CHAPTER

2

Methodological Considerations for the Study of Adult Development and Aging

Stuart W.S. MacDonald[1] *and Robert S. Stawski*[2]

[1]**Department of Psychology, University of Victoria, Victoria, BC, Canada** [2]**School of Social and Behavioral Health Sciences, Oregon State University, Corvallis, OR, USA**

OUTLINE

K.W. Schaie & S.L. Willis (Eds)
Handbook of the Psychology of Aging, Eighth edition.

DOI: http://dx.doi.org/10.1016/B978-0-12-411469-2.00002-9

INTRODUCTION

This chapter explores key methodological and analytical considerations for the study of adult development and aging. In particular, we focus on central themes that are routinely encountered in conducting current aging research. We address a range of topics, from design selection and sampling considerations (including novel developmental research designs) to key considerations regarding missing data as well as the impact of attrition and retest on statistical parameter estimates. Given recent advances in research design and statistical modeling of developmental phenomena and their application to the study of the psychology of aging, we overview several analytic procedures and approaches that help to efficiently characterize aging-related change for various phenomena. In particular, we summarize several models for measuring change, explore multivariate approaches for examining correlated and coupled change, as well as compare alternative metrics for parameterizing developmental time. Finally, we conclude by highlighting emerging methodological trends in the study of adult development and aging, including recent emphasis on integrated data analysis and harmonization, as well as adopting an intraindividual variability approach for informing dynamic aging-related processes.

RESEARCH DESIGNS AND SAMPLING CONSIDERATIONS FOR THE STUDY OF ADULT DEVELOPMENT AND AGING

This section overviews two classic research designs for the psychology of aging, contrasts their relative strengths and weaknesses, and concludes with a thorough overview of a specific subtype of longitudinal design (the measurement burst design) and its merits for studying select developmental phenomena.

Cross-Sectional Versus Longitudinal Designs

The theoretical focus of any study, as well as its corresponding research questions, helps to predetermine selection of the most suitable research design. Research designs for studies of adult development and aging reflect a combination of age, cohort, and period effects (Schaie, 2013). Further, for any study of the psychology of aging and underlying developmental processes, it is essential to distinguish between age-related differences and aging-related changes. The following section briefly addresses these issues; rather than an exhaustive overview of possible research designs, we focus in particular on cross-sectional versus longitudinal approaches to the study of adult developmental and aging.

Age Differences Versus Change

Age differences are indexed using cross-sectional research designs and reflect differences in constructs (e.g., cognitive function, wellbeing) across age-heterogeneous groups or samples of individuals measured at *a single point* in time. Comparisons across these individuals or groups would afford insight into *age differences* in level(s) of cognitive function or wellbeing, but provide no information about how these constructs may be changing over time. In contrast, the study of aging-related change is the province of longitudinal research designs. Such designs index changes in constructs by testing a group of individuals over multiple occasions of assessment. By studying the same individuals over time, we are able to derive within-person estimates for the direction and rate of change.

Relative Advantages Versus Disadvantages

Cross-sectional studies offer a number of advantages including efficiency (e.g., less time required to collect data) as well as avoiding select confounds such as retest effects (Salthouse, 2009). However, with regard to the study of adult development and aging, notable weaknesses of cross-sectional designs include an overestimation of age-related performance differences due to cohort effects (Nilsson, Sternäng, Rönnlund, & Nyberg, 2009), as well as an inability to address arguably the most *important aim* of aging research—whether *aging-related change* is occurring (Hofer & Sliwinski, 2001). As cross-sectional assessments are conducted at a single point in time, such designs necessarily confound age and cohort effects. Consequently, it is not possible to differentiate whether observed group differences are due to developmental age processes or to shared experiences characterizing cohort effects. In contrast, longitudinal studies facilitate the direct estimation of within-person change, as well as the possibility of investigating individual differences in change (Hofer & Sliwinski, 2001). Most if not all research questions and theories in adult development and aging are interested in such effects. To be sure, longitudinal designs also entail a number of limitations including the cost (both in terms of added expense for longitudinal collections and the time required to conduct repeated assessments to study aging processes that typically span years rather than months), as well as design considerations and analytic complexities (Curran & Bauer, 2011; Hertzog, Lindenberger, Ghisletta, & Oertzen, 2006; Hoffman & Stawski, 2009). Furthermore, longitudinal designs confound age and time of measurement—observed changes in outcomes of interest may be due to age- and/or cohort-related processes (if an age-heterogeneous cohort is being studied longitudinally), or to events at the time of measurement that exerted a pervasive influence on all individuals.

When contrasting patterns and magnitude of effects, decades of research have demonstrated differences in results between cross-sectional (i.e., age-related differences) and longitudinal (i.e., aging-related changes) designs. More specifically, estimates of longitudinal aging-related changes are routinely smaller than estimates of cross-sectional age-related differences. Cross-sectional age-related differences are often greatly influenced by cohort effects between the age groups under study, such as societal shifts in formal education and the corresponding impact on cognitive performance (Nilsson et al., 2009). Longitudinal aging-related changes are often influenced by selective attrition from longitudinal follow-up (e.g., more frail individuals discontinue participation), as well as practice or retest effects (with repeated exposure/assessments tending to obscure true age-related decline). The topics of attrition and retest are reviewed in detail in the subsequent section concerning key methodological considerations for the study of aging.

Which Design Is Best Suited for the Study of Aging?

A recent special issue in *Neurobiology of Aging* (Volume 30, 2009) focused on an

enduring question in research on the psychology of aging—"*When does age-related cognitive decline begin?*" Perhaps better than any description we can offer, this collection of articles directly addresses the conundrum regarding whether cross-sectional versus longitudinal designs are best suited for the study of aging. Despite consistently reported negative associations between age and cognitive function in cross-sectional studies (cf. Salthouse, 2009), many theorists and methodologists alike posit that the study of *aging-related change* necessitates longitudinal data. Indeed, Molenaar (2004) has forcefully argued that inferences about longitudinal aging-related change can only be drawn from studies of cross-sectional age-related differences when very strict (and often unrealistic) assumptions are met (also see Curran & Bauer, 2011; Hoffman & Stawski, 2009). Similarly, Hofer and Sliwinski (2001) contend that aging is a *within-person phenomenon*, and that longitudinal research designs are requisite for evaluating aging-related theories and propositions in particular. A central tenet of their argument is that the study of aging is a process that transpires within-persons over time, and can only be observed through the study of change. Moreover, as findings have clearly shown (MacDonald, Hultsch, Strauss, & Dixon, 2003; Sliwinski & Buschke, 1999), the correspondence between age-cognition trends for between-person versus within-person variance and covariance estimates is often modest at best.

One might question the relative importance of this issue and why it matters. To address this, consider an example regarding the process of forgetting from the episodic memory literature, where for decades, general consensus was that rates of forgetting were invariant across persons, despite known individual differences in encoding and retrieval processes (cf. MacDonald, Stigsdotter-Neely, Derwinger, & Bäckman, 2006). The generally accepted interpretation was that rates of acquisition and forgetting are asymmetrical, rather than processes anchoring disparate ends of a memory continuum. However, a competing explanation as to why individual differences in forgetting were rarely identified may be based upon this literature's more typical reliance on between- as opposed to within-person designs and estimates. With regard to forgetting, it is tenuous to assume that mean group differences will exhibit identical patterns to individual differences (Hofer & Sliwinski, 2001). For example, a negative correlation between learning and forgetting reported at the between-participants level (those individuals who learned more will also forget less) does not guarantee that a similar negative association will be observed at the within-person level (for any given individual, learning information at a faster rate will be associated with a slower rate of forgetting over time). Such discrepancies have been long described by the ecological fallacy (Robinson, 1950), stating that mean (group-level) findings can differ in both magnitude and valence relative to individual results (Molenaar, 2004).

Of direct relevance to the question regarding which research design is best suited to the study of aging, the *aggregation bias* just described represents perhaps the most critical weakness of cross-sectional designs. Specifically, due to considerable between-subject age heterogeneity (e.g., samples that span 50–90 years of age) at the single point of assessment, associations between measures (e.g., memory and sensory function) observed in cross-sectional designs are positively biased due to the confounding influence of population average age trends. Virtually any variables that exhibit cross-sectional age differences on average (e.g., poorer memory function and auditory acuity for those in the ninth versus seventh decades of life) will result in a positively biased association at the between-person level *even if* corresponding within-person associations for rates of change for the very same measures are nonsignificant or inversely associated (for further discussion, see

Hofer & Sliwinski, 2001). This bias introduced in cross-sectional studies due to population mean confounds is particularly troubling for hypotheses and theories predicated largely upon cross-sectional data. For example, evidence from cross-sectional studies consistently provided strong support for the processing speed hypothesis (Salthouse, 1996), indicating that age-related differences in higher order cognitive function could be explained by age-related decreases in processing speed. However, when examined using longitudinal data, evidence for this hypothesis was modest at best (MacDonald et al., 2003; Sliwinski & Buschke, 1999; Stawski, Sliwinski, & Hofer, 2013). Whereas cross-sectional studies routinely reported that greater than 90% of age-related differences in cognitive function could be accounted for by processing speed, the use of identical constructs and measures in longitudinal designs found that *change* in perceptual speed accounted for only 20% (or less) of change variance in other cognitive outcomes. Such a discrepancy provides an important example of the cross-sectional fallacy—within-person aging-related changes spanning longitudinal segments of time cannot be necessarily inferred from cross-sectional age-related differences indexed at any single point in time (cf. Schaie, 2009).

Summary

Beyond the mere passage of time, understanding how the aging process unfolds requires research designs that incorporate between-person differences, within-person rates of change, as well as individual differences in change. There is a long history in the study of human development, and adulthood development and aging in particular, advocating for longitudinal designs in keeping with key foci including the study of performance change over time as well as an idiographic emphasis (Nesselroade & Baltes, 1979). With regard to the study of aging-related change, we side with many other aging scholars who advocate for the use of longitudinal designs (Ferrer & Ghisletta, 2011; Schaie, 2009).

Longitudinal Designs: Select Subtypes

Whereas longitudinal designs provide a vehicle for directly examining aging-related changes, simply collecting longitudinal (or repeated measures data), without consideration of the temporal cadence of the phenomena under study, may offer relatively limited theoretical and empirical yield. For example, whether one's focus concerns ontogenetic versus microgenetic forms of within-person change will necessitate selection of a specific subtype of longitudinal design. Thus, if the focus concerns aging-related changes in cognitive function, such characteristically slow(er) and more enduring within-person change reflects processes that transpire across months, years, or decades, with a typically employed longitudinal design characterized by single assessments separated by months or years (cf. Nesselroade, 1991). In contrast, more labile (i.e., transient, fluctuating) phenomena (e.g., neuroendocrine or emotional responses to stressful experiences, trial-to-trial variability in response times (RTs)) require indexing change across much shorter time periods (e.g., seconds, minutes, days, or weeks). Failure to consider the (hypothesized) temporal interval of the process or phenomena of interest and design a longitudinal study accordingly could lead to results and conclusions that are misaligned with theory and process (Neupert, Stawski, & Almeida, 2008). Employing longitudinal research designs (e.g., multiple time points with well-reasoned retest intervals) and corresponding analytic techniques (e.g., linear mixed models) represent critical considerations when attempting to study processes in their appropriate time courses in service of the study of aging. In particular, the measurement burst design facilitates the study of dynamic aging processes that unfold across distinct temporal intervals. The

following subsection briefly overviews the longitudinal intensive measurement burst design (Nesselroade, 1991; Rast, MacDonald, & Hofer, 2012; Sliwinski, 2008) and its utility for studying select aging processes.

Intensive Measurement Burst Design

The measurement burst design incorporates data sampling across distinct temporal intervals: bursts of intensive repeated assessments within a relatively short duration (e.g., spanning hours, days, or weeks), with these bursts repeated longitudinally across longer temporal intervals (e.g., months, years). A cross-sectional study conducting assessment for a single point in time confounds trait-like (e.g., stable characteristics of a person such as intelligence or personality), state-like (e.g., a person's momentary state characterized by stress, fatigue, or anger), and developmental (e.g., developmental metastates such as pre- vs. postretirement, pre- vs. postdisease state) influences. Single assessment designs simply cannot distinguish among these competing sources of variance. By blending intensive repeated measures designs (e.g., ecological momentary assessment, daily diaries) within traditional longitudinal designs (e.g., annual retests), the measurement burst design attempts to address these shortcomings.

There are numerous advantages of the measurement burst design, including: (i) the use of multiple assessments within a short period of time offering improved measurement properties of variables and for the detection of change, (ii) the ability to disambiguate shorter-term and transient fluctuations (i.e., intraindividual variability) from longer-term and durable changes (i.e., intraindividual change), and (iii) the ability to formally examine how faster-moving processes, reflected in intraindividual variability, influence slower-moving processes reflected in intraindividual change (Nesselroade, 1991; Ram & Gerstorf, 2009; Rast et al., 2012; Stawski, MacDonald, & Sliwinski, in press; Stawski, Smith, & MacDonald, 2015).

Of particular note is the third point above—that the measurement burst design represents an invaluable methodological tool for the study of dynamic processes that unfold across both near- and long-term intervals, as well as how these processes influence one another (see related discussion in later section on *intraindividual variability*). As with standard longitudinal studies, the sampling timescale of the measurement burst design must be carefully matched to the particular aging process under study. However, in contrast to traditional longitudinal designs that only need to consider the interval between successive assessments, measurement burst designs require consideration of the temporal interval of the intensive burst of assessments, as well as the temporal interval over which these successive bursts of assessments will be repeated. Such decisions should be informed on both theoretical and empirical grounds. For example, for cognitive processes like memory, a well-suited decision might entail a series of short-term assessments spanning days or weeks as well as longer-term follow-up assessments spanning years, with the former elucidating intraindividual variability in memory processes (e.g., learning) and the latter informing more durable, developmental change. In contrast, processes such as emotional reactivity to stressors that transpire over much shorter timescales will need to consider the appropriate interval of assessment and design accordingly. By conducting assessments within (e.g., ecological momentary assessment) and/or across (e.g., daily diary designs) days, such designs are particularly effective at capturing dynamic processes. In some instances, employing variation in the spacing of assessments may be particularly advantageous, both within bursts (e.g., random or event contingent sampling for ecological momentary assessment) as well as across bursts (e.g., more frequent assessments for at-risk populations—such as 6-month retests for those in the early stages of dementia versus every few years for

otherwise healthy older adults). To be sure, it should be emphasized that different timescales are not necessarily interchangeable (Neupert et al., 2008), and that variance in processes observed across these distinct timescales is not necessarily a function of the same causes or correlates (Hoffman & Stawski, 2009; Sliwinski, Smyth, Stawski, & Wasylyshyn, 2005).

KEY THREATS TO THE VALIDITY OF LONGITUDINAL DESIGNS

Although longitudinal designs have many definitive advantages for addressing central research questions in the study of adult development and aging, to be sure, there are some notable limitations that must be considered including attrition, retest effects, and missingness. In this section, we provide an overview of these limitations, as well as offer some basic guidelines for researchers to consider when analyzing data influenced by these factors.

Attrition

Selection processes including non-representative initial sampling and attrition pose important concerns for drawing inferences from our data. The potential impact of incomplete data is invariably first encountered during the participant recruitment phase. At this initial stage, attrition due to refusal to participate or failure to respond to the invitation is often discounted as an important source of sampling bias (Ferrer & Ghisletta, 2011). However, to the extent that the initial sample in a longitudinal study is less or non-representative of the target population, then parameter estimates and corresponding inferences drawn about longitudinal change may be biased or inaccurate (Hofer & Sliwinski, 2001). Participant attrition in a longitudinal study may be due to illness (self or other), lack of interest, adverse reactions to testing, relocation, or death. In addition, it is not uncommon for participants to selectively complete certain tasks in the measurement battery, and to avoid attempting others. Such observed attrition within a longitudinal study represents an internal validity threat to the research study design (Hultsch, Hertzog, Dixon, & Small, 1998). Of even greater concern is the issue of whether attrition is non-random. If there is a systematic relationship between attrition, missing an entire retest assessment, or failing to complete specific measures, such non-random or selective attrition is likely to systematically bias patterns or rates of change, with the most pronounced effects of attrition usually occurring between the first and second measurement occasions (Hultsch, Hertzog, Dixon, & Small, 1998). Individuals who remain in longitudinal studies often tend to be more select, exhibiting better health and cognitive functioning (Radler & Ryff, 2010). In addition to threatening internal validity, attrition may also result in diminished statistical power (Ferrer & Ghisletta, 2011; Gustavson, von Soest, Karevold, & Roysamb, 2012). Longitudinal studies provide opportunity to explore the impact of a given selection process (e.g., dropout, death) as well as to incorporate such processes into the model to improve our inferences about change based on tenable assumptions regarding the underlying attrition process (Graham, 2009; Little & Rubin, 1987).

Retest Effects

For some time, practice or retest effects have been recognized as a threat to the internal validity of longitudinal studies (Salthouse, 2009). In the case of cognitive function, the process under study (e.g., episodic memory) may be directly influenced by repeated exposure to memory tasks, thereby benefitting performance on subsequent occasions. Any systematic association between the process under study (i.e., aging-related declines in episodic memory) and the repeated longitudinal assessment (i.e., retest or practice-related improvements)

exert opposing influences on performance and (potentially) bias observed developmental trajectories (Hultsch et al., 1998; Schaie, 2013). The degree of retest effect bias is influenced by several determinants including the amenability of the construct under study to practice, the length of time spanned by the retest interval, and the number of longitudinal assessments (Ferrer & Ghisletta, 2011; McArdle & Woodcock, 1997). Certain attributes such as measures of biological function (e.g., markers of blood chemistry, pulmonary function) can remain largely uninfluenced by repeated assessments, whereas other abilities are far more amenable to practice (e.g., developing strategies for successfully completing cognitive tasks). In the study of adult development and aging, cognitive functions are putatively the most susceptible to retest effects (Hultsch et al., 1998; McArdle & Woodcock, 1997; Schaie, 2013). Similarly, longer retest intervals (e.g., >5 years between retest intervals) are suggested to exert a more modest effect on patterns of change (Rönnlund, Nyberg, Bäckman, & Nilsson, 2005; Schaie, 2013), with the most marked retest effects observed between the first two repeated assessments and the positive benefits of retest diminishing for three or more assessments (Hultsch et al., 1998; McArdle & Woodcock, 1997; Rabbitt, Diggle, Smith, Holland, & McInnes, 2001).

Concerns about retest effects in longitudinal studies include the possibility that they mask aging-related declines due to the benefits conferred by prior test experience, and may in part account for the oft-reported discrepancies between trends reported in cross-sectional versus longitudinal studies (cf. Rönnlund et al., 2005; Salthouse, 2009). Retest effects may result in the systematic underestimation of rates of aging-related change, or may even enhance performance, for various reasons including recall of the correct response when exposed to the very same task, the reflection upon and development of generalized strategies for completing tasks, or the diminishment of anxiety during follow-up testing occasions (Hultsch et al., 1998). In order to have confidence in such inferences drawn, it is necessary to disambiguate estimates of change by attempting to differentiate sources due to developmental shifts versus retest effects.

Assessing the Impact of Repeated Practice on Trajectories of Age-Related Change

Gauging the impact of retest effects is commonly accomplished in one of several traditions, either via the *sampling approach* (research design) or *quantitative model parameterization* (statistical control). The sampling approach involves retaining a randomly sampled select subset of participants who are not administered any measures that are to be assessed for retest effects (Baltes, Reese, & Nesselroade, 1977). Other than not being tested on the target measures that represent the focus of study for assessing practice effects, this reserve sample is identical to the parent longitudinal sample. The magnitude of practice effects are evidenced by comparing the time 2 performance of the longitudinal sample (tested on two occasions) to the time 2 performance of the reserve sample (tested only at time 2), with observed performance differences between groups reflecting retest effects (Schaie, 2013). Problems with this approach include: (i) attrition in the longitudinal sample that may positively bias both individual differences and change in performance; as well as (ii) the fact that the refreshment sample is drawn at a different time of measurement and is thus subject to changes over time in selection effects including population change, sampling methods, and volunteering behaviors (Hultsch et al., 1998). Recent research using an age-heterogeneous sample and sampling-based approach revealed very modest evidence for retest effects on performance level for two of five cognitive outcomes assessed (Thorvaldsson, Hofer, Berg, & Johansson, 2006). Thus, sampling-based approaches to assessing

retest effects can provide invaluable insights, but can be time-consuming and expensive.

Many longitudinal studies of aging do not include reserve or refreshment samples, but rather conduct repeated longitudinal assessments for a single cohort. As such, quantitative modeling approaches have been developed to distinguish the effects of within-person change from repeated exposure (Ferrer, Salthouse, Stewart, & Schwartz, 2004; McArdle & Woodcock, 1997; Rabbitt et al., 2001). Ferrer et al. (2004), for example, employed a statistical approach to estimate separate effects for retest and within-person age-related change for select measures of cognitive function. Of particular note, when analyses were conducted that excluded the parameterization for practice effects, the estimates for age-related cognitive decline were underestimated. However, a profound issue related to quantitative modeling approaches for assessing retest effects involves estimating the *separate* effects of retest and within-person developmental change in the same model. This requires the inclusion of specific time parameterizations—one per effect. However, the time structures underlying processes of retest and change (maturation) are not independent (Nilsson et al., 2009). In order for such models to converge and provide estimates for both retest and developmental change, it has been suggested that they rely upon between-person age differences to estimate effects of repeated testing (Thorvaldsson et al., 2006), and as such are susceptible to population mean confounds discussed earlier (cf. Hofer & Sliwinski, 2001).

In concluding this subsection, it should be noted that retest effects are not solely applicable to longitudinal studies. Cross-sectional studies that employ testing batteries comprised of multiple indicators of the same construct (e.g., various measures of executive function) are also susceptible to retest, and may require counterbalancing the order of task administration (Ferrer & Ghisletta, 2011). Further, regardless of whether a design-based or quantitative approach is adopted for indexing retest effects, additional confounds may influence estimates. For example, with either the design-based or statistical approach, cohort effects (e.g., history-graded influences) may bias retest estimates as it is assumed that the groups being compared differ primarily in terms of the number of repeated assessments. If the samples also differ as a function of cohort effects, this confounds interpretation of any observed retest effect differences (Hultsch et al., 1998; Schaie, 2013).

Missingness: Causes, Consequences, and Potential Solutions

As first introduced in the section on participant attrition, missing data due to various sources—from initial sampling selectivity, to dropout, to non-random completion of tasks in the test battery—can adversely bias parameter estimates, particularly in studies of adult development and aging. This section overviews how patterns of missingness are classified, the corresponding implications, and outlines approaches for effectively addressing missingness.

Classifications of Missingness

Prior to analyzing data, it is imperative for the adult development and aging researcher to assess whether data missingness, due in particular to non-random factors, is present. A greater degree of non-random dropout begets greater concern about the representativeness of a given sample. The nature of missingness can exert influences ranging from the relatively benign (reducing statistical power) to those eliciting great concern (e.g., resulting in the substantial bias of parameter estimates). This awareness led to formal classifications identifying three distinct patterns or classes of missingness, each with different implications for interpreting one's data (Graham, 2009; Little & Rubin, 1987; McKnight, McKnight, Sidani, &

Figueredo, 2007; Rubin, 1976,1987; Schafer & Graham, 2002): missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Each class refers to the probability of missing data values given information about the dependent variable(s) of interest, other associated predictor variables under study, and the hypothetical mechanism thought to underlie the missing data (Enders, 2010; McKnight et al., 2007).

Data are classified as MCAR if the missing data occur by virtue of a random process. In such instances, the reason for the missing data is unrelated to observed or unobserved variables in a study, the mechanism underlying missingness is ignorable, and the missing data can be safely ignored. Data are classified as MAR if the missing data for a given variable occur by virtue of a random process after taking other observed variables in the study into account. That is, the mechanism of underlying missingness has been accounted for based on associations with other measured variables and any potential threat has been negated. Data are classified as MNAR if the reason for the missing data on a particular variable is directly attributable to the construct that variable reflects. That is, data are missing because the variable (or outcome) of interest carries information about why the variable is missing in the first place. These three distinct classes or mechanisms reflect the degree to which missingness may bias any statistical analysis; from MNAR through MAR to MCAR, in order of greatest to least concern. The missing data mechanism is considered ignorable for MCAR or MAR, but is nonignorable for MNAR (Rubin, 1976). In actual practice, data are rarely MCAR, with the primary distinction between MAR and MCAR reflecting whether additional variables under study are associated with missing data for a given variable (Rubin, 1976); it is difficult to distinguish between MAR and MNAR (McKnight et al., 2007). Fortunately, considerable advancements have been made with respect to statistical analysis in the presence of missing data (Enders, 2010).

Approaches for Dealing with Missing Data: A Brief Overview

Methodologists have developed modern statistical approaches that facilitate obtaining unbiased model estimates for incomplete datasets. Over the past few decades, imputation approaches have emerged as a popular approach for addressing missingness. Initially, approaches like mean or regression-based imputation were adopted. Mean imputation entails replacing a missing observation for a given variable with the sample mean, or with a person-level mean if longitudinal data are available. Although often employed, there are many concerns with this approach including the systematic reduction of observed variance for the mean-imputed variable, as well as biased parameter estimates. Although regression-based substitution represented an improvement, it is still a single imputation procedure that systematically underestimates variance. Such limitations led to the development of multiple imputation (MI) approaches that replace missing data with multiple possible values (5–10 or more; Schafer, 1999). There are many advantages to the MI approach including unbiased and precise estimation of parameters as well as its easy implementation in many modern statistical software packages. In contrast to single imputation approaches, the MI approach entails generating a distribution of estimates to replace missing values (for further details, see Allison, 2002; McKnight et al., 2007; Rubin, 1987; Schafer & Graham, 2002). The optimal number of MI estimates ranges from 3 to 10, with the estimates iteratively derived based on observed between- and within-person sources of variance. For example, if ten new estimates are derived via MI to replace missing values for a variable, then a corresponding number of new datasets

(i.e., 10) is generated—one new dataset per imputed value. Analyses of interest are then computed for each of these imputed datasets, with the corresponding parameter estimates obtained subsequently combined to derive a single best estimate. Whereas single imputation approaches tend to reduce variance in the observed variable and underestimate standard errors for parameter estimates, the multiple estimates involved in the MI approach permit more accurate estimates of standard errors and reduce Type I errors (McKnight et al., 2007).

Another approach for analyzing incomplete data involves likelihood-based estimation procedures, such as full information maximum likelihood (FIML) estimation. Unlike imputation-based approaches, FIML derives parameter estimates based upon all available information as opposed to complete (e.g., listwise deleted) or imputed data. Further, FIML will preferentially weight cases with greater numbers of observations (less missing data). Benefitting from a number of desirable statistical properties, maximum likelihood estimates are known to be consistent (are unbiased and converge on unknown true values of population parameters) and efficient (yield smaller standard errors), with normally distributed sampling distributions (Singer & Willett, 2003). In contrast however to MI approaches, likelihood-based approaches do require a correctly specified model to explain the structure of the data (Ferrer & Ghisletta, 2011), and are most appropriately employed on larger sample sizes (Singer & Willett, 2003). Despite the presence of missing data, approaches such as FIML use all available data (including all partial data) to produce estimates for various population parameters that maximize the probability of having observed patterns (e.g., aging-related rates of change in cognitive function) for the given sample under study. Maximum likelihood derived estimates of population parameters require the computation of a likelihood function to characterize the probability of observing associations in the sample data as a function of unknown model parameters (for further details, see Dempster, Laird, & Rubin, 1977; Singer & Willett, 2003). The process proceeds iteratively, with competing estimates compared until estimates are identified that maximize the log-likelihood function (i.e., the final estimates yield the greatest probability of having been observed given the sample data under study). When the difference between competing successive estimates is sufficiently small (i.e., the model converges), the final model estimates are identified. FIML assumes that missing data are MCAR or MAR (Rubin, 1987; Schafer, 1999), and thus requires valid inferences about the reasons for missingness. Thus, it is critical to examine differences between those individuals with complete versus missing data. Key questions to be addressed include whether any observed group differences are systematically related to variables under study. Further, regardless of whether an MI or FIML approach is employed, the inclusion of *auxiliary variables* can reduce: (i) bias by facilitating a closer approximation of the MAR assumption, (ii) marked variability in the imputed values, and (iii) standard errors of estimates derived for the final model (Allison, 2012). Auxiliary variables are not intended for inclusion in the final model, but are rather selected based upon their association with model-based variables with missing data. By including auxiliary variables in the imputation or modeling process, the resulting imputation or model-based estimates are conditioned upon the reasons for missingness (i.e., the auxiliary variables are associated with other variables under study that are related to missingness), thereby increasing the tenability of the MAR assumption and improving the quality of parameter estimates (Allison, 2012; Ferrer & Ghisletta, 2011; Graham, 2009). Virtually all quantitative analysis software packages include

likelihood-based estimation algorithms, which make them an accessible and attractive option for researchers.

Because longitudinal studies on adult development and aging typically involve attrition and missing data, both imputation- and likelihood-based estimation procedures are frequently employed. Either modern approach has proven superior to more traditional methods of listwise deletion or single imputation regression methods. However, although both MI and likelihood-based approaches benefit from similar statistical properties and make similar assumptions, some important differences should be noted. Allison (2012), for example, notes that MI approaches yield a distribution of results predicated upon the multiple random draws that are central to the MI process. How varied this distribution of results is depends upon the number of new MI datasets created. Whereas MI requires a decision about the number of random draws to be made, the maximum likelihood approach yields a single deterministic result. MI also requires a logical consistency between your analysis model and your imputation model; nuances in one model (e.g., interaction terms, transformed variables) should be reflected in the other (Allison, 2012). In contrast, FIML employs a single model, which may improve generalizability of findings.

Planned Missingness

To this point, we have introduced some of the analytic-based solutions for dealing with missing data from longitudinal studies that have already been conducted. Recently, Little and Rhemtulla (2013) have offered a design-based complement for missing data in longitudinal studies. Planned missingness designs involve the a priori specification of a study design such that participant data will be incomplete or "missing," but this missingness is determined in an a priori fashion and controlled by the researcher. Such designs are attractive as they reduce participant burden as well as the total volume of data collection and resources needed to field longitudinal studies. Recent research has provided empirical support for the successful use of planned missingness designs in developmental research with minimal loss of fidelity or statistical efficiency (Rhemtulla, Jia, Wu, & Little, 2014). Combined with the contemporary and advanced analytic techniques for accommodating missing data (e.g., MI and maximum likelihood approaches), planned missingness designs can be a powerful, efficient and attractive option for longitudinal research in aging and human development in general (Little, Jorgensen, Lang, & Moore, 2014).

Section Summary: Key Methodological Considerations for Incomplete Data

To summarize, reasons for missingness range from sampling selectivity during initial recruitment to attrition in longitudinal studies due to health or mortality. In order to minimize threats to internal validity, as well as to maximize both efficiency and consistency in the computation of model-based parameter estimates, the analyst should attend to several basic considerations. A step that is often ignored involves assessing patterns of missingness in one's data, as well as contemplating the feasibility of MCAR and MAR assumptions vis-à-vis the appropriateness of a specific analytic technique. With regard to assumptions regarding missingness classifications, some statistical packages (e.g., SAS, SPSS) have incorporated basic statistical tests, such as Little's (1998) MCAR test. A significant chi-square value associated with Little's test indicates that the data are not MCAR. Imputing missing data using MI approaches requires careful consideration of the imputation model and its correspondence with the planned statistical model. Similarly, likelihood-based approaches require that the model be appropriately specified and based

upon a sufficient number of cases to yield consistent and efficient estimates.

MODELING CHANGE IN STUDIES OF AGING

As a corollary to the discussion on longitudinal research designs, a corresponding increase in attention has been devoted to accompanying statistical models that examine the dynamic nature of both growth and decline associated with various aging processes. In the following section, we overview some basic analytic approaches for modeling both continuous and categorical outcomes, differentiate correlated from coupled change as foci in developmental analyses, and discuss the modeling of change based upon alternative parameterizations of developmental time.

Select Statistical Models for Change

Multilevel and Latent Growth Curve Approaches for Continuous Outcomes

Until several decades ago, most studies of developmental change for longitudinal panel data employed balanced research designs and general linear model (GLM) approaches such as repeated measures ANOVA. The experimental tradition at the time often resulted in longitudinal studies that failed to detect change due to limited sample size, the inclusion of few measurement occasions, compromised statistical power for detecting differences, and a differential focus on between-group differences as opposed to within-participant change (MacDonald et al., 2006). Among the shortcomings, these initial GLM approaches for assessing change focused on mean estimates aggregated across individuals, with the assumption that all individuals from a specific group were characterized by the very same pattern of (mean) change over time, and any deviation from this average assumed to reflect error.

Several vastly improved approaches are now typically employed to analyze change for continuous outcomes in adult development and aging (e.g., aging-related change in cognitive function). Both multilevel or linear mixed models of change (Raudenbush & Bryk, 2002; Singer & Willett, 2003) as well as latent growth curve (LGC; Willett & Sayer, 1994) approaches are commonly employed. These approaches consider both intraindividual change over time and interindividual differences in change over time (Baltes & Nesselroade, 1979). In addition to the linear analysis of continuous change, multilevel and growth curve models are also particularly well suited to the study of discontinuous developmental processes (Ram & Grimm, 2007; Singer & Willett, 2003). For example, in research on aging, it is of particular theoretical interest to contrast patterns of change both prior to and following critical events, such as the onset of menopause to gauge the impact of estrogen depletion on cognitive function (Thilers, MacDonald, Nilsson, & Herlitz, 2010), to differentiate normal from pathological cognitive aging by identifying the inflection point thought to indicate the onset of the prodromal phase of dementia (Thorvaldsson et al., 2011), or to disambiguate rates of longitudinal change in outcomes attributable to aging- versus mortality-, disease-, or disablement-related processes (Fauth, Gerstorf, Ram, & Malmberg, 2014; Gerstorf, Ram, Lindenberger, & Smith, 2013). Patterns of change prior to and following such critical events might be characterized quite differently, with both differences in the magnitude of change as well as the transition point for such differences of particular research interest (Cohen, 2008; Cudeck & Klebe, 2002). Thus, contemporary modeling frameworks provide considerable flexibility for examining developmental and other time-dependent processes.

Modern approaches have notable statistical advantages for the assessment of change (Hertzog & Nesselroade, 2003; MacDonald

et al., 2006). First, they do not assume equality of slopes across individuals, but rather empirically test this notion by including variance terms for various fixed effects (including change slopes) in the model. Another advantage is the ability to examine change despite heterogeneity in retest schedules. Further, both the multilevel and LGC approaches yield parameter estimates using FIML based upon all available information, assuming that missing data are MAR.

The mutlilevel and LGC approaches are similar in that both provide estimates of individual differences and change in performance, and indeed can be structured to be equivalent and to yield identical estimates (Curran, 2003; Ghisletta & Lindenberger, 2004). However, important differences should also be noted. For example, time is treated differently between the multilevel and LGC models, introduced as a level 1 predictor yielding a fixed effect in the former case, and incorporated into the model via the factor loadings for the latent slope variable for the latter. This represents a fundamental distinction: the treatment of time is univariate for multilevel models (time is parameterized as distinct observations for the same variable) versus multivariate for LGC models (each time point represents a distinct variable; Stoel, Van den Wittenboer, & Hox, 2003). Other advantages of LGC models including more flexible specifications of residual covariance structures, as well as simple extensions of LGC estimates of change to other outcomes within a broader SEM framework (Stoel et al., 2003). In contrast, multilevel models are advantageous for incorporating higher levels of nested structures (e.g., three-level structures common in measurement burst designs such as weekly sessions within annual retests within persons). On balance, the differences between the multilevel and LGC approaches are modest, with many modern software packages (e.g., Mplus; Muthén & Muthén, 2012) seamlessly estimating both statistical models of change.

Generalized Linear Mixed and Survival Models for Categorical Outcomes

Research applications for the psychology of aging are also based upon longitudinal responses that are not continuous (e.g., presence or absence of a disease process over time, counts of specific event occurrences such as stressors, etc.). As such, this requires separate models including *generalized linear mixed effects models* (Fitzmaurice, Laird, & Ware, 2004) as well as *survival models* (Singer & Willett, 2003). Generalized linear mixed effects models represent an extension to linear mixed models of continuous data where longitudinal categorical outcomes can be examined by transforming the mean response using a *link function* and then relating the transformed outcome to predictors. Appropriate selection of the link function transformation of the non-normal outcome depends upon the distribution of the outcome data (e.g., a logit transform for binary data characterized by a Bernoulli distribution). The transformed outcome can then be predicted by covariates of interest using the familiar GLM; effectively, continuous models with normal distributions are simply special cases of GLMs (see Fitzmaurice et al., 2004, for further reading).

Survival (or event history) analysis models the risk of a particular event occurrence (e.g., disease onset, death) as a function of specific predictors in your model. In a longitudinal analysis, this risk of event occurrence is referred to as a hazard—the probability that an individual will experience the event within a period of time. A key feature of survival models is the ability to consider both event occurrence and time-to-event occurrence. In psychological aging research, survival models have been commonly used for the study of disease risk, as well as for the study of terminal decline that examines an accelerated decline or drop in cognitive function in proximity to death (MacDonald, Hultsch, & Dixon, 2011). The terminal decline hypothesis has been examined using both conventional survival methods (Schaie, 1989), as

well as modern analytic approaches that combine the statistical analysis of change (e.g., linear mixed models) with survival models (Ghisletta, McArdle, & Lindenberger, 2006). Such joint modeling approaches have also been employed to examine how individual differences in levels and rates of change in cognitive function from LGCs are related to onset of Alzheimer's disease (McArdle, Small, Bäckman, & Fratiglioni, 2005).

Correlated and Coupled Change

The analysis of change goes beyond a simple decision regarding which type of analytic approach to employ. Indeed, the primary research question itself has an important bearing on the nature of change examined. Multilevel and LGC approaches offer flexibility for examining change in one outcome, and potential moderators or sources of individual or group differences in rates of change (i.e., interindividual differences in intraindividual change). However, such approaches can be expanded to consider scenarios where researchers might be interested in how two or more variables may be changing together over time. As such, multivariate approaches allow the consideration of how variables and rates of change in these variables are related over time.

To facilitate a more stringent test of a developmental hypothesis, a researcher might be interested in examining whether two processes (cognitive and physiological function) change together within an individual over time. In order to examine the time-varying covariation between these two processes, a researcher could explore either *correlated* or *coupled* change between physiological and cognitive function (Sliwinski & Mogle, 2008). These two approaches actually address disparate questions. By way of example, one could correlate two separate slopes of aging-related change—one for cognitive and one for physiological function. Such an analysis would yield insight regarding *correlated change*—the extent to which individuals' whose cognitive function is changing at a faster rate is also exhibiting faster rates of change in physiological function. Alternatively, one could examine the time-varying covariation of cognitive and physiological function after taking the longitudinal trends for each into account. Such an analysis would yield insights into *coupled change*—the extent to which an individual's level of cognitive function at a particular sampling occasion is related to their level of physiological function at the same sampling occasion. Here, it is important to note that correlated change involves the examination of individual differences (or between-person associations), whereas coupled change involves the examination of intraindividual differences (or within-persons associations). The approaches often yield similar estimates, but may diverge due to aggregation bias in longitudinal research (Sliwinski & Buschke, 1999).

Although analyses of correlated and coupled change address complementary questions about how variables are related over time, such approaches are ultimately correlational and preclude inferences about causation or lead or lagged effects. Models such as the bivariate dual change score model (McArdle & Hamagami, 2001) represent an analytic alternative that incorporates both the longitudinal modeling of two variables, as well as lead and lag parameters to allow for the rate of change in one variable to be prospectively predictive in the other variable (and vice versa). Such an approach affords researchers the ability to examine individual differences in and correlations between rates of change among variables, as well as how individual differences in rates of change among variables can be antecedent to each other to reveal unidirectional and/or bidirectional causal influence among variables.

Developmental Parameterizations of Time

A critical issue for longitudinal studies on adult development and aging concerns how we

define the time continuum used for characterizing change. In this section, we highlight examples from the literature that demonstrate how employing various parameterizations of time can influence results observed in longitudinal investigations (cf. Morrell, Brant, & Ferrucci, 2009).

Is Chronological Age the Only Metric?

Despite considerable advances in research designs (Stawski et al., in press) and statistical procedures (McArdle, 2009) for the study of the psychology of aging, chronological age perseveres as arguably the most used predictor and developmental time metric for charting performance differences and changes (MacDonald, DeCarlo, & Dixon, 2011). Despite this popularity, the weaknesses of age as a developmental index have been well documented (Birren, 1999; Dixon, 2011). Specifically, rather than a causal mechanism underlying cognitive and functional decline, chronological age is said to merely reflect a temporal dimension along which causal factors (e.g., biological, environmental, health, and neurological) operate (MacDonald, Karlsson, Fratiglioni, & Bäckman, 2011). Consequently, observing that chronological age is associated with performance decline (e.g., in cognition) does not inform the specific or general mechanisms underlying age-related cognitive impairment—rather, age is likely an indirect reflection (i.e., a proxy) of true mechanistic changes (e.g., accumulated biological and environmental factors) that influence cognition across time.

Beyond these theoretical concerns with the use of chronological age as the primary developmental time metric for charting change, the selection of a specific time parameterization is known to influence the interpretation of results in longitudinal studies. In particular, for longitudinal studies characterized by considerable age heterogeneity in the sample at baseline assessment, opting to model long-term change using chronological age as the time basis without accounting for differences in age at study entry (i.e., different age cohorts were sampled) assumes the equivalence (or convergence) of cross-sectional and longitudinal aging effects (Morrell et al., 2009; Sliwinksi, Hoffman, & Hofer, 2010). The advantage of such a model is that a single trajectory of change spanning the entire observed age range (e.g., 60–90 years) can be estimated—a combination of age information spanning various cohorts that is much larger than the range measured over the longitudinal follow-up (e.g., three retests spanning a 6-year period; cf. Singer & Willett, 2003). Of course, such failures of age convergence—the assumption that cross-sectional age differences and longitudinal age changes converge onto a common trajectory—are well documented in the adult development and aging literature (Hoffman & Stawski, 2009; Sliwinksi et al., 2010). The impact that between-person difference in age at baseline can exert upon appropriate inferences regarding within-person change has led researchers to consider additional parameterizations of developmental time.

Alternative Parameterizations of Time

In reviewing recent literature in adult development and aging, it is not uncommon to observe the use of various time parameterizations for modeling within-person change in various aging processes. Most longitudinal research parameterizes developmental time using three *time basis structures*: chronological age (e.g., years since birth), measurement occasion (e.g., 0, 1, 2), and time in study (years from baseline assessment) (Morrell et al., 2009).

The *age-as-time* parameterization estimates within-person change as a function of chronological age. However, as mentioned in the preceding section, such an approach not only assumes age convergence, but may also fail to capture important sources of heterogeneity. Variance due to underlying health conditions such as cardiovascular disease, for example, may be misattributed to chronological

age (Spiro & Brady, 2008). Additionally, using age-as-time can reveal complex non-linear trends that potentially reflect cross-sectional mean differences introduced by heterogeneity in age at baseline (Hofer & Sliwinski, 2001).

Employing measurement occasion as a time basis is common, but this approach fails to capture important individual differences in time sampling. Specifically, such an approach fails to capture important variation in individual retest intervals; regardless of whether the first retest interval spanned 6 months for some individuals versus 16 months for others, the time as measurement occasion parameterization would treat these as equivalent. Such an assumption may be entirely reasonable for certain processes and populations under study, but may be grossly inefficient in other contexts. When examining changes in cognitive function during the prodromal phase of dementia, for example, changes across even relatively short retest intervals can be meaningful (i.e., individual differences of even several months matter).

To improve precision, many opt to parameterize time using a *time-in-study* metric. This approach charts time elapsed from the beginning of the study, including a precise incorporation of individual differences in retest intervals. Morrell et al. (2009) directly compared such competing time parameterizations, and report that accurate inference about within-person change in longitudinal studies is best accomplished by parameterizing time as *time-in-study* as well as by including a between-subject term indexing age heterogeneity at entry into the study (baseline assessment). Moreover, this approach allows for explicit examination of whether rates of aging-related longitudinal change vary as a function of cross-sectional age-related differences at baseline (e.g., is the rate of decline in memory faster among individuals who were older upon entry into the study?).

Recently, *time-to-event* (or time-as-process) parameterizations have been proposed as another promising developmental metric for indexing change (Sliwinski & Mogle, 2008). A time-to-event approach indexes change in relation to the onset of central events or processes. Biological age represents one such alternative marker of developmental time, with recent empirical findings and theorizing linking biological processes (e.g., vascular health) to age-related cognitive decline (DeCarlo, Tuokko, Williams, Dixon, & MacDonald, 2014; MacDonald et al., 2011). Other examples include examining change in cognitive function in relation to specific processes including dementia progression or menopause. For the former, one recent study explored within-person change in cognition as a function of years following dementia diagnosis (MacDonald et al., 2011)—understanding why some individuals progress through the dementia stages more rapidly than others could inform efforts to lessen caregiver burden or to target effective drug trials. For the latter, another study examined cognitive change as a function of time prior to and following menopause (Thilers et al., 2010). Structuring cognitive change in relation to the menopause event failed to support claims that estrogen depletion for postmenopausal women leads to cognitive decline.

Beyond cognitive domains, time-to-event approaches have been employed to examine longitudinal changes in mental and physical health with respect to disablement- (Fauth et al., 2014), and mortality- (Gerstorf et al., 2013) related processes. Echoing the findings of terminal decline in cognitive function using time-to-event approaches, these recent research findings indicate that mental and physical health exhibit accelerated decline proximal to specific health-related events. This suggests that time-to-event-based examinations are incredibly valuable for articulating processes other than aging that impact mental, physical and cognitive health during the later years, and there is considerable promise for applying such approaches for understanding multiple

dimensions and processes in the context of later life (Gerstorf & Ram, 2013).

EMERGING METHODOLOGICAL TRENDS FOR THE STUDY OF AGING

In this final section, we overview several trends that are exerting considerable influence on the current scope of adult development and aging research—the integrated data analysis and intraindividual variability approaches.

Select Approaches to Integrated Data Analysis

Over the past decade, a clear trend in research on the psychology of aging has seen innovative efforts to comprehensively integrate data *across* studies. Such integration affords a number of advantages including improved statistical power, improved precision of parameter estimation, and, perhaps most importantly, the testing of whether findings from single sample studies are generalizable. Several approaches to integrated data analysis are overviewed here.

Meta-Analysis

The meta-analytic approach to integrated data analysis involves synthesizing various summary statistics (effect sizes, regression coefficients, probability values) across individual studies that share a number of important similarities (Cumming, 2012). In effect, a meta-analysis is an aggregate study about a number of prior studies. This accumulation of evidence yields a quantitative summary of findings, including an omnibus measure of effect size (e.g., Hedges' *g*) as well as the identification of key variables that moderate this effect and explain variation between studies. The degree to which summary statistics from individual studies influence the final measure of effect can be weighted according to important factors (e.g., sample size). Key steps involved in conducting a meta-analysis include deciding upon the target research questions, deciding upon the parameters governing the literature search and choosing studies that clearly meet criteria, requesting data from researchers as required, addressing incomplete data, and data analysis (Sternberg, Baradaran, Abbott, Lamb, & Guterman, 2006). There are important assumptions when conducting a meta-analysis, including the need to ensure the psychometric comparability of pooled constructs, measures, and measurement scales. Advantages of the meta-analytic approach over individual studies include the improvement of statistical power and precision, a direct means of addressing equivocal findings within a research field (e.g., are opposing findings due to systematic between-study differences), as well as drawing inference regarding generalizability of findings (Cumming, 2012). However, meta-analytic approaches are also subject to notable concerns including the impact of publication bias (as many nonsignificant findings are not represented in the literature) as well as the ecological fallacy, where inaccurate inferences about individuals are based upon population mean trends (Hofer & Sliwinski, 2001; Sternberg et al., 2006).

Mega-Analysis

In contrast to a meta-analytic approach, mega-analytic studies derive similar quantitative benefits through the actual pooling of raw data across many studies/samples (Cooper & Patall, 2009; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). The distinguishing feature between meta- versus mega-analysis concerns the type of information that is concatenated across studies. For meta-analysis, various summary statistics are compiled, whereas the actual raw data are concatenated across study for the mega-analysis. In this sense, a mega-analysis shares greater similarity with a conventional research study where the researcher collects and analyzes

person-level data from multiple research contexts. Advantages of the mega-analytic approach include the ability to conduct a meaningful investigation even when few studies are available (a distinct problem for meta-analysis), improved reliability and precision of model-based parameter estimates, and perhaps most importantly an increased flexibility with regard to research questions that can be pursued and analytic techniques that can be employed (Sternberg et al., 2006). While there are limitations involved with the mega-analytic approach, such as the need for identical measures across studies, it is certainly a powerful and attractive approach for large-scale examinations of research questions.

Data Harmonization

Data harmonization, whether retrospective or prospective, represents another approach to pooled data analysis. In contrast to a mega-analysis, a *retrospective* harmonized dataset reflects more than the mere pooling of raw data across studies; rather, the harmonized dataset derives novel variables and constructs based upon complex harmonization procedures (cf. Anstey et al., 2009). Not only are raw data concatenated across studies, but entirely new variables are generated to index constructs of interest. For example, in the Dynamic Analyses to Optimize Ageing (DYNOPTA) project, Anstey et al. (2009) created a retrospectively pooled dataset spanning nine Australian Longitudinal Studies of Ageing, and applied harmonization procedures to create new variables to facilitate comparison with clinically meaningful scores (e.g., harmonized data on physical activity to reflect national recommendations for weekly participation levels).

Recently, efforts have been championed toward *prospective* harmonization across studies, where across-study consensus guidelines are adopted that govern *new* data collection and measurement. For example, some international research on dementia has adopted prospective guidelines for the selection of standardized measures (e.g., the measurement of biomarkers like CSF, common neuroimaging protocols) across studies (Frisoni, 2010). Similarly, the U.S. National Institutes of Health has developed the NIH Toolbox, a common battery of performance measures (cognition, emotion, motor, and sensory function), as a standardized testing platform that can be administered across studies of human development and aging (Bauer & Zelazo, 2014). Although promising, the process of data harmonization is laborious, with numerous challenges including common across-study variation in how constructs have been measured, the development and application of harmonization methods to facilitate comparability, and the retrospective versus prospective nature of the harmonization initiative (Erten-Lyons, et al., 2012).

Coordinated Analysis with Replication

The coordinated and integrated analysis of original data from multiple studies can augment scientific knowledge through the *replication* and *extension* of key findings. For the study of aging, the process involves identifying central research questions, conducting *parallel analyses* across multiple studies to ascertain whether the effect(s) of interest can be replicated, and interpreting similarities (or differences) across patterns of results to further inform generalizability and theory development (Curran & Hussong, 2009; Hofer & Piccinin, 2009). With particular regard to longitudinal studies, data pooling approaches can be problematic. For example, meta-analytic approaches can be quite limited by the body of longitudinal research published on particular research questions, as well as the types of research designs and analyses employed. Similarly, pooled analysis approaches (mega-analyses) are often limited by the lack of overlap of specific measures across studies, requiring more involved harmonization efforts. In these instances, a coordinated analysis

platform is particularly advantageous, such as the Integrative Analysis of Longitudinal Studies of Aging (IALSA) project consisting of over 40 longitudinal studies of aging (Hofer & Piccinin, 2009). IALSA facilitates access to member studies data, analysis scripts, and output, with the strengths (e.g., immediate replication of novel findings in the literature and/or consideration of alternative hypotheses, generalizability of findings, improved statistical power) of such coordinated efforts having furthered our knowledge of the psychology of aging and its associated theories.

An Intraindividual Variability Approach

Beyond the First Order Moment

Recently, aging theorists and developmentalists alike have demonstrated a renewed interest on approaches for studying intraindividual variability for a host of domains in aging and across the lifespan (for a comprehensive overview, see Diehl, Hooker, & Sliwinski, 2015). The reemergence of variability derives from a growing body of evidence demonstrating that short-term fluctuations often reflect more than random error or measurement unreliability, are systematically associated with numerous developmental outcomes, and are informative vis-a-vis theories of processing dynamics (MacDonald & Stawski, 2015; Ram & Gerstorf, 2009; Stawski et al., 2015). Intraindividual variability, however, does not necessarily reflect a psychological primitive (e.g., processing speed) per se, but rather *an approach* to the study of adult development and aging that facilitates the examination of dynamic fluctuations in function that confer *meaning beyond mean* and static considerations. To be certain, the examination of mean remains a central focus, contributing essential information for characterizing behavior over time. However, recent findings have demonstrated that this knowledge should be supplemented by also asking how variable this performance is over time. Do trajectories of performance reflect mean stability characterized by modest or substantial variability? Variability not only contributes unique information independent of mean (cf. MacDonald & Stawski, 2015), but it also improves our understanding of the dynamic nature of the developmental process under study (Ram & Gerstorf, 2009). The following subsections overview examples of variability research from the adult development and aging literature.

RT Inconsistency Across Response Latency Trials

Growing consensus from various scientific disciplines including lifespan psychology, cognitive neuroscience, neuropsychology, and mathematical modeling suggests that theoretically interesting aspects of cognitive function are not completely captured by mean performance (Garrett, Kovacevic, McIntosh, & Grady, 2010). RT inconsistency, as defined by trial-to-trial fluctuations in RT latencies on performance-based measures of cognition, has emerged as one index thought to capture important features of behavioral and systemic integrity (MacDonald & Stawski, 2015), including mental noise (Van Gemmert & Van Galen, 1997), transient lapses of attention (West, Murphy, Armilio, Craik, & Stuss, 2002), and a more enduring behavioral signature of compromised brain and neural function (MacDonald, Karlsson, Rieckmann, Nyberg, & Bäckman, 2012) and cognitive status (Dixon et al., 2007). Numerous indices may be computed to index intraindividual variability across response latency trials. Among the simplest, the intraindividual standard deviation (ISD) can be computed within persons and across trials to index fluctuations in response latencies. In order to adjust for potential confounds (e.g., individual differences in average level of performance, response speed, or systematic learning over time), effects associated with age, disease status, and trial can

be partialled (Hultsch, Strauss, Hunter, & MacDonald, 2008). Other operationalizations of variability that may be considered include the coefficient of variation (each person's SD/M), high versus low percentiles from RT distributions (Hultsch et al., 2008), as well as approaches that simultaneously model mean and variability (Schmiedek, Lövdén, & Lindenberger, 2009).

Disambiguating state- and trait-like variation in dynamic characteristics and processes within measurement burst designs. Measurement burst designs (Rast et al., 2012; Sliwinski, 2008; Stawski et al., in press), which involve assessing individuals intensively over shorter intervals (e.g., across trials or over days), and repeating this intensive assessment longitudinally over longer intervals (e.g., over months or years), are ideal for examining and quantifying variability within persons across shorter or longer time horizons, or between persons. Evidence of significant intraindividual variability and interindividual differences in RT inconsistency would suggest that RT inconsistency exhibits both state-like and trait-like variation, and that this variation is potentially related to time-varying and time-invariant predictors (Hoffman & Stawski, 2009). Most research has focused on RT inconsistency as a trait-like characteristic (e.g., changes in CNS, changes in brain structure/function or underlying disease/pathology) for differentiating individuals and/or subgroups, ignoring potentially important variation that may exist within persons across shorter intervals such as days, weeks, or months that may reflect variation in a person's context (e.g., increased stress, distress, diminished sleep, attention).

In future research, the analysis of indices of variability for various processes (e.g., cognition, gait, neural function) in the context of measurement burst designs represents a novel empirical approach to examining both the state-like and trait-like modulators of performance fluctuation. Empirical decomposition of variation in RT inconsistency, for example, will help better understand the utility of RT inconsistency as a behavioral indicator of cognitive, brain and CNS function, and may facilitate identification of risks (e.g., falls, delirium) for individuals with dementia. Consistent with this proposition, recent research on daily stress employing a measurement burst design has shown that among older adults, only 25% of the variability in emotional reactivity to daily stressors reflects individual differences or dispositional variation. This suggests that dynamic processes, in and of themselves, may be susceptible to vicissitudes of other time-varying processes or influences that operate at difference timescales (Schmiedek, Lovden, & Lindenberger, 2013), and underscores the need to examine factors beyond individual and group differences as important sources of variation in dynamic processes.

Linking intraindividual variability to long(er)-term outcomes using intensive repeated measures and measurement burst designs. The vast majority of research in cognition and aging has focused on the utility of intraindividual variability as a more static indicator and proxy for dynamic processes that are reflected in behavior/performance. Intraindividual variability has a rich history in other domains, particularly affect and emotion, whereby intraindividual variability is thought to reflect, in part, the systematic impact of contextual and experiential forces such as stressful experiences (Almeida, 2005). Thus, intensive repeated measures designs (e.g., ecological momentary assessment and daily diaries) can be exploited to examine the time-varying covariation of stressors and affect as a way to examine individuals' emotional reactivity to stressors (and individual differences therein) as dynamic phenomena. Importantly, these intensive repeated measures design protocols can be repeated at periodic intervals, effectively yielding a measurement burst study. Such an approach is attractive as it allows for examining a dynamic process (e.g., emotional reactivity to stressors), how that process

changes developmentally, and how that process impacts change in other outcomes of interest.

For example, theoretical accounts of the impact of stress on long-term health have emphasized the importance of dynamic, micro-level processes including stressor reactivity as the mechanism underlying the stress–health link (Cacioppo, 1998). In the daily stress literature, links between micro-level stress reactivity processes and long-term health outcomes have only recently been explored. Mroczek et al. (2013) showed that greater emotional reactivity to daily stressors was associated with an increased likelihood of mortality among older men. Similarly, individual differences in emotional reactivity to daily stressors have been linked to increased distressed affect and self-reported affective disorders (Charles, Piazza, Mogle, Sliwinski, & Almeida, 2013) and increased risk of chronic health conditions (Piazza, Charles, Sliwinski, Mogle, & Almeida, 2012) 10 years later. Additionally, Sliwinski, Almeida, Smyth, and Stawski (2009) reported that emotional reactivity to daily stressors increases longitudinally across 2.5 and 10 years from two separate measurement burst studies of midlife and old age. These recent findings from the stress and affect literature exemplify the promise measurement burst designs hold for examining the longitudinal dynamics and impact of fast-acting processes.

CONCLUSIONS

In writing this chapter, our goal was to selectively highlight methodological considerations and concerns that characterize current research on the psychology of aging. The overview of sampling and design considerations emphasized missing data considerations and retest effects, as well as their corresponding impact on model-based parameter estimates and (in) accuracy of inferences drawn. In particular, we emphasized the strengths of the measurement burst design. Such intensive measurement designs hold real promise for improving our understanding of dynamic aging-related processes, including current trends such as whether intraindividual variability reflects both state-like and trait-like influences. We reviewed common analytic approaches for analyzing change in both continuous (LGC, multilevel models) and categorical (survival) outcomes, as well as emphasized the need to carefully consider alternative parameterizations of developmental time to chronological age. Finally, we concluded by exploring some emerging trends in the study of the psychology of aging, including the promise of integrated data analysis for informing the key scientific issues of generalizability and theory development. The advances in design and analysis and their corresponding recent applications have given rise to an exciting time for research in the psychology of aging, as we strive to further our understanding of dynamic developmental processes.

Acknowledgments

This chapter was supported by grants from the National Institutes of Health/National Institute on Aging (R21AG045575) and the Natural Sciences and Engineering Research Council of Canada (418676-2012). Stuart MacDonald also acknowledges support from the Michael Smith Foundation for Health Research (Career Investigator Scholar Award).

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Allison, P. D. (2012). Handling missing data by maximum likelihood. *SAS Global Forum 2012 Proceedings*, 312–2012.

Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*, *14*(2), 62–68.

Anstey, K. J., Byles, J. E., Luszcz, M. A., Mitchell, P., Steel, D., Booth, H., et al. (2009). Cohort profile: The dynamic analyses to optimize ageing (DYNOPTA) project. *International Journal of Epidemiology*, *39*, 44–51.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study*

of behavior and development (pp. 1–39). New York: Academic Press.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Lifespan developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole.

Bauer, P. J., & Zelazo, P. D. (2014). The national institutes of health toolbox for the assessment of neurological and behavioral function: A tool for developmental science. *Child Development Perspectives*, *8*, 119–124.

Birren, J. E. (1999). Theories of aging: A personal perspective. In V. L. Bengtson & K. W. Schaie (Eds.), *Handbook of theories of aging*. New York, NY: Springer.

Cacioppo, J. T. (1998). Somatic responses to psychological stress: The reactivity hypothesis. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science: Biological and cognitive aspects* (Vol. 2, pp. 87–112). Hove, England: Psychology Press/Erlbaum (UK).

Charles, S. T., Piazza, J. R., Mogle, J., Sliwinski, M. J., & Almeida, D. M. (2013). The wear and tear of daily stressors on mental health. *Psychological Science*, *24*(5), 733–741.

Cohen, P. (Ed.). (2008). *Applied data analytic techniques for turning points research*. New York, NY: Routledge.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, *14*, 165–176.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*, 41–63.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Curran, P. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavior Research*, *38*(4), 529–569.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

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

DeCarlo, C. A., Tuokko, H. A., Williams, D., Dixon, R. A., & MacDonald, S. W. S. (2014). BioAge: Toward a multidetermined, mechanistic account of cognitive aging. *Ageing Research Reviews*, *18*, 95–105.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Diehl, M., Hooker, K., & Sliwinski, M. (Eds.), (2015). *Handbook of intraindividual variability across the life span*. New York, NY: Routledge.

Dixon, R. A. (2011). Enduring theoretical themes in psychological aging: Derivation, functions, perspectives, and opportunities. In K. W. Schaie & S. L. Willis (Eds.), *Handbook of the psychology of aging* (pp. 3–23) (7th ed). San Diego, CA: Elsevier.

Dixon, R. A., Garrett, D. D., Lentz, T. L., MacDonald, S. W. S., Strauss, E., & Hultsch, D. F. (2007). Neurocognitive resources in cognitive impairment: Exploring markers of speed and inconsistency. *Neuropsychology*, *21*(3), 381–399.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Erten-Lyons, D., Sherbakov, L. O., Piccinin, A. M., Hofer, S. M., Dodge, H. H., Quinn, J. F., et al. (2012). *Alzheimers and Dementia*, *8*, 584–589.

Fauth, E. B., Gerstorf, D., Ram, N., & Malmberg, B. (2014). Comparing changes in late-life depressive symptoms across aging, disablement, and mortality processes. *Developmental Psychology*, *50*, 1584–1593.

Ferrer, E., & Ghisletta, P. (2011). Methodological and analytical issues in the psychology of aging. In K. W. Schaie & S. L. Willis (Eds.), *Handbook of the psychology of aging* (pp. 25–39) (7th ed). Amsterdam, the Netherlands: Elsevier.

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging*, *19*, 243–259.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Frisoni, G. B. (2010). Alzheimer's disease neuroimaging initiative in Europe. *Alzheimer's and Dementia*, *6*(3), 280–285. [PubMed: 20451877]

Garrett, D. D., Kovacevic, N., McIntosh, A. R., & Grady, C. L. (2010). Blood oxygen level dependent signal variability is more than just noise. *The Journal of Neuroscience*, *30*(14), 4914–4921.

Gerstorf, D., & Ram, N. (2013). Inquiry into terminal decline: Five objectives for future study. *The Gerontologist*, *53*, 727–737.

Gerstorf, D., Ram, N., Lindenberger, U., & Smith, J. (2013). Age and time-to-death trajectories of change in indicators of cognitive, sensory, physical, health, social, and self-related functions. *Developmental Psychology*, *49*, 1805–1821.

Ghisletta, P., & Lindenberger, U. (2004). Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology*, *50*, 12–16.

Ghisletta, P., McArdle, J. J., & Lindenberger, U. (2006). Longitudinal cognition-survival relations in old and very old age: 13-year data from the Berlin Aging Study. *European Psychologist*, *11*, 204–223.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Gustavson, K., von Soest, T., Karevold, E., & Roysamb, E. (2012). Attrition and generalizability in longitudinal studies: Findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health*, *12*, 918.

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. v. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*, 244–252.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639–657.

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, *14*, 150–164.

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

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, *6*, 97–120.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged*. New York, NY: Cambridge University Press.

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 491–556, 3rd ed). New York, NY: Psychology Press.

Little, R. J. A. (1998). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Little, R. T. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. (2014). On the joys of missing data. *Journal of Pediatric Psychology*, *39*, 151–162.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*, 199–204.

MacDonald, S. W. S., DeCarlo, C. A., & Dixon, R. A. (2011). Linking biological and cognitive aging: Toward improving characterizations of developmental time. *The Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*, *66B*(S1), i59–i70.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2011). Aging and the shape of cognitive change before death: Terminal decline or terminal drop? *Journal of Gerontology: Psychological Sciences*, *66*, 292–301.

MacDonald, S., Hultsch, D., Strauss, E., & Dixon, R. (2003). Age-related slowing of digit symbol substitution revisited: What do longitudinal age changes reflect? *Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*, *58B*, P187–P194.

MacDonald, S. W. S., Karlsson, S., Fratiglioni, L., & Bäckman, L. (2011). Trajectories of Cognitive decline following dementia onset: What accounts for variation in progression. *Dementia and Geriatric Cognitive Disorders*, *31*, 202–209.

MacDonald, S. W. S., Karlsson, S., Rieckmann, A., Nyberg, L., & Bäckman, L. (2012). Aging-related increases in behavioral variability: Relations to losses of dopamine D1 receptors. *The Journal of Neuroscience*, *32*(24), 8186–8191.

MacDonald, S. W. S., & Stawski, R. S. (2015). Intraindividual variability—an indicator of vulnerability or resilience in adult development and aging?. In M. Diehl, K. Hooker, & M. Sliwinski (Eds.), *The handbook of intraindividual variability across the life span* (pp. 231–257). New York, NY: Routledge.

MacDonald, S. W. S., Stigsdotter Neely, A., Derwinger, A., & Bäckman, L. (2006). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology: General*, *3*, 368–390.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605.

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

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.

McArdle, J. J., Small, B. J., Bäckman, L., & Fratiglioni, L. (2005). Longitudinal models of growth and survival applied to the early detection of Alzheimer's Disease. *Journal of Geriatric Psychiatry and Neurology*, *18*, 234–241.

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest design to include developmental time-lag components. *Psychological Methods*, *2*, 403–435.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: The Guilford Press.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218.

Morrell, C. H., Brant, L. J., & Ferrucci, L. (2009). Model choice can obscure results in longitudinal studies. *Journal of Gerontology: Medical Sciences*, *64A*, 215–222.

Mroczek, D. K., Stawski, R. S., Turiano, N. A., Chan, W., Almeida, D. M., Neupert, S. D., et al. (2013). Emotional reactivity and mortality: Longitudinal findings from the VA normative aging study. *Journals of Gerontology* http://dx.doi.org/10.1093/geronb/gbt107.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palermo (Eds.), *Visions of Aesthetics, the Environment, and Development: The Legacy of Joachim F. Wohwill* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nesselroade, J. R., & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. New York, NY: Academic Press.

Neupert, S. D., Stawski, R. S., & Almeida, D. M. (2008). Considerations for sampling time in aging research. In S. M. Hofer & D. F. Alwin (Eds.), *The handbook of cognitive aging: Interdisciplinary perspectives* (pp. 492–505). Thousand Oaks, CA: Sage Publications.

Nilsson, L.-G., Sternäng, O., Rönnlund, M., & Nyberg, L. (2009). Challenging the notion of an early-onset of cognitive decline. *Neurobiology of Aging*, *30*, 521–524.

Piazza, J., Charles, S., Sliwinski, M., Mogle, J., & Almeida, D. (2012). Affective reactivity to daily stressors and long-term risk of reporting a chronic physical health condition. *Annals of Behavioral Medicine*, *45*, 110–120.

Rabbitt, P., Diggle, P., Smith, D., Holland, F., & McInnes, L. (2001). Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. *Neuropsychologia*, *39*, 532–543.

Radler, B. T., & Ryff, C. D. (2010). Who participates? Accounting for longitudinal retention in the MIDUS National Study of Health and Well-Being. *Journal of Aging and Health*, *22*, 307–331.

Ram, N., & Gerstorf, D. (2009). Time structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*, 778–791.

Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching method to theory. *International Journal of Behavioral Development*, *31*, 303–316.

Rast, P., MacDonald, S. W. S., & Hofer, S. M. (2012). Intensive measurement designs for research on aging. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, *25*, 45–55.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2014). Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development*, *38*, 423–434.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review*, *15*, 351–357.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L. -G. (2005). Stability, growth, and decline in adult life-span development of declarative memory: Cross-sectional and longitudinal data from a population-based sample. *Psychology and Aging*, *20*, 3–18.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Salthouse, T. A. (1996). General and specific speed mediation of adult age differences in memory. *Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*, *51B*, P30–P42.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*, 507–514.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, *8*, 3–15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schaie, K. W. (1989). The hazards of cognitive aging. *The Gerontologist*, *29*, 484–493.

Schaie, K. W. (2013). *Intellectual development in adulthood: The seattle longitudinal study* (2nd Ed). New York, NY: Oxford University Press.

Schaie, K. W. (2009). When does age-related cognitive decline begin? Salthouse again reifies the cross-sectional fallacy. *Neurobiology of Aging*, *30*, 528–529.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging*, *24*(4), 841–857.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2013). Keeping it steady: Older adults perform more consistently on cognitive tasks than younger adults. *Psychological Science*, *24*(9), 1747–1754.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass*, *2*, 245–261.

Sliwinski, M., & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, memory and processing speed. *Psychology and Aging*, *14*, 18–33.

Sliwinksi, M., Hoffman, L., & Hofer, S. M. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Research in Human Development*, *7*, 45–60.

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 of cognitive aging: Interdisciplinary perspectives* (pp. 477–491). Thousand Oaks, CA: Sage.

Sliwinski, M. J., Almeida, D. M., Smyth, J. M., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: Findings from two diary burst studies. *Psychology and Aging*, *24*, 828–840. PMCID: PMC2857711.

Sliwinski, M. J., Smyth, J. M., Stawski, R. S., & Wasylyshyn, C. (2005). Stress and working memory: Between-person and within-person relationships. In R. W. Engle, G. Sedek, U. von Ecker, & D. McIntosh (Eds.), *Cognitive limitations in aging and psychopathology: attention, working memory, and executive functions* (pp. 73–93). Cambridge, MA: Cambridge University Press.

Spiro, A., III, & Brady, C. B. (2008). Integrating health into cognitive aging research and theory: Quo vadis?. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 260–282). Thousand Oaks, CA: Sage.

Stawski, R. S., MacDonald, S. W. S., & Sliwinski, M. J. (in press). Measurement burst design. In: S.K. Whitbourne (Ed.), *Encyclopedia of adulthood and aging*. New York, NY: Wiley.

Stawski, R. S., Sliwinski, M. J., & Hofer, S. M. (2013). Between-person and within-person associations among processing speed, attention switching, and working memory in younger and older adults. *Experimental Aging Research.*, *39*, 194–214.

Stawski, R. S., Smith, J., & MacDonald, S. W. S. (2015). Intraindividual variability and covariation across domains in adulthood and aging: Contributions for understanding behavior, health and development. In M. Diehl, K. Hooker, & M. Sliwinski (Eds.), *The handbook of intraindividual variability across the life span* (pp. 258–279). New York, NY: Routledge.

Sternberg, K. J., Baradaran, L. P., Abbott, C. B., Lamb, M. E., & Guterman, E. (2006). Type of violence, age, and gender differences in the effects of family violence on children's behavior problems: A mega-analysis. *Developmental Review*, *26*, 89–112.

Stoel, R. D., Van den Wittenboer, G., & Hox, J. J. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodologia de las Ciencas del Comportamiento*, *5*, 21–42.

Thilers, P. P., MacDonald, S. W. S., Nilsson, L. -G., & Herlitz, A. (2010). Accelerated postmenopausal cognitive decline is restricted to women with normal BMI: Longitudinal evidence from the Betula Project. *Psychoneuroendocrinology*, *35*, 516–524.

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*, 348–354.

Thorvaldsson, V., MacDonald, S. W. S., Fratiglioni, L., Winblad, B., Kivipelto, M., Laukka, E. J., et al. (2011). Onset and rate of cognitive change before dementia diagnosis: Findings from two Swedish population-based longitudinal studies. *Journal of the International Neuropsychological Society*, *17*, 154–162.

Van Gemmert, A. W. A., & Van Galen, G. P. (1997). Stress, neuromotor noise, and human performance: A theoretical perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1299–1313.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition*, *49*(3), 402–419.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.