National Institute of Statistical Sciences

ANNUAL REPORT

July 1, 2022 - June 30, 2023

N I S S

Webinars

Research

Affiliates ∞ Early Career

Interpretable/ Explainable Machine Learning

National Institute of Statistical Sciences

July 1, 2022 - June 30, 2023

From the Leadership…

This report describes the year from July 1, 2022 through June 30, 2023. NISS emerged from the pandemic challenges stronger in some ways ; NISS is now a hybrid organization with new virtual connections internationally and new opportunities for in-person connections and engagement as part of the NISS family. NISS's online presence continues to grow through high impact webinars, lecture series and career fairs that are archived on our YouTube channel: 110 and counting!

This year of transition began with the August 2022 announcement at JSM in Washington DC that David S. Matteson was named the incoming director of NISS – he initially took the formal title of Deputy Director, pending appointment of Director. David's vision calls for a coordinated network of NISS Hubs that serve as regional centers of activity at select NISS Affiliates, affiliate-wide themed semester programming, expanded post- and pre-doc training opportunities, and a new NISS Co-Laboratory for seeding new long-term collaborations. Sadly, our long-term CFO Alan Lee died suddenly in August 2022, which created a vacuum just as we were preparing for audit and applying for renewal contracts. However, the affiliate program continues to provide a diverse array of activities spanning academia, government, and industry with many sponsored activities reported in this annual report.

Research collaboration with the National Agricultural Statistical Services (NASS) continued to thrive with new innovations for gathering, analyzing, interpreting, and communicating statistical data from the Department of Agriculture. The contract from the National Center for Education Statistics (NCES) in the Department of Education ended as this year began. The enormous success of NISS researchers is documented by articles appearing in refereed journals, by presentations at national/international professional society meetings, and by agency awards for their contributions, as detailed in this report.

In October 2022, Nell Sedransk retired as the director of the DC office and Brian Habing assumed leadership in pursuing research support for educational activities funded through ESSIN for the past 10 years. NISS contributions culminated in a series of expert panels and directed research that provided the foundation for NCES to holistically reimagine the design process for its multiple surveys. The ESSIN program was replaced by PRESTO, which issued a request for proposal. NISS submitted a proposal in Fall 2022 and received approval in February 2022; NISS was selected by NCES as a preferred ID/IQ provider for category 1 RFQs. Now NISS awaits calls for proposals through this successor program (PRESTO). NISS expansion into data visualization, initially requested by NCES to support their reports and data publication, has launched other new projects and potential collaborations for NISS.

NISS emphasis on engagement with junior academic colleagues is reflected in an energetic Graduate Student Network (GSN) that grew out of the affiliate program and continues to organize activities focused on graduate study and experiences and career development. Successful virtual events on soft skills and career planning have been augmented by networking activities. The third annual NISS GSN Conference was held May 20-21, 2023. NISS Career Fairs, sponsored by GSN and by Affiliates Committees continue to be among the most widely attended virtual events. The 2022 Writing Workshop at JSM was partly virtual in August 2022.

The financial statements from 2022-23 highlight a difficult fiscal year, primarily due to the continued expenses committed during the previous fiscal year under the expectation that NCES funding would continue at its previous level without interruption. Unfortunately, the delay in PRESTO resulted in over a year with continuing expenses for affiliated research staff without income from NCES. However, as the year ended with a revised and balanced budget, NISS is now poised to transition to new challenges and exciting plans for the new year.

James L. Rosenberger, PhD
Director

David S. Matteson, PhD
Deputy Director

Table of Contents

North Carolina

The NISS Building in Research Triangle Park is currently partially leased. The original NISS building, and the upper floor of the new wing are occupied by Teledyne with a 10-year lease ending 31 Oct 2028. The NISS building is currently marketed through CBRE, Inc. (Raleigh).

On the Cover

L to R: Jerry Reiter (Duke), Jenny Thompson (US Census Bureau), Tom Krenzke (Westat), Amy O'Hara (Georgetown University), Jingchen Monika Hu (Vassar College), and Claire Bowen (Urban Institute). Bin Yu (University of California, Berkeley) & Cynthia Rudin (Duke University).

AWARDS

Jerome Sacks Award for Outstanding Cross-Disciplinary Research

XIHONG LIN

NISS is pleased to recognize Xihong Lin with the 2022 Jerome Sacks Award. Dr. Lin is Professor and former Chair of the Department of Biostatistics, Coordinating Director of the Program in Quantitative Genomics at the Harvard T. H. Chan School of Public Health, and Professor of the Department of Statistics at the Faculty of Arts and Sciences of Harvard University, and Associate Member of the Broad Institute of Harvard and MIT.


National Institute of Statistical Sciences
Research Triangle Park, NC
Jerome Sacks Award
for Cross-Disciplinary Research
Presented on August 8, 2022 by action of the Board of Trustees to
Xihong Lin
For her exceptional cross-disciplinary research and contributions to our profession, for leadership in developing methodology, including causal inference for complex clinical trials. She is an outstanding mentor, an influential leader, who has significantly advanced statistics, public health, and medicine.

Citation

"For her exceptional cross-disciplinary research and contributions to our profession, for leadership in developing methodology, including causal inference for complex clinical trials. She is an outstanding mentor, an influential leader, who has significantly advanced statistics, public health, and medicine."

Distinguished Service Award

JAMES G. BOOTH


National Institute of Statistical Sciences
Research Triangle Park, NC
Distinguished Service Award
Presented by action of the Board of Trustees on August 8, 2022 to
James G. Booth
Honoring Professor James (Jim) Booth for his dedicated service on the NISS board of trustees, serving three terms as the IMS Member of the Corporation, chairing the Awards and Nomination Committee, and since 2021 serving as chair of the board of trustees as an elected board member.

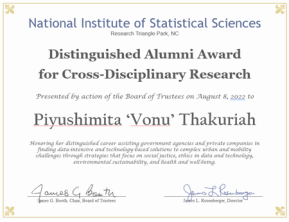
NISS is pleased to present the 2022 Distinguished Service Award to James Booth, Professor; Director of Graduate Studies PHD, Department of Statistics and Data Science at Cornell University, one of three departments in Computing and Information Science since 2004. Professor Booth is a Fellow of the American Statistical Association and has been an active member of the NISS Board of Trustees since 2012 and Chair since 2021.

Citation

"Honoring Professor James (Jim) Booth for his dedicated service on the NISS board of trustees, serving three terms as the IMS Member of the Corporation, chairing the Awards and Nomination Committee, and since 2021 serving as chair of the board of trustees as an elected board member."

Distinguished Alumni Award

PIYUSHIMITA 'VONU' THAKURIAH


National Institute of Statistical Sciences
Research Triangle Park, NC
Distinguished Alumni Award
for Cross-Disciplinary Research
Presented by action of the Board of Trustees on August 8, 2022 to
Piyushimita 'Vonu' Thakuriah
Honoring her distinguished career assisting government agencies and private companies in finding data-intensive and technology-based solutions to complex urban and mobility challenges through strategies that focus on social justice, ethics in data and technology, environmental sustainability, and health and well-being.

NISS is pleased to present the 2022 Distinguished Alumni Award to Piyushimita 'Vonu' Thakuriah, Distinguished Professor, Rutgers University-New Brunswick and Director, Rutgers Urban & Civic Informatics Lab. Vonu came to NISS as a post-doctoral fellow in 1994, becoming life-long friends with Jerry Sacks.

After her postdoc at NISS, Dr. Thakuriah found a home at The University of Illinois at Chicago (Professor), then moved to The University of Glasgow, United Kingdom serving as Director, UK ESRC. Urban Big Data Centre and Chair of Transport and Professor. Since 2018, she has expanded her brilliant career Rutgers University as professor and dean.

Citation

"Honoring her distinguished career assisting government agencies and private companies in finding data-intensive and technology-based solutions to complex urban and mobility challenges through strategies that focus on social justice, ethics in data and technology, environmental sustainability, and health and well-being."

RESEARCH

NISS RESEARCH PROGRAM OVERVIEW

United States Department of Agriculture
National Agricultural Statistics Service

NISS research responds to problems where data is the key to science and the foundation for evidence-based decision-making and policy. Its research results are theoretical, computational, and impactful.

In keeping with its roots in cross-disciplinary research, NISS continues to reinterpret its mission, best viewed now in 2023 in the context of data science. NISS maintains a three-fold active research program including: 1) independent research at NISS or via collaboration with university faculty & students, 2) as part of onsite or virtual joint research at federal agencies, and 3) as a neutral resource for technical expertise to address pressing issues at federal statistical agencies, and beyond.

During the past year, three full-time NISS researchers - Lu Chen, Luca Sartore, and Ruiyi Zhang - and Senior Research Fellow, Bruce Craig, have conducted research at the National Agricultural Statistics Service (NASS) within the US Department of Agriculture. NISS funded research with the National Center for Education Statistics (NCES) in the US Dept. of Education ended June 30, 2022, after 10 years of funding from ESSIN. NISS was successful in being selected as an ID/IQ contractor for new projects through the PRESTO program. NISS researchers Ya Mo, Haley Jeppson, and NISS Associate Director Brian Habing are poised to respond to new RFPs to continue research in educational areas where NISS has expertise.

What follows is a snapshot of NISS research activities over the past year.

RESEARCH AT USDA - NATIONAL AGRICULTURAL STATISTICS SERVICE

USDA National Agricultural Statistics Service

IMAGES: Integrated Modeling and Geospatial Estimation System

IMAGES is one of the top-ten priority projects of USDA (United States Department of Agriculture). The aim of the project is to develop statistical methodology leveraging machine learning algorithms, remote sensing data, administrative and survey data to produce early acreage estimates (and measures of uncertainty) for major crops in the speculative regions. However, due the adverse impact of anomalous rainfalls during the planting season, or other extreme weather phenomena, NASS (National Agricultural Statistics Service) is studying more advanced approaches to blend all available and useful data into a unique framework to accurately produce state-level weekly planted acreage estimates.

We studied several datasets have been studied to inform a field-level model for predicting the next crop rotation. We compared different models based on DANN (Deep Artificial Neural Networks) by looking at the average predictive performance computed on administrative data. We studied and improved on the use of the state-level PCDL (Predictive Cropland Data Layers) and related entropy layers in providing guidelines for data collection and imputation over 48 states in the CONUS (CONterminoUS United States). We studied different estimation models to integrate survey data, weather variables and geospatial information. Furthermore, we evaluated the coverage of Farm Service Agency (FSA) administrative data and developed a new raster data format to accelerate computational performances.

We have identified auxiliary information such as weather data, farm acres, and many others to improve model performance.

We find that a spatio-temporal model is better than the temporal model and generally provides better results than the survey alone.

We validated model performance internally and externally.

Through effective leadership and collaboration we guided teams in implementing these enhancements.

In addition, we actively worked and collaborated with colleagues within RDD as well as MD and SD to get insights and data for district-level modeling.

We investigated the June Area Survey record level data to generate benchmarked district level-data and variance estimates for the ASD level model.

The new estimates show great improvement over the proportional ones proposed in earlier simulation studies.

CASH RENTAL RATE PROJECT

Estimates of agricultural cash rental rates at different geographic levels (state-level and county-level) have taken on increased importance in recent years due to their application in rental agreement formulations, farm program administration and related activities. The USDA's NASS implemented a model-based approach for estimation of rental rates at the county level that uses two years (not necessarily consecutive) of data from its survey. The state-level small area models were also implemented in 2022.

We focused mainly on developing and implementing state-level and county-level cash rental rate estimates of interest in the Cash Rent project, working closely with the operation team, including MD/SD, to support this research. We authored R code for this project and applied it to generate all needed estimates in support of the 2022 production from June to August. We continue investigating additional methods to identify influential records in counties of concern.

CENSUS OF AGRICULTURE ESTIMATES AND VARIANCES

The aim of this project is to develop a new capture-recapture model for the Census of Agriculture (COA). The capture-recapture model is applied to estimate adjustment weights for Census farms. The new capture-recapture model includes parameters for the estimation of the probability of response. This is joint work with Grace Yoon and team lead by Habtamu Benecha.

CENSUS IMPUTATION TEAM (previously known As "Active Status 34")

The Census Mailing List (CML) consists of several agricultural operations that may or may not satisfy the definition of a farm. Records previously labeled as Active Status 34 (AS34) are small operations that may provide some information with the potential to better capture the diversity of US farmers as well as increase the overall response rate of the Census. NASS is hence studying classification models that target AS34 records and are more likely to be farms. These approaches, when combined with follow-up interviews, imputation and Dual-System Estimation (DSE), provide adequate information for also improving inference for underrepresented groups.

We developed Deep Neural Network and Support Vector Machine models, and tested them on the 2017 US Census of Agriculture data to better understand their performance when compared with other models (such as a generalized linear model with logistic link and tree-based models, like random forest). We developed a resampling-permutation technique to evaluate the accuracy and precision of the proposed models, and a new imputation methodology.

Furthermore, we documented guidelines for developing software able to operate in a secure environment along-side existing techniques developed in-house by NASS.

COUNTY-LEVEL VARIANCE SMOOTHING

Survey-based estimates are usually affected by a degree of variability caused by several factors (such as sampling and data processes); however, when sample sizes tend to be small, the variance assessment is less precise - leading to unrealistic results. To address this issue, we investigate different approaches to smooth the anomalous (either too small or too large) variances obtained at the county-level.

We developed a Bayesian method to smooth the anomalous variances obtained at the county-level. We also implemented alternative methods using yield data from crops county estimates. We published our results in "Stats" a leading open-access peer-reviewed Statistics journal.

CROPS COUNTY ESTIMATES

The US Department of Agriculture's (USDA) National Agricultural Statistics Services (NASS) publishes county-level estimates of planted acreage, harvested acreage, yield, and production annually. These official statistics are important to farmers, ranchers, and other federal agencies for planning and decision making. NASS transitioned from the traditional ASB (Agricultural Statistics Board) process to data-based modeling for the first time in 2020 (the 2020 production was from December 2020 to April 2021). NASS also conducts one-year revisions at the end of each following year. These modeling approaches are also implemented into one-year revisions beginning in 2020.

In 2022, we explored extending the harvested model for certain states with only corn-for-grain to model both corn-for-silage and corn-for-grain even though the county-level estimates for corn-for-silage are not required for the county estimates program. This was an E2 recommendation by an RFO that could potentially improve the county-level estimates of harvested acres. The team developed and tested code using 2019 and 2020 data as examples. In 2022 production, we successfully extended the corn-for-silage model in crops county estimates to other states based on the RFO suggestion and our crops county estimates team recommendation. This information helps the harvested model for corn-for-grain as well. We provided key technical support on extended development of the crop's county-level estimates program. In addition, to improving cotton modeling results, we coordinated with RFO staff and HQ staff for additional feedback. The resulting production first model works better for various states, and we provided additional modeled estimates for these during the 2022 production.

DUAL-SYSTEM ESTIMATION OF THE U.S. CENSUS OF AGRICULTURE

The United States Census of Agriculture is conducted every five years and provides a complete enumeration of farms and ranches in the US. However, under-coverage of the NASS list frame, nonresponse to the Census questionnaire, and misclassification of farms and non-farms occur in the Census, and each record receives a weight to adjust for these factors. The weights are estimated using a DSE (Dual-System Estimation) methodology. This method is based on standard logistic models that are separately estimated on a tiny percentage of the original data. This often results in unstable estimates that require additional labor to assure the quality of the estimates. We are studying a unified framework that allows the estimation of the model parameters using all available Census data by constraining the resulting DSE weights to be within reasonable ranges.

We suggested using an alternative likelihood that accounts for 1) coverage, 2) non-response and 3) farm-status classification. Thus, the likelihood contribution of each record combines the three logistic models allowing for a single estimation algorithm that uses all Census data. We proposed weights that better represent the Census data collection. We developed a model for using misclassification in the weighting system. We refined preliminary results, resulting in excellent performance compared to the existing DSE methodology.

FARM LABOR

USDA NASS conducts the Farm Labor Survey to produce estimates of the numbers of workers, duration of workweek, and wage rates for all workers. We incorporated new models for these into the official Farm Labor publication starting from May 2020.

We provided documented R routines to the Methodology Division, who execute all models during production. We continue to support production for 2022 November and 2023 May publication dates, with emphasis on model calibration. We presented this work to the calibration board and provided technical support related to the model during production. We wrote a manuscript detailing our new methodology and its application on this project and submitted it for peer-review.

INTEGER CALIBRATION FOR NASS SURVEYS

Integer Calibration (INCA) weighting was developed at NASS in collaboration with NISS. This approach was originally designed to operate with Census data and integer weights with specific properties. However, it was developed to operate in a very specific environment, but it soon became a necessity it operate more broadly in different environments. Integer Calibration (INCA) weighting was developed at NASS in collaboration with NISS. This approach was originally designed to operate with Census data and integer weights with specific properties and was developed to operate in a very specific environment. It was more broadly useful for addressing issues inherent in extreme weights that characterize select surveys, and was made publicly available for R (on CRAN as "inca"). This required the ability to operate in different environments. Changes in recent R versions required changing the package to continue working both outside and within the NASS environment.

We have considered introducing new classes of objective functions and penalties that have been unsuccessfully tested on 2012 US Census of Agriculture data because these so-called “unsuccessful” functions and penalties have been performing well on other surveys (e.g., the 2015 and 2020 Local Food Marketing Practices Data and the most recent Labor Surveys). The initial structure of a new R package is under development. We described this work at JSM 2023, and prepared a draft companion article to be submitted for peer-review.

JUNE AREA AUTOMATIC IMPUTATION

The June Area Survey (JAS) is among the most extensive surveys NASS conducts each year, encompassing the first two weeks of June. This survey is based on an area-frame sample where selected segments are inspected by enumerators that contact the farmers operating the land for in-person interviews. However, due to the COVID-19 pandemic, the JAS data collection mechanism has partially changed allowing for phone interviews, which increased the number of nonrespondents. Consequently, NASS personnel in the RFOs (Regional Field Offices) are required to impute more records manually. Therefore, to reduce the imputation efforts, a standardized procedure that automatically imputes acreage (by crop) needs to be developed and fully tested.

We investigated the potential uses of PCDL (Predictive Cropland Data Layers) and its entropy layer for automatic imputation of the JAS. We also consider CSBs (Crop Sequence Boundaries) and tree-based alternative models (using XGBoost regression) to predict acreage that was not reported. We further found that the arithmetic average between the acreages predicted using the PCDL and CSBs outperforms surveys when compared to administrative FSA (Farm Service Agency) data.

OUTLIER DETECTION AND MACHINE LEARNING

NASS datasets are often revised manually to assure which data items need further edits to improve the consistency of estimates provided to the ASB (Agricultural Statistics Board). To partially automate this process NASS has historically relied on rule-based methods to identify outliers. However, these approaches have not been fully successful in reducing manual labor. Thus, our ML (Machine Learning) group within NASS has started a series of competitions among several teams to collect and test a wide variety of ML methods to identify outliers.

As a result, we have improved the R-C interface of a fuzzy logic system that operates using robust approaches to identify tail, historical, and “relational” outliers, providing better anomaly scores that account for geographical stratification. This algorithm has been presented at two conferences and it has been detailed in a draft manuscript for a possible publication in a peer-reviewed journal: Classification of Open-Ended Comments from Multiple Surveys Based on XGBoost and Natural Language Processing”.

STATISTICAL DISCLOSURE LIMITATIONS

To prevent the disclosure of sensitive information NASS has adopted a cell-suppression approach for its US Census of Agriculture. However, new computational advances are potentially weakening the efficacy of cell-suppression techniques in preserving the privacy of Census respondents. Therefore, NASS has started to investigate the adoption of differential privacy methods or alternative methods to improve the protection of the respondent privacy.

We studied new algorithms for mixed-integer programming to efficiently identify which complementary cells to suppress. We also expanded our focus on newly developed tabular and unit-level methods to provide stochastic solutions for the privacy protection of Census respondents.

STATISTICAL DISCLOSURE LIMITATION PROJECT

NASS seeks to develop an innovative statistical disclosure method for the 2027 Census of Agriculture (COA). This is a cross-divisional project including team members from RDD, MD, CSD, FO West, FO East, SPBSD and SD. This project aims to explore three mainstream methods: (1) cell suppression, (2) tabular model, and (3) synthetic data and differential privacy. We lead the cell suppression development and are deeply involved with the other two.

We reviewed the NASS disclosure system and source code, and find that the heuristic algorithm commonly results in over-suppression and under-suppression. A new disclosure system with optimal/near-optimal solutions to the (mixed) integer programming that can eliminate under-suppression and reduce over-suppression would be desirable. We have explored various disclosure software and linear programming solvers for a new NASS disclosure system. The candidates include the R sdcTable package, tau-Argus, CPLEX, Gurobi and an open-source linear programming solver based on C++. We have run the NASS disclosure program on real Michigan State data as well as some simulation data to accurately quantify the over-suppression and under-suppression of the new program.

We have reviewed related the literature on cell suppression methodology, have recently presented an “Overview of the Cell Suppression Problem” at JSM 2023, and have a review paper in preparation. Our synthetic data and differential privacy and tabular model sub-teams joined a pilot study to produce synthetic data with Tumult.

STATISTICAL DISCLOSURE LIMITATION RESEARCH TEAM

We created a Statistical Disclosure Limitation (SDL) Research Team to organize and conduct systematic research on the disclosure methodologies based on the SDL recent seminars and group discussions. The initial focus includes disclosure methodologies such as suppression with general linear programming, controlled tabular adjustment (CTA), and a synthetic data approach. Other methods are under consideration by our team. Based on our initial analyses, we will make recommendations on an SDL approach to the NASS Business council and senior leadership team by 2025. The new SDL methodology method and system should be fully operational in 2026 for the 2027 Census of Agriculture.

USDA MACHINE LEARNING PILOT PROJECT

USDA has promoted an inter-agency effort to develop machine learning tools to address a set of research questions where AI (Artificial Intelligence) and ML (Machine Learning) models and methods can be applied. Among the several questions, AMS (Agricultural Marketing Service) presented a project where ML and AI models should be developed to identify adulterated honey. The aim of this project is to develop a new automated system that may have superior performance over the current approach based on thresholds.

Using data derived by the spectral signature of the chemical composition of the samples, we developed a set of models based on neural networks, a modified Mahalanobis distance, and other nonlinear methods. The results of these several methods can be combined to identify adulterated samples using a summary statistic indicating the propensity of having an adulterated sample.

USDA MACHINE LEARNING PILOT PROJECT

The USDA Artificial Intelligence Center of Excellence (AI-COE) organizes machine learning projects. One project aims to detect outliers in Child Nutrition data reported by Food and Nutrition Service (FNS). FNS receives financial claims data and data about program operations from states that operate the Child Nutrition Programs. FNS staff spend time researching what appear to be outliers to determine whether claims and data are legitimate. A machine learning model may feasibly utilize both internal and external data to present a more complete picture of what impacts the claims and data submissions for each state. Our objective is to provide a comprehensive method for identifying instances of reporting error in Child Nutrition administrative forms and data sets.

So far, we have led time series modeling and investigated how to apply the FuzzyHRT algorithm. We identify three types of outliers: historical anomalies, tail anomalies, and relational anomalies. Our algorithm provides output including a single anomaly score ([0,1]) for each data item and the outlier flag based on a user-defined contamination percentage. We lead biweekly meetings to provide model updates to the entire team and actively work with FNS researchers to calibrate models to further improve performance.

VARIANCE ESTIMATION OF THE U.S. CENSUS OF AGRICULTURE

As reported above, the US CoA (Census of Agriculture) is conducted every five years and provides a complete enumeration of the farms and ranches in the US. However, because Census records receive weights to account for under-coverage, nonresponse, and misclassification, estimated Census totals and medians are affected by some level of error. We previously applied a new resampling methodology for the 2017 US CoA that was designed to improve precision of the variances of Census totals. However, NASS internally computes several weighted medians, and a better methodology to assess the variability of such estimates is needed. To address variance estimation for both Census totals and medians, we are studying several approaches with the potential to result in a single methodology valid for both totals, medians, and other statistics.

Our proposed likelihood for the DSE (Dual-System-Estimation) has been revised to account for replication schemes (such as delete-a-group jackknife and bootstrap approaches). Furthermore, we found that delete-a-group jackknife has been providing more accurate results. However, to better account for the uncertainty due to calibration, a new program was developed for calibrating replicate weights in parallel. This program has shown a substantial reduction of the computational burdens within the current production environment.

IMPUTATIONS, DETERMINISTIC EDITS, AUTOMATION AND LOGIC (IDEAL) PROJECT

Our IDEAL project aims to modularize and automate the editing and imputation process for Blaise surveys in a generalized system to decrease staff time spent editing and to increase data quality. The NISS team works in a subgroup, led by Michael Laird, including Enrique Perez, Robert Emmet, Scott Matthews, Jennifer Maiwurm and Megan Lipke. Our focus is to test our phase one IDEAL product, Jimmy, which is a user interface developed to manage/share edit logic and track the status of imputed values.

Our continuing role is to help to improve the Jimmy system for auto-imputation by comparing its output and the manual imputation (ground truth) from the 2022 September Crops CE, to report the difference, and to investigate the potential reasons for the difference.

Zhang, R., Chen, L. , Cheng, Y., Jacobsen, M. *Overview of the Cell Suppression Problem*. (in progress)

Abernethy, J., Beeson, P., Boryan, C., Hunt, K., & **Sartore, L.** (2023). Preseason Crop Type Prediction Using Crop Sequence Boundaries. *Computers and Electronics in Agriculture*, 208, 107768. Available at: https://doi.org/10.1016/j.compag.2023.107768.

Chen, L., **Sartore, L.**, Benecha H, Bejleri V, Nandram B. (2022). Smoothing County-Level Sampling Variances to Improve Small Area Models' Outputs. *Stats*, 5(3):898-915. https://doi.org/10.3390/stats5030052.

Chen, L. and Nandram, B. (2022). Hierarchical Bayesian Model for County-Level Cash Rental Rates. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association, 1823-1839.

Chen, L., Cruze, N. B., and Young, L. J. (2022). Model-based Estimates for Farm Labor Quantities. *Stats*, 5(3), 738-754. https://doi.org/10.3390/stats5030043.

Chen, L., Nandram, B. and Cruze, N. B. (2022). Hierarchical Bayesian Model with Inequality Constraints for US County Estimates. *Journal of Official Statistics*, 38(3):709–732. https://doi.org/10.2478/jos-2022-0032.

Chen, L., Sartore, L., Benecha, H., Bejleri, V., & Nandram, B. (2022). Smoothing County-Level Sampling Variances to Improve Small Area Models' Outputs. *Stats*, 5(3), 898-915. Available at: https://www.mdpi.com/2571-905X/5/3/52/pdf.

Hyman, M., **Sartore, L.** and Young, L.J. (2022). Capture-Recapture Estimation of Characteristics of US Local Food Farms Using a Web-Scraped List Frame. *Journal of Survey Statistics and Methodology*, 10 (4), 979-1004. Available at: https://doi.org/10.1093/jssam/smab008.

Nandram, B., Cruze, N.B., Erciulescu, A.L. and **Chen, L.**. (2022). Bayesian Small Area Models under Inequality Constraints with Benchmarking and Double Shrinkage. *Research Report RDD-22-02*, National Agricultural Statistics Service, USDA. Available at: https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/ResearchReport_constraintmodel.pdf. Access date: July 2022.

Sartore, L., Boryan, C., Dau, A., & Willis, P. (2023). An Assessment of Crop-Specific Land Cover Predictions Using High-Order Markov Chains and Deep Neural Networks. *Journal of Data Science*, 21(2), 333-353. Available at: https://doi.org/10.6339/23-JDS1098.

Sartore, L., Boryan, C., Willis, P. (2022) Developing entropies of Predictive Cropland Data Layers for crop survey imputation. *IGARSS 2022-2022* IEEE International Geoscience and Remote Sensing Symposium, 1404–1407, Available at: https://doi.org/10.1109/IGARSS46834.2022.9884059.

Young, L. J. and **Chen, L**. (2022) Using Small Area Estimation to Produce Official Statistics. *Stats*, 5(3):881-897. https://doi.org/10.3390/stats5030051.

RESEARCH

PRESENTATIONS (PREVIOUS YEAR)

Chen, L. (2021, August). Model-based Estimates for Farm Labor Quantities. *2021 Joint Statistical Meetings*.

Chen, L. (2021, October). Model-based Estimates for Farm Labor Quantities. *2021 Women in Statistics and Data Science Conference.*

Sartore, L., Benecha, H. (2021, August). Estimation of Power Transformations in Capture-recapture Models. *2021 Joint Statistical Meetings*

Sartore, L., Boryan, C., Dau, A., Willis, P. (2022). Predicting Crop Specific Land Cover using Transition Probabilities, Deep and Quantum-inspired Neural Network Models. *2022 Statistics and Data Science Symposium*, Pittsburgh, PA.

PRESENTATIONS

Abernethy, J., **Sartore, L**., Hunt, K., Boryan, C. (2022). A Computationally Efficient Model for Large Scale Crop Type Forecasting (JSM) in Washington, DC. USA.

Benecha, H., **Sartore, L**., Yoon, G., Craig, B. A., Abreu, D. A., Young, L. J. (2022). Extending the Dual-System Estimation for the Census of Agriculture (JSM) in Washington, DC. USA.

Chen, L. and Nandram, B. (2022, October). Model for County-Level Cash Rental Rates. Presentation at the Women in Statistics and Data Science.

Chen, L. and Nandram, B. (2022 August). A Bayesian Small Area Model for State-level Cash Rental Rates (JSM) in Toronto. Canada.

Chen, L., Cruze, N. B., and Young, L.J. (2022 October). Model-based Estimates for Farm Labor Quantities. Presentation at 2022 FCSM.

Chen, L., **Sartore, L.**, Abernethy, J., Young, L.J. (2023 May). Integration of Survey and Non-Survey Data in the Estimation of Planted Acreage (ICAS IX) in Washington, DC. USA.

Cheng, Y., **Chen, L.**, Datta, G., and Nandram, B. (2023 August). Mixture Model and Its Application (JSM) in Toronto. Canada.

Murphy, T., Rosales, A., **Sartore, L.**, Abreu, D. A. (2022 August). Automatic Imputation for an Area Survey (JSM) in Washington, DC. USA.

Pallotta N., **Chen, L.**, Sarkar, B., Winterowd J., and Dau, A. (2023 May). Classification of Open-Ended Comments from Multiple Surveys Based on XGBoost and Natural Language Processing in the 78th AAPOR in Philadelphia, PA, USA.

Sartore, L., Benecha, H., **Chen, L.**, and Bejleri, V. (2022 August). Uncertainty Assessment of Finite-Population Medians Under Complex Sampling Designs. Presentation at the meeting of Joint Statistical Meetings, Washington D.C.

Sartore, L., Boryan, C., Dau, A., Willis, P. (2022). Predicting Crop-Specific Land Cover Using Higher Order Markov Chains (Virtual Workshop of the North American Tripartite Committee on Agricultural Statistics).

Sartore, L., **Chen, L.** (2023 August). Consistency of Survey Estimates through Adjusted Integer Weights (JSM) in Toronto. Canada.

Sartore, L., **Chen, L.**, van Wart, J., Dau, A., Bejleri, V. (2023 May). Fuzzy Detection of Cellwise Anomalies in Agricultural Surveys (ICAS IX) in Washington, DC. USA.

Zhang, R, Chen L., Cheng Y., and Jacobsen, M. (2023 August). Overview of Cell Suppression Methods (JSM) in Toronto. Canada.

Zhang, R. (2023 August). Overview of the Cell Suppression Problem (JSM) in Toronto. Canada.

USDA, Research and Development Division Achievement Award

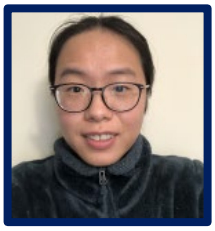
To **Ruiyi Zhang** for the contributions to NASS missions for the period of 2022-10-01 to 2023-02-03. (March 2023)

USDA Cash Rents Team Award

To **Lu Chen**, Nathan Cruz, Lori Harper, Michael Mathison, Balgobin Nandram, Jennifer Rhorer, **Luca Sartore**, Tenopra Shepphard for terrific teamwork and technical expertise in implementing new Cash Rents statistics with an updated outlier identification process and a new modeling process.

USDA Cash Rents Team Award

To **Lu Chen**, Nathan Cruz, Marisa Reuber, Jennifer Rhorer, **Luca Sartore**, Sterling Shrader for their tremendous efforts in implementing new and high-quality Farm Labor statistics in service to U.S. agriculture.



Ruiyi Zhang
Research Associate
NISS / NASS

NAGGLE - Internal Competition of the NASS Focus Group on Machine Learning and Artificial Intelligence (July 2022)

To **Lu Chen** & **Luca Sartore** for developing the best multivariate anomaly detection techniques to focus attention on reports that need further revisions.

USDA NASS Achievement Award (Feb. 2023 & June 2023)

To **Lu Chen for** implemented the state-level Bayesian models showing impressive accuracy and achieving outstanding outcomes.

Lu Chen
Research Associate
NISS / NASS

Cropscape for June Area Land Tool Award (Dec. 2022)

To Claire Boryan, Andrew Dau, Matthew Deaton, Michael Echtenkamp, Avery Nagle, Arthur Rosales, **Luca Sartore**, Patrick Willis for producing timely and useful data for the editing and imputation stages of the 2021 June Area Survey by imagining and building the June Area Land Tool.

USDA NASS Second Quarter Award (Feb. 2023)

To **Luca Sartore** for the contributions to the development of the new Census estimation methodology.

Luca Sartore
Sr. Research Associate
NISS / NASS

Farm Labor Calibration and Outlier Team Award (Dec. 2022)

To **Lu Chen**, Nathan Cruz, Marisa Reuber, Jennifer Rhorer, **Luca Sartore**, Sterling Shrader for tremendous efforts in implementing new and high-quality Farm Labor statistics in service to U.S. agriculture.

RESEARCH AT USDED - NATIONAL CENTER FOR EDUCATION STATISTICS

From ESSIN to PRESTO

Much of the work done on contract with the National Center for Education Statistics (NCES) is done through the companies and organizations that qualify for decade long ID/IQ (Indefinite Delivery/Indefinite Quantity) programs. For the past decade as part of the (ID/IQ) ESSIN program (Education Statistics Services Institute Network) at NCES, NISS has provided its expertise on tasks such as conducting statistical research on topics of interest to NCES and forming expert panels that examined technical issues facing NCES.

The final task for NISS on ESSIN was an examination of the issues surrounding accurately reporting, quantifying and analyzing decision-making for school participation in NCES surveys. The report for the "Expert Panel on School Survey Participation and Burden," held in May and June 2022, is available in the complete collection of technical panel reports in our NCES Report Library at NISS.org.

Following the expiration of the ESSIN program in the Department of Education, the fall of 2022 involved navigating the two-part proposal system to qualify for its replacement program, PRESTO (Procurement of Research Evaluation and Statistics Task). NISS was notified in February 2023 that it had been accepted as one of the organizations that qualified for PRESTO and is now eligible to bid on specific task orders as these are announced.

Effective Data Visualization for Education Data

Even while NISS's work for the National Center for Education Statistics (NCES) is in transition, NISS continued its work on visualizations and interactive graphics for education data and government reports – laying the groundwork to expand those types of offerings to other organizations. This included continuing to develop tools that can inform style guides and recommendations and include the latest statistical and visualization best practices, accessibility standards, and legal requirements. A particular focus was aimed on guidelines for visualizing uncertainty and context-dependent color scales for optimal scientific literacy. Two presentations (see below) were made at JSM 2022 and a training session on "Visualizations and Interactive Graphics using R" by **Haley Jeppson** and **Brian Habing** was accepted for the annual meeting of the National Council on Measurement in Education in April 2023.

Scholar: He**ike Hofmann**, Iowa State University; Research Associate: Hale**y Jeppson**, National Institute of Statistical Sciences

Appropriate Analysis of Secondary Data

The various large scale national educational surveys and exams by the National Center for Educational Statistics are intentionally made available for other researchers to analyze (so-called Secondary Data). The particular questions they raise include what new insights can be gained using new statistical and data science tools (such as tree-based methods) and what practical conclusions can be drawn about questions of substantive importance (the effect of writing prompt structure on bias against English language learners), and what needs to be done to use more classical methods with the data given (such as the appropriateness of different conditioning variables in searching for item bias/differential item functioning). The first of the three topics has now appeared in the *Journal of Data Science* (see below).

Scholars: **Ya Mo**, Boise State University and NISS) ▪ **Brian Habing**, National Institute of Statistical Sciences ▪ **Nell Sedransk**, National Institute of Statistical Sciences

Assessing and Adjusting for Non-Response: An Educational Achievement Case Study

Nonresponse bias is a widely prevalent problem for data collection. Exemplars are developed to guide nonresponse bias analysis (NRBA) in cross-sectional and longitudinal studies, respectively, with an application to the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011). Wave nonresponse with attrition yields a monotone missingness pattern, and the missingness mechanism can be missing at random (MAR) or missing not at random (MNAR). Weighting adjustments are effective for NRBA when constructed weights are correlated to whichever variable is of interest, whereas multiple imputation (MI) allows the inclusion of auxiliary predictive variables into the imputation model if it yields more efficient estimates. Adding offsets in MI results assesses MNAR deviations from MAR as a sensitivity analysis. Multilevel models with maximum likelihood estimation and marginal models estimated using generalized estimating equations can also handle incomplete longitudinal data. Bayesian methods introduce prior information and potentially improve model estimation. Application of these different methods to the ECLS-K:2011 data yielded only minor changes in substantive conclusions. The strength of NRBA evidence depends on the strength of the relationship between the variables in the adjustment and the key survey outcomes. Part of this work has now appeared in the *Journal of Educational and Behavioral Statistics* (see below). The second paper is under review.

Scholars: **Rod Little**, University of Michigan ▪ **Yajuan Si**, University of Michigan ▪ **Ya Mo**, Boise State University and NISS ▪ **Nell Sedransk**, National Institute of Statistical Sciences

PRESENTATIONS

Combining Visualization Best Practices with Agency Standards for Statistical Graphics in Government Reports

This roundtable discussed the challenges arising from simultaneously adhering to graphical standards for government reports and data visualization best practices. It centered on the issues of drafting a one-size-fits-all set of recommendations for style guides for statistical graphics. The main topic of interest was the proper navigation of 508-compliance and accessibility and the hurdles faced when developing a flexible yet concise set of color scales suitable for various data types.

Haley Jeppson, National Institute of Statistical Sciences; **Brian Habing**, National Institute of Statistical Sciences; **Heike Hofmann**, Iowa State University;

Student Insights into Graphical Innovation for Educational and Government Reports

This poster presented the winning entries from the "**NISS Statistically Accurate Interactive Displays ... in Graphics**!" student competition that was held in the first-half of 2022 and explored what it showed for current Educational and Government reports. The poster format allowed for discussion with the statistical team members of several government agencies.

Brian Habing, National Institute of Statistical Sciences; **Haley Jeppson**, National Institute of Statistical Sciences

Generalized Mosaic Plots in the ggplot2 Framework

Haley Jeppson and **Heike Hoffman**
The R Journal - https://journal.r-project.org/articles/RJ-2023-013/RJ-2023-013.pdf

Graphical methods for categorical variables are not well developed when compared with visualizations for numeric data. One method available for multidimensional categorical data visualizations is mosaic plots. Mosaic plots are an easy and powerful option for identifying relationships between multiple categorical variables. Although various packages have implemented mosaic plots, no implementation within the grammar of graphics supports mosaic plots. We present a new implementation of mosaic plots in R, ggmosaic, that implements a custom ggplot2 geom designed for generalized mosaic plots. Equipped with the functionality and flexibility of ggplot2, ggmosaic introduces new features not previously available for mosaic plots, including a novel method of incorporating a rendering of the underlying density via jittering. This paper provides an overview of the implementation and examples that highlight the versatility and ease of use of ggmosaic while demonstrating the practicality of mosaic plots.

Tree-Based Methods: A Tool for Modeling Nonlinear Complex Relationships and Generating New Insights from Data

Ya Mo, **Brian Habing**, and **Nell Sedransk**
Journal of Data Science 20 (30) 259-379

Our paper introduces tree-based methods, specifically classification and regression trees (CRT), to study student achievement. CRT allows data analysis to be driven by the data's internal structure. Thus, CRT can model complex nonlinear relationships and supplement traditional hypothesis-testing approaches to provide a fuller picture of the topic being studied. Using Early Childhood Longitudinal Study-Kindergarten 2011 data as a case study, our research investigated predictors from students' demographic backgrounds to ascertain their relationships to students' academic performance and achievement gains in reading and math. In our study, CRT displays complex patterns between predictors and outcomes; more specifically, the patterns illuminated by the regression trees differ across the subject areas (i.e., reading and math) and between the performance levels and achievement gains. Through the use of real-world assessment datasets, this article demonstrates the strengths and limitations of CRT when analyzing student achievement data as well as the challenges. When achievement data such as achievement gains in our case study are not linearly strongly related to any continuous predictors, regression trees may make more accurate predictions than general linear models and produce results that are easier to interpret. Our study illustrates scenarios when CRT on achievement data is most appropriate and beneficial.

A Case Study of Nonresponse Bias Analysis in Educational Assessment Surveys

Yajuan Si, **Roderick J.A. Little**, **Ya Mo**, and **Nell Sedransk**
Journal of Educational and Behavioral Statistics - https://doi.org/10.3102/10769986221141074

Nonresponse bias is a widely prevalent problem for data on education. We develop a ten-step exemplar to guide nonresponse bias analysis (NRBA) in cross-sectional studies and apply these steps to the Early Childhood Longitudinal Study, Kindergarten Class of 2010–2011. A key step is the construction of indices of nonresponse bias based on proxy pattern-mixture models for survey variables of interest. A novel feature is to characterize the strength of evidence about nonresponse bias contained in these indices, based on the strength of the relationship between the characteristics in the nonresponse adjustment and the key survey variables. Our NRBA improves the existing methods by incorporating both missing at random and missing not at random mechanisms, and all analyses can be done straightforwardly with standard statistical software.

RESEARCH

NCES Report Library

For the past decade, NISS has convened several panels each year of technical experts on high priority topics for the National Center for Education Statistics (NCES) within the Institute for Education Sciences (IES) at the US Department of Education. Following review of the panels' reports, these are included in the NCES Library on the NISS website where they are publicly available. See: https://www.niss.org/nces-report-library.

Topics have ranged from technical statistical issues such as survey design, non-response bias adjustment and data visualization to data issues such as release of keystroke process data and building data foundations for policy-making to quantitative methodology. The Library currently contains 38 panel reports; 23 of these panels met within the last decade.

NCES National Center for Education Statistics

HOME > NCES REPORT LIBRARY

NCES Report Library

Year	Month	Title	Document Type	Direct Links to Texts
2022	June	School Survey Participation and Burden	Expert Panel	Full Text Executive Summary
2022	May	Coordinating Designs for Multiple Surveys	Technical Forum	Full Text Executive Summary
2021	November	Connecting the Dots, I & II: Integrated Sampling Approach for Multiple Surveys	Expert Panel	Full Text Executive Summary
2021	March	Setting Priorities for Federal Data Access to Expand The Context for Education Data	Expert Panel	Full Text Executive Summary
2021	January	Innovative Graphics for NCES Online Reports	Expert Panel	Full Text Executive Summary Micromap Video
2020	December	Post COVID Surveys	Expert Panel	Full Text Executive Summary
2020	March	Release of Process Data to Researchers	Expert Panel	Full Text

Ingram Olkin Forum (IOF) Series

Statistical Serving Society (S^3) Forums

The forums' goals are to bring innovations and statistics to new research, accelerate development of innovative approaches, develop statistical action items to inform public policy and generate reliable evidence, and facilitate new collaborations between statisticians and stakeholders. During this forum we hope we achieved these goals in our breakout session, and plan to continue developing working groups after the webinar. You can see past forums and their outcomes for other examples on how we did this here: https://www.niss.org/ingram-olkin-forums.

Advancing Demographic Equity with Privacy Preserving Methodologies (Hybrid)

Attend

1/12-13/2023 — 60

Moderators: **Jingchen (Monika) Hu** (Vassar College) & **Claire Bowen** (Urban Institute)

Speakers: **James Rosenberger** (NISS), **Amy O'Hara** (Georgetown University), **Terrance Savitsky** (Bureau of Labor Statistics), **Ellen Galantucci** (Federal Maritime Commission), **Jeremy Seeman** (Penn State University), **Quentin Brummet** (NORC), **Daneil Susser** (Penn State University), **Joerg Drechsler** (Institute for Employment Research (IAB), Germany), **Christopher Dick** (DA Advisors / Georgetown University), **Daniel Goroff** (Sloan Foundation), **Joshua Snoke** (RAND), **John Abowd** (Census Bureau), **Victoria Bryant** (Internal Revenue Service), **Madeline Pickens** (Urban Institute), **Darius Singpurwalla** (National Center for Science and Engineering Statistics), **Matthew Williams** (RTI International), **Don Jang** (NORC), **Saki Kinney** (RTI International), **Darius McDaniel** (Drexel University), **Aaron Williams** (Urban Institute**), Jae Jun Lee** (Georgetown University) , **Tom Krenzke** (Westat), **Lisa Mirel** (Center for Disease Control and Prevention), **Crescent Martin** (Center for Disease Control and Prevention), **Sesa Slavkovic** (Penn State University), **Wendy Martinez** (Census Bureau) , **Jennifer Andre** (Urban Institute), **Rob McClelland** (Urban Institute), **Graham MacDonald** (Urban Institute), & **Leonard Burman** (Urban Institute).

Statistical Methods for Combatting Human Trafficking

Attend

3/15/2023 — 126

Forum Chair: **David Banks** (Duke University)

Speakers: **Margaret Henderson** (University of North Carolina at Chapel Hill), **Nancy Hagan, PhD** (North Carolina Human Trafficking Commission), **Tyler McCormick** (University of Washington), **Rowland Seymour** (University of Birmingham), **Daniel Manrique-Vallier** (Indiana University) & **Megan Price** (Human Rights Data Analysis Group).

DAVID BANKS
Duke University (Forum Chair)

MARGARET HENDERSON
University of North Carolina

NANCY E. HAGAN
North Carolina Human Trafficking Commission

TYLER MCCORMICK
University of Washington

ROWLAND SEYMOUR
University of Birmingham

BERNARD SILVERMAN
Oxford University

FOCUS on EARLY CAREER STATISTICIANS

NISS Graduate Student Network (GSN)

Graduate students at many of the NISS Academic Affiliates have come together to establish a lively network with programs designed to meet their education and career goals. The overall goal of the NISS Graduate Student Network is to create connections among graduate students from different academic institutions within the NISS Affiliates Program. During the last year students organized a diversity of well attended activities ranging from networking to career-oriented topics to technical presentations. Less formal activities include partnering to share graduate student experiences, informal peer mentoring, meetups, and social events.

National Institute of Statistical Sciences

Graduate Student Network

3rd NISS Graduate Student Network Research Conference, May 20-21, 2023

The GSN Research Conference was a multifaceted event, featuring an array of activities that aimed to foster academic excellence within the graduate student community. The centerpiece of the conference was the graduate student presentations, where students showcased their research findings and insights through either oral or poster presentations. These presentations were grouped into three categories: Original Research, Literature Research, and Literature Review.

In addition to the captivating student presentations, the conference offered a Scientific Writing Short Course, presented by **Don Richards** from the Department of Statistics at Penn State. This course aimed to equip attendees with invaluable insights into refining their scientific writing skills, enabling them to communicate their research effectively to a broader audience.

As a highlight of the event, two Invited Talks featured esteemed scholars from renowned institutions. Day 1 welcomed **Dan Jeske** from the Department of Statistics at UC Riverside on "*A Neutral Zone Classifier for Three Classes with an Application to Text Mining*", while Day 2 featured the brilliant **Xihong Lin** from Harvard University on her talk "*A Multi-Dimensional Integrative Scoring Framework for Predicting Functional Variants in the Human Genome*" Both speakers shared their expertise and knowledge, adding significant value to the conference.

One of the most awaited moments of the conference was the recognition and reward of exemplary graduate students. Four best presentation awardees were selected based on the content and quality of their presentations. The 1st and 2nd Tier winners received a prestigious prize of $150 USD each, while the 3rd Tier presentation winner was granted $100. Additionally, an outstanding poster award recipient also received $100. The awardees were not only celebrated during the conference but will also receive recognition at the NISS Reception and Awards Ceremony during JSM 2023 in Toronto, Canada, further amplifying the significance of their achievements.

1st Tier Winner - **Rashmi Ranjan Bhuyan**, University of Southern California
2nd Tier Winner 1 - **Durbadal Ghosh**, Florida State University
2nd Tier Winner 2 - **Pengfei Lyu**, Florida State University
Poster Winner - **Priyanjali Bukke**, George Mason University

NISS GRADUATE STUDENT NETWORK (GSN)

The goals of NISS GSN are to create connections among graduate students from different institutions and to devise programs to meet their contemporary needs. Within this network, activities are organized to help students tackle challenges of graduate programs and to help with planning their future careers. Among these challenges is securing grants, which has become integral to research and in some cases career advancement. Preparing a successful grant proposal was the topic of an online webinar organized and presented by the NISS Academic Affiliates Committee; the webinar brought together four accomplished senior statisticians who have extensive experience in both writing and reviewing grant proposals.

First NISS Graduate Student Network Meetup at JSM — Attend

8/7/2022 — **Sumantu Basu**, Assistant Professor & Faculty GSN Steering Committee Member, Cornell Univ. — ≈75

GSN Executive Committee Members:

Manqi Cai, University of Pittsburgh

Rebecca Kurtz-Garcia, UC Riverside

Hannah Waddel, Emory University

The NISS Graduate Student Network held its own reception at JSM this year and it is certain to become an annual event! The overall goal of the NISS Graduate Student Network is to create connections among graduate students from different academic institutions within the NISS Affiliates Program. Over the last year a diverse assortment of valued and well attended activities were organized to help students tackle challenges of graduate programs and help with their future careers. Speakers have been selected to share with students their work experience or internships they were involved in.

NISS Graduate Student Network Socials

11/4/2022 and 2/10/2023

GSN Socials allow graduate students to share their experiences and hear from other grads in our NISS Academic Affiliate departments to help form connections and build your network!

NISS Graduate Student Internship Opportunities Webinar (Online Webinar) — Attend

11/16/2022 — 150+

Rob Baker, Procter & Gamble

Qing Ji, Procter & Gamble

Jianchang Lin, Takeda Pharmaceuticals

Melinda Thielbar, Research Triangle Analysts, Fidelity Investments

Grayson Dill, Bank of America

Peter Henstock, Pfizer

Nathan Lally, Hartford Steam Boiler/Munich Re Group

Moderator: **Manqi Cai**, PhD student, Biostatistics, University of Pittsburgh

Panelists shared information about internship opportunities from their industries and also discuss the job description/qualifications/ requirements for these types of positions. In addition, this event was organized such that it occurs before the deadlines for internship applications from these companies.

NISS-GSN Short Course - QUARTO — Attend

4/7/2023 — **Mine Çetinkaya-Rundel**, Duke University — 55

Quarto is the next generation of R Markdown for publishing, including dynamic and static documents and multi-lingual programming language support. With Quarto you can create documents, books, presentations, blogs or other online resources. This workshop introduced authoring Quarto documents using RStudio Visual Editor, which provides a user interface for editing all of Pandoc markdown. You also got a brief look into different formats (such as slides), publishing Quarto documents to QuartoPub, and writing scientific manuscripts with Quarto.

FOCUS ON EARLY CAREER STATISTICIANS cont.

NISS Events - Career Fairs

Academic Career Fair		Attended
10/12/2022	Speakers: **Steve MacEachern**, Ohio State University **Bhramar Mukherjee**, University of Michigan **Sujit Ghosh**, North Carolina State University	Moderator: **Sharmistha Guha**, Texas A&M	106

Government Career Fair		Attended
12/14/2022	Speakers: **Cha-Chi Fan**, Bureau of Transportation Statistics **Ashley Remik**, Energy Information Administration **Stephen Nalley**, Energy Information Administration **Jonah L. Wong**, Census Bureau	Moderator: **Danny Ying**, Master of Applied Statistics student, UCLA	82

Industry Career Fair		Attended
1/18/2023	Speakers: **Jonathan Legare**, Fidelity Investments **Tim Hesterberg**, Instacart **Madhumita (Bonnie) Ghosh-Dastidar**, RAND Corporation	Moderator: **Hang Nguyen**, SMU	121

'Careers Beyond Tech' Career Fair		Attended
3/22/2023	Speakers: **Rebecca Ferrell**, MS, Starbucks Seattle **Daniel Jeske**, PhD, University of California, Riverside **Thanh Nguyen**, MS, Moffit Cancer Center **Jim Harmon**, PhD, Liberty Mutual	Moderator: **Wei Vivian Li**, UC Riverside	80

Biopharmaceutical Industry Career Fair		Attended
5/10/2023	Speakers: **José Pinheiro**, Johnson & Johnson **Rebbecca Wilson**, Janssen Pharmaceutical Companies **Yu Du,** Eli Lilly and Company **Yenny Webb Vargas**, Genentech	Moderator: **Christian Geneus**, Procter & Gamble	113

EVENTS

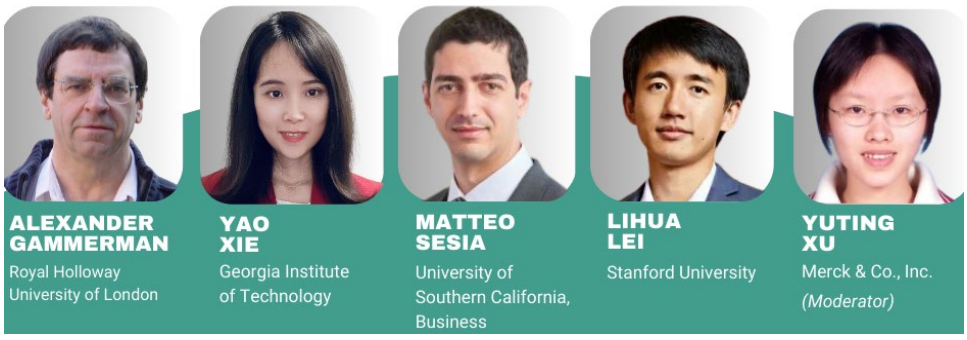
NISS-Merck Meet-Ups

NISS/Merck Meetup: Interpretable/Explainable Machine Learning			Attended
9/28/2022	Speakers: **Cynthia Rudin**, Duke University *Do Simpler Machine Learning Models Exist, and How Can We Find Them?* **Bin Yu**, University of California, Berkeley *Interpreting deep neural networks towards trustworthiness*	Moderator: **Junshui Ma**, Merck	329

Interpretable/explainable Machine Learning (ML) is an emerging field to address the black-box nature of complex models obtained by many popular ML methods. Its success can remove one of the major obstacles that prevent ML from having more impact on areas (i.e., healthcare), where human understanding of how a data-driven model works is crucial for many reasons. In this meetup, two leading researchers shared their efforts, contributions, and visions about this rapidly developing field.

NISS/Merck Meetup: Conformal Inference: Advancing the Boundaries of Machine Learning			Attended
4/19/2023	Speakers: **Alexander Gammerman**, Royal Holloway University of London **Yao Xie**, Georgia Institute of Technology **Matteo Sesia**, University of Southern California, Marshall School of Business **Lihua Lei**, Stanford University	Moderator: **Yuting Xu**, Merck & Co.	217

Conformal inference is a statistical framework that provides a way to make predictions with prediction intervals and associated confidence levels in machine learning. Unlike traditional point estimates, this approach constructs intervals for individual data points, allowing for more robust and reliable predictions. The confidence levels provided by the prediction intervals enable users to assess the uncertainty of the predictions and make informed decisions based on the level of confidence. Notably, prediction intervals are produced without assumption on the prediction algorithms and data distribution, making conformal inference a flexible and powerful tool in machine learning. Conformal inference is used for both classification and regression and is currently an active area of research in statistics and computer science. This framework has the potential to advance the boundaries of machine learning by enabling more accurate and trustworthy predictions with applications in a variety of fields, including finance, medicine, and environmental science.



NISS Panels

Integrating Statistics & Data Science Environments in Academic Departments	Attend
11/30/2022 **Heike Hofmann**, Iowa State University **Funda Güneş**, Duke University **Rachel Levy (Ray)**, North Carolina State University Data Science Academy **Chong Ho Yu (Alex)**, Azusa Pacific University Moderator: **Daniel Jeske**, University of California, Riverside	167

Should Statistics Departments in academic institutions integrate or change their title to Data Science? This panel aimed to reset the conversation of how statistics and data science might co-exist in academic departments. Key topics discussed: Industry trends, integration applications, student enrollment, changing of department name, and focus tracks implementing both Data Science and Statistics.

Panel on Navigating Massive Layoffs, Hiring Freezes, and Tech Industry Careers in 2023	Attend
2/15/2023 **Bernard Armah**, Wells Fargo **Shuonan (Shannon) Dong**, UW Boeing Advanced Research Center **Martin Tingley**, Netflix **Dhanushi Wijeyakulasuriya**, Microsoft Moderator: **Lingzhou Xue**, Penn State/NISS	106

Many students studying statistics and data science degrees dream to enter the kind of jobs that the tech industries provide, is this still possible given the drawbacks of positions in recent months? As an early-career statistician and data scientist ready for finding your next position, how can you secure your position in this industry? What opportunities currently exist for those job searching right now?

Panel on the Role of Biostatistics in an Increasingly Big Data/Data Science World	Attend
5/3/2023 **Xihong Lin**, Harvard University **Lance Waller**, Emory University **Jeff Goldsmith**, Columbia University **Yu Shyr**, Vanderbilt University Moderator: **David Benkeser**, Emory University	157

What is the role of biostatistics in an increasingly big data/data science focused world. There is so much attention these days on machine learning and AI in biomedical research. What are the opportunities for synergies between biostatistics and other data science disciplines?

COPSS-NISS Leadership Webinar Series

The COPSS-NISS Leadership Webinar Series is co-organized by the Committee of the Presidents of Statistical Societies (COPSS) Emerging Leaders in Statistics and the National Institute of Statistical Sciences (NISS). The purpose of the webinar series is to promote leadership skills for members of the statistical societies at any stage in their careers. The series features conversations with leaders throughout the discipline, including leaders from major academic and government institutions, and companies. Invited speakers share their leadership stories and answer questions about their experiences. Each webinar is moderated by a member of the COPSS Emerging Leaders in Statistics program.

Leadership in Academia		Attend
1/24/2023	**Sally Morton**, Arizona State University **Ron Brookmeyer**, University of California, Los Angeles Moderator: **Natalie Dean**, Emory University	105

Join us for our inaugural webinar on the topic of leadership in academia. Hear from statisticians leading schools and universities, strategizing about the future of education and research.

Leadership in Social Justice & Community Leadership		Attend
2/8/2023	**Mary Gray**, American University **Melody Goodman**, New York University **Natalie Dean**, Emory University **Patrick Ball**, Human Rights Data Analysis Group Moderator: **Kristian Lum**, University of Chicago	64

"I think of myself more as a generalist, I do a lot of community engaged, community-based and participatory research and I'm interested in letting those communities drive the research agenda so that my work is focused on questions that are important to people that that I can help in ways as a statistician." - Melody Goodman

Leadership in Statistical Research		Attend
3/28/2023	**Susan Murphy**, Harvard University **Trevor Hastie**, Stanford University Moderator: **Lingzhou Xue**, Penn State/NISS	206

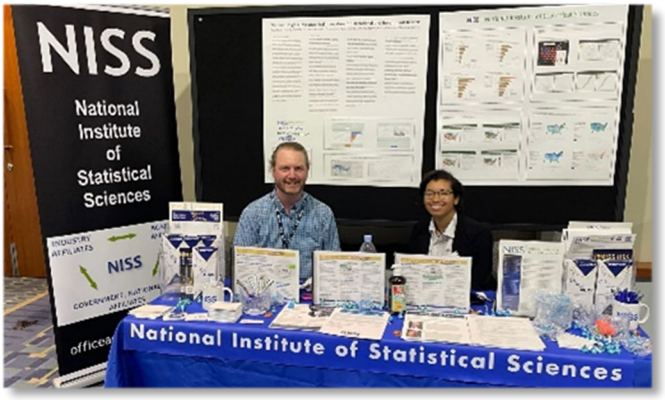
The webinar was an engaging and thought-provoking event that delved into the realm of leadership in statistical research.

Leadership in Government		Attend
4/28/2023	**Denice Ross**, U.S. Chief Data Scientist, White House Office of Science and Technology Policy **Robert L. Santos**, Director, U.S. Census Bureau Moderator: **Claire Bowen**, The Urban Institute	87

We were privileged to hear from two of the leading statisticians who strategized on the future of statistics in government, sharing their leadership journeys and providing their insights in discussions.

On Linkedin, this topped NISS' highest generated impressions ever at 1,982!

NISS



SUNDAY, AUGUST 7

- **NISSearchers** - Day 3
- **NISS Affiliate Luncheon**
- **NISS Graduate Student Network** & **NISS Writing Workshop for Junior Researchers Meet-Up** from 2020, 2021 & 2022!

MONDAY, AUGUST 8

- **Time Series Methods and Applications**, "*A Computationally Efficient Model for Large Scale Crop Type Forecasting*". J. Abernethy, USDA-NASS; **Luca Sartore**, **NISS**/USDA-NASS; K. Hunt, USDA-NASS; C. Boryan, USDA-NASS.
- **NISS Reception** with presentations for the 2022 **Jerome Sacks Award for Outstanding Cross-Disciplinary Research**, **NISS Distinguished Service Award** and the **NISS Distinguished Alumni Award**.

TUESDAY, AUGUST 9

- **Government Statistics Section A.M. Roundtable Discussion**, "*Combining Visualization Best Practices with Agency Standards for Statistical Graphics in Government Reports*", **Haley Jeppson**, **NISS**; **Brian Habing**, **NISS**; **Heike Hofmann**, Iowa State Univ.
- **Cross-Disciplinary Research on Health Data Science — Invited Papers**, **Lingzhou Xue**, Organizer/Chair & **James L. Rosenberger**, Discussant (**NISS Associate Director & Director**/PSU)
- **Statistical Methods Under Preferential and Informative Sampling**
- "*Spatial Kriging in the Presence of Informative Sampling Designs*", **Erin Schliep**, Univ. of Missouri; **Christopher K. Wikle**, Univ. of Missouri; R. Daw, Univ. of Missouri.
- "*An Empirical Evaluation of Alternative Approaches to Adjusting for Attrition When Analyzing Longitudinal Survey Data*", **Yajuan Si**, Univ. of Michigan.
- **Weighting and Variance Estimation in Complex Samples**, "*Uncertainty Assessment of Finite-Population Medians Under Complex Sampling Designs*". **Luca Sartore**, **NISS**/USDA-NASS; H. Benecha, USDA-NASS; V. Bejleri, USDA-NASS; **Lu Chen**, **NISS**/USDA-NASS.
- **Alignment, Accuracy, Precision: Comparing and Combining Data from Multiple Sources**, "*A Bayesian Small Area Model for County-Level Cash Rental Rates*". **Lu Chen**, **NISS**/USDA-NASS; B. Nandram, Worcester Polytechnic Institute.
- **Contributed Poster Presentations: Government Statistics Section**, "*Student Insights into Graphical Innovation for Educational and Government Reports*", **Brian Habing**, **NISS**; **Haley Jeppson**, **NISS**.

WEDNESDAY, AUGUST 10

- **Social Statistics Section A.M. Roundtable Discussion**. "*Gun Violence Research Designs Needed to Combat Myths About Police Shootings and Safety"*, **James Landis Rosenberger**, N**ISS Director** & Penn State.
- **Causal Estimand in Clinical Trials**, "*Estimands & Their Estimators for Clinical Trials Impacted by the COVID-19 Pandemic*", K. Van Lancker, Johns Hopkins, BSPH and Ghent Univ., Belgium; S. Tarima, Medical College of Wisconsin; J. Bartlett, Univ. of Bath; M. Bauer, Keck School of Medicine, Univ. of Southern California; B. Bharani-Dharan, Novartis Pharmaceuticals; F. Bretz, Novartis; N. Flournoy, Univ. of Missouri; H. Michiels, Ghent Univ.; C. Olarte Parra, Univ. of Bath; **James L. Rosenberger, NISS Director** & Penn State; S. Cro, Clinical Trials Unit, Imperial College London.
- **Record Linkage and Auxiliary Data Sources**
- "*Automatic Imputation for an Area Survey"*, T. Murphy, USDA-NASS; A. Rosales, USDA-NASS; **Luca Sartore**, **NISS**/USDA-NASS; D. A. Abreu, USDA-NASS.
- "*Extending the Dual-System Estimation for the Census of Agriculture*", H. Benecha, USDA-NASS; **Luca Sartore**, **NISS**/USDA-NASS; G. Yoon, USDA-NASS; **Bruce A. Craig**, Purdue Univ.; D. A. Abreu, USDA-NASS; L. J. Young, USDA-NASS.**S Writing Workshop for Junior Re**

NISS Affiliate Luncheon and Reception at JSM

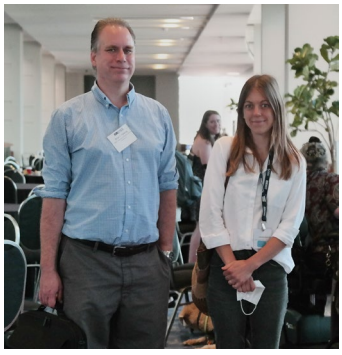
Member liaisons of our NISS Affiliate institutions gather together for a luncheon each year at JSM. The luncheon this year gave our members a chance to get together to reconnect, converse, plan, and get to know new people. Activities that our affiliates are involved in were shared, especially those highlighting the cross-disciplinary nature of challenges involved and the benefits that the synergy of statisticians from industry, governmental institutions and academia bring. Andy White from the National Center for Education Statistics (NCES) and Linda Young from the National Agricultural Statistical Services (NASS) spoke about what their organizations are involved in and recent research they are involved in. Others also were given the opportunity to introduce themselves and their affiliation with NISS. William Brenneman from Procter & Gamble was thanked for his support of the NISS GSN Network, and he volunteered to provide support again for the GSN research conference next year. Nancy Flournoy gave an update of the various Ingram Olkin Forums that are underway as well as a call for proposals on new topics for future Forums.

Jerome Sacks Award for Outstanding Cross-Disciplinary Research winner, **Dr. Xihong Lin**.

NISS Distinguished Alumni Award winner, **Dr. Piyushimita 'Vonu' Thakuriah**.

Jim introduces the crowd to **David S. Matteson**, the new NISS Director. **Jerry Reiter** honors Jim's work as Director.

NISS Around JSM: A.M. Roundtable Discussion



Brian Habing & Haley Jeppson

Jim Rosenberger

A.M. Roundtable Discussion

NISS Annual Writing Workshop

On July 29, August 5 and 7, 2022, the **National Institute of Statistical Sciences** once again held the annual **NISS Writing Workshop for Junior Researchers**. A successful workshop since its inception in 2007, this year's workshop was no different! NISS gathered a host of senior authors , editors, grant writers/ reviewers to share their advice and experiences. Each participant in the group of 30 junior faculty and early career researchers worked one-on-one with a senior mentor to critique and improve their manuscript.

2022 NISS Writing Workshop for Junior Researchers (hybrid)	Attend
July 29, August 5 and 7, 2022	30

In 2020 and again in 2021 NISS scrambled, prepared for and continue to maintain very successful workshops despite the pandemic wreaking havoc with everyone's ability to gather and converse. This year, as the American Statistical Association decided to hold the Joint Statistical Meetings in-person in Washington, D.C. NISS decided to take a hybrid approach splitting the conference, as usual, into three days but holding the first two days weeks in advance before coming together for the last day at JSM. By maintaining an online aspect to the workshop NISS was able to reach out to junior researchers who might not have been able to travel to JSM, and especially reach out to those in foreign countries. Once again, the response from senior individuals to take time out their schedule to talk with attendees about their writing was overwhelming. Below is the generalized agenda:

Tutorial 1: Writing Well as an Act of Kindness for Our Students and Colleagues
Tutorial 2: Choosing Where to Publish
Panel 1: Statistics and Data Science Journals
Tutorial 3: Reviewing and Revising
Mentor and Mentee one-on-one meetings in Breakout Rooms
Tutorial 4: How to Write a Collaborative Paper
Panel 2: Grant Writing
Tutorial 5: Ethical Issues and Reproducibility
Happy Hour, hang out, swap email and contact information
Tutorial 6: How to Get Started and Avoid Writer's Block
Panel 3: Speaking from Experiences and Career Development
Discussion, Evaluation and Certificates
Networking Event including virtual Writing Workshop attendees from 2020 & 2021!

NISS says "THANK YOU!" to Mentors & Speakers

Jim Albert, Bowling Green State University
Naomi Altman, Penn State University
Angela Dean, Ohio State University
Cheryl Eavey, SBE/MMS, National Science Foundation
Susan Ellenberg, University of PA Perelman School of Medicine
Xuming He, University of Michigan
Gabriel Huerta, Sandia National Laboratories
Peter Imrey, Cleveland Clinic
Nicholas Jewell, UC Berkeley; London School of Hygiene & Tropical Medicine
Jiming Jiang, University of California, Davis
Galin Jones, University of Minnesota
Karen Kafadar, University of Virginia
Nicole Lazar, Pennsylvania State University
Shuangge Steven Ma, Yale School of Public Health
Regina Nuzzo, Gallaudet University
Edsel Pena, National Science Foundation; University of South Carolina
Donald Richards, Pennsylvania State University
David Rocke, University of California, Davis
James Rosenberger, Director, National Institute of Statistical Sciences
William Rosenberger, George Mason University
Ali Shojaie, University of Washington
Judy Wang, George Washington University
Lingzhou Xue, Penn State University; National Institute of Statistical Sciences

SOCIAL PRESENCE

NISS by the Numbers

NISS Affiliates, through the Affiliates Committee planned 2 Graduate Student Network events, 1 meet-up, 1 panel, 1 webinar, and 1 workshop, and 2 virtual career fairs hosted by NISS. NISS also co-sponsored 9 events hosted by NISS Affiliates. The reach of NISS grows as the NISS Parameters Newsletter & Affiliates Update grows to boast an over 8,000 person circulation! NISS Social Media numbers are:

LinkedIn	Twitter	Facebook	YouTube
83 Posts	3696 Followers	44 Posts	17 New Videos (114)
2200 + Followers	98 New followers	754 Friends	18,288 Views
≈ 35.5K Views	1,761 Engagements	≈ 425 Page Visits	387 Subscribers
≈ 400 Views / Post	93 Total Tweets	3000 + Followers	

Career Fairs → Academic 6 → Industry 9 → Government 12 → Internship 1

On the NISS HTTPS://WWW.**YOUTUBE**.COM/C/NISSCOMMUNICATIONS Channel

- 6 COPSS-NISS COVID-19 Data Science Webinar Series
- 27 NISS Academic and Research Webinars
- 14 NISS-Merck Meet-Ups
- 23 Ingram Olkin Forum: *Statistics Serving Society* videos
- 8 Data Science for Business

NISS Job Announcements & NISS Careers web pages (https://www.niss.org/careers)

Some NISS Academic and Research Webinars, NISS-Merck Meet-ups, NISS Virtual Career Fairs as well as many other session videos that NISS has hosted or sponsored, can also be found here:

HTTPS://WWW.NISS.ORG/MEET-RECORDINGS

Sponsorships

11/10/2022 Data Science Conference

Tian Zheng, Columbia University; **Mike West**, Duke University; **Jun Liu**, Harvard University; **Donald B. Rubin**, Harvard University; **René Vidal**, Johns Hopkins University; **Adji Bousso Dieng**, Princeton University; **Rong Chen**, Rutgers University; **Edoardo M. Airoldi**, Temple University; **Ron Anderson**, Temple University; **Alan F. Karr**, Temple University; **Ken McAlinn**, Temple University; **Ed George**, University of Pennsylvania; **Chris Harvey**, Wells Fargo.

11/9/2022 Myles Hollander Distinguished Lecture

Trevor Hastie, Stanford University

10/28/2022 Georgia Statistics Day (UGA)

Nilanjan Chatterjee, Johns Hopkins University
Liang Liu, University of Georgia
Benjamin Risk, Emory University
Yao Xie, Georgia Institute of Technology

10/11/2022 1st International Day of Women in Statistics and Data Science

Norma Bargary, University of Limerick, Ireland; **Rashida Bokhari**, Federal Bureau of Statistics, Islamabad; **Saleha Habibullah**, Federal Bureau of Statistics, Islamabad; **Katie Buchhorn**, Queensland University of Technology, Australia; **Miguel de Carvalho**, University of Edinburgh, UK; **Isabella Deutsch**, University of Edinburgh, UK; **Upasana Roy Chowdhury**, Deloitte India; **Gerda Claeskens**, OR-STAT and Leuven Statistics Research Center, Belgium; **Rosa M. Crujeiras**, University of Santiago de Compostela, Spain; **Nairanjana Dasgupta**, Washington State University; **Josée Dupuis**, McGill University, CA; **Alexandra Schmidt**, McGill University, CA; **Katherine Ensor**, Rice University; **Ivette Gomes**, Centre for Statistics and Applications, UL; **Claire Cameron**, University of Otago, New Zealand; **Gabrielle Davie**, University of Otago, New Zealand; **Jill Haszard**, University of Otago, New Zealand; **Ella Iosua**, University of Otago, New Zealand; **Alice Kim**, University of Otago, New Zealand; **Ritika Jain**, Centre for Development Studies, Trivandrum; **Sayantee Jana**, Indian Institute of Technology, India; **Shuvashree Mondal**, Indian Institute of Technology, India; **Bei Jiang**, University of Alberta, CA; **Jessica Kohlschmidt**, Caucus for Women in Statistics/Ohio State University; **Eun-Kyung Lee**, Ewha Womens University, Korea; **Man-Suk Oh**, Ewha Womens University, Korea; **Tarim Lee**, Korea National Open University, Korea; **Shili Lin**, Ohio State University; **Sarah Lotspeich**, Wake Forest University; **Vanda Lourenco**, NOVA University of Lisbon, Portugal; **Sumithra Mandrekar**, Mayo Clinic; **Maya Mathur**, Stanford University; **Aya Mitani**, University of Toronto, CA; **Paula Moraga**, King Abdullah University of Science and Technology, Saudi Arabia; **Tomi Mori**, St. Jude Children's Research Hospital; **Arzu Onar-Thomas**, St. Jude Children's Research Hospital; **Kerrie Nelson**, Boston University; **Megan Othus**, Fred Hutchinson Cancer Center; **Sohee Park**, Yonsei University, Korea; **Micaela Parker**, Academic Data Science Alliance; **Kendra Plourde**, Yale University; **Federica Zoe Ricci**, University of California, Irvine; **Nokuthaba Sibanda**, Victoria University of Wellington, New Zealand; **Haritini Tsangari**, University of Nicosia, Cyprus.

Sponsorships

10/7-9/2022 International Conference on Advances in Interdisciplinary Statistics and Combinatorics (AISC 2022)

Wendy Martinez, Bureau of Labor Statistics
Srinivasa Varadhan, NYU, Courant Institute
Richard Davis, Columbia University
Bobby Gramacy, Virginia Tech
Eiren Jacobson, University of St Andrews, UK
Karen Kafadar, University of Virginia
Regina Liu, Rutgers University
Jerry Reiter, Duke University

10/6-8/2022 Women in Statistics and Data Science (WSDS 2022)

Laura Wendelberger, North Carolina State University
Barbara Wendelberger, Berry Consultants
Joanne Wendelberger, Los Alamos National Laboratory
Cynthia Phillips, Sandia National Laboratories
Kristin Potter, National Renewable Energy Laboratory
Kary Myers, Los Alamos National Laboratory
Rasitha Jayasekare, Butler University
Madeline Neely, Butler University
Margaret Betz, Purdue University

9/2,7,16,23/2022 Distinguished Theme Seminar Series 2022: Recent Advances in Statistical Inference

Michael I. Jordan, University of California, Berkeley
Emmanuel Candes, Stanford University
Xiao-Li Meng, Harvard University
Rina Barber, University of Chicago

9/19-20/2022 Second Penn Conference on Big Data in Biomedical and Population Health Sciences

Joseph Hogan, Brown University; **Yuanjia Wang**, Columbia University; **Linda Valeri**, Columbia University; **Ying Guo**, Emory University; **Tianxi Cai**, Harvard University; **Eimear Kenny**, Ichan School of Medicine at Mount Sinai; **Genevieve Stein-O'Brien**, Johns Hopkins University; **Colin Begg**, Memorial Sloan Kettering Cancer Center; **Jordan Bisanz**, Penn State University; **Kari North**, University of North Carolina, Chapel Hill; **Hongzhe Li**, University of Pennsylvania; **Enrique Schisterman**, University of Pennsylvania; **Jinbo Chen**, University of Pennsylvania; **Ravi Parikh**, University of Pennsylvania; **Jing Huang**, University of Pennsylvania; **Sarah Tishkoff**, University of Pennsylvania; **Mingyao Li**, University of Pennsylvania; **Taki Shinohara**, University of Pennsylvania; **Despina Kontos**, University of Pennsylvania; **Qi Long**, University of Pennsylvania; **Rebecca Hubbard**, University of Pennsylvania; **Nandita Mitra**, University of Pennsylvania; **Eric J. Tchetgen Tchetgen**, University of Pennsylvania; **Yu Shen**, University of Texas MD Anderson Cancer Center, Houston; **Yang Xie**, UT Southwestern; **Eleanor Pullenayegum**, University of Toronto; **Christina Kendziorski**, University of Wisconsin at Madison; **Nico Dosenbath**, Washington University; **Eran Segal**, Weizmann Institute of Science.

Sponsorships

9/1/2022 2022 Joint UGA-Clemson Seminar at UGA
Dennis K. J. Lin, Purdue University

1/9-11/23 International Conference on Health Policy Statistics: Upgrading the Pipeline from Health Data to Health Policy
Sherry Glied, New York University
Ziad Obermeyer, University of California, Berkeley
Arlene Ash, University of Massachusetts Chan Medical School
J. Chris Carey, University of Oklahoma
David Etzioni, Mayo Clinic in Arizona
Amanda Jean Stevenson, US Census Bureau

2/3/2023 Annual Lange Symposium: Genomics and Computational Statistics
Genevera Allen, Rice University
David Hunter, Penn State University
Oscar Padilla, UCLA
Marc Suchard, UCLA
Jason Xu, Duke University

4/3/2023 Annual Clemson-UGA Joint Colloquium #74
Michael R. Kosorok, University of North Carolina at Chapel Hill

4/13-14/23 UCLA Synthetic Data Workshop
Aloni Cohen, UChicago Computer Science; **Xiao-Li Meng**, Harvard; **Kalyan Veeramachaneni**, MIT LIDS, DataCebo; **Roman Vershynin**, UCI Mathematics; **Nikita Aggarwal**, UCLA School of Law; **Ali Golshan**, Gretel.ai; **Tobias Hann**, MOSTLY AI; **Vamsi Krishna Potluru**, J.P. Morgan AI Research; **Alexandra Wood**, Harvard Berkman Klein Center; **Belinda Zeng**, Amazon; **Ryan McKenna**, Google; **Ching-Yun (Irene) Ko**, MIT EECS; **Yunkai Zhang**, UCB IEOR; **Lucas Rosenblatt**, NYU Center for Responsible AI; **Jimeng Sun**, UIUC Computer Science; **Steven Wu**, CMU Computer Science; **Ying Nian Wu**, UCLA Statistics; **Alessandro Achille**, Amazon; **Nanyun Peng**, UCLA Computer Science; **Bolei Zhou**, UCLA Computer Science; **Andrés Felipe Barrientos**, FSU Statistics; **Joerg Drechsler**, IAB in Germany; **Joshua Snoke**, RAND; **Aaron Williams**, Urban Institute.

4/14/2023 Bradley Lecture 2023: Protocols for Observational Studies: Methods and a Gun Violence Prevention Study
Dylan Small, University of Pennsylvania, Wharton School

Sponsorships

4/17/2023 15th Annual Conference on Statistical Issues in Clinical Trials: Advances in Time-to-Event Analyses in Clinical Trials

Terry Therneau, PhD, Mayo Clinic
Lu Tian, ScD, Stanford University
Zhenzhen Xu, PhD, Food and Drug Administration
Fan Li, PhD, Yale University
Devan Mehrotra, PhD, Merck and Co.
Lu Mao, PhD, University of Wisconsin
Anne Eaton, PhD, University of Minnesota
Richard Cook, PhD, University of Waterloo
Ionut Bebu, PhD, George Washington University
Rebecca Betensky, PhD, New York University
Michael Fay, PhD, National Institute of Allergy and Infectious Diseases
Pralay Mukhopadhyay, PhD, Otsuka Pharmaceutical
Douglas Schaubel, PhD, University of Pennsylvania
Mei-Cheng Wang, PhD, Johns Hopkins

5/1-3/23 Minitab Insights Global Conference: Discover. Predict. Transform. Achieve.

Michelle Ruehl, Executive Consultant, USAF
Connor McIntyre, Six Sigma Quality Champion, Amazon

5/5-7/23 The 8th Workshop on Biostatistics and Bioinformatics: Learning and Implementing Bayesian Adaptive Designs

Nilanjan Chatterjee, Johns Hopkins University

5/16-18/23 Interdisciplinary Workshop on Weather and Climate Extremes

Francis Zwiers, University of Victoria
Michael Wehner, Lawrence Berkeley National Laboratory
Kenneth Kunkel, North Carolina State University
Richard Smith, University of North Carolina, Chapel Hill
Brian Reich, North Carolina State University
Ben Shaby, Colorado State University
Likun Zhang, University of Missouri
Robert Lund, University of California - Santa Cruz
Thomas Wahl, University of Central Florida
Dan Cooley, Colorado State University
Alex Cannon, Environment and Climate Change Canada
Gabriele Villarini, University of Iowa
Mitchell Krock, Argonne National Laboratory
Ryan Sriver, University of Illinois at Urbana-Champaign
Xuebin Zhang, Environment and Climate Change Canada
Rao Kotamarthi, Argonne National Laboratory
Mihai Anitescu, Argonne National Laboratory
Jiali Wang, Argonne National Laboratory
Emily Hector, North Carolina State University

Sponsorships

5/18-20/23 IRSA Conference - The Fast and the Curious: Modern Markov Chain Monte Carlo
Yves Atchade; **Stephen Berg**; **Riddhiman Bhattacharya**; **Alexandre Bouchard-Côté**; **Yuansi Chen; Andrew Chin**; **Murat A Erdogdu**; **Murali Haran**; **Andrew J. Holbrook**; **Nianqiao Phyllis Ju**; **Kshitij Khare**; **Florian Maire**; **Jeffrey Negrea**; **Arman Oganisian**; **Vivekananda Roy**; **Aaron Smith**; **Aixin Tan**; **Dootika Vats**; **Guanyang Wang**; **Lu Zhang**; **Quan Zhou**.

5/19-20/23 Statistical Network Science with Applications Workshop
Reka Albert, Penn State University
Shankar Bhamidi, University of North Carolina at Chapel Hill
Yuguo Chen, University of Illinois Urbana-Champaign
Bruce Desmarais, Penn State University
Yang Feng, New York University
Tracy Ke, Harvard University
Jing Lei, Carnegie Mellon University
Martina Morris, University of Washington
Sarah Rajtmajer, Penn State University
Purnamrita Sarkar, University of Texas at Austin
Ji Zhu, University of Michigan

5/22-24/23 Statistical Methods in Imaging Conference 2023
Mingyao Li, University of Pennsylvania
Kelvin Lim, University of Minnesota
Ranjan Maitra, Iowa State University

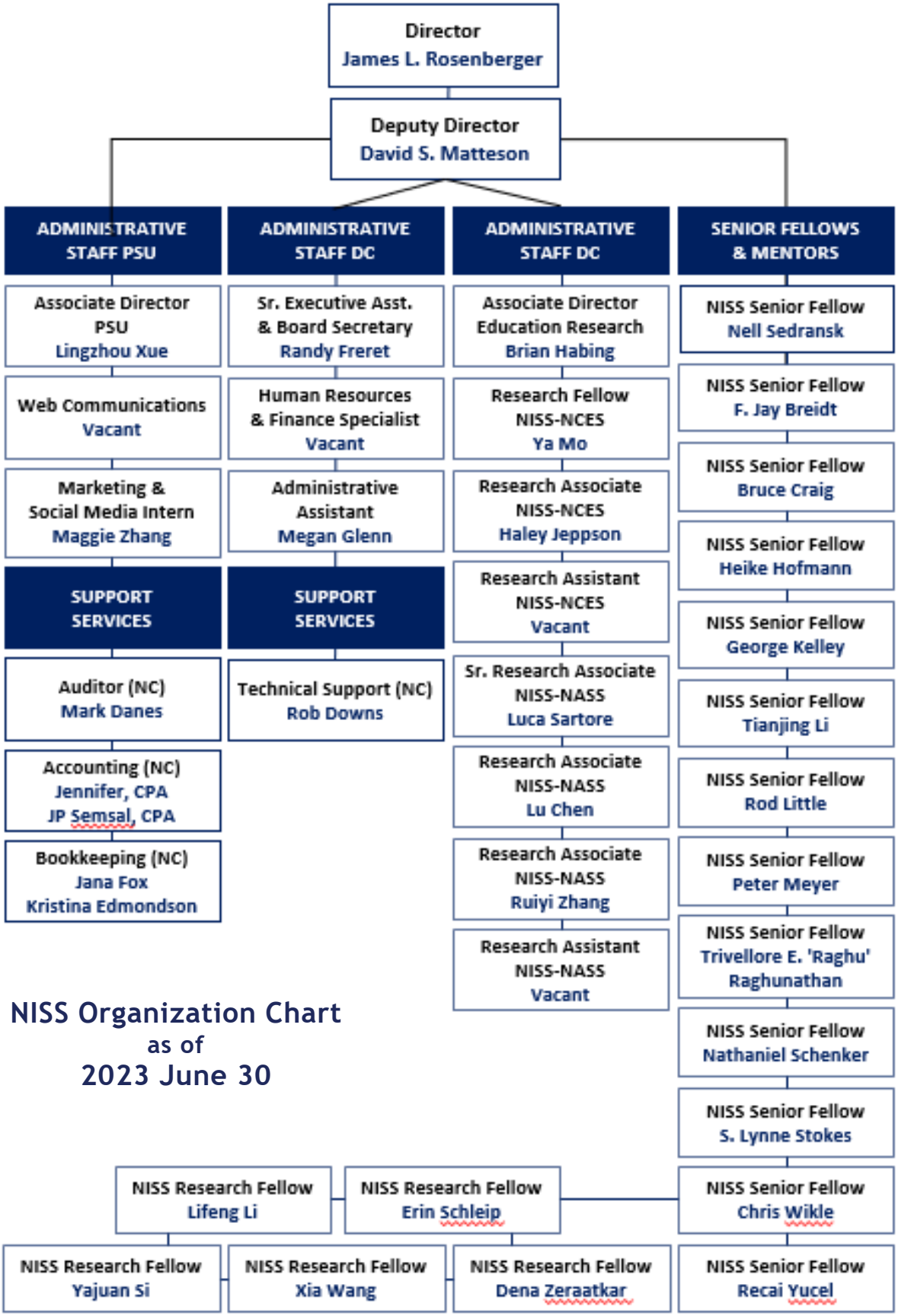
5/30-6/3/23 USCOTS 2023
Regina Nuzzo, Gallaudet University
Lawrence (Larry) Lesser, The University of Texas at El Paso
Kelly Spoon, San Diego Mesa College
Sara Stoudt, Bucknell University

6/6-8/23 QPRC 2023: Quality and Productivity Research Conference
Nicole Radziwill, Ultranauts Inc. (formerly Ultra Testing)
Yu Ding, Texas A&M University
Hung-da Wan, University of Texas at San Antonio
Fugee Tsung, The Hong Kong University of Science and Technology (Guangzhou)

6/19-21/23 2023 ASA-BIOP Nonclinical Biostatistics Conference
Madhumita (Bonnie) Ghosh Dastidar, RAND Statistics Group
Ajaz S. Hussain, freelance
Luwis Diya, Janssen
Will Landau, Eli Lilly
Thomas Mathew, UMBC

ORG CHART

As of June 30, 2023



FINANCIAL STATEMENT

NISS STATEMENT OF ASSETS, LIABILITIES AND NET ASSETS

	Total	
	As of Jun 30, 2023	As of Jun 30, 2022
ASSETS		
Current Assets		
Bank Accounts		
FNB Checking	362,289	561,328
FNB Money Market	101,551	0
Total Bank Accounts	**463,840**	**561,328**
Accounts Receivable		
Affiliate Program Receivables	4,000	-24,000
Contract Receivables	82,343	233,941
Total Accounts Receivable	**86,343**	**191,701**
Other Current Assets		
Other Recievable	0	20,000
Prepaid Expenses	8,022	102
Sales Tax Receivable	3,421	753
Total Other Current Assets	**11,444**	**20,855**
Total Current Assets	**561,626**	**773,884**
Fixed Assets		
Accumulated Amortization	-94,416	-73,401
Accumulated Depreciation	-3,486,058	-3,303,058
Building - Addition	4,342,915	4,342,915
Building - Construction	2,456,778	2,456,778
Building - Improvements	183,811	183,811
Equipment	69,921	69,921
Furniture & Fixtures	7,326	7,326
Land	1,350,000	1,350,000
Lease Closing Costs	205,480	205,480
Website	97,000	97,000
Total Fixed Assets	**5,132,757**	**5,336,772**
Other Assets		
Security Deposits	5,400	5,400
Total Other Assets	**5,400**	**5,400**
TOTAL ASSETS	**5,699,783**	**6,116,057**
LIABILITIES AND EQUITY		
Liabilities		
Total Accounts Payable	**19,483**	**15,328**
Accrued Vacation Leave	26,443	26,000
Total Other Current Liabilities	**27,747**	**26,779**
Total Current Liabilities	**47,336**	**43,471**
Long-Term Liabilities		
Construction Loan - 2.5m Re-fi	2,134,889	2,279,085
Deferred Affiliate Rev. (Rest.)	120,695	147,403
Deferred Base Rent Revenue	-156,310	-194,687
Total Long-Term Liabilities	**2,099,273**	**2,231,801**
Total Liabilities	**2,146,609**	**2,275,272**
Equity		
Net Assets	3,579,833	3,579,833
Unrestricted Net Assets	260,952	308,666
Net Income	-287,610	-47,715
Total Equity	**3,553,174**	**3,840,785**
TOTAL LIABILITIES AND EQUITY	**5,699,783**	**6,116,057**

FINANCIAL STATEMENT

NISS STATEMENT OF ACTIVITIES

National Institute of Statistical Sciences

As of 30 June 2023

Accrual Basis

	Jul 2022 - Jun 2023	Jul 2021 - Jun 2022
Income		
Total BUILDING REVENUE	**425,114**	**418,501**
GRANT AND CONTRACT REVENUE	865,211	1,634,285
OTHER REVENUE	3,019	-
Affiliate Revenue (Unrestricted	228,615	128,329
Dividend Income	707	696
Interest Income	551	-
NISS Fund / Other Revenue	37,635	1,983
NISS Workshop Registration	25,033	12,835
Total OTHER REVENUE	**308,561**	**178,843**
Total Income	**1,599,885**	**2,231,629**
Expenses		
Total BUILDING EXPENSES	**192,910**	**197,868**
CONSULTANTS / PROFESSIONAL FEES		
Accounting Fees	16,120	26,255
Consultants	214,578	392,522
Consultants - ED and Research	1,208	-
Honorarium	1,200	10,750
Other Fees	17,588	16,087
Professional Services	11,585	46,179
Web Support	2,370	3,535
Total CONSULTANTS / PROFESSIONAL FEES	**264,650**	**495,329**
Total MISCELLANEOUS	**21,958**	**6,593**
Total OFFICE EXPENSES	**110,752**	**163,830**
Total Depreciation Expense	**183,000**	**177,498**
Total OTHER EXPENSES	**183,000**	**177,498**
Total SALARIES and FRINGE BENEFITS	**1,005,296**	**1,158,333**
Total TRAVEL EXPENSES	**51,590**	**13,060**
Total Expenses	**1,830,156**	**2,212,639**
Net Operating Income	**(230,270)**	**18,990**
Total Other Expenses - Interest	**57,340**	**66,705**
Net Income	**(287,610)**	**(47,715)**
Unaudited - Oct 26, 2023 - Accrual Basis		

National Institute of Statistical Sciences
1750 K Street, NW, Suite 1100
Washington, DC 20006-2036
officeadmin@niss.org
www.niss.org