Nicholas J. Seewald
Research Statement

I develop statistical methodology for the design and analysis of randomized trials which yield complex longitudinal data for precision health. Specifically, I focus on methods that aid in the construction of decision rules which specify for whom to provide what treatment and when. I take a broad approach to this, seeking to make an impact in both statistics and domain sciences from a project's inception to the dissemination of results. This approach allows me to

1. work closely with domain scientists in health and education to deeply understand theoretical models guiding treatment and policy decision-making,
2. use this knowledge to formulate appropriate scientific questions (estimands) regarding development of tailored intervention sequences in health and education,
3. develop novel experimental designs (approaches to data collection) and associated analytic methods to address those scientific questions, and
4. create dissemination tools, including workshops and web apps, to place methods I develop directly in the hands of domain scientists.

In general, my research goals are driven by the needs of domain scientists to address novel scientific questions which offer opportunities for statistical innovation. I have a history of successful collaborations across a wide array of disciplines, from substance use to physical activity to oncology, and I look forward to opportunities to continue collaborative work.

Sequential, Multiple-Assignment Randomized Trials

Inerventionists across domains are increasingly interested in developing sequences of treatments which are able to adapt to an individual's changing needs, much like standard clinical practice. These sequences, commonly called dynamic treatment regimes (DTRs), can be developed using a sequential, multiple-assignment randomized trial (SMART). SMARTs consist of multiple stages, each corresponding to a treatment recommendation made by a DTR.

SMARTs often have longitudinal outcomes which are collected over the course of multiple stages of the trial. I developed the first easy-to-use sample size method for two-stage SMARTs with continuous, longitudinal outcomes in which the primary aim is to compare the mean end-of-study outcomes of two embedded DTRs which recommend different first-stage treatments [2]. The sample size formula is based on a regression analysis which leverages the longitudinal outcome. It can be decomposed into the product of three terms: (1) the standard sample size formula for a two-arm randomized trial, (2) a deflation factor to account for within-person correlation in the longitudinal outcome, and (3) an inflation factor to accommodate the SMART design. In that paper, I also introduced method-of-moments estimators for the working covariance components of the GEE-type estimating equations used in longitudinal SMART analyses, and developed a new way to simulate data from a SMART.

Micro-Randomized Trials

To capitalize on today's prevalence of smartphones, behavioral interventionists are increasingly interested in mobile health and just-in-time adaptive interventions (JITAIs). These allow for the delivery of highly-tailored brief intervention components with the goal of improving near-term health outcomes (e.g., step count). The micro-randomized trial (MRT) is a novel experimental tool for testing the efficacy of components of a JITAI. In an MRT, participants may be randomized hundreds or thousands of times to receive (or not receive) a brief mobile health intervention.

I was the primary data analyst on HeartSteps, the first ever MRT [1]. In the trial, participants were randomized five times daily to receive a push notification encouraging them to engage in physical activity, with the goal of seeing an increase in their step count in the 30 minutes following randomization. Serving as the analyst on this trial allowed me to formulate best practices for designing research-grade automated data collection systems for MRTs [3, 5]. The expertise I developed in managing complex datasets has also proven valuable in statistical consulting settings.

Dissemination and Impact

Statisticians are uniquely positioned to impact the scientific process, and disseminating methods is a large component of that. I create tools which can reduce the barriers to entry to the methods I develop, thus allowing me to see the impact of my work on the scientific community in a relatively short period of time.

I have written and deployed two comprehensive, widely-used R Shiny web apps for sample size calculation: one for SMARTs with both binary and continuous end-of-study outcomes, and one for micro-randomized trials [4]. The SMART tool allows investigators to describe their proposed trial design by answering a series of questions, and elicits necessary parameters in an intuitive way. This is in sharp contrast to many other online sample size tools, which often rely on concepts and notation which are difficult to understand. I developed the MRT app as part of my work on analyzing data from HeartSteps. As in the SMART tool, the user is guided through a series of steps in which they provide information about their trial design and the hypothesized evolution of the proximal treatment effect over time. Both applications are designed to be simple and helpful, with the goal of making these methods accessible to the widest possible audience.

Another component of my dissemination work is to facilitate trainings for methodologists and domain scientists interested in employing SMART designs in their research. After finishing my M.S. in biostatistics, I co-facilitated a workshop for epidemiologists and scientists in an HIV/AIDS research group at the University of California San Francsico. Since then, I and have been asked to lead or co-lead multiple such trainings. Currently, I am developing a half-day session on analyzing data from a SMART as part of a three-day federally-funded statistical training sponsored by my research group. These opportunities allow me to leverage both my teaching and consulting experience, and I have received strong positive feedback for my roles.

Future Research Goals

A natural extension of the sample size method introduced in [2] is to accommodate sample size calculations for estimands other than a difference in end-of-study outcomes between two embedded DTRs. Other common comparisons may be of area-under-the-curve or time-averaged differences, again between two DTRs. The sample size formulae for these comparisons are not as interpretable as that for the end-of-study comparison, and parameter elicitation may be more challenging; therefore, I plan to develop a software tool for this, rather than a set of formulae.

A second extension of this work is to address balancing the number of measurement occasions with the sample size, subject to a cost constraint. The addition this constraint recognizes the practical realities of running a trial; this project, like the rest of my methodological contributions, will develop a useful tool for domain scientists and applied statisticians to more easily implement novel experimental designs which address pressing questions in their field.

In the long term, I plan to pursue methods to handle intensive longitudinal data (ILD) arising from a SMART. Current methodology for longitudinal-outcome SMARTs is largely based on generalized estimating equations (GEE). In an ILD setting, in which data collection is potentially burdensome to participants, it may be difficult to achieve a large enough sample size for the asymptotic properties of GEE to hold. Additionally, the relative inflexibility of GEE may not be able to address questions about the dynamics involved in tailored transitioning between stages in a SMART. This method will enable scientists to assess, in more detail, treatment trajectories and delayed effects that may arise as a consequence of sequencing interventions in a DTR.

I look forward to continue developing novel, impactful experimental design tools and to use them in collaboration with scientists to address pressing questions across domains.

References

[1] Predrag Klasnja, Shawna Smith, Nicholas J Seewald, Andy Lee, Kelly Hall, Brook Luers, Eric B Hekler, and Susan A Murphy. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. *ann. behav. med.*, page 10, 2018.

[2] Nicholas J Seewald, Kelley M Kidwell, Inbal Nahum-Shani, Tianshuang Wu, James R McKay, and Daniel Almirall. Sample size considerations for comparing dynamic treatment regimens in a sequential multiple-assignment randomized trial with a continuous longitudinal outcome. *Statistical Methods in Medical Research*, 2019.

[3] Nicholas J. Seewald, Shawna N. Smith, Andy Jinseok Lee, Predrag Klasnja, and Susan A. Murphy. Practical Considerations for Data Collection and Management in Mobile Health Micro-randomized Trials. *Statistics in Biosciences*, January 2019.

[4] Nicholas J. Seewald, Ji Sun, and Peng Liao. MRT-SS Calculator: An R Shiny Application for Sample Size Calculation in Micro-Randomized Trials. *arXiv:1609.00695 [stat]*, September 2016. arXiv: 1609.00695.

[5] Shawna N. Smith, Andy Jinseok Lee, Kelly Hall, Nicholas J. Seewald, Audrey Boruvka, Susan A. Murphy, and Predrag Klasnja. Design Lessons from a Micro-Randomized Pilot Study in Mobile Health. In James M. Rehg, Susan A. Murphy, and Santosh Kumar, editors, *Mobile Health*, pages 59–82. Springer International Publishing, Cham, 2017.