ADDRESSING COMPLEX PATTERNS OF MISSING DATA AND MEASUREMENT ERROR IN EHR-BASED RESEARCH

REBECCA A. HUBBARD
University of Pennsylvania, Philadelphia, PA, USA
e-mail: rhubb@mail.pennmedicine.upenn.edu

Electronic Health Records (EHR) include a wide variety of clinical and administrative data that can be used to study exposures and health outcomes in a real-world setting. However, because these data were not collected for research purposes, the availability and quality of data are typically much poorer than data collected using traditional observational study designs. Although EHR data provide information about patient exposures and outcomes, many factors affect data quality and availability, including institutional coding conventions, financial incentives, and record-keeping systems. For instance, which biomarkers are measured and how frequently, medication prescribing patterns, and use of specific diagnosis codes may vary systematically between medical providers. Statistical methods that ignore this complex data generating mechanism may suffer from substantial bias. We propose a Bayesian latent variables approach, harnessing existing evidence and expert opinion on the data generating process, to facilitate estimation of patient outcomes and exposure/outcome associations. Through simulation studies we demonstrate improvements in bias and efficiency associated with using this approach compared to alternative missing data approaches. We applied our new approach in a study of pediatric diabetes, using EHR data from a network of children’s hospitals.