Hoffman Statement of Purpose for Biostatistics PhD programs

*corresponding to **this blog post** on applying to PhDs*

This is my Statement of Purpose (SOP) written in Fall 2022 for the Fall 2023 PhD matriculation application cycle. I've put the SOP text in full below, and then **added an identical version with comments below it**.

Caveat: this SOP is by no means perfect. However, I did receive input from 3-4 biostatistician colleagues and friends who have served on admissions committees at different institutions, so I think this final version probably hits many points that admission committees are looking for. I have posted it publicly because I struggled in writing this quite a bit and was grateful to have people in my life to ask for advice. I hope it will help anyone else who is similarly struggling to understand what to include and how to emphasize various accomplishments in their own SOP.

This should go without saying for PhD applicants, but do not paste exact sentences or phrases from my SOP into your own. 🙂

STATEMENT WITHOUT COMMENTS:

Statement of Purpose

University of Washington – Phd in Biostatistics

Katherine (Kat) Hoffman

I am applying to the PhD program in Biostatistics at the University of Washington to accomplish my goal of becoming a leader in the development and application of modern statistical methodology for healthcare research. My passion for biostatistics is rooted in a deep interest in medicine and a lifelong fascination in mathematics and its use in the pursuit of knowledge. As my academic training, publication, teaching, and leadership record demonstrate, I have the drive and competence to make impactful contributions to statistics and public health.

My interest in medicine started early in life. My mother was a registered nurse and instilled in me a curiosity about the origins of disease and an interest in improving the way we take care of our sick. As early as elementary school, I studied her books on epidemics and evidence-driven medicine, and in college, I cared for hospitalized patients with intentions of attending medical school. However, I also derived an innate enjoyment from mathematics, which led me to pursue classes in both pure and applied mathematics. When I discovered biostatistics during my final year of undergraduate studies, I was drawn to a subject that would allow me to integrate my two passions, mathematics and medicine, and to use mathematics to achieve my goal of improving people's health.

This interest led me to pursue a master's degree in biostatistics at the University of Michigan. Through my coursework, I built a strong theoretical foundation, which deepened my passion for the subject. Outside of my coursework, I sought opportunities to apply my knowledge, earning research assistantships in the neuroscience, computational biology, and biostatistics departments, and volunteering for applied projects with Statistics in the Community. Through my research projects and background in patient care, I discovered a natural aptitude for critically thinking through healthcare applications.

Upon graduation, I began working as a biostatistician researcher at Weill Cornell Medicine (WCM), a top-tier academic research institution in New York City. Through short- and long-term collaborations, I became well-versed in study design, data management, efficient programming, data visualization, statistical modeling, and communicating results. I thrived in the self-driven environment of academia, and my contributions led to co-authorship on over 40 published manuscripts, including four papers on statistical methods and four with first or co-first authorship. My position at WCM also allowed tremendous opportunities for growth. I worked closely on several projects with Dr. Iván Díaz, an emerging leader in causal inference, and he introduced me to a rapidly-growing subfield of semi-parametric estimation for causal inference. I immediately understood that causality is at the root of scientific progress, and I was eager to learn all I could about statistical methods to investigate causation.

The five years I spent as a biostatistician researcher led me to realize that many of the conclusions we reach in clinical research are based on studies with restrictive assumptions and imperfect statistical analyses. At the same time, my independent study of causal inference showed me that we often do have methodology and software to more appropriately answer complex research questions. Unfortunately, there remains a large communication and knowledge gap between methodologists, analysts, and domain-expert researchers, preventing these methods from being routinely used in practice. I see this disconnect as directly harmful to patients, and I am highly motivated to contribute to addressing these challenges. My ultimate career goal is to become an academic statistician at the intersection of methods development and applications, and to significantly improve the implementation of statistics in medicine.

I am already working towards accomplishing this goal. This year, I first-authored a research paper implementing a modern method for causal inference to estimate the effect of corticosteroids for COVID-19 patients on mortality in *JAMA Network Open*. We compared this modern approach (i.e., a longitudinal target trial emulation with a dynamic treatment rule and doubly robust estimator) to other methods common in the clinical research literature (e.g., defining the treated group as "receiving corticosteroids within 5 days of hospitalization" and fitting a Cox proportional hazards model). This application is an example where standard research practice has repeatedly failed; despite multiple randomized trials showing a protective effect, results of analyses on observational data are inconclusive and plagued with study design and estimation biases (e.g. poorly defined treatment windows, inappropriate adjustment for time-dependent confounding). I discussed my results in an oral presentation at the American Causal Inference Conference. The enjoyment I derived from this process and its potential impact on clinical research made me realize that a PhD is the right path forward for me to pursue my passions.

In terms of methodology, I am keen to contribute to semi-parametric methods for causal inference, for example for the design and analysis of longitudinal studies from observational data. I have already made some progress in this direction; I co-authored a longitudinal estimation paper with Dr. Díaz in the *Journal of the American Statistical Association* and recently first-authored a paper submitted to *Epidemiology* describing the same method for applied researchers. I am confident I have the drive, training, and resiliency necessary to generate novel ideas, manage my responsibilities, and complete research endeavors during a PhD. I am excited about the prospect of using my training time as an opportunity to acquire a deeper understanding of statistical theory, which will empower me to pursue independent methodological research.

Washington's highly reputed Department of Biostatistics is particularly suited for my methodological research interests. I am open to all research ideas, but I am most motivated by the prospect of developing novel, flexible methods and estimators for causal inference and applying them to the cutting-edge data sources available at Washington. I am especially interested in working with faculty at the intersection of non-parametric statistics and causal inference, such as Drs. Marco Carone, Alex Luedtke, and Peter Gilbert. It would be a privilege to extend my knowledge of statistical theory in a top-tier environment, and I know that if given the opportunity, the education and training I would receive at Washington would optimally prepare me for my future endeavors and goals. On a personal note, I am interested in pursuing a doctoral degree in Seattle due to its proximity to many outdoor activities, which is how I spend much of my free time.

As eager as I am to contribute to new statistical methodology at the University of Washington, I am similarly excited to contribute to the future of statistics education. Teaching and mentoring is a lifelong passion of mine, and I greatly enjoyed my time supervising and mentoring biostatisticians at WCM. I have written dozens of statistics and programming blog posts on my website, KHstats, and given numerous talks at meetings and conferences. For many of my scientific communication modes, I create detailed illustrations explaining the concepts, such as a series of visual guides explaining causal inference estimation methods to applied researchers. I also helped teach seven semesters of Biostatistics and Data Science MS courses at WCM, generating materials, running labs, and guest lecturing. I genuinely enjoy thinking about the way we teach statistics, and I am enthusiastic for opportunities to continue educating during and after a PhD.

My record shows a long history of starting and finishing projects and of successfully holding leadership positions. Growing up I was frequently elected to leadership roles, and my senior year of college I was selected captain of my Division I softball team. During my master's degree, I was president of University of Michigan's Biostatistics Student Association, and while at WCM, I hosted our computing club and managed a causal inference learning group. I look forward to the opportunity to bring my research, education, and leadership skills to Washington while learning the statistical theory necessary for this next stage of my career.

STATEMENT WITH COMMENTS:

General advice I received: The SOP is not about qualifications, that's what the CV is for. There will be many qualified people applying and qualifications alone do not mean success in academia. You need drive and motivation. The SOP is your opportunity to tell the committee about that drive and motivation.

Statement of Purpose

University of Washington – Phd in Biostatistics

Katherine (Kat) Hoffman

I tried to structure my entire statement around: answering the questions of 1. why I want to pursue a PhD and 2. why I will be a successful PhD student.

I am applying to the PhD program in Biostatistics at the University of Washington to accomplish my goal of becoming a leader in the development and application of modern statistical methodology for healthcare research. My passion for biostatistics is rooted in a deep interest in medicine and a lifelong fascination in mathematics and its use in the pursuit of knowledge. As my academic training, publication, teaching, and leadership record demonstrate, I have the drive and competence to make impactful contributions to statistics and public health.

Advice I received: keep the lifelong interest in science and/or medicine and/or statistics details to a minimum. It is more-or-less assumed that everyone applying for a PhD program in biostat has an interest or aptitude for these subjects. Admissions committees only care about your background if it is somehow informative as to whether you'll succeed in the program.

My interest in medicine started early in life. My mother was a registered nurse and instilled in me a curiosity about the origins of disease and an interest in improving the way we take care of our sick. As early as elementary school, I studied her books on epidemics and evidence-driven medicine, and in college, I cared for hospitalized patients with intentions of attending medical school. However, I also derived an innate enjoyment from mathematics, which led me to pursue classes in both pure and applied mathematics. When I discovered biostatistics during my final year of undergraduate studies, I was drawn to a subject that would allow me to integrate my two passions, mathematics and medicine, and to use mathematics to achieve my goal of improving people's health.

In general you should start with your "why"s (what your purpose in pursuing a PhD is) and then only use the "what"s (prior accomplishments) to demonstrate that why. A nice paragraph structure to follow is:

1. *Topic sentence: state accomplishment/activity/period in your life that you're going to discuss*
2. *Body sentences: Explain what you did* ***and why that's prepared you for some aspect of graduate school***
3. *Conclusion sentence: Summarize why this particular aspect of your life supports your overall statement of purpose (I.e. why did it prepare you to pursue a PhD in Biostatistics)*

This interest led me to pursue a master's degree in biostatistics at the University of Michigan. Through my coursework, I built a strong theoretical foundation, which deepened my passion for the subject. Outside of my coursework, I sought opportunities to apply my knowledge, earning research assistantships in the neuroscience, computational biology, and biostatistics departments, and volunteering for applied projects with Statistics in the Community. *[I received feedback on a version of my SOP very close to this that if I wanted to stay closer to word limits, I could use less sentences describing which courses and colleges I went to. I opted to leave this in (see comments at end of document about word limits).]* Through my research projects and background in patient care, I discovered a natural aptitude for critically thinking through healthcare applications.

Upon graduation, I began working as a biostatistician researcher at Weill Cornell Medicine (WCM), a top-tier academic research institution in New York City. Through short- and long-term collaborations, I became well-versed in study design, data management, efficient programming, data visualization, statistical modeling, and communicating results. I thrived in the self-driven environment of academia, and my contributions led to co-authorship on over 40 published manuscripts, including four papers on statistical methods and four with first or co-first authorship. *[I* ***hate*** *writing in this tone, but as far as I can tell, this is how SOPs are written. In every sentence it is important to use emphatic sentences and convey to the application committee that you know what a PhD in Biostatistics will entail and to leave them with no doubt that you are capable of succeeding in it.]* My position at WCM also allowed tremendous opportunities for growth. I worked closely on several projects with Dr. Iván Díaz, an emerging leader in causal inference, and he introduced me to a rapidly-growing subfield of semi-parametric estimation for causal inference. I immediately understood that causality is at the root of scientific progress, and I was eager to learn all I could about statistical methods to investigate causation. *[I received advice from one or two people to not sound so decisive about what I want to study (causal inference), in part because someone reviewing my application might not agree with me or would interpret it as me being inflexible. I decided to ignore this advice, but passing on anyways!]*

This next paragraph is the most important one in my statement -- it is describing my primary motivation for pursuing a PhD. The "why" paragraph could fit into many different locations in your statement depending on your specific background/story, but it should be a central part of your SOP.

The five years I spent as a biostatistician researcher led me to realize that many of the conclusions we reach in clinical research are based on studies with restrictive assumptions and imperfect statistical analyses. At the same time, my independent study of causal inference showed me that we often do have methodology and software to more appropriately answer complex research questions. Unfortunately, there remains a large communication and knowledge gap between methodologists, analysts, and domain-expert researchers, preventing these methods from being routinely used in practice. I see this disconnect as directly harmful to patients, and I am highly motivated to contribute to addressing these challenges. My ultimate career goal is to become an academic statistician at the intersection of methods development and applications, and to significantly improve the implementation of statistics in medicine.

Application committees highly value research experience. If you've led any of your own research or played key roles in ongoing research, make sure you elaborate on how that has impacted you and shows your scientific potential.

I am already working towards accomplishing this goal. This year, I first-authored a research paper implementing a modern method for causal inference to estimate the effect of corticosteroids for COVID-19 patients on mortality in *JAMA Network Open*. We compared this modern approach (i.e., a longitudinal target trial emulation with a dynamic treatment rule and doubly robust estimator) to other methods common in the clinical research literature (e.g., defining the treated group as "receiving corticosteroids within 5 days of hospitalization" and fitting a Cox proportional hazards model). This application is an example where standard research practice has repeatedly failed; despite multiple randomized trials showing a protective effect, results of analyses on observational data are inconclusive and plagued with study design and estimation biases (e.g. poorly defined treatment windows, inappropriate adjustment for time-dependent confounding). I discussed my results in an oral presentation at the American Causal Inference Conference. The enjoyment I derived from this process and its potential impact on clinical research made me realize that a PhD is the right path forward for me to pursue my passions.

[Make sure you talk about your research interests within biostatistics!] In terms of methodology, I am keen to contribute to semi-parametric methods for causal inference, for example for the design and analysis of longitudinal studies from observational data. I have already made some progress in this direction; I co-authored a longitudinal estimation paper with Dr. Díaz in the *Journal of the American Statistical Association* and recently first-authored a paper submitted to *Epidemiology* describing the same method for applied researchers. I am confident I have the drive, training, and resiliency necessary to generate novel ideas, manage my responsibilities, and complete research endeavors during a PhD. I am excited about the prospect of using my training time as an opportunity to acquire a deeper understanding of statistical theory, which will empower me to pursue independent methodological research.

General advice for SOPs is to have at least one school-specific paragraph that name-drops who you'd want to work with in the department (and even more specifically, what line of work they're in, which should align with your previously stated research interests). In one or two of the other SOPs I read, the writer went into much more detail about each professor's research... but my SOP was already quite long so I kept this part short.

Washington's highly reputed Department of Biostatistics is particularly suited for my methodological research interests. I am open to all research ideas, but I am most motivated by the prospect of developing novel, flexible methods and estimators for causal inference and applying them to the cutting-edge data sources available at Washington. I am especially interested in working with faculty at the intersection of non-parametric statistics and causal inference, such as Drs. Marco Carone, Alex Luedtke, and Peter Gilbert. It would be a privilege to extend my knowledge of statistical theory in a top-tier environment, and I know that if given the opportunity, the education and training I would receive at Washington would optimally prepare me for my future endeavors and goals. On a personal note, I am interested in pursuing a doctoral degree in Seattle due to its proximity to many outdoor activities, which is how I spend much of my free time. *[I added at least a sentence to each statement about why I liked the location of the program. I also added why I liked the University and/or hospital system.]*

If it were up to me I would've centered the entire statement around how blogging and teaching shows my passion for biostatistics. However, I was told by many that admissions committees don't care much about this. As one mentor told me, "your altruistic reasons for a PhD reason are nice, but they are not what admissions committees are looking for." Thus, I kept my teaching and blogging enthusiasm contained to this later paragraph and recommend the same to anyone with a similar background.

As eager as I am to contribute to new statistical methodology at the University of Washington, I am similarly excited to contribute to the future of statistics education. Teaching and mentoring is a lifelong passion of mine, and I greatly enjoyed my time supervising and mentoring biostatisticians at WCM. I have written dozens of statistics and programming blog posts on my website, KHstats, and given numerous talks at meetings and conferences. For many of my scientific communication modes, I create detailed illustrations explaining the concepts, such as a series of visual guides explaining causal inference estimation methods to applied researchers. I also helped teach seven semesters of Biostatistics and Data Science MS courses at WCM, generating materials, running labs, and guest lecturing. I genuinely enjoy thinking about the way we teach statistics, and I am enthusiastic for opportunities to continue educating during and after a PhD.

I used this concluding paragraph to briefly highlight previous leadership roles on my CV and tie up my first sentence of the entire statement ("I desire to be a leader in the field of biostatistics"). These were accomplishments I could have spent more time talking about, but I could not find a way to directly relate to the main messages of the statement.

My record shows a long history of starting and finishing projects and of successfully holding leadership positions. Growing up I was frequently elected to leadership roles, and my senior year of college I was selected captain of my Division I softball team. During my master's degree, I was president of University of Michigan's Biostatistics Student Association, and while at WCM, I hosted our computing club and managed a causal inference learning group. I look forward to the opportunity to bring my research, education, and leadership skills to Washington while learning the statistical theory necessary for this next stage of my career.

My SOP is quite long (>1200 words). Some schools had word limits as short as 500, but I submitted this version (switching out school-specific details) to all schools and ignored the word limit. I am not sure I would recommend that tactic unless you've had several other people review your SOP and confirmed that you don't have unnecessary / repetitive details.

***

These are all my comments. Good luck writing your own SOP and please email me if anything is unclear!

–Kat