Specific Aims

National health databases and electronic health records (EHR) are inherently clustered by procedure, provider, service, institution and geography. Their rich spatial and temporal organization is most realistically captured in multilevel hierarchical statistical models, but efficient model fit and computational implementation are still difficult. We propose to further develop Stan, a novel, probabilistic programming language, to make advanced hierarchical modeling more readily accessible to data scientists for outcome and health services research.

Flexibility and robustness of hierarchical modeling could transform EHR based outcomes research. We 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. Typically we are interested to investigate differences in therapeutic effects or to predict poor outcomes to prevent them. (1) By estimating individual effects for each provider or procedure, random effects can help to control for potentially confounding differences in quality of care by different teams. (2) Spatial clustering of adherence behavior , e.g by different services, can be represented by multilevel modeling. (3) Especially in subgroups with sparse data, partial pooling can improve parameter estimates; for example prediction of poor health outcomes can be improved by exploiting the implied correlations using information from different but related subsets. Failure to account for the highly structured and correlated nature of health care delivery may lead to incorrect statistical inferences.

Hierarchical modeling and related diagnostics should be readily accessible to clinical data scientists. At present, it still takes a computationally sophisticated statisticians to write complex hierarchical models, transform the data or parameters to facilitate model convergence and index the group indicators in a multilevel hierarchical model error free. We lack intuitive diagnostic, visual and exploratory tools to address convergence and model fit. Model convergence diagnostics and troubleshooting algorithms are still underdeveloped.

Classical approaches and software packages often lack flexibility for multilevel hierarchical modeling. Available algorithms are slow to converge even on advanced workstations, if they converge at all. Stan, our flexible general-purpose modeling language facilitated much more widespread multilevel modeling statistics in biostatistics, epidemiology, public health and political science and pharmacokinetic modeling. Stan's novel Hamiltonian Monte Carlo algorithm converges faster by orders of magnitude and is ideally suited to fit even advanced and unorthodox statistical models. Stan's development was funded through the NSF; therefore Stan and its algorithms are open source; they are implemented on multiple platforms (Python, Julia, R/Rstudio).

Robust, efficient, expressive and accessible software promotes Big Data outcomes research. We propose to further simplify complex multilevel model building for clinical and health services research by developing additional and more user friendly open source software packages around Stan, incorporating interactive, intuitive visual and novel statistical tools to facilitate principled model optimization and checking.

Specific aims

Aim 1: To develop simple and user-friendly software packages around Stan in the open-source statistical computing environment R/Rstudio in collaboration with applied clinical data scientists and biostatisticians. To make complex hierarchical modeling of EMR computationally efficient and readily accessible to average clinical data scientists with a simple standardized function call to a representative class of multilevel models.

Aim 2: To develop an interactive diagnostic software package to analyze and visually explore the convergence and output of complex hierarchical models and to develop novel principled diagnostic algorithms and utilities to diagnose and troubleshoot non-convergence, detect multidimensional co-linearity and accelerate computational implementation of advance hierarchical models.

Aim 3: To explicate, document and disseminate complex hierarchical modeling and its advanced computational implementation in collaboration with the clinical Big Data science community with hands on workshops, e-books, online tutorials and electronic resources. To solicit the Big Data community feedback, engage new software developers and to incorporate corrections into our software through online user and developer groups.

Research Strategy

Significance

The nested structure of health care delivery and electronic health data

Clinical data scientists are faced with an abundance of useful electronic health data, but limitations of existing statistical inference tools constrain the scientific hypotheses they can be explore and evaluate. Electronic health related data sets are not only growing exponentially in number of units of observations or variables observed, but this growth implies increasingly complex interactions and correlations.

National Anesthesia Clinical Outcomes Registry is an example of hierarchically structured health data

We will illustrate the clustered, nested data structure of contemporary health care delivery and electronic health data capture with the example of the National Anesthesia Clinical Outcomes Registry (NACOR), maintained by the Anesthesia Quality Institute and funded by the American Society of Anesthesiologists.

Perioperative health care delivery and data capture are nested and clustered.

Anesthesia is an illustrative case of the how health care is increasingly electronically documented, facilitating the continuous registration of multiple simultaneous physiological data and therapeutic interventions during critical interventions. To support billing, this electronically captured data set is jointed with surgical and anesthesia procedure codes, International Classification of Disease (ICD) codes, provider identifiers, patient perioperative risk, outcome and provider compliance assessment. Participating institutions upload this comprehensive file from their anesthesia information management systems (AIMS) directly to NACOR. NACOR contains at present over 30 million individual electronic records of anesthesia care provided; like similar databases, NACOR is growing exponentially. This data mine invites health services and outcomes research.

Care delivered and outcomes achieved depend on procedure, providers and patient characteristics.

For example, the particular anesthetic a given individual receives will depend on the surgical procedure the patient is undergoing and under which service, but also on the local institutional culture and indeed the individual anesthesia provider and his or her qualifications and preferences[1]; We seek to substantiate the nested structure of outcome data in health care with two examples from (1) labor analgesia and (2) spine surgery :

   (1) Anesthesiologists may feel more or less inclined or competent to offer regional anesthesia techniques; provision of epidural labor anesthesia varies widely across the nation and within an institution and is predicted by socioeconomic and racial patient characteristics[2,3].

   (2) Bleeding during spine surgery is substantially less, if performed by a neurosurgeons versus by an orthopedic surgeon; while true on average, an individual gifted orthopedic surgeon may outperform the average neurosurgeon with regards to surgical blood loss.

Hierarchical models capture contemporary health care practice realistically

Hierarchical modeling could transform electronic health records based outcomes research, because the evident rich spatial and temporal organization of electronic health records is most realistically captured in multilevel hierarchical statistical models. However, efficient model fit and computational implementation are still difficult. Additional depth of data (simply more units of observation) would increase power and make our clinical data analysis easier. Rather, there is more breadth to the data: more subgroups, locations, provider or time granularity than is currently being modeled, more partial, incomplete and noisy measurements that cannot easily be incorporated into standard models, more related studies available for meta-analysis[4,5].

Modeling the multifaceted correlations in EHR is reflecting actual clinical practice

To realistically model the multifaceted correlations seen in actual clinical practice, when we fitted a regression model to investigate predictors of anesthesia quality in the NACOR database, we wanted regression coefficients to vary by provider, providers again nested by service or by hospital[1]. We needed to control for the surgical procedure type as random effect as well, with thousands of different surgical procedures performed in the NACOR population. The number of parameters to estimate grows very quickly and so do the potential interactions. Even with very large data sets, the sample size in each subgroup will shrink rapidly; estimates using least squares or maximum likelihood will become noisy and thus often become essentially useless. One solution lies in hierarchical modeling, where we estimate hyper-parameters and hyper-hyper-parameters (Figure **??**), to represent how lower level parameters vary across different groupings[6].

Hierarchical models provide efficient inferences with partial pooling

Inference based on partial pooling outperforms (a) the No-pooling and (b) the complete-pooling approaches, as can be shown mathematically[7] or via cross-validation[8].

- (a) Using the No-pooling approach, we would estimate the model for each specific subset of interest separately. But this leads to far too many sub-classifications, e.g. one model for each type of surgery, thus too small samples in any given subgroup for useful inferences, if we fully explore the complexity and granularity, the richness of the EMR data.

- (b) Employing complete pooling or structural modeling constitutes the other extreme of the spectrum, but the implied hard constraints on the coefficients in different groups may lead to bias, negating the obvious known differences in the data, for example between patients undergoing tracheotomy versus cesarean section: we gloss over such detail and lose granularity.

We choose the middle ground: for our richly organized NACOR data set, inference using partial pooling or hierarchical modeling is especially effective , because the estimate of each individual parameter is simultaneously informed by data from all the other patients in our cohort, improving inferences in particular for subgroups with sparse data.[9]. Effron explained this apparent paradox well to non-statisticians in the Scientific American[10;11].

Meta-analysis for evidence based clinical care

Clinicians are familiar with systematic reviews and meta-analysis[12]. Evidence synthesis (a more accurate term than meta-analysis) is a powerful tool to pool clinical trials to guide evidence based clinical care[13]. Rigorous evidence synthesis is considered the highest level of evidence to support clinical decision making[14].

Variance in study design and outcome reporting hamper evidence synthesis

However, studies on perioperative outcomes tend to vary in design and reported outcomes[15], making evidence synthesis challenging with classical or frequentist statistical models[16]; not least, because often only dichotomous aggregate results are reported[17]. The integration of dichotomous outcomes with continuous outcomes often requires access to individual patient data, rarely available for evidence synthesis[4]. Meta-regression of effect dose dependence can explain substantial between-study variance in outcomes reported to reconcile study findings[4;18;19]. Different study designs can make it difficult to perform meta-analysis or meta-regression with classical statistical methods and standard systematic review software[20]. Classical meta-analysis may also underestimate the between-study-variability for small numbers of trials[21;22].

The antimony of complete versus no-pooling also limits meta-analysis of perioperative outcomes

Perioperative outcomes are often recorded at different follow intervals in different studies; some studies report repeated measures, others only a single terminal observation; This leads to the same issue of (a) complete pooling versus (b) No-pooling also in evidence synthesis[17].

- (a) Complete pooling irrespective of follow up-time. Evidence synthesis of all effect estimates at different time points is only appropriate if the effect estimates does not depend on when it was observed. This is obviously often untenable assumption.

- (b) No-pooling, but conducting meta-analyses at each time point: performing separate meta-analysis for each follow up time point separately, would drastically reduce sample size and hence power and precision, undermining the main strength of meta-analysis.

Clustering can affect inferences in meta-analysis

In addition to a possible influence on the point estimate of the measure of effect, an even more important influence can be the transformation of a significant effect into a non-significant effect by changing of the confidence interval of the pooled effect estimate. For example assume across all studies we estimated an overall effect of RR of 0.8. A 95 percent confidence interval of 0.55-1.15 would lead to the inference of "no statistically significant effect" and thwart further attempt to study this intervention, while a confidence interval of 0.7-0.9 may lead to widespread adoption of this therapeutic approach. Interpretation of meta-analyses, and especially Cochrane Reviews, in clinical practice are unfortunately often reduced to "shows a significant effect" vs. "shows no significant effect", differences that have huge clinical impact.

Ecological, disease and geographic study level characteristics can influence inferences

Besides clustering by reported time endpoint, there are other forms of ecological bias or clusters to be considered in meta-analysis; certain geographical or historical settings, similar surgical procedures or diseases will lead to correlated outcomes in patient cohorts[19;15;4;17]. Such the same antimony of complete pooling versus no-pooling limits meta-analysis for perioperative outcomes and hinders evidence based medicine. Besides the effects estimates themselves, their precision, inter and within study variability may differ considerably contingent on disease, procedure or other study level characteristics[15;4;17].

To illustrate this let us consider study findings more prominent in the included cohort and not as strong in other related populations. This would especially be true if average effect sizes, precision of pooled effect estimates, inter- and within-study variability differ considerably from the study cohort used to obtain aforementioned findings. An example of this would be meta-analyses of the treatment effect in different but related diseases like diabetic, HIV-related, traumatic, and idiopathic chronic painful neuropathy[4]. Information in one subset or study population can and should inform estimates in other similar populations at least to some extend in a hierarchical meta-analysis. The principle applies in other fields of medicine, e.g. critical care as well[17].

Partial pooling improves evidence synthesis of studies with variable follow up intervals and endpoints

Hierarchical models may hence be a useful tool to pool heterogeneous perioperative outcome data from long-term studies with varied design to better inform clinical decisions[23;24]; complex multilevel models can be difficult to fit with standard software and should be more accessible. We build a multilevel hierarchical model to pool individual patient data with continuous and dichotomous aggregate study level data, clustered at the study level by followup interval and at the study level by surgical intervention performed (Figure **??**).

Integrating missing data and clustered EHR modeling

Too often data scientists either (1) do complex model but limit the analysis to complete cases, ignoring the missing data or (2) impute the missing data but build overly simplified models.

EHR data are not missing at random.

physicians chose which test to get to inform specific therapeutic decisions and different types of data are recorded in different clinical setting, e.g. arterial lines may not be permissible on the floor, vitals are recorded in greater detail and more frequently in high dependency units like the ICU.

Clustering can bias estimation of confidence intervals.

To correctly estimate confidence intervals even in a simple student t-test, we have to take into account if the data are observed in the same patient repeatedly, i.e. if they are clustered and correlated. Failure to take into account correlations in clustered observation may lead to incorrect inferences. Based on the empirical (robust) (weighted jag knife) methods, the confidence intervals are correct, but such approaches are often ignored in EHR research and they depend on convergence characteristics that may not hold in faceted EHR data.

Punchline: Hierarchical models are best suited to reflect the clustered structure of contemporary health care delivery realistically.

Difficulty to fit and explore complex hierarchical models

It is challenging to build integrated complex models with available software, taking both missing data and clustering into account. This is compounded by the dearth of tools to explore the models and the enormous data stream generated in Markov chain Monte Carlo (MCMC) simulations to fit advanced models.

Innovation

$rstanarm$ and $sinystan$ will address the dearth of accessible software to build complex multilevel hierarchical models, assess and troubleshoot model convergence, confirm congruence of model and data and improve statistical inference. In the process, we will advance statistical methods and algorithms for convergence diagnostics and integrate missing data imputation with advanced hierarchical modeling in a single software package.

Accessible advanced hierarchical modeling for EHR

While there are simple and accessible software packages to implement basic hierarchical models both in the frequentistist and the Bayesian paradigm, multilevel modeling is restricted to data scientist and statisticians with advanced computational and statistical expertise. Even for these, the step learning curve of advanced probabilistic programming languages like Stan constitutes a significant barrier to harness the full potential of hierarchical modeling for electronic health records based outcomes research.

$rstanarm$ is the first very simple and accessible, yet ultra-fast and flexible package to allow even very advanced hierarchical modeling for large data sets. It is open source, and free to the public and our project will make analysis reproducible and reliable enough even for federal regulatory processes.

Fast and flexible hierarchical modeling for realistic EHR outcomes research

Even for the initiated and sophisticated data scientist able to writing complex code to implement complex multilevel models for Big Data, the process of sequential model building and testing is cumbersome and slow. The complexity of programming long convoluted code makes the process error prone. Typically scientist have to explore and compare different angles and approaches to modeling large EHR data sets. Typically researchers need to update and tweak the model to make it more and more complex, to fit the data better or to facilitate convergence of the algorithm. With existing software, like OpenBugs or Jags building and updating hierarchical models sequentially for large data sets becomes prohibitively expensive in computer resources or too slow to be feasible. The challenges of fast and flexible computational implementation such limit the model sophistication; researchers cannot fit multilevel models which reflect the actual hierarchical structure of clinical care delivered.

$rstanarm$ is based on the Hamiltonian Monte Carlo algorithm implemented in Stan, which is orders of magnitude faster than existing MCMC algorithms. $rstanarms$ simple yet flexible function calls facilitate dynamic model building and updating, and allows the fitting and testing of complex models even for larger datasets. $rstanarm$ allows researchers to fit the model they believe to best reflect the clinical question they are investigating with EHR.

Improve tools for graphical exploration of hierarchical models and MCMC output

Accessible tools and utilities, new visual methods to explore sophisticated model and their output would accelerate the model development, reduce modeling errors and fascinates inferential reasoning. $shinystan$ will fill in the void where there is currently a dearth of conceptual approaches, tools and utilities to analyze the very large data generated with Marcov Monte Carlo simulations themselves. $shinystan$ will provide powerful and easily accessible tools for posterior predictive checks to improve congruence of the fitted model with observed data.

Advanced, refined and accessible model convergence diagnostics to enhance model building

If a complex model fails to converge, knowing why is crucial to troubleshooting. $shinystan$ is already unmatched in combining ease of interactive graphical exploration with sophisticated graphical rendering to explore visually and parameter specific co-linearity, auto correlation, tree depth of Hamiltonian Monte Carlo algorithms and many other convergence diagnostics. We will not only further develop $shinystan$ to make it more robust, scale it to work faster also for larger data sets; beyond, we will develop novel graphical methods to assess model convergence and define and refined algorithms to troubleshoot model convergence with a special focus on multilevel hierarchical modeling.

Integrate missing data imputation with hierarchical modeling to advance outcomes research

One major limitation of EHR based outcomes research is imprecise or missing data. We will integrate multiple imputation with advanced hierarchical modeling in $rstanarm$ to facilitate the use of multilevel models for large clinical data sets. We will integrate of both approaches, multiple imputation of missing data with advanced ultra fast complex hierarchical modeling in (a) compatible unified software package(s). This will open a new dimension to clinical data scientist, who were hitherto limited either to complete case analysis if they wanted to use complex multilevel models or to simple models if they utilized multiple imputation of missing data; both with tremendous shortcomings.

$rstanarm$ and $shiystan$ innovation is in also process integration. They will allow clinical scientists to impute missing data from EHR, build and troubleshoot realistic hierarchical multilevel models and explore and analyze the output and results graphically and interactively.

Approach

This project (software development and dissemination) will be guided by our multidisciplinary project team. The team will conduct its work through regularly scheduled weekly meetings, and continued online collaboration via Github and email.

Multidimensional statistical and computational methods for analyzing, inspecting, displaying, representing, parsing, and searching high-dimensional data

Preliminary work

Algorithm, software and prototype development and their application to biomedical data

Funded through several mechanisms including the National Science Foundation (NSF SES-1205516), the team submitting this proposal already developed and implemented many related algorithms and software packages and applied them to biomedical problems funded through the National Institute of Health (5R01GM074806, 5KL2TR001071). The proposed work is a direct continuation of the below described preliminary work of developing and implementing novel ground breaking algorithms for hierarchical modeling of complex data.

Prior work in hierarchical and complex modeling in medicine

Dr. Andreae published several systematic reviews and meta-analyses[15;4;25]. Drs. Andreae, Goodrich and Hall used the prototype of *rstanarm* EHR based health disparities and outcomes research[1] Dr. Goodrich

Dr. Hall is nationally recognized for the development and application of change point models in epidemiology and surveillance. Dr. Gong is leading an NIH funded trial to predict and improve respiratory outcomes after intubation based on real time electronic medical records. Drs. Andreae, Goodrich and collaborators used the software prototype *rstanarm* and shinystan to build a multilevel hierarchical model to investigate health care disparities and quality of anesthesia delivery in the large National Clinical Outcomes Registry maintained by the American Society of Anesthesiology. Dr. Gelman is internationally recognized as a leader in hierarchical modeling with past and present funding and publication in pharmacodynamic modeling, ...

Previous experience in software development The team has ample experience in the development of complex statistical software and in the visualization of model parameters, co-variance matrices and statistical data. Drs. Goodrich and Gelman developed several widely cited and used software packages including the probabilistic programming language Stan[26], the basis for the proposed work. Dr. Goodrich and Andreae work together on the preliminary software packages shinystan and *rstanarm*. Below we outline the trajectory that led to the current project proposal:

A novel algorithm for Bayesian inference: Hamiltonian Monte Carlo

Dr. Gelman, Betancourt and collaborators developed Hamiltonian Monte Carlo (HMC) methods, a novel approach to computationally implement complex hierarchical Bayesian inference through Monte Carlo simulation.

$Stan$**: Open source computational implementation of HMC for diverse interfaces.**

Drs. Gelman, Betancourt and Goodrich developed Stan, an open source multipurpose probabilistic programming language to build complex Bayesian models in several open source and commerical software including so far Stata, Mathlab, Python, Julia and R/Rstudio.

$rstan$**: Implementation in the open source software environment R/Rstudio: rstan**

Drs. Goodrich, Gelman, Betancourt and collaborators developed Rstan, a software package to use Hamiltonian Monte Carlo algorithms the open source statistical software enviroment R/Rstudio.

$shinystan - rstanarm$**: Prototype software development**

Drs. Goodrich, Andreae, Gelman and Betancourt and collaborators developed two additional prototype software package for the open source statistical software environment R/Rstudio, called (1) $rstanarm$ and (2) $shinystan$.

(1) $rstanarm$ **is build to make advanced hierarchical Bayesian models accessible to data scientist** without the need for an understanding of the complexities underlying Hamiltonian Monte Carlo. This software package uses the same notation for model description as other widely accepted software packages for mixed modeling in R/Rstudio like lme4.

(2) $shinystan$ **is an interactive tool to explore and diagnose the output of Monte Carlo simulations** for Bayesian inference. $shinystan$ is a graphical user interface for interactively exploring virtually any Bayesian model fit using a Markov chain Monte Carlo algorithm. Also a package for R/Rstudio, shinystan provides multidimensional statistical, graphical and computational tools for any analyzing, inspecting, displaying, representing, parsing, and searching high-dimensional MCMC output, but is optimized for HMC.

Project scope and goals

The project proposes to develop further develop these two prototypes (1) $rstanarm$ and (2) $shinystan$ into solid, reliable, tested software packages. Both packages will serve as appendage to the $rstan$ package (we developed), which enables the most common applied regression models to be estimated using novel Hamiltonian Monte Carlo algorithms. We will makes our new two packages $rstanarm$ and $shinystan$ available to researchers and the general public via the software repository CRAN.

(1) $rstanarm$

The software package $rstanarm$ will enable the estimation of advanced applied general linear regression models using the existing probabilistic programming language $Stan$. $rstanarm$ implements full Bayesian statistical inference; however, $rstanarm$ will allow users to specify their complex hierarchical models using simplified syntax. Clinical data scientist can therefore take advantage of the more efficient inference of the cutting edge algorithms implemented in Stan, without knowledge of the underlying programming language or the parameter reparametrizations useful to achieve faster model convergence.

Model functions   The prototype of $rstanarm$ already implements many modeling functions; the classical generalized linear regression model functions $stan\_glm$ and $stan\_glmer$ are illustrated below to demonstrate simplicity of model formulation.

Bayesian generalized linear models with and without group-specific terms via Stan are already implemented in $rstanarm$. The $stan\_glm$ and $stan\_glmer$ functions are similar in syntax to the R software functions glm and $glmer$ as used in the R package $lme4$, respectively, but rather than maximum likelihood estimation of generalized linear models, full Bayesian estimation is performed via MCMC.

Model specification   Models are specified using customary R modeling syntax, e.g. analogously to *lme4*, *rstanarm* uses a two-sided linear formula describing both the fixed-effects and random-effects part of the model; the dependent response variable, for example y on the left of a   operator and the independent terms $x_1, x_2...$, separated by $+$ operators, on the right. Random-effects terms are distinguished by vertical bars | separating expressions for design matrices from grouping factors, for example:

$y \sim x1 + (x2|x3)$

In Bayesian inference, priors incorporate subjective beliefs or existing information about a parameter as discussed under significance. All Bayesian models need priors and the prototype of *rstanarm* already adds independent weakly informative priors on the coefficients of the generalized linear model, but prior can be specified by the user. The specification of the prior can such be left to the default implementation in *rstanarm* or the user can choose from a broad array of distribution, e.g. for the intercept one might choose the Student t distribution, which approaches the normal distribution as the degrees of freedom approach infinity and as the degrees of freedom are one, the Cauchy distribution, leaving the user the option of robust priors to allow for outliers.

(2) $shinystan$

We propose to develop *shinystan* as a the graphical user interface for interactively exploring virtually any Bayesian model fit using a Markov chain Monte Carlo algorithm. We will implement *shinystan* in *Shiny*, an R package for interactive web applications. The graphical rendering of our proposed package *shinystan* is building on the superb graphical package ggplot by Hackley Wickham. Many graphical functions of *shinystan* will eventually be ported directly into *rstan* to allow users to integrate the graphical and numerical output of *shinystan* directly into reports generated with the markdown in R.

Our motivation for *shinystan* was the dearth of

Best practices and proven methods for software design, construction, and implementation

Mechanisms for incorporating user reported corrections into the software.

Products of this project

The goal of this project is to encourage wider adoption of complex hierarchical modeling for Big Data by clinical data scientists and to support the creation and reuse of open-source extensions and applications.

Software sharing plan

The software developed for this grant will all be incorporated into Comprehensive R Archive Network of the open-source R Project for Statistical Computing. To facilitate its unimpeded utilization, our source code and documentation will be distributed under the least restrictive open-source licensing terms possible: R's licens-

Year	rstanarm	methodological innovations	shinystan
First	organize project team, submit rstanarm package to CRAN, start online user group, disseminate rstanarm and revamp shinystan		
Second	additional levels to rstanarm generalized linar model, implement basic meta-analysis functions	develop dependence plot, posterior predictive checking and other new visual tools for graphical exploration of model convergence and model fit	
Third	flexible multilevel meta-analysis functions, integrate rstanarm algorithms in multiple imputation package		
Forth	develop and implement basic missing data integration into rstanarm, start implementing advanced multilevel change point models		
Fifth	scale missing data integration to larger datasets		

Table 1: The above time line delineates the planned project progress by year. After setting up the project team in the first year and submitting the rstanarm package to CRAN, in year two and three, we will sequentially add additional model level to existing functions and start to develop new functions for meta-analysis and change point models, scaling also these to multilevel models. In the forth project year, we will begin to integrate missing data algorithms into rstanarm and finally scale the packages to large dataset in year five.

ing is under the GNU General Public License. The benefits are that our software products and subsequent developments

- will be freely available to researchers and the general public, including for commercial use;

- may be freely extended, customized, and incorporated into the other tools;

- can be maintained in the event of the original developers not being willing or able to;

- can be enhanced based on user-provided feedback for bug-fixes, examples, and enhancements; using GitHub pull requests with integration testing

Any documentation is released under the same license as Wikipedia, the Creative Commons Attribution/ShareAlike 4.0 license (CC BY-SA 4.0)[27]

Potential problems

Timeline

test

References

[1] M. Andreae, R. White, J. Gabry, and C. Hall. Antiemetic medication as a marker of healthcare disparities in anesthesia: A Bayesian hierarchical model using the National Anesthesia Clinical Outcomes Registry. *Clinical and Translational Science*, 8(3):180, 2015.

[2] G. Rust, W. N. Nembhard, M. Nichols, F. Omole, P. Minor, G. Barosso, and R. Mayberry. Racial and ethnic disparities in the provision of epidural analgesia to Georgia Medicaid beneficiaries during labor and delivery. *Am J Obstet Gynecol*, 191(2):456–62, 2004.

[3] L. G. Glance, R. Wissler, C. Glantz, T. M. Osler, D. B. Mukamel, and A. W. Dick. Racial differences in the use of epidural analgesia for labor. *Anesthesiology*, 106(1):19–25; discussion 6–8, 2007.

[4] M. H. Andreae, G. M. Carter, N. Shaparin, K. Suslov, R. J. Ellis, M. A. Ware, D. I. Abrams, H. Prasad, B. Wilsey, D. Indyk, M. Johnson, and H. S. Sacks. Inhaled Cannabis for Chronic Neuropathic Pain: A Meta-analysis of Individual Patient Data. *J Pain*, Sep 2015. PMID: 26362106.

[5] M. H. Andreae and D. A. Andreae. Local anaesthetics and regional anaesthesia for preventing chronic pain after surgery. *Cochrane Database Syst Rev*, 10:CD007105, 2012. PMID: 23076930.

[6] J. Bafumi and A. Gelman. Fitting Multilevel Models When Predictors and Group Effects Correlate. SSRN Scholarly Paper ID 1010095, Social Science Research Network, Rochester, NY, September 2007.

[7] B. Efron and C. Morris. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319, 1975.

[8] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.

[9] A. Gelman. *Red state, blue state, rich state, poor state: why Americans vote the way they do*. Princeton University Press, 2009.

[10] B. Efron and C. N. Morris. *Stein's paradox in statistics*. WH Freeman, 1977.

[11] B. Efron and C. N. Morris. Stein's paradox in statistics. *Scientific American*, 263(5):119–127, 1977.

[12] D. L. Sackett, W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson. Evidence based medicine: what it is and what it isn't. *BMJ*, 312(7023):71–72, Jan 1996. PMID: 8555924.

[13] D. Ashby and A. F. Smith. Evidence-based medicine as Bayesian decision-making. *Stat Med*, 19(23):3291–3305, Dec 2000. PMID: 11113960.

[14] D. J. Cook, C. D. Mulrow, and R. B. Haynes. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*, 126(5):376–380, Mar 1997. PMID: 9054282.

[15] M. H. Andreae and D. A. Andreae. Regional anaesthesia to prevent chronic pain after surgery: a Cochrane systematic review and meta-analysis. *Br J Anaesth*, 111(5):711–720, Nov 2013. PMCID: PMC3793661.

[16] D. J. Spiegelhalter, J. P. Myles, D. R. Jones, and K. R. Abrams. Bayesian methods in health technology assessment: a review. *Health Technol Assess*, 4(38):1–130, 2000. PMID: 11134920.

[17] D. Roth, B. Heidinger, C. Havel, and H. Herkner. Different mortality time-points in critical care trials - Current practice and influence on effect estimates in meta-analyse. *Crit Care Med*, 2015 [accepted].

[18] S. G. Thompson and J. P. T. Higgins. How should meta-regression analyses be undertaken and interpreted? *Stat Med*, 21(11):1559–1573, Jun 2002. PMID: 12111920.

[19] F. Abroug, L. Ouanes-Besbes, F. Dachraoui, I. Ouanes, and L. Brochard. An updated study-level meta-analysis of randomised controlled trials on proning in ARDS and acute lung injury. *Crit Care*, 15(1):R6, 2011. PMID: 21211010.

[20] J. Deeks, J. Higgins, and D. Altman. Chapter 9–Analysing Data and Undertaking Meta-analyses: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0 [updated March 2011]. *Cochrane Handbook for Systematic Reviews of Interventions*, 5, 2011.

[21] F. Song, A. Clark, M. O. Bachmann, and J. Maas. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol*, 12:138, 2012. PMID: 22970794.

[22] J. E. Cornell, C. D. Mulrow, R. Localio, C. B. Stack, A. R. Meibohm, E. Guallar, and S. N. Goodman. Random-effects meta-analysis of inconsistent effects: a time for change. *Ann Intern Med*, 160(4):267–270, Feb 2014. PMID: 24727843.

[23] M. Andreae, M. Johnson, and H. Sacks. Bayesian responder meta-analysis of regional anesthesia to prevent chronic pain after iliac crest bone graft harvesting. *Reg. Anesth. Pain Med.*, 38(1):77–88 (A1), 2013.

[24] D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles. *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons, 2004.

[25] G. M. Carter, D. Indyk, M. Johnson, M. Andreae, K. Suslov, S. Busani, A. Esmaeili, and H. S. Sacks. Micronutrients in HIV: a Bayesian meta-analysis. *PLoS One*, 10(4):e0120113, 2015. PMID: 25830916.

[26] Stan Development Team. Stan: A C++ Library for Probability and Sampling, Version 2.5.0, 2014.

[27] Creativecommons.org. Creative Commons - Attribution-ShareAlike 4.0 International - CC BY-SA 4.0, 2015.