Career Aspirations

My research goal is to be a producer of excellent research on quantitative methods in the social sciences, in particular, to develop and study new methods for the study of human development. I see enormous potential in structural equation modeling and related emerging fields to allow researchers to examine nuanced questions about the forces that shape development. New modeling techniques will allow the study of complex interactions among environmental and biological predictors and outcomes over time; when applied correctly, these methods will be a powerful tool for examining human development in its complexity. My background in developmental psychology combined with expertise in quantitative methods has positioned me uniquely to develop and study these methods. My graduate training gave me a certain level of expertise in quantitative methodology, as well as research experience in conducting simulation studies of research methods in the social sciences. To truly achieve expertise in the study of these methods, I am seeking a post-doctoral experience that will allow me to learn from a unique group of faculty who have extensive experience as researchers, consultants, and teachers.

In order for new methods to gain traction among applied researchers, these methods need to be developed, implemented in usable software, studied under realistic data conditions, and explained in accessible ways to substantive researchers. Social scientists at every institution benefit from having colleagues who study methodological developments, publish articles and chapters about them, and work with substantive researchers to guide their research methods in novel and appropriate directions. Without such guidance, few substantive researchers dare attempt new methodological directions. In addition to publishing new methodological research, I aim to contribute to the academic community in my institution and beyond by collaborating on diverse research projects in many substantive areas and providing high quality training opportunities for students and faculty alike.

There is a dearth of high-level training in quantitative methods within the social sciences. Among North American universities, most are unable to offer courses on newly-developed methods for the social sciences, and most graduate students are unfamiliar with the vast majority of specialized analytic techniques for analysing the sorts of data that are prevalent in their fields (Aiken, West, & Millsap, 2008, *American Psychologist*). The development of new methods will be of little use if students do not learn to use them. As such, my teaching goal is to educate new generations of researchers to be knowledgeable users of high-level statistical methods. I aim to help students to really understand the methods they use and to be comfortable knowing how and when those methods can be extended to new research situations. They should be comfortable to explore and teach themselves new methods and to evaluate the research methods of their peers.

A few universities across North America have begun to develop quantitative research and training programs in the social and behavioural sciences. My goal is to be involved in the inception of such a program at a major Canadian university. At the moment, students seeking excellent quantitative training are drawn to a handful of schools in the United States that have large and established quantitative areas. I hope to receive my training at one of these unique programs (the Quantitative Training Program at the University of Kansas) and bring my skills and experience with this program back to a Canadian institution.

Choice of Institution

Although there are excellent methodology researchers at many universities in Canada, few Canadian schools yet have a dedicated quantitative methods program, and fewer have more than one or two faculty members who are primarily engaged in practicing and mentoring quantitative research. The University of Kansas has one of the biggest concentrations of world-class researchers who study methodology for the social sciences in the world. Their combined research interests and class offerings span all areas of modern research methods including structural equation modelling and longitudinal designs (Todd Little), multilevel modelling, multilevel structural equation modelling, factor analysis, and nonparametric analysis (Kristopher Preacher), methods for the analysis of categorical variables, time-series analysis, dynamical systems, and analysis of variance (Pascal Deboeck), regression, propensity score matching, and missing data (Wei Wu), item response theory and differential item functioning (Carol Woods), and simulation studies (Paul Johnson). These scholars are specialists in quantitative methods research. They have vast combined experience in developing, exploring, and testing new methods.

I have proposed to pursue postdoctoral training under the supervision of Todd Little, because of his unparalleled experience in methodology research, collaborative substantive research, attaining grants, and his excellence in teaching and mentorship. My experience with Dr. Little to date has convinced me that he is the most qualified individual to give me well-rounded training in every aspect of academic excellence.

In pursuing postdoctoral research at KU, I will have the opportunity to collaborate with and learn from the many faculty members and affiliated researchers at KU. I will also have the chance to develop my teaching and mentorship skills by delivering lectures in graduate-level methodology classes (mentored by award winning instructors) and by collaborating with undergraduate and graduate students on original research projects.

Research Objectives

Though the idea of deliberately inserting missing data into a research design may sound perverse, *planned missing data designs* allow researchers to reduce the testing burden on participants, leading to higher-quality data with less unplanned missingness and smaller fatigue and practice effects. Twenty years ago, methodologists first promoted the use of planned missing data designs in the study of development (McArdle & Hamagami, 1991). Since then, with the increased availability of modeling software, developmental researchers are increasingly able to ask complex questions about the interrelationships among cognitive, behavioural, and educational variables over time. One of the most useful aspects of modern modeling techniques is that they allow for the presence of missing data. Where once researchers would omit whole cases with any missing data, missing data techniques embedded in newer modeling software have enabled them to recover much of the missing information, resulting in unbiased model estimates and increased power to test their hypotheses. With the increased availability and ease of use of missing data estimation techniques such as full information maximum likelihood estimation (FIML) and multiple imputation (MI), planned missing data designs are poised to achieve widespread use in the developmental sciences. To date, however, there is little guidance for researchers on how to design such studies: which variables can have missingness, what percent of data can be missing, how large a sample is needed, and what the remaining model power will be. The present proposal aims to *provide comprehensive guidance to researchers for using planned missing research designs in developmental research.* The planned missing designs that I will examine will include planned missing data across assessment occasions in longitudinal designs. Through a series of simulation studies, I will vary the critical features of planned missing designs that will *inform developmental researchers on when, how, and why to use planned missing elements in their research designs and demonstrate the conditions under which validity and generalizability are enhanced when such designs are appropriately utilized.*

Missing Data Mechanisms and Estimation Techniques

Missing data can arise from three mechanisms: a truly random process, a measured/predictable process, and an unmeasured/ unpredictable (but not random) process. These three mechanisms are respectively known as *missing completely at random* (MCAR), *missing at random* (MAR), and *missing not at random* (MNAR; see Figure 1). If the missing data are MCAR, then the reason for missingness is unrelated to either the observed data or the missing data. In this ideal situation, the missing data estimation procedures provide unbiased estimates and recover much of the power lost due to attrition. Unfortunately, unplanned MCAR is probably the least common missing data process, particularly for longitudinal data. If the missing data are MAR, the missingness is predictable from other variables in the observed data (e.g., if girls tended to skip a particular question but boys did not, then the missingness would be predicted by sex). In the MAR case, again, *the missing data estimation procedures can recover much of the lost information and power,* to the extent that the missing values are correlated with the remaining observed variables. Finally, if the missing data are MNAR, then the missingness is predicted by the missing variables (e.g., if children with poor math skills tended to skip the math questions on an assessment, and there were no other observed indicators of math ability). In the MNAR case, there is no way to recover the missing information. A critical design feature, therefore, is the collection of data on variables that are likely to predict the missingness. The degree to which modern approaches are able to recover the missing data process depends on (a) whether correlates of the missing data mechanism are measured and (b) the strength of the linear relationship of this known process with the missing data. In longitudinal research, for example, attrition is frequently associated with variables such as SES, sex, and parental involvement. If these are measured and included in the estimation process, some degree of the missing information will be recovered.

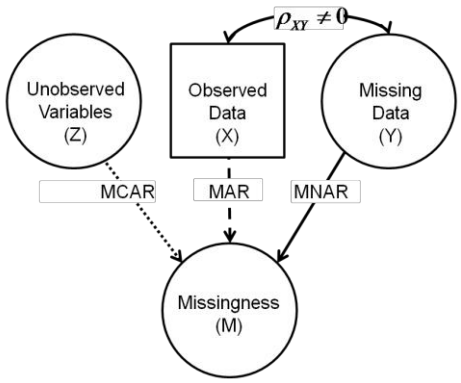

Figure 1. Missing Data Mechanisms. The key distinction between MCAR, MAR, and MNAR is whether the reason for missingness is related to the missing data themselves (MNAR), related to other observed data (MAR), or unrelated to both (MCAR). As increases, more of the missingness can be predicted by X, and MNAR becomes MAR.

There are two recommended modern missing data estimation techniques. The first is full information maximum likelihood (FIML), which uses all of the available data to produce the most accurate estimates of means, variances and covariances, parameter estimates and standard errors in a single step. FIML is implemented in most structural equation modelling software, but it can be used for a wide range of simple and complex analyses including multiple regression. The second is multiple imputation (MI) which is a two-step approach. In the first step, *m* plausible values (e.g., *m* = 100) for each missing data point are imputed based on information from the complete data. These values are substituted for the missing value to create *m* complete datasets. In the second step, the analysis is carried out on each dataset, and the results are combined across datasets to achieve

parameter estimates and standard errors that are typically as accurate as those produced by FIML. With both methods, it is possible to include key *auxiliary variables* in the estimation procedure, that may or may not be relevant to the theoretical model but are thought to be related to the missingness (e.g., SES, sex). Measuring and including the right auxiliary variables can turn unplanned MNAR missingness into MAR, by turning unpredictable missingness into predictable missingness. Planned missing designs have the potential to reduce both MAR and MNAR missingness by reducing the testing burden on participants. When data are missing by design, participants are *randomly assigned* to have particular patterns of missing data, resulting in MCAR missingness. Model parameter estimates are unbiased when missingness is MCAR.

Planned Missing Designs in the Literature

Unplanned missingness can be a serious problem for research generalization. Planned missingness, on the other hand, has tremendous (and for the most part unrealized) potential. Planned missing data designs provide efficient ways to manage cost, improve data quality, reduce participant fatigue, and conserve statistical power to detect effects of interest. Although the concept of planned missing data designs is not new, these designs have only recently become practical with the emergence of advanced missing data techniques such as MI and FIML (Graham, Hofer, & MacKinnon, 1996), and the implementation of these techniques in popular software.

Previous studies have examined the performance of planned missing designs for cross-sectional data (Bunting et al., 2002; Graham et al., 1996; Graham et al., 2006). The 3-form design, for example, allows researchers to randomly remove 25% of questionnaire items and collect data on 1/3 more questions than when all items are given to all people. Importantly, researchers have control over which items are missing: With a planned missing design, researchers may choose to have every participant complete questions that are particularly central to the hypothesis, as well as demographic variables that are likely to be related to unplanned missingness. By giving each participant a smaller number of items to complete, participants are less likely to skip items due to fatigue, reducing the amount of MAR or MNAR missingness. Graham et al. (2006) also proposed a 2-method design that supplements a small amount of data collected on an expensive measure (e.g., fMRI) with a large amount of data collected on a cheaper measure (e.g., questionnaire). These studies have found that modern missing data estimation methods, and FIML in particular, produce unbiased estimates of the relations between variables in cross-sectional planned missing designs.

In cross-sectional studies, planned missing designs have the benefit of optimizing power for a given cost or number of participants. For **longitudinal models**, on top of these benefits, planned missing designs have the potential to reduce retest or practice effects that arise from repeated testing. The goal of much developmental research is to assess the degree of growth or change over time. Practice effects can severely hinder this goal by adding a second source of unwanted change: In most developmental studies, it is impossible to differentiate the effects of development from the effects of practice that stem from repeated testing. Even when alternate forms of a measure are used, retest effects can be quite strong (Kulik et al., 1984; Salthouse & Tucker-Drob, 2008). Meta-analysis results revealed an average retest effect size of .26 in cognitive development, which rose to .46 when identical forms were used across testing occasions (Hausknecht et al., 2007). By reducing the number of testing occasions for individual participants, longitudinal missing data designs will both decrease the amount of practice that participants have (thus reducing the size of retest effects) and allow researchers to disentangle the change due to retest effects from the change due to true developmental processes. Independent from the gains in power and efficiency, the ability to partial out retest effects will be a boon for developmental researchers.

Planned missing designs have appeared occasionally in developmental studies for many years. In particular, using a design where different groups of participants were tested at different time-lags, researchers treated each group as though the data from every other possible time-lag were treated as MCAR data (McArdle et al., 2002; McArdle & Woodcock, 1997; McArdle et al., 2009). Although these designs have been used, little empirical research has studied how well they work. Graham et al. (2001) presented one of the first simulation studies that systematically examined planned missing data designs for longitudinal data. They investigated the power of designs with 1-3 missing timepoints per participant for a longitudinal study with 5 measurement occasions given a fixed sample size. They found that the planned missing data design offered a better combination of cost and power than a complete data design. Of particular note, these authors compared versions of the planned missing design that contained the same percent of missing data but distributed the missingness differently across the 5 timepoints. Reasoning that the rate of developmental change would be easier to detect with more data at the first and last timepoints, they tested a design with more missing data concentrated in the middle timepoints, and found that it had greater power to detect a modest effect size than a design with the same amount of missingness distributed evenly across the 5 timepoints. One goal of the present research is to investigate more closely the effect of the distribution of missing data across timepoints.

Although planned missing data collection procedures offer many potential benefits, few studies have explored their performance across a broad range of realistically simulated conditions or systematically examined the

performance of FIML and MI with planned missing designs. *No research has yet explored the potential for planned missing longitudinal designs to attenuate retest and attrition effects in developmental research*.

Research Question

When data are missing completely at random (MCAR), parameter estimates will be unbiased as the sample size becomes large enough; however, the number corresponding to "large enough" is unknown: We do not know how large samples must be, given a certain percent or distribution of missing data, before estimates will be reliable. We also do not know how the estimation techniques will perform under violations of the assumption of normality that are common in psychological variables. We also do not know how the presence of planned MCAR missingness will affect longitudinal designs that may already have a high level of unplanned MAR and MNAR missingness. In short, simulation work is essential to provide guidance to researchers as to when and how planned missing data procedures can be used as well as the conditions under which caution should be exercised.

Method

The proposed simulation studies will follow 4 steps:

1) Specify the true population model, which will be a two-group (experimental and control), 5-timepoint longitudinal growth-curve model. Two outcome variables will be modelled, X and Y (these could be any cognitive, social, or personality variable that might be expected to increase over time). Each outcome will be generated using an intercept (representing children's initial level of the variable) and a slope (representing their rate of change). The intercepts and slopes of the two variables will be correlated, indicating a common source of developmental change (e.g., Rhemtulla & Tucker-Drob, under review; Salthouse, 1998; see Figure 2).

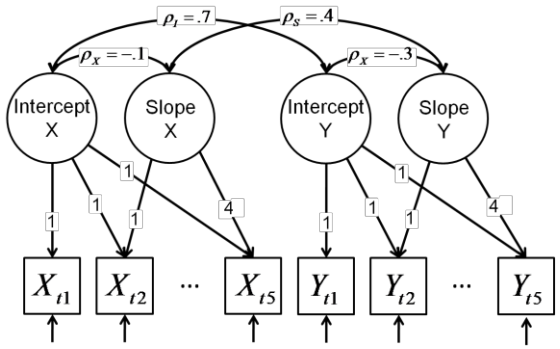

Figure 2. Data Generating Model. This model will be used to simulate data on a set of outcome variables. Note that only one group (control/experimental) is represented here.

To simulate a situation where the experimental group shows quicker growth, the slope of the experimental group will correspond to a larger effect size (mean slope = .7) than that of the control group (mean slope = .5). In addition to the outcome variables, the simulated demographic variables socioeconomic status, sex, and ethnicity will correlate with the outcome variables, and will be used in step 3 to impose unplanned missingness. Finally, we will simulate a test-reactivity variable that will be used in step 3 to impose retest effects on the sample data. Because all models are wrong to some degree, even in the population, I will cross all the conditions and comparisons described here with a parallel set of population data that also includes a small amount of population misfit. Data will be simulated in R (R Development Core Team, 2009).

2) Draw 2000 random samples of a given size for each condition. On average these random samples will contain the precise effects specified in the true population model. Sampling variability will result in some deviation from the population model. The model will be estimated on these complete datasets that have neither test reactivity nor missing data. The parameter estimates from these models provide the baseline of performance against which the missing data analyses will be compared.

3) Impose planned and unplanned missingness and test-reactivity in R. MCAR missingness will be imposed according to one of three planned missing designs (see Table 1), and one complete data design. In each case, 30% of the total data are missing, but the distribution of those data vary across designs A-C. In **Design A**, 10% of participants have complete data across all 5 timepoints, and the remaining 90% of participants are divided into 15 groups representing every combination of 1 or 2 missing timepoints. In **Design B**, 10% of participants have complete data across all 5 timepoints, 60% of participants are missing data on 2 of the middle 3 timepoints, and 30% of participants are missing data on 1 of the middle 3 timepoints. This design maintains complete data at the first and last timepoint, which may allow more power to test the slope parameter (Graham et al., 2001). In **Design C**, 25% of participants have complete data, and the remaining participants are each missing data from 2 of the 3 middle timepoints. In the **Complete Data Design**, all participants will have complete data at all time points, but the sample size will be 70% of the size of the missing data designs (in other words, it too will have 30% missing data, in the form of entire missing cases).

In addition to the unplanned missingness, unplanned MAR data will be imposed using the demographic variables to predict the rate of dropout. On each testing interval after the first, the dropout rate range from 10-30% of the remaining participants (Table 2). After all missingness is imposed, a degree of test-reactivity will be imposed on every testing interval after the first on which a participant is measured. Test-reactivity will be compounded for each additional time that a participant is measured (e.g., a participant who has been measured once before will show less reactivity than one who has been measured twice).

4) Estimate the model twice, using FIML and MI. Estimation will be done in M*plus* and R.

5) Compare estimates. The true parameter estimates are the ones used to generate the data in step 1. To the extent that the missing data estimation methods perform well, they should accurately and efficiently reproduce the model parameter values. **Standardized parameter bias** will be use to evaluate parameter accuracy; standardized bias = 100 × (average parameter estimate – population value)/empirical SE of parameter estimate, where the empirical SE is the standard deviation of the distribution for the parameter estimate across 1000 replications (Collins, Shafer, & Kam, 2001; Graham, 2009). **Standard error bias** reflects the degree to which standard errors accurately reflect the standard deviation of parameter estimates. SE bias = 100 × (average SE estimate – empirical SE of parameter estimate)/empirical SE of parameter estimate. **Confidence Interval Coverage** is a statistic that jointly reflects bias in both parameter and standard error estimates. Coverage is the percent of 95% confidence intervals (C.I.s) around a parameter estimate (out of 1000 replications) that contain the true parameter value. If parameter estimates are unbiased and estimated standard errors are accurate, then 95% of the observed C.I.s should contain the parameter estimate. To the extent that coverage is higher or lower than 95%, this reflects inaccuracy in parameter or standard error estimates.

Table 1. *Longitudinal Planned Missing Designs.*

	measurement occasion					design		
group	**1**	**2**	**3**	**4**	**5**	**A**	**B**	**C**
1	**D,** X	X	X	X	X	10%	10%	25%
2	**D,** X	–	–	X	X	6%	20%	25%
3	**D,** X	–	X	–	X	6%	20%	25%
4	**D,** X	X	–	–	X	6%	20%	25%
5	**D,** X	–	X	X	X	6%	10%	–
6	**D,** X	X	–	X	X	6%	10%	–
7	**D,** X	X	X	–	X	6%	10%	–
8	**D,** X	–	X	X	–	6%	–	–
9	**D,** X	X	–	X	–	6%	–	–
10	**D,** X	X	X	–	–	6%	–	–
11	**D,** X	X	X	X	–	6%	–	–
12	**D,** –	X	X	X	X	6%	–	–
13	**D,** –	–	X	X	X	6%	–	–
14	**D,** –	X	–	X	X	6%	–	–
15	**D,** –	X	X	–	X	6%	–	–
16	**D,** –	X	X	X	–	6%	–	–

Note. D = demographic battery; X = developmental outcome variables. Dashes indicates where the outcome measures are not administered (i.e., planned missing data). Participants are randomly assigned to group according to the percentages indicated for each design.

Table 2. Factors in Simulation Design

Design	N	Retest Effect	Attrition
planned missing A	100	*d* = .15	10%
planned missing B	350	*d* = .25	20%
planned missing C	500	*d* = .35	30%
complete (70% N)	650		

Note. Each factor will be fully crossed to create 96 simulation conditions, each of which will have 1000 replications and be analysed 4 times: with FIML and MI, and using the correct model and the misspecified model (slope = 0). In all, results from 384 conditions will be compared. Attrition refers to the percent of participants who leave the study at each wave; *d* is a standardized effect size measure.

As well as the quality of parameter estimates, we are also interested in the quality of the test statistic, in terms of both observed Type-I error and Power. **Type-I error** is evaluated by computing the percent of replications (out of 1000) where the chi-square test statistic is less than a specified alpha level, e.g., .05. If the chi-square statistic is accurate, then the model should be rejected on about 5% of replications. To evaluate **power**, we must choose a misspecified model that is deviant enough that we care to distinguish it from the correct model. In developmental models, much of the interest is in the mean and variance of the slope factor, because it describes the developmental trajectory. For each replication, we will run a second model on the same data with the mean and variance of the slope factor constrained at 0 (see Graham et al., 2001). The two models will be compared using a nested chi-square comparison. The percent of replications on which the chi-square test returns a significant result (i.e., suggests that the model with no slope fits significantly less well than the model with slope included) is the power of the model to detect that meaningful model misspecification.

Contributions and Dissemination

The proposed research will make substantial progress in identifying the features of longitudinal planned missing designs that will guide developmental researchers on when, how, and why to use planned missing elements in their research designs. Because these designs can increase validity by reducing unwanted influences such as reactivity, fatigue, and test-retest effects, this research has the potential to have a profound influence on the quality of future developmental research. The results will be published in high calibre peer-reviewed journals that are accessible to applied researchers, for example, in *Psychological Methods* and *Child Development*. The results will also be published as a chapter in a peer-reviewed monograph in the Guilford series on Methodology in the Social Sciences. The published papers will help developmental researchers in the social sciences to design and use planned missing techniques, thereby impacting the future of best-practice methodology in developmental research. In addition to demonstrating the merits of these designs and the conditions under which validity is enhanced, the proposed simulations will explore the boundaries of planned missing designs to understand the circumstances under which these designs are and are not advisable, and to understand how to best implement them in the context of real-world problems such as high drop-out rates and strong retest effects.