Shannon **Gallagher**

Data Scientist · Statistician · Ph.D.

SEI Building, 4500 Fifth Ave., Pittsburgh, PA 15213

(724) 504-8990 | skgallagher19@gmail.com, skgallagher@sei.cmu.edu | skgallagher.github.io | skgallagher

Current

Carnegie Mellon University — *Pittsburgh, PA*

CERT Division | Software Engineering Institute — *2021-Present*

Data Scientist

Education

Carnegie Mellon University — *Pittsburgh, PA*

Ph.D. in Statistics — *2014-2019*

M.S. in Statistics — *2014-2015*

B.S. in Mathematical Sciences (University and College Honors) — *2010-2014*

Doctoral Dissertation: "Catalyst: agents of change. Integration of compartment and agent-based models for use in infectious disease methodology." Advisor: William F. Eddy

Selected Publications and Reports

Gallagher, S.K., Ross, D., Mellon J., and Bernaciak, C. "A deepfake detection pipeline for machine learning methods on human faces." In prep., 2022.

Gallagher, S.K., Whisnant A., Hristozov, Anton D., and Vasudevan, Amit. "Reviewing the role of machine learning and artificial intelligence for remote attestation in 5G+ networks." Under review at IEEE Future Networks World Forum, 2022.

Gallagher, S.K., Klieber, W. E., and Svoboda, David. "LLVM intermediate representation for code weakness identification." In prep., 2022.

Gallagher, S.K., Wang, J., Lombard, K., Dodd, L., and Proschan, M. "Using Kaplan Meier estimates in non-inferiority hypothesis tests." To appear in *Statistics in Medicine*, 2022.

Gallagher, S.K. and Follmann, D. "Branching process models to identify risk factors for infectious disease transmission." *Journal of Computational and Graphical Statistics*, 2022.

Bohl, J. [and others including Fintzi, J., **Gallagher, S.K.**, and Manning, J.] "Discovering disease-causing pathogens in resource-scarce Southeast Asia using a global metagenomic pathogen monitoring system." *PNAS*, 2022.

Paules, C.I.†, **Gallagher, S.K.**†, Rapaka, R.R., [and others including Dodd, L.E., and Benson, C.A.] († co-first author). *Clinical Infectious Diseases*, 2021.

Gallagher, S.K., Frisoli K., and Luby, A. "Opening up the court (surface) in tennis grand slams." *Journal of Quantitative Analysis in Sports*, 2021.

Azasi, Y.†, **Gallagher, S.K.**†, [and 11 others including Fay, Michael P., Miura, K., and Miller, Louis H.] († co-first author). "Bliss' and Loewe's additive and synergistic effects in Plasmodium falciparum growth inhibition by AMA1-RON2L, RH5, RIPR and CyRPA antibody combinations." *Scientific Reports*, 2020.

Gallagher, S.K. "Discussion of 'An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China.' " *Journal of Data Science*, 2020.

Gallagher, S.K., Chang, A., Eddy, W.F. "Exploring the nuances of R_0: Eight estimates and application to 2009 pandemic influenza." Pre-print available at: `https://arxiv.org/abs/2003.10442`, 2020.

Gallagher, S.K. and Eddy, W.F. "A hybrid compartment/agent-based model for infectious disease modeling." ENAR Epidemiology section student paper contest submission, 2019.

Gallagher, S.K., Richardson L.F., Ventura S.L., and Eddy, W.F. "SPEW: Synthetic Populations and Ecosystems of the World." *Journal of Computational and Graphical Statistics*, 2018.

Selected Presentations and Posters

Joint Statistical Meetings — *Washington, DC*
Speed Presentation — *August 2022*
"A Statistical Framework for Deepfake Detection."

Military Operations Research Society 90th Symposium — *Quantico, VA*
Tutorial — *June 2022*
"Intro to Data Science for Digital Forensics."

ISACA Huntsville Chapter Meeting — *Virtual Meeting*
Invited Presentation — *May 2022*
"Deepfake Generation & Detection: How easy are they to make and detect?"

Global Platforms Austin Workshop: SESIP's Role in the IoT Ecosystem. — *Virtual Conference.*
Presentation — *April 2022*
"Trust Assessment Models for Next Generation Networks."

5th Annual DoD Artificial Intelligence/Machine Learning TEM — *Virtual Conference*
Presentation — *October 2021*
"Using Transformer Machine Learning to Identify Code Weaknesses."

George Washington University — *Washington D.C.*
Guest Lecture — *February 2020 & 2021*
"A brief survey of statistical models to analyze the transmission of infectious diseases."

Los Alamos National Laboratory — *Virtual Seminar*
Invited Presentation — *August 2020*
"Use of multiple covariates in branching processes for modeling the spread of Tuberculosis."

Carnegie Mellon University — *Virtual Class*
Guest Lecture — *September 2020*
"A foray into infectious disease modeling: compartment models."

Joint Statistical Meetings — *Virtual Conference*
Poster/Presentation — *August 2020*
"Branching processes with covariates to model the spread of Tuberculosis."

ENAR — *Virtual Conference*
Presentation — *March 2020*
"A Hybrid Compartment/Agent-Based Model for Infectious Disease Modeling."

Dissertation Defense — *Pittsburgh, PA*
Presentation — *July 2019*
"Catalyst: agents of change. Integration of compartment and agent-based models for use in infectious disease epidemiology."

Carnegie Mellon Sports Analytics Conference — *Pittsburgh, PA*
Presentation – **Honorable Mention** — *October 2018*
"Opening up the (court) surface in tennis grand slams." Joint work with Kayla Frisoli and Amanda Luby.

International Conference on Synthetic Populations — *Lucca, Italy*
Presentation – **Invited Speaker** — *February 2017*
"Generating Synthetic Ecosystems: A Tutorial" Joint work with Lee Richardson, Samuel Ventura, and William Eddy.

MIDAS National Conference — *Washington D.C.*
Presentation — *May 2016*
"Services for the MIDAS Network: Visualization and Synthetic Ecosystems." Joint work with Lee Richardson, Samuel Ventura, and William Eddy.

UP-STAT — *Buffalo, NY*
Presentation – **2nd place** — *March 2016*
"From forecasting the Flu to Predicting the 'Next' Disease." Joint work with Roni Rosenfeld, Ryan Tibshirani, Lee Richardson, Samuel Ventura, and William Eddy.

Honors & Awards

2022	**Nomination**, Nominated for Newcomer of the Year at SEI	*Pittsburgh*
2022	**Certificate**, Awarded CERT Spotlight for Outstanding Achievement on a project	*Pittsburgh*
2020	**Selected Carnegie Mellon University Student Speaker for 2020 Commencement**, Conferral of Degrees. Full program available at: `https://www.cmu.edu/commencement/`.	*Virtual*
2018	**Honorable Mention**, Carnegie Mellon University Sports Analytics Conference Reproducible Paper Competition. $\$1,000$ award.	*Pittsburgh, PA*
2018, 2014	**Honorable Mention**, Gertrude M. Cox Scholarship. ASA Committee on Women in Statistics and the Caucus for Women in Statistics.	
2018	**Scholarship Recipient**, Summer Institute in Statistics and Modeling. Tuition and travel stipend.	*Seattle, WA*
2017	**Selected Presenter**, AT&T Labs Graduate Student Symposium. One of fourteen PhD students out of 79 applicants selected. Awarded $800 in travel funding.	*New York, NY*
2016	**Hackathon Champion**, MIDAS MISSION Public Health Hackathon. Awarded $3,000 prize.	*Pittsburgh, PA*
2016	**2nd place**, Student presentation at UP-STAT conference.	*Buffalo, NY*
2014	**Judith A. Resnik Award for Outstanding Women in the Sciences**, Carnegie Mellon University.	*Pittsburgh, PA*
2013	**Phi Beta Kappa Honor Society**, Fall induction.	*Pittsburgh, PA*

Software

2022	`Deepfake Detection Pipeline`, **Gallagher, S.K.**, Ross, D., Mellon, J., and Bernaciak, C. Internal SEI repository.
2019-2021	`EpiCompare`, **Gallagher, S.K.** and Leroy, B. Software for simulation and analysis of disease data via ternary plots. Available at `https://skgallagher.github.io/EpiCompare/index.html`.
2021	`InfectionTrees` **R package**, **Gallagher, S.K.** and Follmann, D. Analysis, vignettes, and code for studying transmisison trees and incorporating information from individuals. Available at `https://skgallagher.github.io/InfectionTrees/articles/getting-started.html`.
2019-2020	`loewesadditivity` **R Package**, **Gallagher, S.K.** and Fay, M. P. Software for modelling synergy, antagonism, or Loewe additivity between varying dose combinations of different compounds. Available at `www.github.com/skgallagher/loewesadditivity`.
2019	`catalyst`, **Gallagher, S.K.**. Software for simulation, testing, and analysis of compartment and agent-based models. Available at `www.github.com/skgallagher/catalyst`.
2018	`spew`, Richardson L., **Gallagher, S.K.**, Ventura, S., and Eddy, W.F. `R` package for synthetic ecosystem generation. Available at `www.github.com/lrichardson/spew`.
2016	`spewview`, **Gallagher, S.K.** and Richardson L. `R` Shiny application for infectious disease visualization. Available at `www.github.com/skgallagher/hackathon`.

Research, Teaching, and Work Experience

Data Science Team, CERT, Software Engineering Institute, Carnegie Mellon University — *Pittsburgh, PA*

Data Scientist — *2021-Present*

- PI for deepfake detection internal research project that implements a pipeline of reproducible statistical and ML models
- Explored machine learning models and existing data for code weakness identification
- Consulted Department of Defense (DoD) agencies in AI/ML use in emerging technology
- Presented work at DoD conferences

National Institute of Allergy and Infectious Disease — *Rockville, MD*

Post-Doctoral Fellow — *2019-2021*

- Worked with Dean Follmann to analyze the effect of smear status on spread of Tuberculosis
- Devised and implemented statistical model to analyze the synergy of antibody pair combinations for Malaria vaccine efforts
- Analyzed survival probability under different loss to follow up scenarios
- Helped lead a retrospective analysis of the COVID-19 ACTT-1 Clinical Trial to identify subgroups that respond to remdesivir

Carnegie Mellon University — *Pittsburgh, PA*

Research and Teaching Assistant — *2014-2019*

- Developed and presented material for the Summer Undergraduate Research Experience in Statistics
- Generated high-resolution synthetic ecosystem of the U.S. and 70+ countries for use in agent-based models for transmission of disease
- Oversaw lab for 100 students, organized and led review sessions for a variety of statistics and mathematics classes including Epidemiology, Statistical Computing, Intro to Probability, Advanced Undergraduate Research, Concepts of Mathematics, and Multi-dimensional Calculus

PNC *Pittsburgh, PA*

Graduate Intern *2015*

- Scraped and analyzed social media data for sentiment analysis
- Parallelized code via Hadoop

Professional Service

2017-PRESENT	**Member**, American Statistical Association
2016-PRESENT	**Reviewer**, Statistics in Medicine, Clinical Infectious Diseases, Statistics and Public Policy, Journal of Quantitative Analysis in Sports, Journal of Data Science
2020-2021	**Mentor to students**, National Insitute of Allergy and Infectious Diseases • Statistical mentor for a student's research study to complete her Masters of Health Sciences in Clinical Research Training Program at Duke School of Medicine • Undergraduate student research co-advisor to a Carnegie Mellon University statistics student along with Professor Joel Greenhouse
2018-2019	**PI**, ProSeed/Crosswalk recipient for $1600 to seed a mentorship program across all levels of students within the Stat&DS community
2018-2019	**President**, Carnegie Mellon University Women in Statistics. • Organized Women in Data Science Pittsburgh @CMU as an Executive Committee Member. Inivted speakers and sponsors, helped organize venue logistics, sent out invitations for for attendance, and created the 2018 website • Maintained the Women in Statistics Website from 2017-2018 • Organized a seminar by former PhD student about her experiences as a post-doc at Harvard Biostatistics (2017) • Organized a panel about applying to graduate school for 30+ undergraduate and masters students (2016) • Organized dinner with new dean of Mellon College of Science (2016)
2016-2018	**Co-Organizer**, Pittsburgh useR. Organized meet-ups for 30+ members on a variety of topics including cross-language coding and integrating `R` with `github`
2016-2017	**Judge and volunteer**, Tartan Data Science Cup - three separate events.
2016-2017	**Vice President**, CMU Women in Statistics.
2016	**Presenter**, Coding for Girls

Relevant Course Work

- Machine Learning I and II (**Grad**)
- Statistical Computing (**Grad**)
- Modern Regression (**Grad**)
- Hierarchical Models (**Grad**)
- Multivariate Methods and Data Mining
- Data Matching and Record Linkage
- Advanced Methods for Data Analysis
- Epidemiology