Qingze Gu

+65 83108375 qingze.gu@outlook.com qingzegu.com 0000-0002-3024-3906 guqingze
Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, 308232

Profile

- Clinical epidemiologist and health data scientist across infection, rare disease, pharmacology and population-scale EHR.
- First or joint author on studies spanning linked national records, hospital EHR, biomarkers and clinical NLP/LLMs.
- Apply longitudinal modelling, causal inference and ML to translate healthcare data into clinical decision support.

Education

University of Oxford **UK**
Doctor of Philosophy in Clinical Medicine (Biomedical Data Science) *Oct. 2020 – Oct. 2024*
Thesis entitled "Exploiting electronic health records to improve infection management". Passed viva with no corrections.

University of Oxford **UK**
Master of Science in Pharmacology; Grade: Distinction *Oct. 2019 – Sep. 2020*

Harbin Institute of Technology **China**
Bachelor of Engineering in Pharmaceutics and Food; GPA: 88/100 (Ranking 2/23) *Sep. 2015 – Jul. 2019*
Six-month exchange programme at Instituto Superior Técnico (Portugal), studying Biological Engineering.

Skills

Programming: R, Python, SQL, PySpark, Databricks, Git.

Statistics: Clinical Epidemiology, Survival Analysis, Causal Inference, Mixed/Latent-Class Models.

Machine Learning & NLP: XGBoost, Transformers/LLMs (BERT, GPT), Clinical NLP, OMOP/RxNorm Mapping.

Research Experience

Nanyang Technological University **Oct. 2025 – Present**
Research Fellow *Singapore*
- Curate and quality-check 100,000-participant (SG100K) research phenotypes and link them with national EHR within TRUST, the secure research environment where this linkage and analyses are carried out.
- Built a reusable, release-versioned endpoint-analysis framework (shared base with CVD- and cancer-specific instantiations) to standardise cohort phenotyping and study outputs.
- Developed LLM and rule-based pipelines normalising 35K+ free-text and self-reported medications to generic drug names and OMOP RxNorm concepts for SG100K and TRUST medication data.
- Conduct statistical and ML analyses on high-dimensional SG100K data (labs, imaging, genomics, EHR), including FibroScan-based liver epidemiology in a multiethnic cohort.

University of Oxford **Oct. 2024 – Sep. 2025**
Postdoctoral Health Data Scientist *Oxford, UK*
- Conducted large-scale rare- and chronic-disease EHR research in the NHS England SDE and CPRD.
- Estimated prevalence of 406 rare diseases & COVID-19 burden across ethnicities using data from 62M individuals.
- Built Databricks/PySpark/R/SQL pipelines for population-scale data processing, standardisation and analyses.
- Analysed hypertension–diabetes multimorbidity trajectories to reveal clinical heterogeneity.

Laboratory of Data Discovery for Health ($D^2$4H) **Jun. 2025 – Oct. 2025**
Consultant in Clinical NLP *Remote/Hong Kong*
- Developed and validated LLM-based pipelines for automated TNM cancer staging from oncology clinical notes.
- Led end-to-end data curation: OCR extraction, annotation-schema design and gold-standard labelling.
- Engineered an evaluation harness that improved M-stage extraction accuracy from 50% to 96%.

University of Oxford **Oct. 2020 – Sep. 2024**
PhD Researcher (Biomedical Data Science) *Oxford, UK*
- Characterised pathogen-specific inflammatory-marker and vital-sign trajectories in suspected sepsis via latent class analysis on five years of EHR data, deriving reference charts to guide infection management.
- Applied transformers and LLMs (BERT, GPT) to antibiotic prescribing indications to infer sources of infection from free-text EHRs, benchmarking against ICD-10 coding.
- Evaluated and refined an institutional vancomycin dosing guideline using regression, survival analysis, and population-pharmacokinetic simulation to optimise drug-level targets.

Industry Experience

IQVIA **Jul. 2024 – Aug. 2024**

Real-World Solutions Intern *Beijing, China*

- Assessed regional EHR databases and built feasibility table shells to support real-world study planning.
- Performed desk research on product phases, clinical trials, market status, and disease burden.
- Conducted literature review on patient reported outcomes, created slides and drafted expert interview guide.

Cerner Enviza, an Oracle company **Sep. 2023 – Mar. 2024**

Real-World Evidence Intern *Shanghai, China*

- Assembled data on treatments and marketed drugs for survey-based studies.
- Assisted in developing case report forms and calculating health economics indicators.
- Drafted proposals & reviewed protocols analysing patient characteristics, disease burden and treatment patterns.

Peer-Reviewed Publications

Gu Q, Yuan K, Wei J, Yoon CH, Danielsen AS, Luk A, Eyre DW, Walker AS. Interplay between C-reactive protein responses and antibiotic prescribing in people with suspected infection. *BMC Infectious Diseases*, 2025. DOI: 10.1186/s12879-025-11381-9. (IF: 3.0).

Yuan K[*], Yoon CH[*], **Gu Q**[*], Munby H, Walker AS, Zhu T, Eyre DW. Transformers and large language models are efficient feature extractors for electronic health record studies. *Communications Medicine*, 2025. DOI: 10.1038/s43856-025-00790-1. [*]Coauthors, contributed equally. (IF: 5.4).

Wei J, Zhou J, Zhang Z, Yuan K, **Gu Q**, Luk A, Brent AJ, Clifton DA, Walker AS, Eyre DW. Predicting individual patient and hospital-level discharge using machine learning. *Communications Medicine*, 2024. DOI: 10.1038/s43856-024-00673-x. (IF: 5.4).

Danielsen AS, **Gu Q**, Fostervold A, Eyre DW, Bjørnholt JV. 'Bloodstream infection': a valuable concept we should keep in our toolbox. *Journal of Infection*, 2024. DOI: 10.1016/j.jinf.2024.106236. (IF: 14.3).

Gu Q, Wei J, Yoon CH, Yuan K, Jones N, Brent A, Llewelyn M, Peto TE, Pouwels KB, Eyre DW, Walker AS. Distinct patterns of vital sign and inflammatory marker responses in adults with suspected bloodstream infection. *Journal of Infection*, 2024. DOI: 10.1016/j.jinf.2024.106156. (IF: 14.3).

Wei J, Uppal A[†], Nganjimi C[†], Warr H[†], Ibrahim Y[†], **Gu Q**[‡], Yuan H[‡], Rahman NM, Jones N, Walker AS, Eyre DW. No evidence of difference in mortality with amoxicillin versus co-amoxiclav for hospital treatment of community-acquired pneumonia. *Journal of Infection*, 2024. DOI: 10.1016/j.jinf.2024.106161. [†‡]Contributed equally. (IF: 14.3).

Gu Q, Jones N, Drennan P, Peto TE, Walker AS, Eyre DW. Assessment of an institutional guideline for vancomycin dosing and identification of predictive factors associated with dose and drug trough levels. *Journal of Infection*, 2022. DOI: 10.1016/j.jinf.2022.06.029. (IF: 14.3).

Preprints

Gu Q, Hasheminasab SA, Pineda-Moncusí M, Chilala C, Thygesen JH, Wu H, Khalid S. Prevalence of 406 rare diseases by ethnicity and their associated COVID-19 infection burden: a national cross-sectional study of 62.5 million people in England. *Preprint, medRxiv*, 2026. DOI: 10.64898/2026.01.13.26344068. On behalf of the CVD-COVID-UK/COVID-IMPACT Consortium.

Academic Activities

Poster Presentation: Yoon CH, Yuan K, **Gu Q**, Munby H, Zhu T, Eyre DW, Walker AS. Using Natural Language Processing on drug indications to predict working sources of infection. *Machine Learning for Healthcare*, New York, USA. August 2023.

ePoster Flash Talk: **Gu Q**, Pouwels KB, Walker AS, Eyre DW. What is a "normal" C-reactive protein response to sepsis? An electronic health records study. *33rd European Congress of Clinical Microbiology and Infectious Diseases*, Copenhagen, Denmark. April 2023.