Ken Kelley's Research Statement

I am a research methodologist with two lines of research that are important in the social and behavioral sciences (e.g., management, psychology, sociology, education, & behavioral medicine). My principal line of research is research design, especially issues of sample size planning. In empirical research, choosing an appropriate number of participants for a study is not a simple task but greatly affects the likelihood of satisfying the particular research goal(s). Using a sample size that is much too large for the particular research goal potentially puts more participants than necessary at risk, delays dissemination of findings, and is not a good use of resources. If a sample size is too small for the particular research goal, however, the research goal may not be answered with enough certainty to add to the literature or ensure that the participants' time was used wisely.

Historically, sample size planning has focused almost exclusively on calculating an appropriate sample size to ensure adequate statistical power, which is the probability of rejecting a false null hypothesis (i.e., finding statistical significance). In the last ten years a paradigm shift has occurred in many disciplines involving an explicit move away from null hypothesis significance testing and toward the formation of confidence intervals. However, most applications of sample size planning and ongoing methodological work on sample size planning have continued to focus on statistical power, which is relevant only for hypothesis testing. The new focus on forming confidence intervals requires a completely new way of conceptualizing an appropriate sample size. My major contribution to the literature has been developing the new approach to sample size planning that I termed *accuracy in parameter estimation* (AIPE; Kelley & Maxwell, 2003), which enables scientists working within the confidence interval paradigm to plan an appropriate sample size.

The AIPE approach focuses on the accuracy of an estimated effect size as an estimate of its corresponding population value, in which an effect size quantifies the magnitude of some phenomenon (e.g., a mean difference, correlation coefficient, regression coefficient, etc.). The AIPE approach is operationalized by obtaining a sufficiently narrow confidence interval for the population parameter of interest, which is the value that is ultimately of interest in research, not the obtained sample estimate. There is widespread scientific agreement that confidence intervals are the best way to quantify the uncertainty of an estimate, but in the social and behavioral sciences their use was rare. An eminent methodologist suggested that the reason researchers often did not report confidence intervals was because the widths were often "embarrassingly large" (Cohen, 1994, p. 1002). The problem of "embarrassingly large" widths is addressed with AIPE by planning an appropriate sample size so that the confidence interval will be sufficiently narrow, thus illustrating the accuracy of the estimated magnitude of the parameter, addressing one of the major reasons confidence intervals were only sporadically used in the social and behavioral sciences.

AIPE is widely seen as an important contribution because of the emphasis on the shift away from null hypothesis significance testing and the dichotomous decision it dictates (i.e., reject or fail-to-reject the null hypothesis) in favor of estimating the magnitude of a parameter via confidence intervals. This shift to confidence intervals is exemplified by the strong wording of the American Psychological Association in its publication manual – "it is almost always necessary to include some measure of effect size in the Results section" and "whenever possible, provide a confidence interval for each effect size reported to indicate the precision of estimation of the effect size" (2010, p. 34). Using effect sizes and confidence intervals to make sense of findings and for reporting results is the future of social and behavioral science research and nicely dovetails with my work on sample size planning to estimate the magnitude of an effect, not simply its existence (i.e., if it reaches statistical significance).

I have developed and published AIPE sample size planning methods for the most common statistical methods used in the social and behavioral sciences. In particular, I have personally or in collaboration with colleagues and students published AIPE sample size planning methods for targeted regression coefficients (Kelley & Maxwell, 2003), the squared multiple correlation coefficient when regressors are random (Kelley, 2008) or fixed (Kelley & Maxwell, 2008), the mean difference (Kelley, Maxwell, & Rausch, 2003), the standardized mean difference (Kelley & Rausch, 2006), the coefficient of variation (Kelley, 2007), for targeted effects in structural equation models (Lai & Kelley, 2011), for the root mean square error of approximation in the structural equation modeling context (Kelley & Lai, 2011), and the group-by-time interaction in randomized longitudinal designs (Kelley & Rausch, in press). Additionally, more AIPE methods have been developed and are undergoing review, specifically unstandardized or standardized contrasts in analysis of (co)variance (Lai & Kelley, under review) and reliability coefficients (Terry & Kelley, under review). The AIPE and power analytic approaches to sample size planning were reviewed in an invited submission for *Annual Review of Psychology* (Maxwell, Kelley, Rausch, 2008), and various edited methodological works (Kelley, in press; Kelley & Maxwell, in press, Kelley & Maxwell, 2008), one of which considers ethical implications of sample size planning (Maxwell & Kelley, 2011).

Although confidence intervals are extremely important for quantifying and attempting to understand various phenomena, historically there was a major impediment to computing confidence intervals for effect sizes that are standardized, which are widely used in the social and behavioral sciences. The process of standardization makes the sampling distributions of the statistics more difficult to deal with, yet has the advantage that the effect size is divorced from the particular measurement scale, making it unit free and therefore more comparable across studies, which is important for meta-analyses. Such confidence interval methods, however, could only be implemented with specialized programming, with which many researchers are unfamiliar. Because I needed a way to calculate confidence intervals for my work on AIPE for unstandardized and standardized effect sizes, I wrote my own computer code to implement such confidence intervals. That code became the MBESS package for the program R, both of which are open source. MBESS was originally released in 2006 with major updates in 2008 and 2010. This computer package was twice peer reviewed (Kelley, 2007a; 2007b) and addresses many of the difficulties that arise when calculating effect sizes and forming confidence intervals for standardized effect sizes and has been used by many researchers, recently receiving its own entry in the *Encyclopedia of Research Design* (Kelley, 2010a; AIPE also received its own entry, Kelley, 2010b). As Hoyt, Imel, and Chan note when reviewing the importance yet difficulty of computing confidence intervals for standardized effect sizes, "an important recent innovation is the development of MBESS….the accessibility of MBESS…represents a large step toward making the ideal of routine reporting of effect sizes and confidence intervals" possible (2008, p. 336).

In addition to research design, my other research stream is longitudinal data analysis, commonly termed the analysis of change. This is an important area because it models multiple individuals as they change over time. Not only can comparisons between different types of individuals be made, the types of trajectories the individuals follow as well as correlates of those trajectories can be studied. In two separate works for different time structures, I have delineated a common misconception relating to how certain longitudinal models were interpreted. In particular, I showed that only a limited set of situations allow one to interpret the slope from a straight-line change model as the *average rate of change*, which is a single number that describes the theoretical mean slope across time for an individual or across individuals. Kelley and Maxwell (2008) discussed the average rate of change for a finite number of measurement occasions, whereas Kelley (2009) expanded the methods to continuous time data, both with special emphasis on models with nonlinear functional forms. My research lines of research design and longitudinal data analysis were combined in Rausch, Maxwell, and Kelley (2003) and Kelley and Rausch (in press).

For most longitudinal data analyses, a homogeneous population is assumed (i.e., all individuals share a common set of parameter values). In some situations, however, latent (unobserved) classes of individuals are thought to exist and have a different set of population values for their change coefficients. The growth mixture model was developed to address unobserved classes (essentially a missing grouping variable) and allows latent classes to exist when modeling change. The growth mixture model, however, is built on a system of linear equations, and cannot adequately address phenomena that are both heterogeneous and nonlinear. In Kelley (2008) I developed the *nonlinear change mixture model* that generalized the ideas of the growth mixture model to heterogeneous populations (i.e., in which latent classes exist) where change follows nonlinear functional forms. Special cases of this model subsume many of the most widely used change models, and thus it is a largely encompassing model for modeling phenomena that change over time.

I have also applied sophisticated longitudinal models to address important substantive research questions. Not only are the applications of models beneficial for the particular research domains, it helps me better understand the limitations of currently developed models for the types of data and the types of research questions that are important in different areas of applied research. I have had much success collaborating with substantive researchers. For example, in a series of works collaborating with a nephrologist, I introduced methods to nephrology more commonly used in the social and behavioral sciences to study kidney disease progression and related factors (Agarwal, R., Metiku, Tegegne, Light, Bunaye, Bekele, & Kelley, 2008; Agarwal, Kelley, Light, 2008; Kelley, Aricak, Light, & Agarwal, 2007). Such cross-fertilization of methods proved very useful and led to an R01 NIH grant and a novel methodological development for modeling and understanding the simultaneously steady increase and cyclic nature of blood pressure between dialysis treatments for individuals with chronic kidney disease (Kelley, Light, & Agarwal, 2007), as well as how physicians and medical researchers should most appropriately interpret the results of clinical trials (Singh, Kelley, & Agarwal, 2008). I have also applied sophisticated models to developmental psychology (Alexander, Johnson, Leibham, & Kelley 2008; Stright, Gallagher, 2008; Anderson, Wright, Kelley, K., & Kooreman, 2008). A different type of longitudinal model is a multiplicative heterogeneous diffusion model that addresses time-to-event data. I applied such a time-to-event model to help understand when and what factors led to hospitals adopting electronic medical records between 1976 and 2004 for approximately 4,000 hospitals, in what we believe is the largest application of this type of model (Angst, Agarwal, Sambamurthy, & Kelley, 2010).

I have also done related methodological work, such as evaluating the effectiveness of confidence intervals for effect sizes when assumptions are violated (Kelley, 2005), issues of classification (Holden & Kelley, 2009; Rausch & Kelley, 2009), various measurement issues (Kelley, 2004; Kelley & Cheng, in press), defining effect sizes for mediation models (Preacher & Kelley, in press), and model diagnosis in structure equation modeling (Yuan, Kouros, & Kelley, 2008).

With regards to the impact of my work, according to the ISI Web of Knowledge, there have been 238 citations to my work with an *H*-index of 12. This excludes citations to MBESS, chapters, some journals that are not indexed by ISI, and misspellings of my name (i.e., Kelly instead of Kelley).