Research Statement

Rebecca Yates Coley, Ph.D.

My primary research interest lies in developing statistical methods for precision medicine. As a postdoc, I've collaborated with urologists through the Johns Hopkins Individualized Health Initiative to build a statistical model that predicts the presence of undetected high grade prostate cancer in patients with initial diagnoses of lower grade disease [3]. I am also a co-investigator on a recently awarded methodology grant from the Patient-Centered Outcomes Research Institute (PCORI): *Bayesian hierarchical models for the design and analysis of studies to individualize health care*. This work is a continuation of my dissertation research on Bayesian frailty models to account for individual-level heterogeneity in HIV prevention trials [4]. Research to advance precision or "individualized" medicine requires both principled extension of existing methods as well as creative development of new methods in several areas of statistics that I find particularly exciting: hierarchical Bayesian modeling, unobserved heterogeneity, latent class analysis, missing data, learning health systems, multi-cohort implementation, and real-time sampling.

Hierarchical Bayesian modeling: Hierarchical models provide a natural statistical framework for individualized medicine, allowing for time-varying health states within a patient as well as for variability across individual patients embedded within a population. A Bayesian approach to hierarchical modeling encourages inclusion of existing medical knowledge at each level. Additionally, hierarchical Bayesian models easily handle several characteristics common to data collected in a clinical setting, such as unobserved heterogeneity and informative missingness. My dissertation research also relied on a hierarchical Bayesian framework, which proved to be both particularly intuitive for modeling heterogeneity in HIV risk and computationally expedient.

Unobserved heterogeneity: My dissertation research focused on developing mechanistically-motivated models for heterogeneity in the context of HIV prevention trials. While the Cox model is typically used to estimate intervention effectiveness, the population-level estimate obtained underestimates the intervention's effectiveness for at-risk individuals when heterogeneity in risk is present. As an alternative, I proposed using the compound Poisson distribution– which reflects the sources of variability in HIV risk and allows for no exposure among some trial participants– to model frailty and enable estimation of individual-level effectiveness. This research was the first to demonstrate Bayesian estimation of a compound Poisson frailty model with univariate survival data and entailed development of a novel data augmentation MCMC algorithm for posterior sampling of hierarchically-structured frailty components [1]. Modeling unobserved heterogeneity is also a central aspect of my recent work in individualized medicine as variability in risk of disease, expression of health states, and effects of interventions is expected.

Latent class analysis: In my dissertation, I proposed a joint latent class approach to frailty modeling in which, first, latent classes were identified via risk-related covariates and, second, individuals within a latent class shared a frailty distribution [2]. Latent modeling of an individual's health state also features prominently in individualized medicine since true health states frequently cannot be directly or precisely observed. In the context of prostate cancer, regular biopsies are used to monitor disease state, but biopsy grading is prone to misclassification error. My prediction model estimates a posterior probability for an individual's latent cancer state via joint modeling of the latent state, repeated biopsies, and longitudinal biomarker measurements.

Missing data: New methods for situations where data are missing not at random (MNAR) are also essential in individualized medicine, as missing completely at random and missing at random assumptions are rarely justified in clinical settings. As part of my prediction model for prostate cancer, I have proposed a novel shared parameter model that accommodates MNAR observations of both the true cancer state and biopsy results. I plan on pursing this research further and developing this approach into a more general framework for broader application.

Learning health systems: Individual medicine is most completely realized in the context of a learning health system, that is, a system with the ability to continuously integrate patient data and medical knowledge to optimize patient care. As an example, my prostate cancer prediction model can be adapted to incorporate advances in scientific understanding or newly available clinical measurements without discarding earlier (now incomplete) data. In future research, I would like to establish best practices for statistical modeling to support learning health systems.

Multi-cohort implementation: The efficient implementation of statistical models for individualized medicine also requires transportability beyond local settings. I am collaborating with prostate cancer researchers at the Fred Hutchinson Cancer Research Center to apply my latent class prediction model to an outside cohort. The hierarchical Bayesian framework offers a natural extension of the existing prediction model to another cohort– parameters expected to differ across groups can be allowed to vary while others are shared. We plan to develop a principled procedure for the continued addition of patient cohorts as they arise with the ultimate goal of facilitating a nation-wide network of disease surveillance for prostate cancer patients.

Real-time sampling: Clinical use of individualized predictions require that posteriors be updated in real-time to reflect new observations. I have advised a graduate student in developing an importance sampling algorithm that enables fast updating of latent state predictions for prostate cancer patients [5]. Our future research will focus on developing a more online learning approach that does rely on periodic re-estimation of the full model via MCMC, a process which will become computationally demanding as new patients and cohorts are added.

References

[1] R Yates Coley and Elizabeth R Brown. Estimating effectiveness in HIV prevention trials with a Bayesian hierarchical compound poisson frailty model. 2015.

[2] R Yates Coley and Elizabeth R Brown. Latent class approach to survival analysis with a compound poisson frailty mixture model. 2015.

[3] R Yates Coley, Aaron J Fisher, Mufaddal Mamawala, H Ballentine Carter, Kenneth J Pienta, and Scott L Zeger. Bayesian joint hierarchical model for prediction of latent health states with application to active surveillance of prostate cancer. *arXiv preprint arXiv:1508.07511*, 2015.

[4] Rebecca Yates Coley. *Bayesian Hierarchical Frailty Models for Heterogeneity in Risk*. PhD thesis, University of Washington, 2014.

[5] Aaron J Fisher, R Yates Coley, and Scott L Zeger. Fast out-of-sample predictions for bayesian hierarchical models of latent health states. *arXiv preprint arXiv:1510.08802*, 2015.