Specific Aims

National health databases and electronic health records (EHR) are clustered by procedure, provider, service, institution and geography and inherently incomplete. Often neglected, this rich spatial and temporal organization is most realistically captured in hierarchical statistical models with cluster-specific parameters, but estimating such models is still difficult for medical researchers using traditional software. Uncollected, incomplete or missing data can further hinder modeling or bias inferences. We propose to further develop the capabilities of the software called *Stan* — which is a probabilistic programming language, mathematical library, suite of estimation algorithms, and ecosystem of supporting interfaces — in order to make advanced hierarchical modeling accessible to data scientists for EHR-based medical outcomes research with incomplete data. Making *Stan* more accessible has the potential to transform the way medical outcomes research is conducted.

Flexibility and robustness of hierarchical modeling could transform EHR based outcomes research. Consider surgery as an illustrative example. Patients in the same hospital undergoing the same surgical intervention by the same team will show similar clinical trajectories and responses. We are interested in investigating differences in therapeutic effects and in predicting adverse outcomes in order to prevent them. Doing so entails (1) estimating individual intercept-shifts for each provider and procedure to control for potentially confounding differences in quality of care by different teams, (2) allowing for spatial clustering of adherence behavior, e.g., by different services, which can be represented by multilevel modeling, and (3) partially pooling estimates to improve precision, especially in subgroups with sparse data. Prediction of adverse health outcomes can be improved by exploiting the implied correlations between different but related subsets of data. Conversely, failure to account for the highly structured and correlated nature of health care delivery or failure to account for the mechanisms by which some data are missing may lead to incorrect statistical inferences, poor predictions, and adverse health consequences. Realistic modeling of the heterogeneity in care delivered may help to identify reasons for variance in performance and may point to ways to improve outcomes.

Classical approaches and traditional software often lack flexibility for hierarchical modeling. Most available software limits the types of hierarchical models that can be estimated because their algorithms are more likely to run into computational problems when estimating more complicated hierarchical models. In contrast, *Stan* is a flexible, general-purpose modeling language and a novel, powerful estimation engine that has facilitated advanced hierarchical modeling in biostatistics, epidemiology, public health, political science, and pharmacokinetics. *Stan's* development has been funded by the NSF, DoD, and other organizations. *Stan* and the interfaces to it (e.g. from *Python, Julia, R*, etc.) are open source and platform-independent.

Good hierarchical modeling should be accessible, transparent and requires good model diagnostics However, building and tuning sophisticated hierarchical models with *Stan* (or other available software) is still very challenging even for the initiated. We lack visual exploratory and diagnostic tools to recognize when a complicated model is logically flawed or fails to fit the data empirically. We propose to enhance *Stan's* ecosystem with more accessible interfaces and statistical tools for principled model checking and re-specification.

Aim 1: To further develop our software package *rstanarm*, a more user-friendly interface to *Stan*, for the open-source statistical language and environment *R*, in order to make it more suitable and accessible for clinical data scientists and biostatisticians who want to estimate hierarchical models for EMR data.

Aim 2: To further develop *shinystan*, our interactive web application to analyze and visually explore the output of *Stan* models and to develop diagnostics to identify and troubleshoot computational and empirical problems with advanced hierarchical models.

Aim 3: To further develop *mi*, our *R* package for multiple imputation of missing data so that it can make use of *Stan* via the models provided by *rstanarm*.

Aim 4: To explicate, document, and disseminate realistic hierarchical models for incomplete data to the clinical data science community with hands on use cases, workshops, journal articles, and online tutorials. To solicit the data scientist community feedback, engage new software developers, and to incorporate improvements to our software.