

# W. Katherine Tan

<https://wlktan.github.io/>

Email : [ktan@flatiron.com](mailto:ktan@flatiron.com)

Mobile : +1-206-825-5489

---

## EDUCATION

- **University of Washington** Seattle, WA
  - *PhD Biostatistics* *Sep 2013 – Dec 2018*
  - Certificate in Technology Entrepreneurship, Foster School of Business
- **Johns Hopkins University** Baltimore, MD
  - *B.S. Applied Mathematics & Statistics; Phi Beta Kappa; GPA: 3.90* *Aug 2009 – May 2013*

---

## ACADEMIC & PROFESSIONAL EXPERIENCE

- **Flatiron Health** New York, NY
  - *Quantitative Scientist* *Jan 2019 - Present*
  - Accelerating cancer research and improving the quality of patient care through generating regulatory-grade real-world evidence derived from electronic health records (EHR) data and rigorously developed statistical methods and designs.
- **University of Washington** Seattle, WA
  - *Graduate Research Assistant* *Jan 2014 - Present*
    - **LIRE pragmatic trial:** Developed and implemented an end-to-end Natural Language Processing (NLP) system to identify 26 common and rare anatomical findings from radiology reports; models achieved an average accuracy (AUC) of 95% and used for automated patient subgroup identification.
    - **NEAT-O Phase II clinical trial:** Modeled the effect of a neuroprotective agent in newborn infants; analyses resulted in a peer-reviewed publication and informing subsequent clinical trials.

---

## PUBLICATIONS, PRESENTATIONS & CONFERENCES (SELECTED)

- **Tan WK**, Heagerty PJ. Surrogate-guided sampling designs for classification of rare outcomes from electronic medical records data. *Under review.*
- **Tan WK**, Hassanpour S, Heagerty PJ et al. Comparison of Natural Language Processing Rules-Based and Machine-Learning Systems to Identify Lumbar Spine Imaging Findings Related to Low Back Pain. *Academic Radiology* 25:11 (2018): 1422-1432.
- Jarvik JG, Gold LS, **Tan WK**, et al. Long-term Outcomes of a Large, Prospective Observational Cohort of Older Adults with Back Pain. *The Spine Journal* 18:9 (2018): 1540-1551.
- Huhdanpaa HT, **Tan WK**, Rundell SD et al. Using Natural Language Processing of Free-Text Radiology Reports to Identify Type 1 Modic Endplate Changes. *Journal of Digital Imaging* (2017): 1-7.
- Wu YW, Mathur AM, Chang T [et al., including **Tan WK**]. High-dose Erythropoietin and Hypothermia for Hypoxic-Ischemic Encephalopathy: a Phase II Trial. *Pediatrics* (2016): e20160191.
- Jarvik JG, Comstock BA, James KT [et al., including **Tan WK**]. Lumbar Imaging With Reporting Of Epidemiology (LIRE) - Protocol for a pragmatic cluster randomized trial. *Contemporary Clinical Trials* 45 (2015): 157-163.
- **Tan WK**. Surrogate-guided sampling designs for biomedical natural language processing applications with rare outcomes. *Joint Statistical Meetings, Chicago, IL, Jul 2016.*

---

## TECHNICAL SKILLS

- **Computing:** R (caret, dplyr, quanteda, rmarkdown), Python (pandas, scikit-learn), SQL, Git, Java, MATLAB, L<sup>A</sup>T<sub>E</sub>X
- **Statistics & Machine Learning:** Bayesian Methods, Classification, Correlated Data, Experimental Design, Ensemble Learning, Regression Methods, Sampling Methods, Sparse Learning, Survival Analysis

---

## ADDITIONAL EXPERIENCE, SERVICE, AND AWARDS

- **Teaching:** Teaching Assistant, Applied Biostatistics; Guest Lecturer, Computing tools for Biostatisticians
- **Consulting:** Statistical Consultant, University of Washington Statistical Consulting Services
- **Professional Memberships:** American Statistical Association (ASA)
- **Awards and Honors:** Department of Biostatistics travel award for research in a domestic conference, University of Washington Business Plan Competition Finalist