BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Andrea Sarah Foulkes

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

POSITION TITLE: Professor of Statistics

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Brown University, Providence, RI	BA	05/1994	Mathematics
Harvard University, Boston, MA	ScD	05/2000	Biostatistics
Harvard University, Boston, MA	Postdoctoral Fellow	2000–2002	Biostatistics

A. Personal Statement

I am the Director of the Biostatistics Center at Massachusetts General Hospital and Professor of Medicine at Harvard Medical School. My research program in statistical methods for precision medicine involves developing, applying and evaluating analytical methods for high-dimensional molecular and cellular level data with applications in inflammation. These methods include model-based and machine learning approaches for genetic and genomic association studies as well as strategies for monitoring repeatedly measured biomarkers. As PI of an active R01 award on Methods for integrated analysis of multi-level omics data, I continue leading the effort to develop and evaluate principled statistical methods for interrogating the mechanistic underpinnings of genetic and genomic associations with complex disease traits.

B. Positions and Honors

Employment

2000–2002 Research Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA
2002–2004 Assistant Professor, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA
2004–2008 Assistant Professor, Division of Biostatistics, University of Massachusetts, School of Public Health and Health Sciences, Amherst, MA
2008–2014 Associate Professor, Division of Biostatistics, University of Massachusetts, School of Public Health and Health Sciences, Amherst, MA
2008–2014 Head, Program in Biostatistics, University of Massachusetts, School of Public Health and Health Sciences, Amherst, MA
2012–2014 Director, Institute for Computational Biology, Biostatistics and Bioinformatics, University of Massachusetts, School of Public Health and Health Sciences, Amherst, MA
2014–2015 Professor (on-leave), Division of Biostatistics, University of Massachusetts, School of Public Health and Health Sciences, Amherst, MA
2014–2019 Professor, Department of Mathematics and Statistics, Mount Holyoke College, S. Hadley, MA
2019–present Director, Biostatistics Center, Massachusetts General Hospital Research Institute, Boston, MA
2020–present Professor, Department of Medicine, Harvard Medical School, Boston, MA
2020–present Professor in Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA

Honors and awards

1991, 1993	President's Community Service Fellowship Recipient, Howard Swearer Center for Public Service, Brown University.
1992	Teaching and Research Assistantship, Department of Chemistry, Brown University.
1993–1994	Resource Scholar, Brown University.
1997–1999	Teaching Fellow, Department of Biostatistics, Harvard University.
1995–2002	Training Grant Recipient, National Institutes of Health
2003	Thomas B. and Jeannette E. Laws McCabe Fund Fellow Award Recipient

Other positions

2008–2019	Adjunct Assistant Professor, Program in Molecular Medicine, University of Massachusetts School of Medicine, Worcester, MA.
2010–2018	Member, Scientific Advisory Board, NIH/NIDDK CKD Biomarkers.
2012–2019	Adjunct Associate Professor, Department of Computer Science, UMass, Amherst, MA.

C. Contributions to Science

1. **Methods for genetic association studies.** Advancing our knowledge of the molecular and physiological underpinnings of complex diseases will deepen our insight into disease etiology, while providing opportunity to further develop novel and targeted interventions and lessen disease morbidity and mortality. My lab has contributed to the advancement of statistical methods and their applications to further elucidate complex associations, including: (a) a generalized linear mixed effects modeling (GLMM) framework for the analysis of multi-locus genotype-trait associations; (b) modeling extensions to address the analytic challenge arising from ambiguous (unobservable) haplotypic phase in population-based studies of unrelated individuals; (c) machine learning methods for population-based genetic association studies and studies of immune modulation; and (d) theoretically derived genetic class-level testing frameworks that leverage publicly available meta-analysis summary level data.
   a) **Foulkes, AS**, Reilly, M., Zhou, L., Wolfe, M. and Rader, D.J. (2005) Mixed modeling to characterize genotype-phenotype associations, Statistics in Medicine, 24:775-789.
   b) **Foulkes, AS**, Yucel, R. and Li, X. (2008) A likelihood-based approach to mixed modeling with ambiguity in cluster identifiers, Biostatistics, 9(4):635-657, PMCID: PMC2536727.
   c) Au, K., Lin, R. and **Foulkes, AS** (2011) Mixture modeling as an exploratory framework for genotype-trait associations, JRSS, Applied Statistics, 60(3):355-375, PMCID: PMC3285383.
   d) Qian, J., Nunez, S. Reed, E., Reilly, M.P. and **Foulkes, AS** (2016) A simple test of class-level genetic association can reveal novel cardiometabolic trait loci, PLoS One, DOI: 10.1371.journal.pone.0148218 PMCID: PMC4747495.

2. **Methods for post-analytic interrogation and integrated analysis of -omics data.** Research in the post-GWAS era focuses on interrogation of the relative contributions of genomic elements and the mechanistic underpinnings of disease associations. My contributions in this area include developing statistical methods for discerning drivers in observed gene-trait associations and integrating -omics data derived from independent cohorts to identify potential mechanistic pathways of associations.
   a) Chen, S., Nunez, S., Reilly, M.P. and **Foulkes, AS** (2016) Bayesian variable selection for post-analytic interrogation of susceptibility loci, Biometrics, 73(2): 603-614.
   b) Zhang H, Xue C, Wang Y, Shi J, Zhang X, Li W, Nunez S, **Foulkes A**S, Lin J, Hinkle CC, Yang W, Morrisey EE, Rader DJ, Li M, Reilly MP. (2017) Deep RNA Sequencing Uncovers a Repertoire of Human Macrophage Long Intergenic Noncoding RNAs Modulated by Macrophage Activation and Associated with Cardiometabolic Diseases, J Am Heart Assoc. 6(11) PMCID: PMC5721798.
   c) Qian, J., Ray, E. Brecha RL, Reilly, MP and **Foulkes, AS** (2019) A likelihood-based approach to transcriptome association analysis, Stat Med 38(8) 1357-1373. PMCID: PMC6907891.
   d) Ray, EL, Qian, J, Brecha, R, Reilly, MP and **Foulkes, AS** (2019) Stochastic imputation for integrated transcriptome association analysis of a longitudinally measured trait, Stat Methods Med Res [Epub ahead of print].

3. **Methods for monitoring repeated biomarkers.** Characterizing the genomic contributors to biomarker trajectories will further enhance understanding of the regulation of dynamic inflammatory processes in acute and chronic complex disease. My contributions in this field include methodological advancements for (a) modeling serial genetics data; (b) genetic association studies of the human pathogenic Plasmodium species and HIV in multiply infected individuals; and (c) prediction based classification (PBC) for prioritizing expensive laboratory testing using cheaper markers for monitoring HIV-infected individuals on ART in resource limited settings. As one example, our findings from application of PBC underscore the clinical usefulness of less expensive markers for prioritizing CD4 testing in resource-limited settings, and the potential impact of applying sophisticated and theoretically rigorous statistical modeling frameworks.
   a) Zhao, J., **Foulkes, AS** and George, E. (2005) Exploratory Bayesian model selection for serial genetics data, Biometrics, 61(2): 591-599.
   b) Azzoni, L*. **Foulkes, AS***, Liu, Y., Li, X., Johnson, M. Smith, C., Kamarulzaman, A., Montaner, J., Mounzer, K., Saag, M., Cahn, P., Cesar, C., Krolewiecki, A., Sanne, I., Montaner, L.J. (2012) Prioritizing CD4 count monitoring in response to ART in resource-constrained settings: a retrospective application of prediction-based classification, PLoS Med 9(4): e1001207. doi:10.1371/journal.pmed.1001207, PMCID: PMC3328436. *Indicates joint first authorship
   c) Qian, J., Nunez, S., Kim, S., Reilly, M.P. and **Foulkes, AS** (2017) A score test for genetic class-level association with non-linear biomarker trajectories, Stat Med, 36(19): 3075-3091, PMCID: PMC6002775.
   d) Abdulhaqq SA, Martinez M, Kang G, Rodriguez IV, Nichols SM, Beaumont 4, Joseph J, Azzoni L, Yin X, Wise M, Weiner D, Liu Q, **Foulkes A**, Münch J, Kirchhoff F, Coutifaris C, Tomaras GD, Sariol C, Marx PA, Li Q, Kraiselburd EN, Montaner LJ. (2019) Repeated semen exposure decreases cervicovaginal SIVmac251 infection in rhesus macaques, Nat Commun, 10(1): 3753. PMCID: PMC6704120.

4. **Open-source and publicly available instructional resources.** The public health impact of analytic developments resulting from our research, as well as the work of others in the field of statistical genomics, will be advanced exponentially through the dissemination of carefully formulated and openly available instructional resources. I have devoted substantial effort to developing and disseminating instruction materials and open-source tools, including: a) a textbook on Applied Statistical Genetics with R, with 40.2 thousand chapter downloads reported by Springer; and b) a fully dynamic, reproducible and extensible tutorial for GWA analysis and post-analytic interrogation and visualization. This tutorial was the top-most downloaded article for Statistics in Medicine in both 2017 and 2018 and is the second most downloaded of all time with a reported 35.8 thousand downloads since publication.
   a) **Foulkes, AS**, Applied Statistical Genetics with R (for Population-based Association Studies) (2009) Springer, Use R, 252p, ISBN: 978-0-387-89553-6.
   b) **Foulkes, AS** and Au, K., R Statistical Tools for Gene Discovery (Chapter 5) in In Silico Tools in Gene Discovery (2011) Springer, Methods in Molecular Biology.
   c) Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M.P. and **Foulkes, AS** (2015) A guide to genome-wide association analysis and post-analytic interrogation. Stat Med, 34(28): 3769–3792/ PMCID: PMC5019244.

A full list of published work can be found at:
https://www.ncbi.nlm.nih.gov/myncbi/andrea.foulkes.1/bibliography/public/

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

R01-GM127862 Foulkes 04/01/2018-03/31/2022

NIH

Methods for integrated analysis of multi-level omics data

To advance and evaluate statistical methods for rigorous interrogation of the regulatory and gene-based dynamic and evoked inflammatory responses to stimulus and evaluation of the relative, integrated contributions of multiple regulatory and gene-level elements.

Role: Principal Investigator

5 UL1T R002541-02 Nadler 05/01/2018-04/30/2023

NIH

Harvard Clinical and Translational Science Center

Provide enriched resources to educate and develop the next generation of researchers trained in the complexities of translating research discoveries into clinical trials and ultimately into practice. Design new and improved clinical research informatics tools for analyzing research data and managing clinical trials. Support outreach to underserved populations, local community and advocacy organizations, and health care providers. Assemble interdisciplinary teams and forge new partnerships with private and public health care organizations.
Role: Statistician

5 P30 CA006516-54 Benz 12/22/2016-11/30/2021

NIH-NCI

Cancer Center Support Grant (Biostatistics CORE)

The major goals of this project are to support CORE activities that represent shared resources used by a variety of investigators in different labs throughout the DF/HCC. This is a subcontract through the Dana-Farber Cancer Institute.
Role: Statistician

DP 2HD 101400 Edlow 09/12/2019-06/30/2024

NIH

A Connectome-Based Clinical Trial Platform to Promote Early Recovery of Consciousness after Traumatic Coma

The goal of this project is to develop and implement a clinical trial platform for personalized, targeted therapy aimed at restoring consciousness in patients admitted to the intensive care unit for traumatic coma.
Role: Statistician