Statistical estimates from survey samples have traditionally been obtained via design-based estimators. In many cases these estimators tend to work well for quantities, such as population totals or means, but can fall short as sample sizes become small. In today’s “information age,” there is a strong demand for more granular estimates. To meet this demand, using a Bayesian pseudolikelihood, we propose a computationally efficient unit-level modeling approach for non-Gaussian data collected under informative sampling designs. Specifically, we focus on binary and multinomial data. Our approach is both multivariate and multiscale, incorporating spatial dependence at the area level. We illustrate our approach through an empirical simulation study and through a motivating application to health insurance estimates, using the American Community Survey.

This is joint work with Paul A. Parker (UC Santa Cruz) and Ryan Janicki (U.S. Census Bureau)