Quantitative Methods in Psychological Aging Research: A Mini-Review

Paolo Ghisletta^{a–c} Stephen Aichele^c

^aFaculty of Psychology and Educational Sciences, University of Geneva, Geneva, ^bSwiss Distance Learning University, Brig, and ^cSwiss National Center of Competence in Research LIVES – Overcoming Vulnerability: Life Course Perspectives, University of Geneva, Geneva, Switzerland

Keywords

Full information maximum likelihood · Multiple imputation · Robust statistics · Nonlinear mixed-effects models · Generalized additive models · Induction trees · Random forests

Abstract

As research on psychological aging moves forward, it is increasingly important to accurately assess longitudinal changes in psychological processes and to account for their (often complex) associations with sociodemographic, lifestyle, and health-related variables. Traditional statistical methods, though time tested and well documented, are not always satisfactory for meeting these aims. In this mini-review, we therefore focus the discussion on recent statistical advances that may be of benefit to researchers in psychological aging but that remain novel in our area of study. We first compare two methods for the treatment of incomplete data, a common problem in longitudinal research. We then discuss robust statistics, which address the question of what to do when critical assumptions of a standard statistical test are not met. Next, we discuss two approaches that are promising for accurately describing phenomena that do not unfold linearly over time: nonlinear mixed-effects models and

(generalized) additive models. We conclude by discussing recursive partitioning methods, as these are particularly well suited for exploring complex relations among large sets of variables.

© 2017 S. Karger AG, Basel

Introduction

The primary aims of psychological research on aging are to describe and explain age-related changes and differences in basic and complex psychological processes (e.g., cognition, emotions, personality), to explain the mechanisms of stabilization or decline in such processes, to understand mechanisms that confer resistance to the negative effects of aging (e.g., plasticity, compensation, external support, lifestyle factors, other resources), and to investigate the dimensionality of changes in age-sensitive processes (such as in cognitive abilities). The quantitative study of psychological aging hence includes methods for exploring and testing observed and hypothesized age-related functions [1]. For reviews of current practices, see, e.g. [2–6]. Recent statistical breakthroughs have ameliorated much of the burden associated with analyzing data generated from both of these modes of investigation.

However, some of these statistical advances are still considered novel in our areas of research. For this mini-review, we selected five methods that have only recently become available in common statistical software and that we believe can contribute to psychological aging research. Some of these methods were chosen because they offer viable solutions to common problems specific to aging research (e.g., incomplete data in longitudinal designs, presence of outliers due, for instance, to normative vs. pathological manifestations). Other methods were selected because they can further our understanding of aging-related processes (e.g., changes that deviate from commonly assumed linear or quadratic trends) or that may help to uncover important associations between variables.

We begin by comparing two methods for the treatment of incomplete data: full information maximum likelihood (FIML) and multiple imputation (MI). Development and adaptation of both methods is ongoing, and the choice to prefer one or the other hinges on the underlying reasons for missingness and also on the chosen analytical framework. We then discuss robust statistics, a well-developed field in statistics that addresses the often neglected question of what to do when some assumptions of a statistical test are not met, thereby jeopardizing all conclusions from such a test. Next, we discuss two statistical approaches that only relatively recently have been implemented in statistical software and that we believe are particularly promising for describing phenomena that do not unfold linearly over time: nonlinear mixed-effects (NLME) models and (generalized) additive models. We will conclude by discussing recursive partitioning methods that are particularly well suited for exploring complex relations among a large set of variables and a given outcome.

Incomplete Data

Research on human aging is often longitudinal. While there are significant benefits to repeated assessment designs, participant attrition over time, such as due to death or dropout, means that a study sample will progressively become less representative of the target population [5]. This can be particularly problematic when an outcome of interest, such as mortality risk or depression, is conflated with missingness itself. It is therefore important that researchers consider potential sources of missingness when planning for data collection – and then leverage statistical methodologies best suited for analysis with missing data.

Whether issues pertaining to missingness can be remedied statistically depends to a large extent on the reasons for missingness and on whether additional information (variables) can be used to account for that missingness. Little and Rubin [7] categorized missing data as three types: (a) missing completely at random (MCAR), which holds when data loss occurs for reasons entirely unrelated to variables in the study, (b) missing at random (MAR), which holds when missingness for a given variable can be accounted for by observed values from other variables in the study, and (c) missing not at random (MNAR), which occurs when missing information for a given variable is primarily contingent upon the values of that variable itself (e.g., daily alcohol consumption among heavy drinkers).

For a long time, common approaches for handling missing data were to simply discard incomplete observations or to impute a single set of replacement scores (e.g., by mean substitution). Even in a best-case scenario, these methods can reduce statistical power, distort the underlying distribution, and/or systematically underestimate variance [8]. In recent years, two techniques, FIML and MI, have gained prominence because they work well when data are MAR [9]. Here, we briefly describe each method and then point out general considerations for deciding which approach may be preferable under different circumstances.

Of the two methods, FIML has been more widely adopted in the behavioral sciences. Under FIML, missing information is handled concurrently with model parameter estimation. To meet the assumption of MAR, variables that account for missingness are usually added directly to the analytic model. However, this is not ideal when the variables that contain information about missingness are not in themselves of substantive interest. For this reason, some statistical software packages, such as Mplus [10], allow inclusion of “auxiliary” variables, which are introduced into the FIML estimation framework so as to account for missingness without directly influencing associations between variables of substantive interest [11, 12].

MI is a three step process: missing values are imputed multiple times (to create m complete datasets), each of these datasets is then analyzed independently of the others, and finally results are pooled across these analyses. Different algorithms may be used for imputing missing values (the first step), but the general idea remains the same: variance within each imputed dataset is indicative of uncertainty in measurement, whereas variance across the datasets is a proxy for uncertainty due to missing in-

formation. This contrasts with earlier methods wherein only a single complete dataset is imputed (e.g., by stochastic regression) – and which therefore do not account for uncertainty related to missingness itself. Another advantage of MI is that variables that explain missingness (MAR assumption) but that are not of substantive interest can be included in the imputation step but then excluded from the analytic model. Van Buuren [8] provides numerous illustrations of key technical and theoretical differences between imputation methods as can be implemented using the R statistical software package MICE (multiple imputation by chained equations).

Several considerations inform decisions on when to prefer FIML or MI. FIML is now implemented in most commercially available software packages and is easy to implement (end users simply need select FIML estimation), whereas MI requires additional knowledge and the overhead that comes with managing multiple datasets. FIML is also statistically deterministic and efficient. In practice, this means FIML will provide reproducible estimates with tighter confidence intervals (smaller standard errors) than those generally obtained under MI. Also, under MI, it is necessary to ensure that the model used for imputing missing values is “congenial” (consistent) with the analytic model. For example, analytic models that include multilevel structures (e.g., patients within treatment types) may not be congenial with imputation models lacking those terms. Congeniality is not an issue for FIML wherein missingness and parameter estimation are done under a single model [13].

On the other hand, MI produces complete datasets that can then be used in analyses for which FIML is not yet, or may ever be, implemented, e.g. certain types of survival analysis, data mining (see below), and/or non-parametric methods. This point is especially salient for researchers in gerontology, who are often interested in nonlinear and nonnormally distributed outcomes (e.g., mortality risk, disease prevalence and/or recurrence rates). Another consideration for researchers faced with a large amount of missing data (due to death or dropout) is that FIML estimation will often fail to converge at higher levels of missingness (e.g., >50%) whereas this is less problematic when missing values are imputed prior to analysis.

We have noted that FIML and MI are both well suited to handling incomplete data under the assumption of MAR. What then when incomplete data are MNAR (nonignorable)? MNAR estimation methods, such as selection and pattern-mixture analyses, jointly model observed and missing information. Importantly, these mod-

els rely on strong, unverifiable assumptions [14]. Sensitivity analysis is therefore required to evaluate changes in analytical outcomes given different MNAR scenarios. MNAR models have been adapted for use with both FIML and MI [e.g., 8, 15]; however, at present sensitivity analysis appears to be more straightforward when using MI [16]. Thus, MI may be better suited for researchers who wish to explore alternative hypotheses about unobserved sources of dropout in longitudinal studies of aging.

Robust Statistics

In gerontological research, it is relatively common to rely on small samples or to examine constructs that are assessed by variables typically displaying asymmetric distributions (e.g., number of strokes, difficulties of daily living). Most mathematical models used in psychology (the general linear model, linear structural equation modeling, linear mixed-effects models) rely on the assumptions that the errors are normally distributed. Moreover, commonly used models to compare groups (Student *t* tests, analysis of variance) assume homoscedasticity (or at least sphericity in the case of repeated measures). These assumptions are often not met in real data, and the consequences of their violations can be very serious in terms of increased type I error and decreased power. Robust statistics address the questions “What to do when some basic assumptions of a statistical model are not met?” and “What to do with outliers?” [17].

Furthermore, the models we frequently use (e.g., linear regression, analysis of variance) are quite sensitive to violations of these assumptions [18]. One solution can at times be to transform variables, in an attempt to “normalize” the data. However, often transformations only alleviate, but do not fully get rid of, the problem, while interpretation of final results generally becomes more arduous. Moreover, outliers with unwarranted influence may bias all parameters’ estimates, thereby invalidating the overall analysis. Modern robust statistics deal with such common research instances, by estimating a model’s parameters even with nonnormal errors, heteroskedasticity, and in the presence of outliers [19]. They thus not only allow reaching valid statistical conclusions when assumptions are violated, they also detect and deal with outliers.

The simplest example of a robust statistic is the median, which reacts to a much smaller degree to outliers than the mean. The mean is said to have a 0% breakdown point, meaning that by altering just one value of a series, the mean will change. The median, on the other hand,

enjoys the highest breakdown point possible, 50%. One can alter as much as half the values of a series and still obtain the same median. This is the reason why the location of inherently skewed distributions, such as income, is characterized by the median rather than the mean. Likewise, robust statistics exist also with respect to distributions' scale parameters. The standard deviation, archetypal scale estimate, is more strongly influenced by outliers than the mean, given that its calculations involves squaring residuals, thereby increasing exponentially the deleterious effect of outliers. The median absolute deviation, defined as the median absolute difference between a data point of a series and that series' median value (i.e., $\text{median}(|X_i - \text{median}(X)|)$), is a much more robust estimate of a distribution's scale. These examples are mainly didactical, but they serve well the message that to characterize a distribution that deviates from familiar probability distributions (e.g., normal) or that includes outliers, classical indicators (e.g., the mean and the standard deviation) will not do, and alternative indices exist.

Robust statistics also allow estimating parameters of well-known statistical models. A popular robust estimation method is the MM-estimation, a generalization of maximum likelihood estimation (commonly used in, e.g., logistic regression, confirmatory factor analysis, and structural equation modeling). MM-estimation determines empirically how many and which extreme cases should be excluded during an analysis, and does so as a function of the observed data (e.g., for a right-skewed distribution data will be excluded especially on the right tail, not on the left or on both). If data are excluded, the inferential conclusions are nevertheless based on the entire sample size, whereas with manual removal of outliers the sample is reduced in inferential calculations, thereby losing efficiency and increasing the risk for type 2 errors. Moreover, MM-estimation may not completely exclude an observation, but rather downweight it, so that observations at the extreme tail are weighted less than those at the center of a distribution during estimation (whereas in traditional estimation all observations have the same weight). Thus, robust statistics also permit the automated detection of outliers and prevent researchers from making arbitrary decisions about what are and what to do with outliers. A classic example of robust regression was provided by Rousseeuw and Leroy [20] and consists of a dataset with only 20 observations of schools' and pupils' characteristics. The outcome variable is sixth graders' verbal score, while the predictors are staff salary, teachers' verbal score, and three indicators of pupils' and parents' socioeconomic status. With ordinary least squares estimation, only two predic-

tors were found to be significant, whereas with MM-estimation all predictors were significant, and the effect size increased from $R^2 = 0.87$ to 0.97. This was because MM-estimation is more efficient than nonrobust ML, even in the context of low power ($n = 20$), and therefore provided greater precision and smaller standard errors.

Robust statistics have influenced not just basic indicators of location and shape of various distributions, but also frequently used statistical models, such as ANOVA and linear regression. Multivariate analyses have also been enriched by robust statistics. Factor analysis [21], structural equation modeling [22, 23], and linear mixed-effects models, for both hierarchical and crossed structures [24–26], can also be estimated within the robust framework.

Like others, we do not suggest that robust statistics should become a de facto replacement for classical statistics. We simply point to the availability of software for estimation using robust statistics (e.g., scripts for SPSS and SAS, libraries for R) and suggest that researchers compare results from classical and robust estimation methods. We believe that in the next decade, more and more statistical models (and related software) will benefit from the option to use robust statistics, and that editors and reviewers of empirical studies will increasingly ask for robust statistics. For introductory, nontechnical readings, see for instance [18, 19]. For a more complete understanding, see for instance [27].

Nonlinear Mixed-Effects Models

There is little doubt that research in psychological aging, and more generally in development, has benefited greatly in the past 3 decades from the use of linear mixed-effects models to estimate changes over time in an outcome of interest [28, 29]. Importantly, this modeling framework partitions variance due to interindividual differences (between subjects) from that due to intraindividual change (within subjects), and thereby allows testing the effects of both subject-specific but time-invariant predictors (e.g., sex) and within-subject time-varying predictors (e.g., well-being). This provides greater flexibility than other repeated-measures approaches (such as repeated-measures analysis of variance) for evaluating theoretically relevant questions. Moreover, under certain assumptions this model can provide unbiased parameter estimates despite incomplete data, and it also can reduce the undue influence of slightly outlying observations. That this methodology is readily available in popular sta-

tistical software also adds to its popularity. Not surprisingly, reviews of current research practices widely used in the broad field of development and aging dedicate much space to discussion of this model [e.g., 3–6].

Probably the two most common specifications of this model include a linear and a quadratic polynomial relation between time and the outcome, so that change (growth or decline) is theorized to be linear (thus constant in its rate of change) or with a quadratic curve, interpreted as acceleration or deceleration. In statistical terms, both specifications refer to a linear model, in the sense that the parameters of the model are associated with the outcome via a linear combination. While linear models of change are widely adopted in the psychological aging literature, there are times when they are theoretically too limiting. Indeed, it is rather hard to believe that aging phenomena are linear or quadratic [30–33]. Growth and decline in various organisms often follow a logistic function, survival in a population may follow a Weibull function, and accumulation of assaults to the central nervous system may follow an exponential function. These are just a few examples for which linear models may not satisfactorily describe the data and badly predict unobserved (future) occurrences.

There can also be statistical disadvantages to using polynomial functions. For instance, Fjell et al. [34] demonstrated that in cross-sectional data of hippocampal volume assessed in over 400 individuals between the ages of 8 and 85 years, the estimated age of maximum volume and the estimated decline from a quadratic function in the oldest participants was affected by whether or not data of children were included in the analysis. We would expect to find similar problems in data that are partially longitudinal over a relatively short period (e.g., 8 years) assessed in an age-heterogeneous sample (e.g., from 10 to 100 years of age).

Unlike linear mixed-effects models, NLME models do not constrain the parameters of the model to relate to the outcome via linear combinations only. That is, the NLME model goes beyond the specification of nonlinear change (i.e., as a function of time) via reparameterization within the linear mixed-effects framework. Indeed, NLME models cover a wide array of functional specifications demonstrated to be effective for describing psychological phenomena. For instance, with experimental data, Ghisletta et al. [35] used a three-parameter exponential function to study motor-perceptual learning in an adult sample (age range 19–80 years). Learning was not well characterized as a linear function, whereas the exponential function resulted in very good model fit and allowed for clearer interpretation of the estimated parameters. Specifically, one

parameter represented initial performance, a second indicated the rate of exponential learning, and a third corresponded to maximum performance. While age predicted final performance and rate of learning, spatial abilities (independent of age) predicted initial performance. Thus, older participants with strong spatial abilities could start as high as younger participants with the same level of spatial ability, but they could not hope to learn as much, or to perform as high at the end, as younger participants.

Advantages of nonlinear functions include: (a) they can be specified with parameters that are directly interpretable in psychological terms, such as learning (positive exponential rate), maximal performance (upper asymptote), and intensity of a signal (amplitude); (b) they can naturally accommodate noncontinuous outcomes (such as counts and dichotomous events, both frequent in psychological research) without having to resort to transformations; and (c) they often require fewer parameters (hence are more parsimonious) than linear models to describe adequately a developmental process. For all these reasons, many statisticians and methodologists increasingly favor the use of nonlinear functions to model psychological processes.

Applications in the psychological aging literature remain limited, but they do exist and are gaining traction. For instance, in the context of longitudinal lifespan data, McArdle et al. [36] described cognitive development in different abilities with a dual exponential model; Riediger et al. [37] used a similar model to chart beliefs about cognitive and social functioning; and Grimm et al. [33] used the Preece-Baines model to study height growth in children. Interested readers can consult Ratkowsky [38] for general understanding of nonlinear regression, Draper and Smith [39] for an introduction to nonlinear estimation, Cudeck and Harring [40] for a general overview of NLME models, Sit and Poulin-Costello [41] for a catalog of nonlinear functions, and Ghisletta et al. [42] for a practical procedure, with a detailed example, for estimating such models with common software. Given this availability of software and research examples using NLME models, we encourage psychological aging researchers to consider this analytic approach whenever they think that linear models are neither theoretically nor statistically satisfying.

Additive and Generalized Additive Models

It is not uncommon that the change phenomenon under investigation is rather novel and consequently does not lend itself easily to confirmatory analyses. In such an

exploratory setting, it is difficult to propose a (linear or nonlinear) model that best describes and facilitates interpretation of the phenomenon. Nonetheless, a researcher may wish to understand how the phenomenon unfolds over time, and how covariates of interest (e.g., group membership) are related to the outcome. For instance, treatment effectiveness in clinical psychological practice can be improved by increased understanding of how intervention effects develop over time, but these longitudinal effects are not easily characterized by known statistical functions [43]. As an experimental example, Schmiedek et al. [44] were interested in assessing a mean performance curve in reaction time across 100 daily assessments in a working memory task. Because performance progressed nonmonotonically across the days (as a result of practice, habituation, possibly fatigue, etc.), it was not possible to specify a known function characterizing mean performance. In both of these above studies, rather than exploring a set of known parametric functions, the authors opted for a statistical procedure that captures patterns in the data with smoothing functions.

Additive models [45] are semi- or nonparametric techniques that extend the general linear model by replacing one or more of the standard linear terms with smoothing functions that enter additively in the prediction model. Smoothing functions are usually linear functions (e.g., cubic polynomials) that attempt to capture patterns (but not noise) in the data. Often moving windows are employed, where a smoothing function is first fit to a limited range of values of the predictor (e.g., ages 5–8 years), and then fit to another (possibly overlapping) range of values (e.g., ages 7–10 years). Care is taken that the functions' values match at overlapping segments of the windows. The fitting is controlled by a smoothing parameter, which is estimated to obtain a good compromise between smoothness (lack of "wiggleness" but possible poor fit to individual observations) and jaggedness (better discrimination, but possible overfit to individual observations and hence to noise). Moreover, the smoothing procedure estimates so-called effective degrees of freedom, which assess the degree of nonlinearity of a smoothing term and can be interpreted as the polynomial order of the smoother.

The end result is a continuous smooth line, which may or may not be linear, across all values of the predictors, according to the patterns in the data. If the true relationship between the predictor and the outcome is parametric (e.g., quadratic), then the procedure will produce a line that follows a quadratic function, and the effective degrees of freedom will reveal that. Generalized additive

models (GAMs) extend this technique to noncontinuous dependent variables via link functions, just like the generalized linear model does for the general linear model. Finally, recent extensions to generalized additive mixed models (GAMMs) allow for multiple sources of variance (just like the linear mixed model does for the general linear model), thereby accommodating longitudinal assessments.

Additive models and GAMs have gained popularity in medicine, biostatistics, biology, and related fields, and are slowly also gaining momentum in psychology [e.g., 34, 46–49]. This increasing popularity may in part be attributed to the book by Wood [50] and the related R package *mgcv* that estimates additive models, GAMs, and GAMMs [51]. This package also allows for estimation of interaction effects between predictors and the parameters of a given smoothing function. Thus, for testing the effectiveness of a therapeutic treatment, the application of an additive model would allow (a) studying how the intervention unfolds over a time and (b) testing whether the time course is different between the intervention and the control group [43].

The same would apply when studying how older individuals perform repeatedly on cognitive tasks and whether group differences (e.g., sex, vascular disease status, presence vs. absence of a particular genetic polymorphism) emerge in performance or in decline. While we contend that it is always best to pair theory and data analysis whenever possible, there are instances in which lack of prior knowledge, or the desire to test a covariate's effect with maximal power, make additive models an excellent alternative to (potentially misspecified) parametric models. We believe that research in psychological aging could benefit greatly from this approach.

Random Forest Analyses

Gerontological research encompasses numerous interrelated processes, and teasing apart important associations often presents both theoretical and statistical challenges. For example, variables that are moderately or strongly correlated, such as cognitive decrements in processing speed and fluid intelligence, may be differentially linked to other variables of interest (e.g., depressive symptoms, health intervention outcomes). This can lead to problems such as "multicollinearity" or "suppression" effects when such associations are evaluated using standard regression models. It is also often not clear when to include variables for the purpose of statistical adjustment

(e.g., to account for differences in socioeconomic status or gender) or when improvement in model fit after variable inclusion is merely an artifact of accounting for sample-specific variance (i.e., “junk associations” that do not replicate across studies). In studies that utilize biomedical data, which can be expensive to collect, it is often the case that there is a small participant pool relative to the number of variables considered (i.e., the “small-*n* large-*p*” problem).

Within the behavioral sciences, statistical approaches toward these problems, such as factor analysis or stepwise regression, are often aimed at reducing the number of variables. However, stepwise approaches are well known for producing idiosyncratic results (due to variable inclusion order effects), and other methods based on standard regression are often limited in their ability to account for higher-order interactions and nonlinear effects in predictor-outcome associations. Fortunately, there now exist “data mining” techniques designed to overcome some of these limitations in estimating predictor-outcome associations. Methods based on recursive partitioning [52] have proven particularly effective: the basic algorithm splits observations such that a predictor variable (and corresponding cut-point) is selected to maximally differentiate the resulting subsamples with respect to an outcome variable (which can be either categorical or continuous). This process is then repeated within the subsamples created by the initial split, and so on within additional nested subsamples. This results in a pattern of predictor-outcome relations that can be represented as an upside-down tree – hence the method is often referred to as induction trees [53].

Induction trees have several advantages over traditional stepwise regression methods. Perhaps most importantly, the branching structure is able to approximate all nonlinear and higher-order interaction terms in an exhaustive data-driven approach, whereas such complex terms are often overlooked or excluded (or, when included, limited to a few possible candidates) in standard parametric models. Induction tree estimates of variable importance (i.e., predictive association) therefore implicitly account for the influence of variables in these higher-order associations in addition to their main effects. Additionally, variable selection may be based either on parametric or nonparametric criteria (with nonparametric measures often preferred). Induction trees can also handle most small-*n* large-*p* scenarios without problem.

However, this methodology remains susceptible to multicollinearity and model overfit. For example, if two predictors are comparably strongly associated with an

outcome, selection of one over the other will be somewhat arbitrary and will also influence selection of predictors at more distal nodes of the tree. Spurious variable selection that occurs closer to the root node (first split) can therefore greatly affect the overall branching pattern of predictors. This is referred to as the instability of single trees [54].

Random forest analysis, or RFA [55, 56], further extends the induction tree methodology to address these shortcomings. Specifically, RFA estimates numerous induction trees (hence forest, which typically consists of 100 or more trees) from randomly selected subsamples drawn from the full pool of observations. This provides a built-in cross-validation mechanism to check against model overfit (such as may occur when a single tree is fit to the sample as a whole). Additionally, during variable selection at each split within each tree, candidate predictors are selected from a subset of the total set of predictors. This similarly provides a check against multicollinearity and spurious variable selection. Variable importance (i.e., relative predictive influence) is then summarized across all trees of the forest.

Applications of RFA within psychological research have only begun to appear within the past decade [e.g. 57–61]. Along these lines, RFA may be most effective when applied alongside more standard parametric methods in an exploratory-confirmatory approach. For example, Aichele et al. [62] used RFA in conjunction with Cox proportional hazards modeling to evaluate 65 predictors of mortality risk, including measures of cognitive performance, in a large sample of community dwelling older adults. Each method was conducted in a separate subsample (i.e., half) of observations randomly selected from the total participant pool. This combined methodology provided a way to assess the relative importance of numerous, interrelated mortality risk factors and also to estimate effect sizes for the most important predictors.

Conclusion

In this mini-review, we took the liberty to briefly discuss innovative or unfamiliar methodological and statistical topics that may be useful for research on psychological aging. Researchers in other disciplines are already familiar with these techniques – probably because existing methods, though well established, are not always adequate for a given dataset (e.g., due to incompleteness or nonlinearity) or theoretical motivation (e.g., to define the cumulative rather than proportional rate of change; to

compare large ensembles of variables as predictive of a given outcome). Obviously, our choice of topics to be discussed was subjective, and we have omitted many other important topics, such as Bayesian estimation, differential item functioning, use of permutation tests, dynamic systems and models of continuous time, etc. Notwithstanding this selective focus, we hope to have shed some light on promising (and perhaps lesser-known) methodological and statistical topics that we believe can aid psychological aging researchers in advancing their understanding of important theoretical issues. Above all, we hope that any chosen method of data analysis is up to the challenge of allowing for thorough theoretical interpretation of the phenomenon under investigation.

Acknowledgments

This publication benefited from the support of the Swiss National Centre of Competence in Research LIVES – Overcoming Vulnerability: Life Course Perspectives, which is financed by the Swiss National Science Foundation. The authors are grateful to the Swiss National Science Foundation for its financial assistance.

The authors thank two anonymous reviewers for their valuable comments.

Disclosure Statement

The authors have no conflicts of interest to disclose.

References

- Wohlwill JF: The age variable in psychological research. *Psychol Rev* 1970;77:49–64.
- Salthouse TA: Methodological assumptions in cognitive aging research; in Craik FIM, Salthouse TA (eds): *The Handbook of Aging and Cognition*. Mahwah, Lawrence Erlbaum Associates, 2000, vol 2, pp 467–498.
- Hertzog C, Nesselroade JR: Assessing psychological change in adulthood: an overview of methodological issues. *Psychol Aging* 2003; 18:639–657.
- Hofer SM, Sliwinski MJ: Design and analysis of longitudinal studies on aging; in Birren JE, Schaie KW (eds): *Handbook of the Psychology of Aging*, ed 6. London, Academic Press, 2006, pp 15–37.
- Ferrer E, Ghisletta P: Methodological and analytical issues in the psychology of aging; in Schaie KW, Willis SL (eds): *Handbook of the Psychology of Aging*, ed 7. London, Academic Press, 2011, pp 25–39.
- MacDonald SWS, Stawski RS: Methodological considerations for the study of adult development and aging; in Schaie KW, Willis SL (eds): *Handbook of the Psychology of Aging*, ed 8. London, Academic Press, 2016, pp 15–40.
- Little RJA, Rubin DB: *Statistical Analysis with Missing Data*, ed 2. New York, Wiley, 1987.
- Van Buuren S: *Flexible Imputation of Missing Data*. Boca Raton, Chapman & Hall/CRC, 2012.
- Schafer JL, Graham JW: Missing data: our view of the state of the art. *Psychol Methods* 2002;7:147–177.
- Muthen LK, Muthen BO: *Mplus User's Guide*, ed 7. Los Angeles, Muthen & Muthen, 1998–2012.
- Enders CK: *Applied Missing Data Analysis: Methodology in the Social Sciences*. New York, Guilford Press, 2010.
- Graham JW: Adding missing-data relevant variables to FIML-based structural equation models. *Struct Equ Modeling* 2003;10:80–100.
- Allison PD: Maximum likelihood is better than multiple imputation: part II. May 15, 2015. <http://statisticalhorizons.com/ml-is-better-than-mi> (cited February 13, 2017).
- Aste M, Boninsegni M, Freno A, Trentin E: Techniques for dealing with incomplete data: a tutorial and survey. *Pattern Anal Appl* 2015; 18:1–29.
- Enders CK: Missing not at random models for latent growth curve analysis. *Psychol Methods* 2011;16:1–16.
- Allison PD: Sensitivity analysis for not missing at random. September 25, 2014. <http://statisticalhorizons.com/sensitivity-analysis> (cited February 13, 2017).
- Hampel F, Ronchetti E, Rousseeuw P, Stahel W: *Robust Statistics: The Approach Based on Influence Functions*. New York, Wiley, 1986.
- Erceg-Hurn DM, Mirosevich VM: Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *Am Psychol* 2008;63:591–601.
- Courvoisier DS, Renaud O: Robust analysis of the central tendency, simple, and multiple regression and ANOVA: a step by step tutorial. *Int J Psychol Res* 2010;3:78–87.
- Rousseeuw PJ, Leroy AM: *Robust Regression and Outlier Detection*. Hoboken, John Wiley & Sons, 1987.
- Yuan KH, Zhong X: Outliers, leverage observations, and influential cases in factor analysis: using robust procedures to minimize their effect. *Sociol Methodol* 2008;38:329–368.
- Yuan KH, Bentler PM: Structural equation modeling with robust covariances. *Sociol Methodol* 1998;28:363–396.
- Savalei V: Understanding robust corrections in structural equation modeling. *Struct Equ Modeling* 2014;21:149–160.
- Copt S, Heritier S: Robust alternatives to the F-Test in mixed linear models based on MM-estimates. *Biometrics* 2007;63:1045–1052.
- Copt S, Victoria-Feser M-P: High-breakdown inference for mixed linear models. *J Am Stat Assoc* 2006;101:292–300.
- Koller M: Package “robustlmm.” <https://cran.r-project.org/web/packages/robustlmm/robustlmm.pdf> (cited April 8, 2017).
- Wilcox R: *Introduction to Robust Estimation and Hypothesis Testing*, ed 4. Burlington, Elsevier Academic Press, 2016.
- Laird NM, Ware JH: Random-effects models for longitudinal data. *Biometrics* 1982;38: 963–974.
- Bryk AS, Raudenbush SW: Application of hierarchical linear models to assessing change. *Psychol Bull* 1987;101:147–158.
- Davidian M, Giltinan DM: *Nonlinear Models for Repeated Measurement Data*. London, Chapman & Hall, 1995.
- Vonesh EF, Chinchilli VM: *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York, Dekker, 1996.
- Pinheiro JC, Bates DM: *Mixed-Effect Models in S and S-PLUS*. New York, Springer, 2000.
- Grimm KJ, Ram N, Hamagami F: Nonlinear growth curves in developmental research. *Child Dev* 2011;82:1357–1371.
- Fjell AM, Walhovd KB, Westlye LT, Ostby Y, et al: When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies. *Neuroimage* 2010;50:1376–1383.
- Ghisletta P, Kennedy KM, Rodrigue KM, Lindenberger U, Raz N: Adult age differences and the role of cognitive resources in perceptual-motor skill acquisition: application of a multilevel negative exponential model. *J Gerontol B Psychol Sci Soc Sci* 2010;65B:163–173.

- 36 McArdle JJ, Ferrer-Caja E, Hamagami F, Woodcock RW: Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Dev Psychol* 2002;38:115–142.
- 37 Riediger M, Voelkle MC, Schaefer S, Lindenberger U: Charting the life course: age differences and validity of beliefs about lifespan development. *Psychol Aging* 2014;29:503–520.
- 38 Ratkowsky DA: *Handbook of Nonlinear Regression Models*. New York, Marcel Dekker, 1990.
- 39 Draper NR, Smith H: *Applied Regression Analysis*, ed 3. New York, Wiley-Interscience, 1998.
- 40 Cudeck R, Harring JR: Analysis of nonlinear patterns of change with random coefficient models. *Annu Rev Psychol* 2007;58:615–637.
- 41 Sit V, Poulin-Costello M: *Catalog of Curves for Curve Fitting*. Victoria, Ministry of Forestry, 1994.
- 42 Ghisletta P, Cantoni E, Jacot N: Nonlinear growth curve models; in Stemmler M, von Eye A, Wiedermann W (eds): *Dependent Data in Social Sciences Research: Forms, Issues, and Methods of Analysis*. Berlin, Springer, 2015, pp 47–66.
- 43 Sullivan KJ, Shadish WR, Steiner PM: An introduction to modeling longitudinal data with generalized additive models: applications to single-case designs. *Psychol Methods* 2015;20:26–42.
- 44 Schmiedek F, Lovden M, Lindenberger U: On the relation of mean reaction time and intra-individual reaction time variability. *Psychol Aging* 2009;24:841–857.
- 45 Stone CJ: Additive regression and other non-parametric models. *Ann Statist* 1985;13:689–705.
- 46 Kenny RA, Coen RF, Frewen J, Donoghue OA, Cronin H, Savva GM: Normative values of cognitive and physical function in older adults: findings from the Irish Longitudinal Study on Ageing. *J Am Geriatr Soc* 2013; 61(suppl 2):S279–S290.
- 47 Bisiacchi P, Cona G, Tarantino V, Schiff S, Montagnese S, Amodio P, Capizzi G: Assessing inter- and intra-individual cognitive variability in patients at risk for cognitive impairment: the case of minimal hepatic encephalopathy. *Metab Brain Dis* 2014;29:945–953.
- 48 Tremblay A, Newman AJ: Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology* 2015;52:124–139.
- 49 Hendrix P, Bolger P, Baayen H: Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *J Exp Psychol Learn Mem Cogn* 2017;43:128–149.
- 50 Wood S: *Generalized Additive Models: An Introduction with R*. Boca Raton, Chapman and Hall/CRC, 2006.
- 51 Wood S: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf> (cited February 8, 2017).
- 52 Morgan JN, Sonquist JA: Problems in the analysis of survey data, and a proposal. *J Am Stat Assoc* 1963;58:415–434.
- 53 Breiman L, Friedman JH, Olshen RA, Stone CJ: *Classification and Regression Trees*. New York, Chapman and Hall, 1984.
- 54 Strobl C: Data mining; in Little TD (ed): *The Oxford Handbook of Quantitative Methods*. New York, Oxford University Press, 2013, pp 678–700.
- 55 Breiman L: Random forests. *Mach Learn* 2001;45:5–32.
- 56 Strobl C, Malley J, Tutz G: An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol Methods* 2009;14:323–348.
- 57 Baca-Garcia E, Perez-Rodriguez MM, Saiz-Gonzalez D, Basurte-Villamor I, Saiz-Ruiz J, Leiva-Murillo JM, de Prado-Cumplido M, Santiago-Mozos R, Artes-Rodriguez A, de Leon J: Variables associated with familial suicide attempts in a sample of suicide attempters. *Prog Neuropsychopharmacol Biol Psychiatry* 2007;31:1312–1316.
- 58 Berk RA: An introduction to ensemble methods for data analysis. *Sociol Methods Res* 2006;34:263–295.
- 59 Luellen JK, Shadish WR, Clark MH: Propensity scores: an introduction and experimental test. *Eval Rev* 2005;29:530–558.
- 60 Marinic I, Supek F, Kovacic Z, Rukavina L, Jendricko T, Kozaric-Kovacic D: Posttraumatic stress disorder: diagnostic data analysis by data mining methodology. *Croat Med J* 2007;48:185–197.
- 61 Rossi A, Amaddeo F, Sandri M, Tansella M: Determinants of once-only contact in a community-based psychiatric service. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:50–56.
- 62 Aichele S, Rabbitt P, Ghisletta P: Think fast, feel fine, live long: a 29-year study of cognition, health, and survival in middle-aged and older adults. *Psychol Sci* 2016;27:518–529.