## BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.

Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME

Gelman, Andrew

eRA COMMONS USER NAME (credential, e.g., agency login) AGELMAN

POSITION TITLE

Professor, Department of Statistics, Columbia University

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

|  |  |  |  |
| --- | --- | --- | --- |
| INSTITUTION AND LOCATION | DEGREE  *(if applicable)* | YEAR(s) | FIELD OF STUDY |
| Massachusetts Institute of Technology | S.B. | 1985 | Mathematics |
| Massachusetts Institute of Technology | S.B. | 1986 | Physics |
| Harvard University | M.A. | 1987 | Statistics |
| Harvard University | Ph.D. | 1990 | Statistics |

## Personal Statement

I am professor of Professor of Statistics and Political Science and Director of the Applied Statistics Center at Columbia University. I will serve as a co-investigator on this project to make advanced hierarchical modeling more accessible for data scientists interested in data driven outcomes research.

My special expertise is in Bayesian Inference, especially the fast computational implementation of hierarchical models in *Stan*, a probabilistic programming language for Bayesian inference written in C++; I lead the team at Columbia University developing the probabilistic programming language *Stan* and its R environment spinoff *Rstan*. We are funded through the National Science Foundation to continue to accelerate the Hamiltonian algorithm and other specialized routines for Bayesian inference.  *Stan* is uniquely suited to implement the computationally challenging hierarchical model for real-time electronic medical records, as proposed as a use case in this project by Drs.  Goodrich and Andreae. Such intricate complicated models often proof intractable with other existing software.

This project proposes to develop additional software programs build on *Stan* and its Hamiltonian Monte Carlo algorithm and the no-U-turn sampler. With Ben Goodrich, co-principal investigator on this project, I developed *mi*, a R software package for data imputation.  I founded the multidisciplinary graduate program (Dr. Andreae and Jonah Galbry graduated from) to increase the quantitative training outside the confines of our discipline.  I teach the Ph.D. course in Bayesian statistics and have given many lectures and workshops to a diverse audience teaching on Stan and hierarchical modeling.  I am ideally suited to serve as co-investigator to advice on the *Stan* model library *rstanarm* calls on.

We have previously fitted complex Bayesian hierarchical models using our probabilistic software *Stan*, both in and outside the medical sciences, pushing the boundaries of computability in collaboration with investigators from many disciplines, including astrophysics, econometrics, ecology, political science, sociology, and health sciences. Indeed, *Stan* is motivated primarily by its practical applications.  As an example for medical sciences, we fitted models for pharmacokinetic and pharmacodynamic data in drug development in collaboration with investigators working for Novartis (Switzerland).  We have overcome computational implementation challenges in fitting hierarchical models to very large datasets before, for example, modeling the hierarchical structure of supernova measurements.

Thanks to Dr. Andreae, who introduced us a year ago, we enjoy a promising collaboration with Dr. Gong, at Montefiore and previously submitted a joint NIH application to continue the development of gradient based algorithms, using Dr. Gong’s pragmatic trial data as a use case.  Obviously, Dr. Andreae’s project is closely interlinked with our long-term joint interests. I am proud to be part of Drs. Goodrich and Andreae assembled a very strong multi-disciplinary team: strong clinical data science experts joined hand with theoretical statistics to develop much needed more accessible tools for medical outcomes research.

**B. Positions and Honors**

**Positions and Employment:**

|  |  |
| --- | --- |
| 1990- 1996 | Assistant Professor, Department of Statistics, University of California, Berkeley |
| Spring 1994 | Visiting Assistant Professor, Department of Statistics, University of Chicago |
| 1996-2000 | Associate Professor, Department of Statistics, Columbia University |
| 2000-present | Professor, Department of Statistics, Columbia University |
| 2000-present | Professor, Department of Political Science, Columbia |
| 2000-present | Founding Director, Applied Statistics Center, Columbia University |
| 2009-2010 | Alliance Visiting Professor, Sciences Po, Paris |

**Selected Recent Awards**

|  |  |
| --- | --- |
| 2012 | Open Source Software World Challenge award for Stan: an R and C++ package for Bayesian sampling. (Andrew Gelman, Bob Carpenter, Matt Hoffman, Daniel Lee, Michael Malecki, Ben Goodrich, Michael Betancourt, Marcus Brubaker, and Jiqiang Guo) |
| 2014 | Elected member, International Statistical Institute |
| 2014 | Statistician of the Year, Chicago chapter of the American Statistical Association |

**C. Contribution to Science**

1. My contribution to science is mostly centered on Bayesian Inference, where I am known in particular for early work with my mentor Dr. Rubin, the our development of general methods for monitoring convergence of iterative simulations also known as R-hat. I also wrote well received standard textbooks on Bayesian and hierarchical modeling.

* Gelman, Andrew, and Donald B. Rubin. "Inference from iterative simulation using multiple sequences." *Statistical science* (1992): 457-472.
* Brooks, Stephen P., and Andrew Gelman. "General methods for monitoring convergence of iterative simulations." *Journal of computational and graphical statistics* 7.4 (1998): 434-455.
* Rue, Håvard, Sara Martino, and Nicolas Chopin. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." Journal of the royal statistical society: Series b (statistical methodology) 71.2 (2009): 319-392.
* Gelman, Andrew, et al. *Bayesian data analysis*. Vol. 2. London: Chapman & Hall/CRC, 2014.
* Gelman, Andrew, and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.

1. Relevant to this project, frustrated with the challenges of computational implementation of what seemed appropriate statistical models, I participated in the development of computational implementation and accessible software packages, early on the *R2WinBUGS* software package for running WinBUGS from R, recently for Stata. With Dr. Goodrich, co-priniciple investigator, I developed the R package *mi* for data imputation. Most importantly, I lead the team at Columbia University developing the probabilistic programming language Stan and its R environment spinoff *Rstan*. Stan is an ultra-fast computational implementation of hierarchical models, written in C++, that this project proposes to further develop with the software packages *rstanarm* and *shinystan* software.

* Gelman, Andrew, Daniel Lee, and Jiqiang Guo. "Stan A Probabilistic Programming Language for Bayesian Inference and Optimization." *Journal of Educational and Behavioral Statistics* (2015): 1076998615606113.
* Bob Carpenter, et al. In press. Stan: A Probabilistic Programming Language. Journal of Statistical Software
* Su, Yu-Sung, et al. "Multiple imputation with diagnostics (mi) in R: Opening windows into the black box." *Journal of Statistical Software* 45.2 (2011): 1-31.
* Hoffman, Matthew D., and Andrew Gelman. "The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." *The Journal of Machine Learning Research* 15.1 (2014): 1593-1623
* Grant, Robert L., et al. "Fitting Bayesian item response models in Stata and Stan." *arXiv preprint arXiv:1601.03443* (2016).

1. Also relevant to my role in advising on the choice of priors for hierarchical models in this project, I contributed significantly to the development of this under-researched field of Bayesian inference. In particular with regards to hierarchical modeling. I am known for recommending the Cauchy distribution and for similar work on regularizing hierarchical models by using informative priors.

* Gelman, A., Hwang, J., and Vehtari, A. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24, 997-1016. 2014
* Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. S. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*. 2008
* Gelman, A. Prior distributions for hierarchical variance parameters. *Bayesian Analysis*. 2006
* Gelman, Andrew. "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)." *Bayesian analysis* 1.3 (2006): 515-534.
* Gelman, Andrew. "Prior distribution." Encyclopedia of environmetrics (2002).

4. Pertaining to my role in dissemination and teaching of advanced hierarchical modeling in this project, I would like to list some work showcasing my engagement in reaching out to diverse audience to promote adapted and suitable models for a wide range of scientific inquiries. My collaborations spanned from political science to pharmacokinetic modeling.

* Gelman, Andrew. *Red state, blue state, rich state, poor state: why Americans vote the way they do*. Princeton University Press, 2009.
* Wang, Wei, et al. "Forecasting elections with non-representative polls." *International Journal of Forecasting* (2014).
* Price, P. N., and Gelman, A. Centralized analysis of local data, with dollars and lives on the line: Lessons from the home radon experience. In Data Science for Politics, Policy and Government, ed. R. Michael Alvarez. Cambridge University Press. 2015
* Gelman, Andrew, Frederic Bois, and Jiming Jiang. "Physiological pharmacokinetic analysis using population modeling and informative prior distributions." Journal of the American Statistical Association 91.436 (1996): 1400-1412.
* Gelman, Andrew, and Deborah Nolan. Teaching statistics: A bag of tricks. Oxford University Press, 2002.

[**Selected Pubmed indexed peer-reviewed publications**](http://www.ncbi.nlm.nih.gov/sites/myncbi/michael.andreae.1/collections/45912675/public/)(out of 219 articles and 11 books)

# D. Research Support Ongoing Research Support

SES-1424962 Phillips (PI) 09/15/14-08/31/17

Using Multilevel Regression and Poststratification to Measure and Study Dynamic

The goal of this project is to develop a ’dynamic Multilevel Regression Poststratification’ that will allow researchers to generate time-varying estimates of public opinion and to improve on ’normal’ Multilevel Regression Poststratification

Role: Co-PI

DE R305D140059 Gelman (PI) 08/01/14-07/31/17

Solving Difficult Bayesian Computation Problems in Education Research Using STAN

The project aims to provide a general, user-friendly, and efficient tool for statistical inference in education research.

Role: PI

CNS-1205516 Gelman (PI) 06/01/12-05/31/15

CI-ADDO-NEW: Stan, Scalable Software for Bayesian Modeling

The project aims to design, code, document, test, disseminate, and maintain Stan, which is an extensible open-source software framework and compiler for efficient and scalable Bayesian statistical modeling. Stan is designed to facilitate scientist and engineers with Bayesian inference on measured or simulated data.

Role: PI

SES-1023176 Zheng (PI) 09/01/10-08/31/15

Methods for social network analysis

This project uses *aggregated relational data* to propose statistical innovations that address questions about network characteristics while making data more accessible to a wide range of social scientists. Role: Co-PI

R01 HS21734 Ma (PI) 9/30/2013 – 9/29/2015

Effects of Missing Data Strageties on Disparities Research Results in HCUP SID

The major goal of this project is to make SID a more useful resource for the study of racial disparity in TKR and other areas of medicine. The project aims to generate multiply imputed datasets that account for missing information including race in the SID, and to use the datasets to measure racial disparity in TKR.

Role: Co-PI

Department of Education IES (Hill) 9/1/2012 – 8/31/2017 Post-Doctoral Training Program

The major goals of this project are to train postdoctoral scholars in statistics and education research. Role: Co-PI

RHF CU11-1145 Garfinkle (PI) 1/1/2011 – 4/30/2015

Measuring Poverty and Disadvantage in New York City

The goal of this project is to create one or more measures of poverty and disadvantage for New York City that provides a better guide to grant making than does the federal and city official measures of poverty.

Role: Co-PI

**Completed Research Support**

SES-1023189 Phillips (PI) 10/01/10-09/30/14

Multilevel regression and poststratification for understanding public opinion and policy Role: Co-PI2R01HD036916-11A1 McLanahan (PI) 07/04/06-06/30/14

Fragile Families and Child Well-Being Role: Co-investigator

ATM-0934516 Gelman (PI) 09/15/09-08/31/13

CMG Research: Reconstructing Climate from Tree Ring Data Role: PI

DE-SC0002099 Gelman (PI) 09/01/09-08/31/13

Petascale Hierarchical Modeling Via Parallel Execution Role: PI

DE R305D100017 Gelman (PI) 03/01/10-07/31/13

Practical Tools for Multilevel Hierarchical Modeling in Education Research Role: PI

R305D090006 Gelman (PI) 03/01/09-05/31/13

Practical Solutions for Missing Data and Imputation Role: PI

H98230-10-1-0184 Gelman (PI) 01/26/10-01/25/12

National Security Agency Weakly Informative Priors Role: PI

CUMC 09050669 Gelman (PI) 07/01/09-06/30/11

Comprehensive Prognostic Modeling for Esophageal Cancer: A Bayesian Approach Role: PI

5P30MH071430-04 Caton (PI) 09/30/05-06/30/10

National Institute of Mental Health

The Columbia Center for Homelessness Prevention Studies Role: Investigator-Methods Core

0532231 Gelman (PI) 08/15/05-7/31/09

Design and analysis of 'How many X's do you know' surveys for the study of polarization in social networks

Role: Lead PI

5R01GM074806-03 Gelman (PI) 07/01/06-06/30/09

Bayesian analysis of serial dilution assays Role: Lead PI

1R21DA021122-01A2 Aidala (PI) 05/21/07-04/30/09

National Institute on Drug Abuse

Drug Abuse, Mental Illness, Homelessness and HIV: Evaluating Models of Care Role: Consultant