# Jason Thomas, PhD

## Medical Data Scientist | Strategist | Informatician

Porfolio
LinkedIn
Github
Google Scholar

Accomplished business leader and innovator at the intersection of data science & artificial intelligence, medical informatics, and organizational data strategy. I have 12+ years experience in medicine ranging from non-profit executive leadership, frontline healthcare delivery, observational & interventional research, analysis/modeling of real-world data to new product introduction R&D. In just 2.5 years of work in the medical devices industry, I have driven innovation and led lean teams to generate outsized business value through the application of data science and AI methods, and have had a similarly outsized impact on shaping broader data strategy with senior management.

## Education

**PhD Biomedical Informatics; Data Science Specialization**
University of Washington - Seattle WA

**Bachelor of Science: Human Physiology | Biology**
University of Oregon - Eugene OR

## Awards

**Innovator Award Finalist (individual award)**
Philips IGTD Excellence Awards | Oct 2023
*Highest individual R&D Award in IGTD*

**Program of the Year Finalist (team award)**
Philips IGTD Excellence Awards | Oct 2023

**Best Business Impact, Natural Language Processing Category**
Philips Global Data & AI Conference | *Short Paper* | Sep 2023

**Editor's Choice, Research & Applications**
Journal of the American Medical Informatics Association
*Manuscript I wrote as first-author chosen* | May 2022

**Biomedical Informatics & Data Science Pre-Doctoral Fellowship**
National Library of Medicine T15 Grant | Sep 2017
*Full tuition waiver and stipend, ~40 new slots/year*

**Top Scholar Top off Award**
University of Washington | Sep 2017
*One-time extra $ to top 2 recruits/year in the BIME program*

**Young Investigator Award Finalist**
International Congress of Electrocardiology (ICE) 2017
*Global electrical heterogeneity in young athletes | 2017*

## Select Skills & Tools

Artificial Intelligence | Gen AI | Prompt Engineering | Machine Learning | Data Strategy | Data Quality | Fit for Purpose Assessment | Software Engineering Best Practices | Cloud Computing | Stats | Natural Language Processing | Data mining | Claims & Hospital Admin Data | Ontologies | Computable Phenotypes | Python | SQL | Pyspark | R | AWS | Amazon Bedrock | LangChain | Plotly | Seaborn | Sonarqube | Git | Quicksight | Sagemaker | Keras | ScikitLearn | OHDSI tools | Palantir Foundry | REDCap |

## Academic Collaborations

**Journal of the American Medical Informatics Association (JAMIA) Editorial Board Member**
*Invited to join by the editor-in-chief. Only PhD student on the full board of ~60 board members total.*
2021-2023

**National Covid Cohort Collaborative (N3C) Synthetic Data Validation Workstream Team Member**
2020-2021

**JAMIA Student Editorial Board Member**
*Six members. 50+ PhD/Postdoc/MD applied*
*Mentored by director of NEI Michael Chiang*
2019-2021

**AMIA Annual Symposium Scientific Program Committee**
*One of ~30 members. Assign reviewers, final decision on science accepted to the conference*
2021

**Cascadia Data Alliance**
2019-2021

**National Research Network for EHR Audit Log Data**
2019-2021

## Select Work & Medical Research History

**Sep 2023 - Present** — **Tech Lead - Senior Data & AI Scientist**
*Philips Image Guided Therapy Devices*
- Lead team of 4 data Scientist to develop solutions that drive business value
- Lead & own IGTD's Galileo iniative. Key advisor on projects involving data & AI
- Collaborate with senior management to define data & AI strategy and integrate data science insights, AI capabilities & SWE best practices into broader company strategies.
- Awarded $400k (6 teams selected of >30 apps) for cross-deptartment Gen AI project

**Feb 2022 - Aug 2023** — **Senior Data & AI Scientist**
*Philips Image Guided Therapy Devices*
- Designed, launched, led program that achieved payback on $1million by generating evidence at scale from hospital admin data supporting NPIs, reg compliance, marketing
- Selected, influenced infra/vendor choices for multiple projects including Quicksight Q
- Conceived of, wrote data strategy proposals adopted by senior leadership.
- Product manager and lead data scientist on generative AI use cases.
- Operate at the intersection of clinical development & R&D, regulatory affairs, business development & marketing, architecture, strategy, software engineering, data science.
- IP Generation: 253 inner-source contributions 2023: 57% commits,9% pull requests

**Sep 2017 - Sep 2021** — **National Library of Medicine Biomedical Informatics & Data Science Pre-Doctoral Fellow**
*University of Washington Department of Biomedical Informatics & Medical Education*
- Predicted dementia status from FramingHam Heart Study Cognitive Aging Cohort data using acoustic, linguistic and clinical data; identified data utility issues with recordings.
- Wrote, co-designed $1.75mill NIH grant on EHR data quality scored top 38 percentile
- Applied EHR data analytics, predictive modeling, OHDSI model-to-data studies
- Assessed fitness for use of synthetic and real EHR and log data for research & hospital operations using UW Medicine and National Covid Cohort Collaborative (N3C) data
- Contributed to ETL of an OMOP UW Medicine COVID-19 Research Data Warehouse
- Participated in the Medical Natural Language Inference subtask during MEDIQA 2019
- Improved state of the art on health questions answering for GARD dataset

**Apr 2015 - Sep 2017** — **Senior Research Assistant**
*Oregon Health & Science University Knight Cardiovascular Institute - Translational electrophysiology lab*
- Data analysis with python, stata, excel; aid lab mission to predict sudden cardiac death
- Coordinated, helped annotate >100,000 electrocardiograms from ARIC & CHS cohorts
- Assessed human-computer Interaction, accessibility barriers to use of ECG patches in home monitoring & creation of patient-generated self-tracking data in clinical studies.
- Translation of grants and research designs into IRB submissions; wrote study designs
- Recruited >350 patients; 50 in 3.5 days, an RCT, US National Alpine Championships
- Predicted SICD eligibility; Published 3 equal-1st author peer-reviewed journal articles.
- Co-author on multiple retrospective, observational peer-reviewed studies & a RCT
- Data collection: ECGs, 6-minute walk, surveys, chart reviews, device interrogations & intracardiac EGMs (including during cath-lab procedures).
- Obtained >100k ECGs from IT; equal-1st-author retrospective observational study

**Oct 2013 - Jul 2016** — **Executive Director**
*Glow XC 501(c)(3)*
- Cofounder 2013, Executive Director 2014+. 300-person race raising $ for rural health EMS
- Total responsibility for P&L, logistics, legal compliance, 5-10 person team. Radio interviews

**Jun 2013 - Mar 2015** — **Clinic Associate & Electrocardiogram Technician**
*ZOOM+Care*
- Worked at >15 different clinics performing electrocardiograms, blood draws, rapid tests, in-person scheduling/billing, training of >10 new employees and process improvements

**Feb 2012 - Jun 2013** — **Volunteer Cardiopulmonary & Respiratory Research Assistant**
*University of Oregon Human Physiology Dept Cardiopulmonary & Respiratory Lab*
- Conducted & recorded results of V02 max exercise tests and altitude chamber studies with human subjects, processed lab specimens, subject recruiting and scheduling, data analysis

**Sep 2010 - Jun 2013** — **Facility Manager**
*University of Oregon Student Recreation Center*
- Managed ~10 direct reports per shift in a 250k ft2 facility, first responder & responsible for safety of all students & staff, developed new hiring process to screen 700 applicants

## Peer-Reviewed Journal Articles

**2022**

- **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). J Am Med Inform Assoc. 2022;29:1350–1365. ***Editor's Choice***

**2021**

- Foraker R, Guo A, **Thomas J**, et al. The National COVID Cohort Collaborative: Analyses of Original and Computationally Derived Electronic Health Record Data. J Med Internet Res. 2021;23:e30697.
- Zhang L, Ngo A, **Thomas JA**, et al. Neuropsychological test validation of speech markers of cognitive impairment in the Framingham Cognitive Aging Cohort. Explor Med. 2021;2:232–252.
- 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;2:374–381.
- 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. 2021:rs.3.rs-279400. **[Thomas JA in et al.]**
- Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021;28:427–443. **[consortial authorship only]***
- Perez-Alday EA, Haq KT, German DM, et al. Mechanisms of Arrhythmogenicity in Hypertrophic Cardiomyopathy: Insight From Non-invasive Electrocardiographic Imaging. Front Physiol. 2020;11:344. **[Thomas JA in et al.]**

**2020**

- **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;76:905–922. [‡Equal First-author]

**2019**

- 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. **[Thomas JA in et al.]**
- **Thomas JA‡,** A Perez-Alday E‡, Junell A, et al. Vectorcardiogram in athletes: The Sun Valley Ski Study. Ann Noninvasive Electrocardiol. 2019;24:e12614. [‡Equal First-author]
- 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;104:127–138. **[Thomas JA in et al.]**

**2018**

- **Thomas JA‡**, Perez-Alday EA‡, Hamilton C, Kabir MM, Park EA, Tereshchenko LG. The utility of routine clinical 12-lead ECG in assessing eligibility for subcutaneous implantable cardioverter defibrillator. Comput Biol Med. 2018;102:242–250. [‡Equal First-author]
- Biering-Sørensen T, Kabir M, Waks JW, et al. Global ECG Measures and Cardiac Structure and Function: The ARIC Study (Atherosclerosis Risk in Communities). Circ Arrhythm Electrophysiol. 2018;11:e005961. **[Thomas JA in et al.]**
- Perez-Alday EA, **Thomas JA**, Kabir M, et al. Torso geometry reconstruction and body surface electrode localization using three-dimensional photography. Journal of Electrocardiology. 2018;51:60–67.

**2017**

- Kabir MM, Perez-Alday EA, **Thomas J**, Sedaghat G, Tereshchenko LG. Optimal configuration of adhesive ECG patches suitable for long-term monitoring of a vectorcardiogram. J Electrocardiol. 2017;50:342–348.
- 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;228:219–224. [‡Equal First-author]

## Peer-Reviewed Conference Proceedings

**2023**

- **Thomas JA**, Hilton DB. *REDACTED*. Philips Global Data and AI Conference. 2023. **[Best Business Impact - NLP Category Award]**

**2019**

- Kearns WR, Lau W, **Thomas JA**. UW-BHI at MEDIQA 2019: An Analysis of Representation Methods for Medical Natural Language Inference. Proceedings of the 18th BioNLP Workshop and Shared Task. Florence, Italy. Association for Computational Linguistics. 2019.

**2018**

- Kearns WR, **Thomas JA**. Resource and Response Type Classification for Consumer Health Question Answering. AMIA Annu Symp Proc. 2018;2018:634–643.

**2016**

- Kabir MM, Sedaghat G, **Thomas J**, Tereshchenko LG. Reproducibility of Heart Rate Variability Characteristics Measured on Random 10-second ECG using Joint Symbolic Dynamics. Comput Cardiol (2010). 2016;2016:289–292.

## Rxiv publications

- 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 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. **[Thomas JA in et al.]**

## Select Presentations, Invited Talks & Panels

**2024**

- *REDACTED*| Team presentation | **US Food & Drug Administration** | 2024

**2023**

- *REDACTED* | Team presentation | **Philips Data & AI Community of Practice** | Dec 2023
- BU Tech Talk: Image-Guided Therapy Devices | Invited talk | **Philips Global Data & AI Conference** | 2023
- *REDACTED* | Conference Short Paper | **Philips Global OCUPAI Conference** | 2023 | **Best Business Impact, Natural Language Processing Category**
- *REDACTED* | Invited talk | **Philips Image Guided Therapy Management Team** | 2023
- Global Electrical Heterogeneity in Young Athletes | Conference Abstract | **International Congress on Electrocardiology 2017** | 2017 | **Young Investigator Award Finalist**

**2021**

- Panel: LB06 Panel - Demonstrations in Synthetic Data and the National COVID Cohort Collaborative (N3C) | Panel | **AMIA Annual Symposium** | 2021
- Assessing Single Sign-on Authentication Behaviors to Inform Customized Solutions Using Real & Synthetic Log Data | Invited talk | **National Research Network for EHR Log Data** | 2021