Cheng-Han Yang

Website: https://cyang728.github.io
Email: cheng-han.yang@yale.edu
GitHub: github.com/cyang728
ORCID: 0000-0002-4161-3140

Research Interests

EHR data analysis, Survival analysis, Bayesian dose-finding design, Causal inference

Education

University of Texas Health Science Center Houston, TX
Ph.D. in Biostatistics and Data Science, Advisor: Ruitao Lin 2021–2025

- Thesis: "Methodological Advances in Early Phase Clinical Trial Design and Analysis."
- Graduate Research Assistant at The University of Texas MD Anderson Cancer Center.

National Tsing Hua University Hsinchu, Taiwan
M.S. in Statistics, Advisor: Yu-Jen Cheng 2015–2017

- Thesis: "Regularized Interaction Boosting with Right Censored Data."

Experience

Yale University New Haven, CT
Associate Research Scientist (Mentor: Dr. Bhramar Mukherjee) Aug 2025 –Present

- Developing joint models to address informative presence and observation missingness mechanism in EHR data.
- Proposing a Missing at Random test for clinical reocording process in EHR data.

The University of Texas MD Anderson Cancer Center Houston, TX
Graduate Research Assistant (Mentor: Dr. Ruitao Lin) Aug 2021 –Aug 2025

- Bridged short-term surrogates and long-term survival in dose-finding by developing "DEMO," a Phase I-II Bayesian design modeling biological markers as causal mediators.
- Mitigated the conservatism of standard BLRM designs by developing "Switch-BLRM" with a dose-switching rule that expands the admissible dose set.
- Improved dose-response estimation under baseline heterogeneity by deriving a covariate-adjusted g-computation estimator with closed-form variance, extending MCP-Mod.
- Enabled robust dose estimation across heterogeneous early-phase trials by developing "REDOMA," a Bayesian meta-analysis framework with spike-and-slab priors.

Merck & Co., Inc. Upper Gwynedd, PA
Biostatistics Intern May 2024 –Aug 2024

- Informed regulatory strategy by evaluating FDA case studies on leveraging external clinical trial data for oncology drug development.
- Optimized externally controlled trials by proposing the Data-Covariate-Outcome (DCO) framework for integrating historical and real-world data.

Publications

Methodology Papers

[1] S McGrath, **CH Yang**, J Kimmelman, O Ozturk, R Steele, and A Benedetti. "Meta-analysis of Median Survival Times with Inverse-Variance Weighting". In: (2026). Accepted by *Statistics in Medicine*.

[2] **CH Yang**, PF Thall, D Marin, SY Belay, and R Lin. "BAR12: Bayesian Autoregressive Phase 1-2 Design for Cell Therapy Trials with Manufacturing Changes". In: (2026). Accepted by *Statistics in Medicine*.

[3] **CH Yang**, PF Thall, and R Lin. "Dose exploration, monitoring, and optimization using biological and clinical outcomes". In: *The Annals of Applied Statistics* 19.4 (2025), pp. 2599–2617.

[4] **CH Yang**, G Cheng, and R Lin. "On the relative conservativeness of Bayesian logistic regression method in oncology dose-finding studies". In: *Pharmaceutical statistics* 23.4 (2024), pp. 585–594.

[5] **CH Yang** and YJ Cheng. "A model-free variable screening method for optimal treatment regimes with high-dimensional survival data". In: *Biometrika* 111.4 (2024), pp. 1369–1386.

[6] **CH Yang**, E Kwiatkowski, JJ Lee, and R Lin. "REDOMA: Bayesian random-effects dose-optimization meta-analysis using spike-and-slab priors". In: *Statistics in medicine* 43.18 (2024), pp. 3484–3502.

Collaborative Papers

[7] DN Yeboa, BT Whitfield, R Lin, CL Ejezie, TA Swanson, TH Beckham, C Wang, B De, S Perni, MC Tom, J Li, SL McGovern, R Harrison, NK Majd, VK Puduvalli, AE Aaroe, M Loghin, BJ O'Brien, AD Patel, CB Patel, JS Wefel, C Altintas Taslicay, M Gule-Monroe, AC Paulino, MF McAleer, DR Grosshans, AJ Ghia, W Jiang, C Chung, M Maor, **CH Yang**, MA Gubbiotti, C Kamiya-Matsuoka, LY Ballester, SP Weathers, and JT Huse. "Prospective phase II clinical trial of molecular glioblastoma (historical grade 2 and 3 IDH wildtype gliomas) preliminary novel exploratory analyses". In: *Journal of Neuro-Oncology* 176.1 (2026), pp. 1–11.

[8] KM Moody, MC Swartz, Z Gresham, M Askins, L Cahalan, C Geistkemper, A Heaton, R Lin, M Melo, A Rajan, M Smith, P Tewari, K Williams, **CH Yang**, and R Robert. "A Virtual Wellness Program to Enhance Well-being for Pediatric Oncology Staff during the COVID-19 Pandemic". In: *Adv Cancer Educ Qual Improv* 1.1 (2025), a18. DOI: 10.52519/ACEQI.25.1.1.a18.

Preprints / Under Review

*Equal contribution

[9] X Chen, **CH Yang**, Z Wu, and B Mukherjee. *Prediction-based Inference in Electronic Health Record (EHR)-linked Biobanks*. 2026. arXiv: 2603.14356.

[10] J Du*, **CH Yang***, X Shi, and B Mukherjee. "Analyzing Longitudinal Outcomes from Electronic Health Records: A Joint Modeling Approach to Handle Informative Visits and Observations". Revised and resubmitted to *Journal of the American Statistical Association*. 2026.

[11] **CH Yang**, X Shi, and B Mukherjee. *Joint Modeling of Longitudinal EHR Data with Shared Random Effects for Informative Visiting and Observation Processes*. 2026. arXiv: 2602.15374.

[12] Y Hou, T Ward, **CH Yang**, E Jernigan, G Caturegli, D Boffa, and B Mukherjee. *Assessing Statistical Practices of Existing Artificial Intelligence (AI) Models for Lung Cancer Detection, Prognosis, and Risk Prediction: A Cross-Sectional Meta-Research Study Supplemented by Human and Large Language Model (LLM)-Directed Quality Appraisal*. 2025. DOI: 10.64898/2025.12.23.25342849v1. medRxiv: 2025.12.23.25342849.

Invited Talks

- **Yang CH**. "A Joint Modeling Approach for Longitudinal EHR Data with Informative Visiting and Observation Processes." Yale Public Health Data Science and Data Equity Seminar, New Haven, CT, January 2026.
- **Yang CH**. "A Joint Modeling Approach for Longitudinal EHR Data with Informative Visiting and Observation Processes." Penn Medicine, University of Pennsylvania, Philadelphia, PA, March 2026.

Poster Presentations

- **Yang CH**, Thall PF, Lin R. "Dose-Finding Design Guided by Early Biological Mediators and Clinical Outcomes." ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop, Rockville, MD, Sep 2025.
- **Yang CH**, Thall PF, Lin R. "Mediator-Supported Design for Exploring and Optimizing Doses in Phase I-II Trials." Statistics in Pharmaceuticals (SIP) Conference, Storrs, CT, Aug 2025.

Awards and Honors

- **Elected Full Member**, Sigma Xi, The Scientific Research Honor Society 2025
- **Inducted Member**, Delta Omega Honorary Society in Public Health 2025
- **Student Travel Grant**, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop 2025
- **Scholarship Award**, North America Taiwanese Engineers' Association (NATEA) 2023
- **Robert H. Bigelow Endowed Scholarship**, UT Health Science Center at Houston 2022, 2024
- **Academic Excellence Award**, National Tsing Hua University 2016, 2017
- **Merit Scholarship**, National Tsing Hua University 2015

Teaching Experience

National Tsing Hua University Hsinchu, Taiwan
Graduate Teaching Assistant 2017 –2021

- STAT661000 Survival Analysis (Spring 2020, Spring 2021)
- STAT519100 Applied Multivariate Analysis (Spring 2017, Spring 2019)
- STAT541000 Linear Model (Fall 2018)

National Tsing Hua University Hsinchu, Taiwan
College Teaching Assistant 2016 –2020

- ECON303300 Econometrics (Fall 2019)
- STAT202000 Statistics (Fall 2016, Spring 2020)