Jason A. Thomas, PhD

Medical Informatician & Data Strategist

Professional Summary

PhD graduate in Biomedical Informatics with 10 years experience in healthcare - both frontline healthcare delivery and analysis/modeling of the resulting data. Expertise in clinical trials and longitudinal cohort studies, digital biomarkers & therapeutics, analysis & data modeling of synthetic and real electronic health records. Research interests: digital therapeutics, mental health, data access & sharing, privacy preserving technologies, data utility & quality.

Work History

Graduate Research Assistant

2020-09 - 2021-09

University Of Washington School Of Medicine, Seattle, WA

- Assessed fitness for use of synthetic and real electronic health records (including log data) for observational research and hospital operations using UW Medicine and National Covid Cohort Collaborative (N3C) data
- Contributed to ETL and overall creation of OMOP-formatted UW Medicine COVID-19 Research Data Warehouse
- Conducted OHDSI model-to-data network studies on COVID-19, found and reported open-source software issues in their R packages on Github

National Library of Medicine Pre-Doctoral Fellow

2017-09 - 2020-09

University Of Washington School Of Medicine, Seattle, Washington

- Used machine learning to predict dementia status from Framingham Heart Study participant interview transcripts, audio recordings, demographics and clinical characteristics; identified data utility issues with recordings.
- Co-wrote and designed $1.75million NIH grant on Electronic Health Record (EHR) data quality scored at top 38th percentile.
- Conducted applied EHR data warehouse research and analytics (e.g. same-day readmissions & data quality issues, machine learning model to predict patient missingness between databases)
- Participated in the Medical Natural Language Inference (MedNLI) subtask of the MEDIQA 2019 shared task
- Aided experimental design, annotation & analysis of ~3k Yahoo! Answers health questions for: Gated Recurrent Unit (GRU) classification over concatenation of PubMed and ELMo embeddings. Improved state of the art on health questions answering for GARD dataset, released Yahoo! dataset.

Senior Research Assistant

2015-03 - 2017-09

Oregon Health & Science University, Portland, Oregon

- Liaison between engineers, sponsors, health workers, ethics board, research subjects, hospital IT to do electrophysiology cardiology research.
- Organization, monitoring of data annotation, cleaning, analysis pipelines
- Assessed Human Computer Interaction and accessibility barriers to 1) use of continuous electrocardiogram patches for home monitoring and 2) production of patient-generated self-tracking data in clinical studies.
- Translation of grants and research designs into IRB submissions.
- Recruitment, scheduling of >350 patients to concurrent clinical trials, including field research alone.
- Published scientific journal articles including logistic regression modeling results to predict eligibility for subcutaneous implantable defibrillators.
- Data collection: electrocardiograms, 6-minute walk tests, surveys, chart reviews, device interrogations (including at cath-lab procedures).

Contact

Address
Seattle, WA, 98115

E-mail
thomasjt@uw.edu

WWW
https://jasonthomas.io/

LinkedIn
https://www.linkedin.com/in/jasonandrewthomas/

Twitter
@realjasonthomas

Education

2017-09 - 2021-09
Ph.D.: Biomedical Informatics Data Science Specialization

University of Washington - Seattle, WA

NLM Pre-Doctoral Fellowship (2 slots available per year at UW), Top Scholar Award

2009-09 - 2013-06
Bachelor of Science: Human Physiology, Biology (minor in Chemistry)

University of Oregon - Eugene, OR

Dean's Scholarship

Software

Advanced knowledge of data visualization (deep familiarity creating interactive or static visualizations using Plotly, Altair, Seaborn, ggplot2, Tableau)

Intermediate knowledge of Python (object oriented programming, unit testing, Jupyter Notebooks, docstrings). Proficient at git version control (merging, branching, etc.), unix command line, Proficient at SQL and R, experience using PySpark, language models (BERT) and natural language processing tools (allenNLP)

Experience with data warehousing and analysis platforms (SQL Server Management Studio, PostgreSQL, Palantir Foundry), Electronic Data Capture (REDCAP), OHDSI querying, phenotyping, & analysis tools (ATLAS, CHARYBDIS, HADES, Data Quality Dashboard), ontology editing tools (Protege), biomedical knowledge bases, vocabularies (UMLS, ATHENA, LOINC, SNOMED, etc.) and data models (OMOP, FHIR)

Experience generating and/or analyzing synthetic data to assess its fitness for purpose (Synthea, MDClone, Synthetic Data Vault) and deep familiarity with

Timeline (left column)

2013-06 - 2015-03

Electrocardiogram Technician & Clinic Associate
ZoomCare, Portland, OR

- Performed electrocardiograms, blood draws, rapid tests, in-person scheduling/billing, training of >10 new employees and creation/updates to standard operating procedure documents.

2012-06 - 2016-06

Executive Director & Race Director
Glow XC Nonprofit, Eugene, OR

From 2014-2016 I served as both Executive Director of the Glow XC 501(c)3 non-profit and as Race Director for our yearly charity 7k. I led an organizing committee of 5-10 members to fundraise, plan & oversee the charity race, and increase statewide awareness for our target cause of rural emergency services responder training grants. One of my two radio interviews speaking about Glow XC on the second highest rated station in Eugene-Springfield, OR area can be found here: https://soundcloud.com/glow-xc. In 2012 I was a sponsorship and marketing volunteer for the race in its previous form, the 'Moonlight Five. In 2013 I was the sponsorship and marketing director as well as a board member for Glow XC.

2012-02 - 2013-06

Volunteer Research Assistant
University Of Oregon Dept Of Human Physiology, Eugene, OR

- Conducted & recorded results of VO2 max exercise tests, hypobaric simulations with human subjects
- Ran blood gas analysis, measured hematocrit, collected plasma

2011-10 - 2012-01

Volunteer Clinic Support Staff
White Bird Medical Clinic, Eugene, OR

- Aided healthcare delivery to mainly low-income & special needs population
- Assembled prescriptions for patients and kept track of inventory
- Assisted clinic-wide conversion of paper medical records to digital files

Publications

2021-06

Foraker RE, Guo A, **Thomas JA**, et al. Analyses of Original and Computationally-Derived Electronic Health Record Data: The National COVID Cohort Collaborative. JMIR Preprints Published Online First: 2021.https://preprints.jmir.org/preprint/30697).

2021-07

Thomas JA, Foraker RE, Zamstein N, et al. Demonstrating an approach for evaluating synthetic geospatial and temporal epidemiologic data utility: Results from analyzing 1.8 million SARS-CoV-2 tests in the United States National COVID Cohort Collaborative (N3C). medRxiv Published Online First: 2021. doi:10.1101/2021.07.06.21259051

2021-06

Zhang L, Ngo A, **Thomas JA**, et al. Neuropsychological test validation of speech markers of cognitive impairment in the Framingham Cognitive Aging Cohort. Exploration of Medicine 2021;:232–52. doi:10.37349/emed.2021.00044

2021-06

Haq KT, Rogovoy NM, **Thomas JA**, et al. Adaptive Cardiac Resynchronization Therapy Effect on Electrical Dyssynchrony (aCRT-ELSYNC): A randomized controlled trial. Heart Rhythm O2 2021;2:374–81. doi:10.1016/j.hroo.2021.06.006

2021-03

Prieto-Alhambra D, Kostka K, Duarte-Salles T, et al. Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using CHARYBDIS. Res Sq Published Online First: 1 March 2021. doi:10.21203/rs.3.rs-279400/v1

2021-03

Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. Journal of the American Medical Informatics Association 2021;28:427–43. doi:10.1093/jamia/ocaa196 [consortial authorship only]*

2020-10

Golozar A, Lai LY, Sena AG, et al. Baseline phenotype and 30-day outcomes of people tested for COVID-19: an international network cohort including >3.32

(right column)

tradeoffs in privacy preserving technologies, data access and sharing strategies (federated learning, model-to-data approaches, privacy budgets etc.)

Courses

Human Physiology & Biology courses (anatomy lab with cadavers, systems neuroscience, molecular biology, physiology, etc.)

Organic & General Chemistry (including labs)

Medical Ethics

Applied Biostatistics (descriptive statistics, survival analyses, statistical tests)

Knowledge Representation (OWL/RDF, SPARQL, semantic web, UMLS)

Scientific Supercomputing (cloud computing, parallel programming)

Machine Learning for Health Data (statistical learning, linear regression, resampling methods, model selection and regularization, tree-based methods, clustering)

Deep Learning for Computational Scientists (model training and hyperparameter tuning, use of high-level libraries like Keras)

Security and Privacy (adversarial AI, attacks, anonymity, authentication, web privacy & security)

Organizations

National Covid Cohort Collaborative (N3C) Synthetic Data Workstream
Current Member

- Evaluated synthetic data utility
- Presented findings to workgroups
- Wrote or co-authored two manuscripts currently undergoing revisions
- Provide strategic guidance on N3C synthetic data analyses and future directions

Journal of the American Medical Informatics Association (JAMIA)
Editorial Board Member 2021-2023

- Reviewed 12 manuscripts (was a student editorial board member 2019-2020) so far
- Participate in shaping the direction of the journal through editorial board meeting discussions

Cascadia Data Alliance
Member

- Conducting research on federated learning between health systems

Observational Data Science and Informatics (OHDSI)
Collaborator

- Conduct observational network studies, contribute to open-source code
- Participate in weekly meetings

National Research Network
Member

- Participate in monthly meetings, present research findings, contribute to

million people tested with real-time PCR and >219,000 tested positive for SARS-CoV-2 in South Korea, Spain and the United States. medRxiv 2020;:2020.10.25.20218875. doi:10.1101/2020.10.25.20218875

2020-04 Perez-Alday EA, Haq KT, German DM, et al. Mechanisms of Arrhythmogenicity in Hypertrophic Cardiomyopathy: Insight From Non-invasive Electrocardiographic Imaging. Frontiers in Physiology 2020;11:344. doi:10.3389/fphys.2020.00344

2020-01 **Thomas JA**, Burkhardt HA, Chaudhry S, et al. Assessing the Utility of Language and Voice Biomarkers to Predict Cognitive Impairment in the Framingham Heart Study Cognitive Aging Cohort Data. J Alzheimers Dis 2020;76:905–22. doi:10.3233/JAD-190783

2019-12 Perez-Alday EA, Bender A, German D, et al. Dynamic predictive accuracy of electrocardiographic biomarkers of sudden cardiac death within a survival framework: the Atherosclerosis Risk in Communities (ARIC) study. BMC Cardiovasc Disord 2019;19:255. doi:10.1186/s12872-019-1234-9

2019-07 Kearns W, Lau W, **Thomas J**. UW-BHI at MEDIQA 2019: An Analysis of Representation Methods for Medical Natural Language Inference. In: Proceedings of the 18th BioNLP Workshop and Shared Task. Florence, Italy: : Association for Computational Linguistics 2019. 500–9. doi:10.18653/v1/W19-5054

2019-05 **Thomas JA**, A Perez-Alday E, Junell A, et al. Vectorcardiogram in athletes: The Sun Valley Ski Study. Ann Noninvasive Electrocardiol 2019;24:e12614. doi:10.1111/anec.12614

2019-01 Perez-Alday EA, Li-Pershing Y, Bender A, et al. Importance of the heart vector origin point definition for an ECG analysis: The Atherosclerosis Risk in Communities (ARIC) study. Comput Biol Med 2019;104:127–38. doi:10.1016/j.compbiomed.2018.11.013

2018-11 **Thomas JA**, Perez-Alday EA, Hamilton C, et al. The Utility of Routine Clinical 12-lead ECG in Assessing Eligibility for Subcutaneous Implantable Cardioverter Defibrillator. Comput Biol Med 2018;102:242–50. doi:10.1016/j.compbiomed.2018.05.002

2018-03 Biering-Sørensen T, Kabir M, Waks JW, et al. Global Electrocardiographic Measures and Cardiac Structure and Function: The Atherosclerosis Risk in Communities (ARIC) Study. Circ Arrhythm Electrophysiol 2018;11:e005961. doi:10.1161/CIRCEP.117.005961

2021-03 Kearns WR, **Thomas JA**. Resource and Response Type Classification for Consumer Health Question Answering. AMIA Annu Symp Proc 2018;2018:634–43.

2018-01 Perez-Alday EA, **Thomas JA**, Kabir M, et al. Torso Geometry Reconstruction and Body Surface Electrode Localization Using Three-Dimensional Photography. J Electrocardiol 2018;51:60–7. doi:10.1016/j.jelectrocard.2017.08.035

2017-06 Kabir MM, Perez-Alday EA, **Thomas J**, et al. Optimal Configuration of Adhesive ECG Patches suitable for Long-term Monitoring of a Vectorcardiogram. J Electrocardiol 2017;50:342–8. doi:10.1016/j.jelectrocard.2016.12.005

2017-02 Junell A, **Thomas J**, Hawkins L, et al. Screening Entire Healthcare System ECG Database: Association of Deep Terminal Negativity of P wave in lead V1 and ECG Referral with Mortality. Int J Cardiol 2017;228:219–24. doi:10.1016/j.ijcard.2016.11.128

2016-09 Kabir MM, Sedaghat G, **Thomas J**, et al. Reproducibility of Heart Rate Variability Characteristics Measured on Random 10-second ECG using Joint Symbolic Dynamics. Comput Cardiol (2010) 2016;2016:289–92. doi:10.23919/CIC.2016.7868736

discussions surrounding hospital log data use and standards.

American Medical Informatics Association (AMIA)
Member
- AMIA 2021 Annual Symposium Scientific Program Committee Member
- Attend conferences, will be on a panel discussing N3C Synthetic data this November

Glow XC 501(c)(3) Charity Race
Previous Executive Director
- Ran day to day fundraising and logistics, personnel management and marketing to raise money for rural emergency services in Oregon State