









---

## Education

- PhD Statistics, [Iowa State U](#), 2016
  - MS Statistics, [Iowa State U](#) 2013
  - BS Math, [U of Chicago](#) 2011
- 

## Publications

- , Niemi, J., and Nettleton, D., "Fully Bayesian analysis of RNA-seq counts for the detection of gene expression heterosis". *Journal of the American Statistical Association*, <https://doi.org/10.1080/01621459.2018.1497496>.
-  (2018), "The drake R package: a pipeline toolkit for reproducibility and high-performance computing". *Journal of Open Source Software*, 3(21), 550, <https://doi.org/10.21105/joss.00550>.
- Niemi, J., Mittman, E., , and Nettleton, D. (2015), "Empirical Bayes Analysis of RNA-seq Data for Detection of Gene Expression Heterosis," *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 1-15. Available at [link.springer.com](http://link.springer.com).
-  and Liu, P. (2013), "Dispersion Estimation and Its Effect on Test Performance in RNA-seq Data Analysis: A Simulation-Based Comparison of Methods," *PLOS One*, 8. Available at [journals.plos.org](http://journals.plos.org).
- Ratliff, B., Womack, C., Tang, X., , Butler, L., and Szpunar, D. (2010), "Modeling the Rovibrationally Excited C2H4OH Radicals from the Photodissociation of 2-Bromoethanol at 193 nm," *Journal of Physical Chemistry*, 114, 4934-4945. Available at [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov).

## Open Source Software

- drake, an R-focused pipeline toolkit for reproducible computation and high-performance computing. Part of [rOpenSci](#).
- downsize, and R package to toggle between the test and production versions of large workflows.
- R packages fbseq, fbseqCUDA, and fbseqOpenMP from dissertation research.
- R utilities eply, grapes, txtq, and wildcard.

## Awards

- 2017 Lilly Innovator Award. Awarded for leading a successful team effort to modernize Lilly's internal process for contributing open source software.
- Student Paper Award, American Statistical Association Section on Statistical Computing, Jan 2016. Awarded for an early draft of the preprint at [REDACTED].
- [Vince Sposito Statistical Computing Award](#), Iowa State University, Aug 2013.
- [GlaxoSmithKline Industrial Scholarship](#), Iowa State University, Sep 2011.
- Alumni Scholarship, Iowa State University, Aug 2011.

## Skills

- Reproducible research, hierarchical models, Bayesian methods, Markov chain Monte Carlo, statistical computing, high-dimensional data analysis, genomics data analysis, exploratory analysis, visualization, linear and nonlinear models, data mining, machine learning, predictive modeling, multivariate analysis.
- High-performance computing, R, R package development, general-purpose graphics processing unit (GPU) computing, CUDA, shell scripting, LaTeX, HTML, CSS.
- Past experience with C/C++, MPI, OpenMP, Python, JavaScript, AWK, Fortran.

## Research statistician

- **October 2016 - Present**
- [Eli Lilly and Company](#)
- Helped design and plan early- and mid-phase clinical trials as the lead statistician for three new autoimmune therapies.
- Supported late-phase clinical trial teams with advanced analytics, including clinical program simulation and tailored therapeutics.
- Developed interactive [R Shiny](#) web applications to facilitate decision-making and simulation.
- Created the [REDACTED] package to enhance reproducibility and high-performance computing in R.

## Research assistant

- **May 2013 - Aug 2016**
- RNA-sequencing Working Group, Department of Statistics, [Iowa State University](#).
- Funded by NIH grant R01GM109458 with Drs. Dan Nettleton and Jarad Niemi.
- Developed a new fully Bayesian analysis method for high-dimensional genomic datasets using hierarchical models.
- Implemented massively parallelized Markov chain Monte Carlo.

- Created the [REDACTED] R package to distribute the analysis method.
- Implemented and distributed parallel computing backends for [CUDA GPUs](#) ([REDACTED]) and [OpenMP](#) ([REDACTED]).
- Created the [REDACTED], [REDACTED], and [REDACTED] packages to manage, ameliorate, expedite, and accelerate computationally heavy reproducible workflows that are under heavy development.

## Seminar instructor

- **Aug - Dec, 2012 and 2013.**
- Department of Statistics, [Iowa State University](#).
- [GPU computing](#) seminar series at [REDACTED].
- Educated faculty and graduate students on massively parallel computing with [general-purpose graphics processing units](#).
- Constructed, curated, and distributed slides, video, and example code at [REDACTED] and on YouTube.

## Course instructor

- **Jan - May, 2012 and 2013.**
- Department of Statistics, [Iowa State University](#).
- STAT 305: Engineering Statistics ([REDACTED]).

## Grader

- **Aug - Dec, 2011.**
- Department of Statistics, [Iowa State University](#).
- STAT 231: Engineering Probability.
- STAT 105: Introduction to Engineering Statistics.

## Leadership at Eli Lilly and Company

- Led a successful team effort to modernize Lilly's internal procedure for contributing open source software.
- Served as a volunteer moderator in the [2017 National Science Bowl \(high school Indiana Regionals\)](#).

# Leadership at Iowa State University

- Founder and leader, Cloud Computing Working Group, Sep - Dec 2015.
- Member, Computation Advisory Committee, Sep 2015 - May 2016.
- Volunteer instructor, Office of Precollegiate Programs for Talented and Gifted ([OPPTAG](#)), Mar 13, 2014.
- Fellow, [Preparing Future Faculty](#), Aug 2013 - May 2014.
- Assistant Coach, Boxing Club, Aug 2013 - Dec 2013.

## References

- Jarad Niemi, PhD advisor and major professor, [niemi@iastate.edu](mailto:niemi@iastate.edu).
- Dan Nettleton, lead principal investigator of the RNA-sequencing Working Group (Iowa State Department of Statistics), [dnett@iastate.edu](mailto:dnett@iastate.edu).
- Peng Liu, MS advisor and major professor, [pliu@iastate.edu](mailto:pliu@iastate.edu).
- Additional references available on request.

## Hobbies

- Climbing, Brazillian Jiu Jitsu, sailing, windsurfing