 2005b;11:468–478. [PubMed: 16164588]
- Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine* 2003;22:3687–3709. [PubMed: 14652869]
- Spiegelhalter, DJ.; Abrams, KR.; Myles, JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: Wiley; 2004.
- Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA* 1994;271:1004–10. [PubMed: 8139057]
- Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418–422. [PubMed: 8094183]
- Sutton, AJ.; Abrams, KR.; Jones, DR.; Sheldon, TA.; Song, F. *Methods for meta-analysis in medical research*. New York: Wiley; 2000.
- Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine* 2008;27:625–650. [PubMed: 17590884]
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999;18:2693–2708. [PubMed: 10521860]
- Thorvaldsson V, Hofer SM, Berg S, Johansson B. Effects of repeated testing in a longitudinal age-homogeneous study of cognitive aging. *Journal of Gerontology: Psychological Sciences* 2006;61B:P348–P354.
- Thorvaldsson V, Hofer SM, Berg S, Skoog I, Sacuiu S, Johansson B. Onset of terminal decline in cognitive abilities in non-demented individuals. Onset of terminal decline in cognitive abilities in individuals without dementia. *Neurology* 2008;71:882–887. [PubMed: 18753475]
- Thorvaldsson, V.; Hofer, SM.; Hassing, L.; Johansson, B. Cognitive change as conditional on age heterogeneity in onset of mortality-related processes and repeated testing effects. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks: Sage Publications; 2008. p. 284–297.
- Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *American Journal of Epidemiology* 2005;161:280–288. [PubMed: 15671260]
- Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *J Royal Statistical Soc Series A* 2009;172:23–47.
- Van Dijk KRA, Van Gerven PWM, Van Boxtel MPJ, Van der Elst W, Jolles J. No protective effects of education during normal cognitive aging: Results from the 6-year follow-up of the Maastricht Aging Study. *Psychology and Aging* 2008;23:119–130. [PubMed: 18361661]
- Weiner JM, Hanley RJ, Clark R, Van Nostrand JF. Measuring the activities of daily living: Comparisons across national surveys. *Journal of Gerontology: Social Sciences* 1990;45(6):S229–237.
- Whitfield, K.; Morgan, AA. Minority populations and cognitive aging. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks: Sage Publications; 2008. p. 384–397.
- Wilkinson L. Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* 1999;54:594–604.
- Wohlwill, JF. *The study of behavioral development*. New York: Academic Press; 1973.

- Wulf WA. The collaboratory opportunity. *Science* 1993;261:854–855. [PubMed: 8346438]
- Zunzunegui MV, Rodriguez-Laso A, Otero A, Pluijm SMF, Nikula S, Blumstein T, Jylha M, Minicuci N, Deeg DJH. CLESA Working Group. Disability and social ties: Comparative findings of the CLESA study. *Journal European Journal of Ageing* 2006;2:40–47.

Figure 1.
Coordinated research process.