Prasad Patil

801 Massachusetts Avenue, Room 320
Boston, MA 02119
patil@bu.edu
prasadpatil.net

EDUCATION

Ph.D. in Biostatistics
Johns Hopkins University, Baltimore, MD, 2011-2016
Advisor: Jeff Leek
Dissertation Title: "Assessing Reproducibility and Value in Genomic Signatures"

B.A. in Mathematics
New York University, New York, NY, 2005-2008
Concentration: Applied Mathematics (Minor in Computer Science)

POSITIONS

Assistant Professor 2019-present
Department of Biostatistics, Boston University

Postdoctoral Fellow 2016-2019
Department of Biostatistics, Harvard University
Department of Data Sciences, Dana-Farber Cancer Institute

Scientific Programmer/Bioinformatics Fellow 2008-2011
Department of Biomedical Informatics, Harvard University

PUBLICATIONS

Student advisees underlined

Zhang Y, **Patil P**, Johnson WE, Parmigiani G (2020). Robustifying genomic classifiers to batch effects via ensemble learning. In press, *Bioinformatics.*

Westerman K, Fernández-Sanlés A, **Patil P**, Sebastiani P, Jacques P, Starr JM, J. Deary I, Liu Q, Liu S, Elosua R, DeMeo DL (2020). Epigenomic Assessment of Cardiovascular Disease Risk and Interactions With Traditional Risk Metrics. *Journal of the American Heart Association.* 9(8):e015299.

Nudel J, Bishara AM, de Geus SW, **Patil P**, Srinivasan J, Hess DT, Woodson J (2020). Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. *Surgical Endoscopy.* 17:1-0.

Ramchandran M, **Patil P**, Parmigiani G (2019). Tree-weighting approaches in constructing cross-study learners. *Pac Symp Biocomput. 2020*; 25:451-462.

Patil P, Peng RD, Leek JT (2019). A visual tool for defining reproducibility and replicability. *Nature Human Behavior* 3, 650–652.

Patil P, Parmigiani G (2018). Training replicable predictors in multiple studies. *Proceedings of the National Academy of Sciences*, 115(11), 2578-2583.

Patil P, Peng RD, Leek, JT (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science*, 11(4), 539-544.

Patil P, Colantuoni E, Rosenblum MA, Leek JT (2016). Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments. *Contemporary Clinical Trials Communications* 3: 48-54.

Patil P, Leek JT (2015). Discussion of "Visualizing statistical models: Removing the blindfold". *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8(4), 240-241.

Patil P, Bachant-Winner PO, Haibe-Kains B, Leek JT (2015). Test set bias affects reproducibility of gene signatures. *Bioinformatics*, btv157.

Hyland PL, Burke LS, Pfeiffer RM, Rotunno M, Sun D, **Patil P**, Wu X, Tucker MA, Goldstein AM, Yang XR (2014). Constitutional promoter methylation and risk of familial melanoma. *Epigenetics*, 9(5), 685-692.

Fusaro VA, **Patil P**, Chi CL, Contant CF, Tonellato PJ (2013). A Systems Approach to Designing Effective Clinical Trials Using Simulations. *Circulation*, 127(4), 517-526.

Fusaro VA, **Patil P**, Gafni E, Wall DP, Tonellato PJ (2011). Biomedical Cloud Computing Using Amazon Web Services. *PLoS Computational Biology.* 7(8):e1002147.

Wall DP, Kudtarkar P, Fusaro VA, Pivovarov R, **Patil P**, Tonellato PJ (2010). Cloud computing for comparative genomics. *BMC Bioinformatics* 11:259.

In Preparation

Guan Z, Parmigiani G, **Patil P** (2019+). Merging versus Ensembling in Multi-Study Machine Learning: Theoretical Insight from Random Effects. *arXiv*:1905.07382. (under revision, *Biometrika*)

Loewinger GC, **Patil P**, Kishida KK, Parmigiani G (2019+). Covariate-Profile Similarity Weighting and Bagging Studies with the Study Strap: Multi-Study Learning for Human Neurochemical Sensing. *bioRxiv*, 856385. (under revision, *Annals of Applied Statistics*)

Deng Z, Ding F, Dwork C, Hong R, Parmigiani G, **Patil P**, Sur P (2020). Representation via representations: Domain generalization via adversarially learned invariant representations. *arXiv*:2006.11478. (rebuttal period, *AISTATS '21*)

Ren B, **Patil P**, Dominici F, Parmigiani G, Trippa L. (2020). Cross-study learning for generalist and specialist predictions. *arXiv*:2007.12807. (submitted, *Annals of Statistics*)

Li T, Zhang Y, **Patil P**, Johnson WE. Overcoming the impacts of two-step batch effect correction on gene expression estimation and inference using ComBat3S. (submitted, *Biostatistics*)

Wu Y, Ren B, Parmigiani G, **Patil P** (2020). Using Knowledge Transfer to Retain Features in Multi-Study Learning.

Wang X, Hochburg NS, Salgame P, Ellner J, **Patil P***, Johnson WE* (2020). Using gene set enrichment scoring to accurately evaluate genomic tuberculosis biomarker performance across heterogeneous patient cohorts.

Wang X, Harper C, Johnson WE*, **Patil P*** (2020). CuratedTBData: 45 harmonized and clinically annotated tuberculosis gene expression patient cohorts.

* - co-senior authors

GRANTS

Active

NIH 1R21AI154387-01, *Signature profiling and staging the progression of TB from infection to disease.* (co-investigator, 2020-2022)

FAA ASCENT 13-C-AJFE-BU, *Community Measurements of Aviation Emissions Contribution to Ambient Air Quality.* (co-investigator, 2020-2022)

NIH 1R01GM122876-01, *Methods to Estimate the Effect of Interventions on the Incidence and Transmission of Tuberculosis.* (co-investigator, 2019-2022)

NIH 1DP2DA051864-01, *Development of a novel community-based high-performance surveillance network for drug use.* (co-investigator, 2020-2024)

Pending

BU IGNITE CTSI, *DiffeRentIal alcohol-related caNcer risK and incidence eXplained by HIV status (DRINKxHIV) study.* (under review, co-PI)

NIH R01, *Deep Learning Methods for Multi-Class Classification of Tuberculosis Progression.* (principal investigator)

NIH R01, *Addressing an inherent bias in neuroprognostication: A collaboration between the US and Brazil to reduce the impact of self-fulfilling prophecy in cardiac arrest (SPARE).* (co-investigator)

NIH R21, *Investigating HIV, Endemic Infections, Hypertension and Atherosclerosis Risk Together In Ghana (I HEART GHANA).* (co-investigator)

Completed

NSF DMS Award #1810829, *Statistical Methods for Multi-Study Predictions.* (funded post-doc activities)

HDSI Competitive Research Fund, *Representation via Representation.* (funded post-doc activities)

NCI 2T32CA009337-3, *Training Grant in Quantitative Sciences for Cancer Research* (post-doc)

NIH R01GM105705, *Statistical models for biological and technical variation in RNA sequencing* (pre-doc)

NCI U54CA151838, *Analysis Tools and Software for Second-Generation Sequencing Data* (pre-doc)

AWARDS

Jane and Steve Dykacz Award Departmental award for best student paper in medical statistics Awarded for “Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments”	2016
Helen Abbey Award Departmental award for teaching, awared to one student per year	2016
JHSPH Student Assembly Teaching Assistant Recognition Award One of two voted on by students across all courses in JHSPH	2015
NYU Honors Scholar	2008
NYU Dean’s List	2006-2007
National Merit Scholarship	2005-2008

MENTORSHIP

BU

Nathaniel Alemayehu, Undergrad Statistics (summer intern); machine learning methods for T-cell clustering in HIV patients

Sarah Szvetecz, MS Applied Biostatistics; machine learning methods for T-cell clustering in HIV patients

Ziyi Wang, MS Applied Biostatistics; multi-study feature selection.

Xutao Wang, PhD Biostatistics; curation of tuberculosis genomic data and biomarker evaluation

Harvard (co-mentored with Giovanni Parmigiani)

Cathy Wang, PhD candidate, Biostatistics; multi-study gradient boosting

Zoe Guan, PhD candidate, Biostatistics; analytic guidelines for multi-study learning.

Maya Ramchandran, PhD student, Biostatistics; cross-study replicability weights for individual trees within random forests.

Gabe Loewinger, PhD student, Biostatistics; bootstrapped study sampling.

Jinpeng Fu, Master's co-op, Bioinformatics, Northeastern University; transferring genomic signatures from microarray to RNA-seq.

Xutao Wang, Master's student, Computational Biology and Quantitative Genetics; multi-study simulation strategies and ensemble weighting comparisons.

Yujie Wu, Master's student, Biostatistics; knowledge transfer for multi-study feature imputation

Benjamin Glass, Undergraduate, Harvard University; Senior thesis: Accounting for Differing Feature Sets when Training Predictors in Multiple Studies.

Thesis committee membership: Maya Ramchandran, Lucas Schiffer, Zach Baucom, Cathy Wang, Ben Sweigart

PRESENTATIONS AND POSTERS

Invited

Patil P. Improving Predictor Generalizability Using Multiple Studies with Differing Feature Sets. ICSA 2020.

Patil P. Using multiple studies to improve the generalizability of genomic signatures for cancer risk prediction. Computational Biomedicine Seminar Series, Boston University, 2020.

Patil P. Current topics in multi-study learning. ENAR 2019.

Patil P. Defining and evaluating reproducibility and replicability. WNAR 2017.

Patil P. A statistical framework for discussing reproducibility and replicability. ENAR 2017.

Patil P. Setting expectations for replication in science. Harvard Catalyst Biostatistics Symposium 2017.

Patil P. Setting expectations for replication in science. AAAS 2017.

Contributed

Sheng H, Dam K, Liang S, Duskin J, Suchdev K, Spears W, **Patil P**, Greer D. Different Target Temperatures after Cardiac Arrest: Neurologic Recovery and the Potential Impact of Sedatives and Paralytics. AAN 2021 [poster].

Wu Y, Ren B, Parmigiani G, **Patil P**. Improving Predictor Generalizability Using Multiple Studies with Differing Feature Sets. IBC 2020 [poster].

Patil P, Parmigiani G. Training replicable predictors in multiple studies. ENAR 2018.

Patil P. Replicating scientific studies. Harvard Biostatistics Cancer Training Grant Working Group 2017.

Patil P, Alquicira J, Leek JT. Measuring the value of GWAS results in a clinical trial setting. ASHG 2015 [poster].

Patil P. What to expect when you're replicating. Hopkins Biostatistics Journal Club 2015.

Patil P, Leek JT. Reproducibility and value of genomic signatures. JSM 2015.

Patil P. On organization. Hopkins Biostatistics Computing Club 2015].

Patil P, Leek JT. Assessing the reproducibility and value of genomic signatures. ENAR 2015.

Patil P, Haibe-Kains B, Leek JT. Cross-platform gene signature development using Top-Scoring Pairs. JSM 2014.

Patil P. Cross-validation in the presence of many features. Hopkins Biostatistics Journal Club 2014.

Patil P, Haibe-Kains B, Leek JT. Cross-platform gene signature development using Top-Scoring Pairs. ENAR 2014.

Chi CL, Fusaro VA, **Patil P**, Crawford MA, Contant CF, Tonellato PJ. An approach to optimal individualized warfarin treatment through clinical trial simulations. Proceedings of IEEE Cairo International Biomedical Engineering Conference (CIBEC) 2010, Cairo, Egypt.

Chi, CL, **Patil P**, Fusaro VA, Kos PJ, Pivovarov R, Contant CF, Tonellato PJ. A Simulation Platform to Examine Heterogeneity Influence on Treatment. Proceedings of the 2010 American Medical Informatics Association Annual Symposium, Washington, DC.

Chi CL, Kos PJ, Fusaro VA, Pivovarov R, **Patil P**, Tonellato PJ. Mining personalized medicine algorithms with surrogate algorithm tags. Proceedings of the First ACM International Health Informatics Symposium 2010 [poster].

Patil P, Heus H, Arnaout R, Tonellato PJ. Refining a method for processing an individual's whole genome to clinical utility. CSHL Personal Genomes Meeting 2010. Cold Spring Harbor, N.Y.

Patil P, Tonellato PJ. Individual Whole Genome Mapping: from NGS reads to clinical variants, American Medical Informatics Association Clinical Research Informatics Summit 2010, San Francisco, CA [poster].

Fusaro VA, Kos PJ, Tector M, Tector A, **Patil P**, Tonellato PJ. Electronic Medical Record Analysis Using Cloud Computing, American Medical Informatics Association Clinical Research Informatics Summit 2010, San Francisco, CA [poster].

Patil P. Clinical algorithms for whole genome data" Partners Health Care Information Systems Research Council Symposium 2009, Harvard Medical School, Boston, MA.

Patil P. Clinical annotation of an individual whole genome assembly, Center for Biomedical Informatics Research Day 2009, Harvard Medical School, Boston, MA.

TEACHING

at Boston University

BS845: Applied Statistical Modeling and Programming in R 2020-Present
Instructor; successfully petitioned for course update to
Data Science and Statistical Modeling in R for 2021.

at Johns Hopkins

BIO622: Statistical Methods in Public Health II 2015
Guest lecturer; conducted daily lecture for 500+ student course
BIO621-623: Statistical Methods in Public Health I-III 2014-2015
Lead TA; prepared and held 2-3 sections per week, 40-60 students each
Beta-tested, proctored, and graded exams
BIO611-612, 615, 621-624, AS.280.35 2012-2014
TA; graded homework and exams, held small sections and office hours

SERVICE

Refereeing: The American Statistician, Biometrics, Biostatistics, Bioinformatics, Canadian Journal of Statistics, Computational Statistics and Data Analysis, Genome Biology, IEEE Sensors, Nature Methods, Nature Neuroscience, PLOS ONE, Psychological Methods, JRSI

Book review: Springer “Use R” series

Organization

Organizer and Chair, JSM 2021 Invited Session *“Replicability and the narrative of scientific research”*	2021
BU Biostatistics Graduate Program Admission Committee	2020-2023
BU Biostatistics Faculty Search Committee	2019
Chair, BU Biostatistics Student Paper Competition	2019-Present
Co-organizer, BU Biostatistics Seminar Series	2019-Present
Organizer, Harvard Biostatistics Cancer Training Grant Working Group	2017
Session chair, ENAR, *Next Generation Sequencing*	2014
Co-organizer, Johns Hopkins Biostatistics Computing Club	2012

SOFTWARE

R Packages:
`scifigure` (https://cran.r-project.org/package=scifigure)
Visualize Reproducibility and Replicability in a Comparison of Scientific Studies
`healthvis` (https://github.com/prpatil/healthvis)
Interactive health visualizations. Built using `d3`, `shiny`, `htmlWidgets`.
`tdsm` (https://github.com/prpatil/tdsm)
Templated Deterministic Statistical Machines. Automated analysis templates and standardized reports that can be edited and compared.

Languages:
R, Javascript, C/C++, Java, Perl, MATLAB, Stata, Hadoop, Shell scripting