1. Cover Page

Title of the Area: Statistical Learning
with Vaccine Efficacy Trials, Medical Studies and Big Data

Target Category of the Submission

Existing and Emerging Excellence

List of Participating Disciplines

Biostatistics
Bioinformatics
Data Science
Mathematics
Public Health
Public Policy
Sociology
Statistics

List of Participating Departments

Department of Mathematics and Statistics, College of Liberal Arts and Sciences
Department of Bioinformatics and Genomics, College of Computing and Informatics
Department of Geography and Earth Sciences, College of Liberal Arts and Sciences
Department of Public Health Sciences, College of Health and Human Services
Department of Sociology, College of Liberal Arts and Sciences
School of Data Science

Lead of Proposal

Yanqing Sun

List of Program Leads/Co-Leads

Program 1: Statistical Methods for Vaccine Efficacy Trials
    – Yanqing Sun (Lead), Department of Mathematics and Statistics

Program 2: Design and Analysis of Biomedical Studies
    – Qingning Zhou (Lead), Department of Mathematics and Statistics

Program 3: Statistical/Machine Learning and Its Application to Big Data
    – Jiancheng Jiang (Lead), Department of Mathematics and Statistics

Key Words

Biostatistics, Distributed computation, Forecasting, Precision medicine, Vaccine efficacy trials

2. Executive Summary

The nomination consists of three programs led by faculty with outstanding contributions in statistical science and its applications. The team also includes some outstanding young faculty. The team's record of collaborations will ensure the success of the programs.

Program 1. Statistical Methods for Vaccine Efficacy Trials. Dr. Yanqing Sun will lead this program. Randomized, controlled clinical trials that rigorously assess the efficacy of candidate vaccines to prevent infection and/or disease are a core research platform for developing improved vaccines. With her collaborators, Dr. Sun has worked on developing innovative statistical methods for the design and analysis of HIV, dengue and malaria vaccine efficacy trials. Dr. Sun has developed statistical methods to understand the relationship between vaccine efficacy and the genetic features of the viruses. Remarkably, most of the statistical approaches built originally for the HIV or dengue vaccine efficacy studies are completely on target for application to COVID-19 vaccine efficacy trials. Dr. Sun has worked with the HIV Vaccine Trials Network (HVTN) Statistical Center for 20 years to contribute statistical methods, through collaboration with the HVTN Statistical Center PI Peter Gilbert, who now leads the NIH-sponsored Coronavirus Prevention Network Statistical Center, which through Operation Warp Speed co-conducts several phase 3 COVID-19 vaccine efficacy trials. Through this long-standing collaboration, Dr. Sun will contribute statistical methods for application to these trials for developing immune marker surrogate endpoints and for understanding how vaccine efficacy depends on immunological phenotype and sequence features of the SARS-CoV-2 virus. The significance of Dr. Sun's research has been recognized with the continuous funding supports from the the National Science Foundation (NSF) for eighteen years and from the National Institutes of Health (NIH) for fourteen years. She has been awarded over two million dollars in research grants since she joined UNCC in 1994. Dr. Sun has successively supervised thirteen Ph.D. students and five Master students. She and her students coauthored twenty five of publications.

Program 2. Design and Analysis of Biomedical Studies. Dr. Qingning Zhou will lead this program. Epidemiological and biomedical studies that relate disease outcomes to individual exposures and other characteristics play a key role in understanding the determinants of diseases in humans. Dr. Zhou has worked on developing novel statistical methods for problems arising from the design and analysis of such studies, including issues on censored time-to-disease, cost-effective sampling designs, and missing or error-prone exposures. The developed methods were applied to analyzing HIV preventive vaccine trials, AIDS clinical trials, and observational studies on diabetes and hypertension. Dr. Zhou has published her work in top-tier statistical journals and received two prestigious paper awards for young investigators. Her research has been supported by an NSF grant as PI since 2019. Dr. Zhou has also collaborated with experts in sickle cell disease to solve complex statistical problems in this area, leading to publications in high-profile medical journals. They have recently submitted an NIH R01 grant proposal. Since her joining UNCC in 2017, Dr. Zhou has supervised three Master students and two undergraduate projects. She is currently supervising two Ph.D. students and one undergraduate honors project.

Program 3. Statistical/Machine Learning and Its Application to Big Data. Dr. Jiancheng Jiang will lead this program. There are many supervising learning models/algorithms, applied in various areas such as financial market (quantitative trading), network (face perception), medical science (medical diagnostics), and industry (robot), etc. Existing model comparison tools for big data are built upon the assumption that the model is true, but all models are wrong. The team will investigate this problem and provide a general nonparametric solution, so that it can be applicable to model comparison between any two supervising learning models, which may be nested, non-nested, overlapping, mis-specified, or correctly specified. The proposed methodology will also be applicable for variable selection with big data arising from genetics and public health, etc. For massive data, since the sample size is huge, one machine cannot implement the proposed algorithm due to memory constraints. Dr. Jiang's team aims to develop some distributed algorithms and establish their asymptotic equivalence to the ideal algorithms running on a machine with no limited memory constraints. Advancing this kind of tools is significant and critical for identifying better and better automatic machine learning algorithms. The significance of Dr. Jiang's research has been recognized with many of his top journal publications and over 10 funding supports from the NSF and other agencies. He has supervised 14 PhD students, 28 Master students and 9 undergraduate senior projects. He and his students jointly published over 20 papers.

3. Evidence of Strength and Excellence

Program 1. Statistical Methods for Vaccine Efficacy Trials. Dr. Sun has worked in a broad area of statistics and biostatistics. She has made outstanding contributions in developing statistical methods to solve many complex problems arising from medical and public health studies. Dr. Sun has developed innovative statistical methods dealing with complex data such as censored data, missing data, and data with measurement errors. She has published over sixty peer-reviewed statistical papers in statistical journals. Dr. Peter Gilbert, Dr. Sun's long-term collaborator over eighteen years, is a leader in biostatistical design and analysis of vaccine efficacy trials and a member of Statistical Leadership of Coronavirus Prevention Network. Collaborating with Dr. Gilbert, Dr. Sun has been working on improvement and development of statistical methods for the design and analysis of HIV, dengue, malaria vaccine efficacy trials, as well as COVID-19 prevention efficacy trials of broadly neutralizing monoclonal antibodies. Close collaboration with Dr. Gilbert provides information and expertise on the evolving science of the vaccine efficacy trials as well as access to the data for analysis. More importantly, the researches conducted by Dr. Sun are in alignment with the priorities of the vaccine trials, can be incorporated into the statistical analysis plans, and thus will have direct impact on the success of the vaccine trials. Some of methods developed by Dr. Sun are currently being processed for analyses of COVID-19 vaccine trials.

The significance of Dr. Sun's research has been recognized with the continuous funding supports from National Science Foundation (NSF) as Principal Investigator (PI) for eighteen years and from National Institutes of Health (NIH) as Co-Investigator (Co-PI) for fourteen years (with Dr. Gilbert as PI). Dr. Sun has been awarded over two million dollars in research grants since she joined UNCC in 1994. Two (NIAID) NIH sponsored projects for which she served as Co-PI, *"Statistical Methods in HIV Vaccine Efficacy Trials" (4/1/2010–3/31/2020)* and *"Statistical Methods for Efficacy Trials of Vaccines and Monoclonal Antibodies Against Genetically-Diverse Pathogens" (07/01/2020 – 06/30/2024)*, received the MERIT (Method to Extend Research in Time) Award (R37), which provides long-term support to investigators with impressive records of scientific achievement in research areas of special importance or promise. Dr. Sun has received other prestigious honors and are serving on the editorial boards of top-tier statistical journals (listed in her Biographical Sketch).

Program 2. Design and Analysis of Biomedical Studies. Dr. Zhou has a broad research interest in biostatistics including both new statistical methodology and novel applications in biomedical studies. Dr. Zhou has published one book chapter, six papers in top-tier statistical journals, and five papers in high-profile medical journals since 2017. Two of her papers received prestigious paper awards for young investigators from the American Statistical Association (ASA) for their innovation in statistical methodology and applications. Dr. Zhou's research on two-phase sampling designs with interval-censored data was funded by an NSF grant as PI since 2019.

Dr. Zhou has collaborated with Dr. Kenneth Ataga, Director of the Center for Sickle Cell Disease at the University of Tennessee Health Science Center, to solve complex statistical problems from observational studies of sickle cell disease (SCD). SCD is an inherited red blood cell disorder with serious complications and is prevalent in sub-Saharan Africa. About 1 in 365 Black or African American babies is born with SCD and about 1 in 13 is born with sickle cell trait. Dr. Zhou and her collaborators have published four papers on SCD and recently submitted an NIH R01 grant proposal on predicting CKD progression in SCD patients using novel statistical and machine learning methods. Precision medicine has received tremendous attention in clinical practice and medical research since President Obama launched the Precision Medicine Initiative in 2015. Unlike the traditional "one size fits all" medical practice designed for "average patient", precision medicine seeks to recommend the right treatment to the right patient at the right time by taking into account individual differences in genes, environments and lifestyles. Dr. Zhou is currently collaborating with Dr. Sun working on problems arising from AIDS clinical trials and vaccine efficacy trials, and with her UNCC colleagues Drs. Pan and Jiang on precision medicine and big data problems in biomedical studies.

Program 3. Statistical/Machine Learning and Its Application to Big Data. Dr. Jiang has worked in various areas of econometrics, (bio)statistics, and data science. He has made significant contributions to statistical modeling and solved important problems arising from economy and statistics. He has published 63 peer-reviewed papers in the above areas, and many of them are in top journals in (bio)statistics and econometrics. Many of his coauthors are top international statisticians and econometricians. He has 14

publications in top journals with Dr. Ait-Sahalia and Dr. Fan at Princeton University, where the former is a leader in econometrics, and the latter is a leader in statistics, data science and econometrics who received numerical honors and awards such as the COPSS President's Award and Guy Medal in Silver. Dr. Jiang's long-term research team is very capable of conducting leading research work of originality and creativity. These collaborations provide wide applications to analyzing complex data from economy, statistics, genetics, and medical sciences.

Dr. Jiang has won a national and international reputation in his academic career. Examples include but not limited to these publications: Test, 2007, **16**, 409-444, with discussion and 105 SCI citations; JASA, 2005, **100**, 890-907, with 177 SCI citations; Ann. Statist., 2011, **39**, 3092-3120, 115 SCI citations. In the annals paper, he developed an algorithm for choosing 25 important genes among 17118 ones for a non-Hodgkin'ss lymphoma dataset. His recent top journal paper (JASA, 2016, **111**, 275-287, 32 SCI citations) pioneers a classification rule and identifies email spams and lung cancers in high dimensional settings. Dr. Jiang has received over 10 funding supports from the NSF and other agencies. Dr. Jiang has received prestigious honors and are serving on the editorial boards of top-tier statistical journals (listed in his Biographical Sketch).

Strength of Research Collaborations

The nomination consists of three key research areas integral to modern medicine and public health: disease prevention, design and modeling, and computing and machine learning. Problems arising from medical studies and public health often lie in the intersection of these areas. The nomination assembled a strong group of UNCC faculty across disciplines who had an excellent record of research and funding histories. The faculty participating in this nomination has a track record of collaborations with UNCC faculty and with internationally well-known experts, and a strong record of collaborations with the top medical institutions. In particular, Dr. Sun's collaboration with Dr. Gilbert on vaccine efficacy trials has resulted in over one million dollar funding to UNCC.

Within UNCC, Dr. Sun (Lead) and Dr. Zhou (Lead) are currently collaborating to solve problems arising from AIDS clinical trials and vaccine efficacy trials to evaluate the treatment effects for an AIDS clinical trial and vaccine effectiveness of several vaccine candidates. Dr. Sun, Dr. Zhou and Dr. Yinghao Pan (Participant) are working to develop innovative approaches to model HIV and COVID-19 infections. Dr. Sun and Dr. Shaoyu Li (Participant) have collaborated on a microbiome association study. Both Dr. Zhou and Dr. Pan were partially supported by Dr. Sun's current NIH grant. These collaborations have resulted in one accepted paper, one paper to be accepted with minor revision, one under review and making significant progress in another. Dr. Zhou also works with Drs. Pan and Jiang on precision medicine and big data problems in biomedical studies. Dr. Jiang (Lead) and Dr. Sun collaborated on a project developing statistical methods examining the risk of failure with missing data. Dr. Jiang has also collaborated with Dr. Michael Grabchak (Participant). Dr. Jiang is currently collaborating with his UNCC colleagues, Dr. Aziz Issaka (Participant), Dr. Zhou and Dr. Jun Song (Participant), on various research topics in data science. Dr. Jiang and Dr. Issaka collaborated with a proposal submission (NSF-PD 18-1269) in 2020. Dr. Jiang also collaborated on another NIH proposal submission recently with Dr. Stephanie Moller (Participant).

Education and Research Training

The team's research programs have generated many research topics for graduate projects. Dr. Sun has successively supervised thirteen Ph.D. students and published twenty five statistical papers with her students. Dr. Jiang is highly sought-after supervisor by graduates, who has supervised twelve PhD students. Dr. Sun and Dr. Jiang are currently each supervising four Ph.D. students. Dr. Zhou is supervising two Ph.D. students, jointly supervising one of them with Dr. Sun. Dr. Sun and Dr. Pan are also jointly supervising one Ph.D. student.

Resources to Expand upon Past Success

The faculty in the statistics program have achieved national recognition with publications and the federal grant fundings. According to the most recent US News Graduate Program Ranking, the Statistics program at UNCC is ranked #97, the Mathematics program is ranked #108 while the overall ranking of UNCC is #227 in the nation. Additional resources that can be used to expand upon past success are crucial. The proposal requests two faculty in the areas of biostatistics and machine learning, and the funding for three research assistantships to support the research programs of the cluster.

4. Alignment with Regional and National Priorities

The proposed research programs in this nomination align very well with the mission of the university and the regional, national and international priorities. Its success has direct impact on medical, health and economical development for the Charlotte region, the nation and the globe.

NIH Strategic Plans and Visions includes **NIH Strategic Plan for HIV and HIV-Related Research** (`https://www.oar.nih.gov/hiv-policy-and-research/strategic-plan`), which provides the a roadmap for the National Institutes of Health (NIH) HIV/AIDS research program, ensuring that funds are allocated in accordance with established NIH scientific `research priorities`. The top three research priorities are *(1) Reduce the Incidence of HIV, (2) Develop Next-Generation HIV Therapies, (3) Conduct Research Toward HIV Cure.* The proposed Programs 1 and 2 in this nomination play a critical role in the efforts to achieve these goals. The leader (Dr. Sun) of Program 1 has been very successful in developing innovative statistical methods for evaluating HIV vaccine efficacy trials (contributing to the priority (1)) and for the AIDS clinical trials (contributing to the priorities (2) and (3)). Two (NIAID) NIH sponsored projects for which she served as Co-PI, "Statistical Methods in HIV Vaccine Efficacy Trials" (4/1/2010–3/31/2020, PI: Dr. Gilbert) and "Statistical Methods for Efficacy Trials of Vaccines and Monoclonal Antibodies Against Genetically-Diverse Pathogens" (07/01/2020 – 06/30/2024, PI: Dr. Gilbert), received the MERIT (Method to Extend Research in Time) Award (R37). Developing effective COVID-19 vaccine is now an urgent national and international priority. The statistical methods developed by Dr. Sun for the HIV, dengue and malaria vaccine efficacy trials can be incorporated into the statistical analysis plans COVID-19 vaccine research, and will have direct impact on the success of the vaccine trials. Some of the developed methods are currently being processed for analyses of COVID-19 vaccine trials through collaboration with Dr. Gilbert who is a member of Statistical Leadership of Coronavirus Prevention Network.

Precision medicine, or personalized medicine, has received tremendous attention in clinical practice and medical research. In his 2015 State of the Union address, President Obama launched the **Precision Medicine Initiative**, further making precision medicine a national priority. Program 2 contributes to this endeavor through developing novel statistical and machine learning methods for discovering optimal individualized treatment rules to maximize the quality of health care. Program 2 also contributes to advancing the research on sickle cell disease (SCD) through the collaboration with well-known experts of the area. SCD is an inherited red blood cell disorder prevalent in sub-Saharan Africa and mainly occurs in Black or African Americans in the U.S. The National Institutes of Health (NIH) launched the **Cure Sickle Cell Initiative** in 2018 and also launched new collaboration to develop affordable, gene-based cures for SCD and HIV on global scale in 2019.

The research in developing statistical methodologies for big data has received tremendous attention, due to increasing demand to adapt to new types of data that drastically increased dimension, volume, and complexity in multidisciplinary fields. Program 3 in this nomination responds to the call of **NSF's BIGDATA program for 21st-century data-capable workforce** (`https://www.nsf.gov/pubs/2019/nsf19039/nsf19039.jsp`) and has developed novel statistical methods for identifying important features in machine learning, testing critical hypotheses about scientific problems, and forecasting and controlling potential risks in economy development. Developing innovative statistical methods for mis-specified high-dimensional models is not only extremely important for correctly harnessing bigdata, but also fundamental for extracting key factors relevant to economy, environment, and public health.

The Department of Mathematics and Statistics has strong graduate programs in Statistics. It has provided a pipeline of workforce to companies in Charlotte and around the country. During the most recent 5 years (2016 to 2020), the statistics program has graduated 17 Ph.D. students and 26 Master students. The program has an excellent track record of students' job placement. The graduates in the Statistics program are placed into high-paying jobs in the region and around the country. Four graduates are placed as tenure track assistant professors in research oriented universities. Most Ph.D. graduates have found good industry jobs working on model development, risk management and statistical analysis in companies such as Wells Fargo (NC), Bank of America (NC), Brighthouse Financial (NC), Lowe's Company (NC), Intermountain Healthcare (UT). The program also has significant engagement with the Charlotte business industry and has a growing demand for statistical consulting. The program is establishing collaborations with local industries such as Bank of America and Lowe's.

5. Supporting Documents

List of leaders and participants as well as their titles and areas of expertise.

Program 1. Statistical Methods for Vaccine Efficacy Trials.

Name	Title	Role	Areas of Expertise
Yanqing Sun	Professor of Statistics	Lead	Clinical Trials, Longitudinal data analysis, Measurement error, Missing data, Survival analysis, Vaccine efficacy, Two-phase sampling
Qingning Zhou	Assistant Professor of Statistics	Participant	Biostatistics, Survival analysis, Two-phase sampling, Missing data, Measurement error
Yinghao Pan	Assistant Professor of Statistics	Participant	Precision medicine, Machine learning, Survival analysis
Shaoyu Li	Assistant Professor of Statistics	Participant	Statistical genetics & genomics, Statistical & computational methodologies, Cell-type specific transcriptional changes associated with human diseases, Statistical design & analysis methods for mobile health studies
Shi Chen	Assistant Professor of Public Health	Participant	Infectious disease epidemiology, Microbiology, Mathematics, Statistics, Public health
Jun-Tao Guo	Professor of Bioinformatics	Participant	Bioinformatics, Computational modeling and Prediction
Jun Song	Assistant Professor of Statistics	Participant	Dimension reduction, Functional data analysis, Machine learning
Jiancheng Jiang	Professor of Statistics	Participant	Econometrics, Data science, Machine learning
Michael Grabchak	Associate Professor of Probability	Participant	Probability & Statistics, Stochastic process
Taufiquar Khan	Professor of Mathematics	Participant	Deep learning, Regularization, Tomography

Program 2. Statistical Methods in the Design and Analysis of Biomedical Studies.

Name	Title	Role	Areas of Expertise
Qingning Zhou	Assistant Professor of Statistics	Lead	Biostatistics, Survival analysis, Two-phase sampling, Missing data, Measurement error
Yanqing Sun	Professor of Statistics	Participant	Clinical trials, Longitudinal data analysis, Measurement error, Missing data, Survival analysis, Vaccine efficacy, Two-phase sampling
Yinghao Pan	Assistant Professor of Statistics	Participant	Precision medicine, Machine learning, Survival analysis
Jiancheng Jiang	Professor of Statistics	Participant	Econometrics, Data science, Machine learning
Jun Song	Assistant Professor of Statistics	Participant	Dimension reduction, Functional data analysis, Machine learning
Shaoyu Li	Assistant Professor of Statistics	Participant	Statistical genetics & genomics, Statistical & computational methodologies, Cell-type specific transcriptional changes associated with human diseases, Statistical design & analysis methods for mobile health studies
Shi Chen	Assistant Professor of Public Health	Participant	Infectious disease epidemiology, Microbiology, Mathematics, Statistics, Public health
Jun-Tao Guo	Professor of Bioinformatics	Participant	Bioinformatics, Computational modeling and Prediction
Taufiquar Khan	Professor of Mathematics	Participant	Deep learning, Regularization, Tomography
Michael Grabchak	Associate Professor of Probability	Participant	Probability & Statistics, Stochastic process

Program 3. Statistical/Machine Learning and Its Application to Big Data.

Name	Title	Role	Areas of Expertise
Jiancheng Jiang	Professor of Statistics	Lead	Econometrics, Data science, Machine learning
Yang Chao	Associate Professor of Sociology	Participant	Economy of China, Public policy, Sociology
Eliana Christou	Assistant Professor of Statistics	Participant	Dimension reduction, Quantile regression, Variable selection
Michael Grabchak	Associate Professor of Probability	Participant	Probability & Statistics, Stochastic process
Juntao Guo	Professor of Bioinformatics	Participant	Bioinformatics, Computational modeling and Prediction
Aziz Issaka	Assistant Professor of Mathematics	Participant	Mathematical finance
Taufiquar Khan	Professor of Mathematics	Participant	Deep learning, Regularization, Tomography
Shaoyu Li	Assistant Professor of Statistics	Participant	Statistical genetics & genomics, Statistical & computational methodologies, Cell-type specific transcriptional changes associated with human diseases, Statistical design & analysis methods for mobile health studies
Stephanie Moller	Professor of Sociology	Participant	Public policy, Sociology
Yinghao Pan	Assistant Professor of Statistics	Participant	Precision medicine, Machine learning, Survival analysis
Jun Song	Assistant Professor of Statistics	Participant	Dimension reduction, Functional data analysis, Machine learning
Yanqing Sun	Professor of Statistics	Participant	Clinical trials, Longitudinal data analysis, Measurement error, Missing data, Survival analysis, Vaccine efficacy, Two-phase sampling
Wenwu Tang	Associate Professor of Geography and Earth Sciences	Participant	Deep learning, Spatial data analysis
Qingning Zhou	Assistant Professor of Statistics	Participant	Biostatistics, Survival analysis, Two-phase sampling, Missing data, Measurement error