W. Katherine Tan

CONTACT INFORMATION

Box 357232
Department of Biostatistics
University of Washington
Seattle WA 98195 USA

E-mail: wlktan [at] uw [dot] edu
Website: http://students.washington.edu/wlktan

EDUCATION

University of Washington, Seattle, Washington, USA

Ph.D. Candidate, Biostatistics **September 2013 - Present**
Technology Entrepreneurship Certificate Student, Foster School of Business

Johns Hopkins University, Baltimore, Maryland USA

B.S., Applied Mathematics & Statistics **May 2013**

Second major: Molecular & Cellular Biology
Minor: Entrepreneurship & Management

ACADEMIC & PROFESSIONAL EXPERIENCE

University of Washington, Seattle, Washington USA

Graduate Statistical Research Assistant, LIRE project **January 2018 - Present**
Work with an inter-disciplinary team of statisticians, clinicians, epidemiologists, and project managers on various projects related to low back pain research.

Graduate Research Assistant, LIRE NLP sub-project **September 2013 - December 2017**
Led a team building an end-to-end Natural Language Processing (NLP) system to identify 26 anatomical findings from 200,000 radiology text reports. Developed labeled dataset collection methodology with medical collaborators. Engineered features using Apache Lucene and ConText. Implemented prediction model using elastic-net logistic regression; models achieved an average accuracy (AUC) of 95%.

Graduate Research Assistant, NEAT-O Phase II clinical trial **September 2015 - March 2016**
Evaluated the effect of High-dose Erythropoietin on Hypoxic-Ischemic-Encephalopathy in newborn infants. Analysis incorporated data from neuroimaging, laboratory tests, and standardized scores. Identified appropriate statistical methods and interpreted results with medical collaborators.

Statistical Consultant, UW Statistical Consulting Services **Sep 2015 - Jan 2016**
Worked with over 10 clients on experimental design, study protocol, sample size calculation, and data analysis. Client background included auditory sciences, bioengineering, forestry, linguistics, and startups.

Flatiron Health, New York, New York, USA

Quantitative Sciences Intern **June 2017 - August 2017**
Evaluated trial-level surrogacy of real-world Progression Free Survival for Overall Survival among cancer patients. Implemented an analytic pipeline with propensity score weighting, survival analysis, and weighted regression models. Developed diagnostics to ensure validity of end results. Interpreted and presented research to an audience consisting of oncologists, quantitative scientists, and product leadership.

Farmers' Market Price Comparison Research Project, Baltimore, Maryland USA

Principal Student Investigator **February 2012 - May 2013**
Wrote project proposal and awarded grant through Dean's Office. Coordinated a team of six research assistants for bimonthly observational price data collection at farmers' markets and supermarkets in Baltimore city. Led the direction of project through data-driven statistical and qualitative analyses.

Johns Hopkins School of Public Health, Kathmandu Nepal

Data Analyst and Management Intern **May 2012 - August 2012**
Conducted quality analysis on flu-like symptoms on data collected by field workers. Identified inconsistent, inaccurate and missing data for three large and ongoing clinical trials through database with over one million data points. Recovered identification numbers of nasal-swab samples based

on linked information in database; transitioned data from Excel to SQL tables after extensive data cleaning.

Johns Hopkins School of Medicine, Baltimore, Maryland USA

Proteomics Research Analyst **August 2011 - December 2011**
Analyzed mass spectrometry data identify most likely proteins from cancer biomarkers. Tested software for N-X-S/T glycosylation identification written by Senior Bioinformaticist.

Johns Hopkins Department of Biology, Baltimore, Maryland USA

Research Assistant **September 2010 - May 2011**
Determined graphically absorbance of highly conjugated biochemical compounds under various optimized conditions from modified protocols.

Teaching Experience

University of Washington, Seattle, Washington USA

Biostatistics Tutor **September 2016 - Present**
Tutored 4 students in applied biostatistics.

Guest Lecturer, Computing Skills in Biostatistics **February 2017**
Delivered a guest lecture on Introduction to Java programming for non-programmers (statisticians).

Teaching Assistant for BIOST 511 **September 2015 - December 2015**
Worked with a team of professor and four TA's to coordinate a class of over 200 graduate students. Duties included weekly office hours, assisting with discussion sections, and grading of homeworks and examinations.

Johns Hopkins University, Baltimore, Maryland USA

The Learning Den Tutor **February 2013 - May 2013**
Led small group (1-6 students) discussion sessions for sophomore and junior level Probability & Statistics courses.

Publications

Tan WK, Heagerty PJ. Surrogate-guided sampling designs for classification of rare outcomes in biomedical text data. (In preparation).

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. (To appear in *Academic Radiology*).

Jarvik JG, Gold LS, **Tan WK**, et al. Long-term Outcomes of a Large, Prospective Observational Cohort of Older Adults with Back Pain (To appear in *The Spine Journal*).

Huuhdanpaa 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.

Presentations & Conferences

Tan WK. (2016) Surrogate-guided sampling designs for biomedical natural language processing applications with rare outcomes. Joint Statistical Meetings, Chicago, Illinois, July 2016.

Tan WK. (2016) Natural Language Processing (NLP): How does it relate to Biostatistics? UW Biostatistics Student Seminar, Seattle, Washington, January 2016.

Tan WK. (2015) Rule-Based Natural Language Processing for Patient Subgroup Identification. Hegearty Working Group, Seattle, Washington, April 2015.

Tan WK, Rice K, Spziro AA. (2014) The Bayesian Logistic Sandwich. Annual Biostatistics De-

partment Retreat, Blaine, Washington, August 2014.

Tan WK. (2014) Natural Language Processing in Radiology Text Reports. Hegearty Working Group, Seattle, Washington, April 2014.

Honors and Awards

Best Data Analysis Hack, Flatiron Health Hackathon — July 2017
Our team hacked together an analytic pipeline involving paragraph vectors and Convolutional Neural Networks (CNN) implemented in Python to analyze uncategorized clinical documents.

Finalist, University of Washington Business Plan Competition — April 2017
I was part of a team of MBAs, JDs, journalists, and data scientists that developed a business plan for a start-up based on using Natural Language Processing (NLP) for arguments in current issues.

Phi Beta Kappa — March 2013

Dean's Undergraduate Research Awards (DURA) in Political Science — April 2012

Professional Memberships

American Statistical Association (ASA)

The Western North American Region of the International Biometric Society (WNAR)

Service

FIUTS graduate facilitator — January 2018 - Present
Graduate student facilitator for the Foundation for International Understanding Through Students (FIUTS) non-profit organization. Lead and facilitate events involving participants (students and community) from diverse international backgrounds.

Alumni Relations Committee — September 2016 - Present
Organize annual alumni career panel targeting current graduate students interested in industry (non-academic) positions.

Communications Committee Liaison — July 2015 - August 2017
Facilitated content creation and management of the new departmental website.

Graduate Student Mentor — August 2016 - May 2017
Organized group social events for first-year graduate students. Regular one-on-one check-ins with students.

Grads Guiding Grads (G3) mentor — February 2016 - June 2016
Mentored junior graduate students in various departments (outside Biostatistics).

Technical

Computing

- Analytics: R (caret, dplyr, quanteda, rmarkdown), STATA, MATLAB, Octave
- Programming: Java, Python (pandas, scikit-learn), SQL
- Design: HTML, CSS, LaTeX
- Other: Git, Apache Lucene, Unix shell scripts

Languages

- Fluent: English, Chinese
- Intermediate: Bahasa, Japanese