OMB No. 0925-0001 and 0925-0002 (Rev. 10/15 Approved Through 10/31/2018)

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: Sheila Gaynor

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

POSITION TITLE: Doctoral Student

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)* | Start Date  MM/YYYY | Completion Date  MM/YYYY | FIELD OF STUDY |
| --- | --- | --- | --- | --- |
| University of North Carolina  Chapel Hill, NC | BPSH, BA | 08/2009 | 05/2013 | Biostatistics, Mathematics |
| Harvard University  Boston, MA | AM | 08/2013 | 05/2015 | Biostatistics |
| Harvard University  Boston, MA | PhD | 08/2013 | 05/2018 | Biostatistics |

**A. Personal Statement**

My career objective is to obtain an academic position where I can identify important scientific problems in the genetics and genomics arenas and develop novel statistical methods to solve them. My extensive biostatistical experience, beginning in my undergraduate studies, demonstrates my continued dedication and potential to achieve this goal.

I entered my undergraduate career with a keen interest in applied mathematics and statistics. Following my first year of courses, I attended the Summer Institute for Training in Biostatistics (SIBS) at Washington University in St. Louis School of Medicine to learn about public health applications. It was during this experience where I was first introduced to biomedical research, professional development, and statistical computing. To further the skills I developed in the SIBS program, I interned with Dr. Jay Piccirillo in the Clinical Research Outcomes group at the University in St. Louis studying different prognostic models for various cancer types during the summer and into the next term. I was also invited back to SIBS the following summer as the head teaching assistant of the program and to further research with Dr. Piccirillo’s group. I worked closely with his senior research associate and biostatistician, Dr. Dorina Kallogjeri, to compare data across institutions on chart-based and claims-based comorbidity coding methods. I practiced extensive data cleaning, coding, and applications of statistical methods for electronic health records.

In my junior year, I was accepted into the undergraduate major of biostatistics and immediately began research in biostatistics. I worked with Dr. Eric Bair and Dr. Bill Maixner on statistical methodology and projects applied to the Orofacial Pain Prospective Evaluation and Risk Assessment Study. I developed a sparse clustering method that identifies clusters of observations with shared features within the data that may be of interest but obscured by nonrelevant, high variance clusters. I also worked on applied projects, such as identifying clusters with similar features that can be intervened upon for treatment in temporomandibular disorder and characterizing pain location and intensity among women who use hormonal contraceptives.

Building on the experience in biostatistical research I obtained in my undergraduate education, I am now developing methods for integrative analysis of multiple genomic data types, such as genotypes and gene expression, as detailed in my research training plan. I aim to model these in a way that represents the underlying biology and captures all information in the data, including measures of error. These integrative models allow for better understanding of the biological mechanisms of disease and relationships between different biological features.

I have also sought out various projects in graduate school such as my visiting research at Boston University while completing coursework to improve both my biostatistical and professional skills. Additionally, I have worked as a teaching assistant and tutor in a variety of biostatistics courses. The diversity in these collaborations illustrate my effectiveness in communicating statistics and science.

In addition to my academic work, I have sought in my personal life to pursue service to others, as is emphasized in the field of public health. I completed over 300 hours of volunteer community service in my undergraduate career and have continued volunteering by serving in the emergency food pantry in my neighborhood. These experiences have further motivated my desire to work for the betterment of the public, and I believe my skills and interests are best suited for doing so through biostatistical research in public health.

**B. Positions and Honors**

| ACTIVITY/  OCCUPATION | Start Date  MM/YYYY | Completion Date  MM/YYYY | FIELD | INSTITUTION/  COMPANY | SUPERVISOR/  EMPLOYER |
| --- | --- | --- | --- | --- | --- |
| Predoctoral Student | 08/2013 | Present | Biostatistics | Harvard University | Xihong Lin, PhD  John Quackenbush, PhD |
| Rotation Student | 01/2014 | 08/2014 | Biomedical Informatics | Harvard University | Peter Park, PhD |
| Teaching Assistant | 08/2014 | 12/2014 | Biostatistics | Harvard University | Kimberlee Gauvreau, PhD |
| Visiting Researcher | 05/2015 | Present | Biostatistics | Boston University | Belinda Borelli, PhD |
| Head Teaching Assistant | 08/2015 | 12/2015 | Biostatistics | Harvard University | Kimberlee Gauvreau, PhD |
| Tutor | 06/2015 | Present | Biostatistics | Harvard University | Ying Wang, EdM |

**Academic and Professional Honors**

Student Coordinator of Harvard Biostatistics & Biomedical Informatics Big Data Seminar, 2016-2017

Student Member of Harvard Biostatistics Student Advising Committee, 2015-2016

Coordinator, Harvard Biostatistics Big Data Seminar, 2015-2016

Chair, ENAR Session on Graphical Modeling, 2015

Student Member of Harvard Biostatistics Resource Committee, 2014-2016

Student Judge, Harvard School of Public Health Poster Day, 2014

National Science Foundation Graduate Research Fellow, 2013

NIH Pre-doctoral Biostatistics AIDS Training Grant Fellow, 2013

Highest Honors in Biostatistics, University of North Carolina, 2013

Highest Distinction in Biostatistics and Mathematics, University of North Carolina, 2013

Delta Omega Public Health Honors Society Undergraduate Award, 2013

Carolina Research Scholar, University of North Carolina, 2013

Buckley Public Service Scholar, University of North Carolina, 2013

Phi Beta Kappa Honors Society, 2012

**Memberships in Professional Societies**

American Statistical Association

International Biometrics Society, Eastern North American Region (ENAR)

**C. Contributions to Science**

**C.1 Clustering methods with applications to pain conditions**

I engaged in undergraduate research during my time at the University of North Carolina and have continued to contribute to such projects. Advised by Dr. Eric Bair in the Departments of Endodontics and Biostatistics, I considered clustering methods under supervised and unsupervised settings. We worked in collaboration with the Orofacial Pain Prospective Evaluation and Risk Assessment Study to identify novel subtypes of temporomandibular disorder, an orofacial pain condition. We developed statistical methodology to identify biologically relevant disease subtypes that may be obscured by other features and evaluated the identification of subtypes by other clustering methods.

[1] **Gaynor, S.** and Bair, E. Identification of biologically relevant subtypes via preweighted sparse clustering. Oral presentation, ENAR Spring Meeting, Baltimore, MD (2014).

[2] Bair, E., **Gaynor, S.**, Slade, G., Ohrbach, R., Fillingim, R., Greenspan, J., Dubner, R., Smith, S., Diatchenko, L., and Maixner, W. Preweighted sparse clustering with applications to temporomandibular disorder. Oral presentation, International Association for Dental Research Epidemiology Forum, Boston, MA (2015).

[3] Bair, E., **Gaynor, S.**, Slade, G., Ohrbach, R., Fillingim, R., Greenspan, J., Dubner, R., Smith, S., Diatchenko, L., and Maixner, W, Identification of Clusters of Individuals Relevant to Temporomandibular Disorders and Other Chronic Pain Conditions: The OPPERA Study, PAIN. 157 (6) (2016) 1266-78.

[4] **Gaynor, S.** and Bair, E. Identification of relevant subtypes via preweighted sparse clustering. *Under review.*

**C.2 Mediation methods for integrative analysis**

My doctoral research focuses on the development of mediation methods for integrative genomic analysis, one approach of which is mediation. This framework allows one to use a scientifically sound structure to model multiple types of genomic features in a single analysis, and identify the important biological paths that lead to a particular outcome. I am currently developing mediation methodology for the setting where one has a common, binary outcome and continue to study multiple mediation settings across lung and psychiatric diseases.

[1] **Gaynor, S.** and Lin, X. Mediation-based integrative genomic analysis. Oral presentation, ENAR Spring Meeting, Miami, FL (2015).

[2] **Gaynor, S.** and Lin, X. A mediation-based integrative genomic analysis of lung cancer. Poster presentation, Harvard Graduate Women in Science and Engineering Symposium, Boston, MA (2015).

[3] **Gaynor, S.** and Lin, X. Mediation-based integrative genomic analysis. Oral Presentation, Joint Statistical Meetings, Seattle, WA, (2015).

[4] **Gaynor, S.** and Lin, X. Integrating epigenetic and genomic analyses via mediation analysis. Poster presentation, Harvard Medical School Epigenetics Symposium, Boston, MA (2015).

[5] **Gaynor, S.** and Lin, X. Mediation methods for case-control settings with applications to genomics. Oral presentation, ENAR Spring Meeting, Austin, TX (2016).

[6] **Gaynor, S.** and Lin, X. Genomic analysis with common binary outcomes via mediation. Oral presentation:

Joint Statistical Meetings, Chicago, IL, (2016).

[7] **Gaynor, S.**, Guffanti, G., Jovanovic, T., Almli, L., Lori, A., Wingo, A., Binder, E., and Ressler, K. Mediation methods applied to post-traumatic stress disorder to identify genomic effects. Poster presentation: American Society of Human Genetics, Vancouver, BC (2016).

[8] **Gaynor, S.** and Lin, X. Causal mediation analysis for common binary outcomes. *In progress*.

[9] **Gaynor, S.** and Lin, X. Integrative copy number and gene expression analysis for cancer tumor subtypes. *In progress*.

**C.3 Network analysis for integrative genomics**

As part of my doctoral research on integrative genomic analysis, I am developing statistical methods for network analysis of multiple genomic data types. The results of this work will allow for measures of uncertainly in metrics used to evaluate networks, leading towards greater reproducibility and assessments of confidence. I am identifying expression quantitative trait loci (eQTL) networks in chronic obstructive pulmonary disease and using perturbation approaches to assess their uncertainty.

[1] **Gaynor, S.,** Fagny, M., Padi, M., Platig, J. andQuackenbush, J. Error quantification in biologically relevant eQTL networks. ENAR Spring Meeting, Washington, DC (2017).

[2] **Gaynor, S.,** Fagny, M., Padi, M., Platig, J. andQuackenbush, J. Error quantification in biologically relevant metrics of eQTL networks. *In progress.*

**C.4 Biostatistical Consulting**

I have worked as a statistical consultant on multiple projects to engage in different collaborative environments and practice various statistical methods in interdisciplinary settings. The first project utilized data from twelve hospitals to compare chart-based and claims-based comorbidity coding schemes on their prognostic ability. The second project investigated the subtypes of smokers who are unmotivated to quit in order to better target smoking cessation approaches. This utilized a latent class analysis of data collected for a smoking study and I was the senior statistician on the project.

[1] Kallogjeri, D., **Gaynor, S.**, Piccirillo, M., Jean, R., Spitznagel, E., and Picirrillo, J. Comparison of

comorbidity collection methods, Journal of the American College of Surgeons. 219 (2) (2014) 245-255.

[2] Borelli, B., **Gaynor, S.**,Tooley, E., Bartlett, K. and Armitage, C. Identification of three difference types of smokers who are not motivated to quit: results from a latent class analysis. *Under review.*

**D. Scholastic Performance**

| YEAR | SCIENCE COURSE TITLE | GRADE | YEAR | OTHER COURSE TITLE | GRADE |
| --- | --- | --- | --- | --- | --- |
| 2009 | General Chemistry 1 with Lab |  | 2009 | Multivariate Calculus |  |
| 2010 | General Chemistry 2 with Lab |  | 2009 | Cultures of Dissent |  |
| 2010 | Principles of Biology with Lab |  | 2009 | Intermediate Latin |  |
| 2010 | Introduction to Scientific Programming |  | 2009 | Lifetime Fitness |  |
| 2011 | Molecular Biology & Genetics |  | 2010 | Discrete Mathemetics |  |
| 2011 | General Physics 1 with Lab |  | 2010 | Spectacle in the Theater |  |
| 2011 | Social & Behavioral Science in Public Health |  | 2010 | Native Americans in Literature |  |
| 2011 | General Physics 2 with Lab |  | 2010 | Greek and Roman Athletics |  |
| 2012 | Environmental Health |  | 2010 | Discrete Mathemetics |  |
| 2012 | Mathematical Methods for Physical Sciences |  | 2010 | Differential Equations |  |
| 2015 | Molecular Biology for Epidemiologists |  | 2010 | Advanced Calculus |  |
| 2016 | Human Genetics |  | 2011 | Linear Algebra for Applications |  |
| 2014 | Population Genetics |  | 2011 | Service Learning: Blood Donation |  |
| 2015 | Advanced Statistical Genetics |  | 2011 | The Developing World |  |
|  |  |  | 2011 | Computers & Society |  |
|  |  |  | 2011 | Introduction to Biostatistics |  |
|  |  |  | 2011 | Introduction to Statistical Computing & Data Management |  |
|  |  |  | 2012 | Principles of Experimental Analysis |  |
|  |  |  | 2012 | Mathematical Modeling |  |
|  |  |  | 2012 | Introduction to Probability |  |
|  |  |  | 2012 | Probability & Statistical Inference 1 |  |
|  |  |  | 2012 | Field Observations in Biostatistics |  |
|  |  |  | 2012 | Principles of Epidemiology |  |
|  |  |  | 2013 | Probability & Statistical Inference 2 |  |
|  |  |  | 2013 | Survey Sampling |  |
|  |  |  | 2013 | Design of Public Health Studies |  |
|  |  |  | 2013 | Introduction to Statistical Computing Environments |  |
|  |  |  | 2013 | Programming |  |
|  |  |  | 2013 | Introduction to Statistical Genetics |  |
|  |  |  | 2013 | Epidemiology 1 |  |
|  |  |  | 2013 | Probability 1 |  |
|  |  |  | 2013 | Statistical Methods 1 |  |
|  |  |  | 2014 | Topics in Clinical Trials |  |
|  |  |  | 2014 | Statistical Inference 1 |  |
|  |  |  | 2014 | Statistical Methods 2 |  |
|  |  |  | 2014 | Probability 2 |  |
|  |  |  | 2014 | Advanced Regression and Statistical Learning |  |
|  |  |  | 2015 | Statistical Inference 2 |  |
|  |  |  | 2015 | Introduction to Data Structures and Algorithms |  |
|  |  |  | 2016 | Analysis of Multivariate and Longitudinal Data |  |