Sunan Gao

sgao57@jh.edu || +1 (667)335-6100 || Baltimore, Maryland,

EDUCATION

***Johns Hopkins University*,** Bloomberg School of Public Heath — Expected 06/2025

Major: MHS in Epidemology (GPA: 3.9/4.0)

- **Courses:** Methods in Biostatistics (I-III), EPI Methods (I-III), Probability/Statistical Inference (II-III), Statistical Programming, Computer Science for Bioinformatics, Data Structures, EPI of Aging

Certificate in Clinical Trial (Course: Clinical Trial Design, Power &Sample Size) — Expected 10/2024

***Renmin University*,** School of Statistics — 09/2022-07/2023

Major: MS in Biostatistics and Epidemology (Transfer, GPA: 3.9/4.0)

- **Courses**: Causal Inference, Advanced Mathermatics Statistics, Statistical Computing, Biostatistics, Generalized Linear Model, Survival Analysis, Advanced Model Selection, Design of Clinical Trial

***Zhejiang University*,** School of Biosystems Engineering & Food Science — 09/2018-06/2022

Major: BS in Food Science and Engineering (GPA: 4.0/4.0) **Minor:** Psychology (GPA: 4.0/4.0)

- **Courses**: Statistical Machine Learning, Probability Theory, Data Mining, Multivariable Calculus, Linear Algebra, Bioinformatics, Molecular Biology, Biochemistry, Psychometrics, Developmental psychology, Social psychology
- **Awards:** 2019-2022 National Scholarship (1/50, 3 times), Outstanding Undergraduate Thesis (1/150), Provincial Innovation Competition Gold Medal (top 1%, twice), Chu Kochen Scholarship Nomination (18/7000, Top 0.2%)

SELECTED WORK (Co-First/First Author: 3; Other Author: 7)

- **Sunan, G.**, Heming D., Shaobo W., ... Yu, W. (2023). Effects of accelerated biological aging on depressive symptoms in a causal reasoning framework. ***J AFFECTIVE DISORDERS****, 339*, 732–741.
- Qing, Y.[1], **Sunan, G.[1]**, Junfen, L., ... Ming, C. (2023). A machine learning-based data mining in medical examination data: a biological features-based biological age prediction model. ***BMC BIOINFORMATICS****, 23*(1), 411.
- Heming D., **Sunan, G**. ***CRAN R Package*** (2023). uotm: Uncertainty of Time Series Model Selection.
- Xuzhi, W., Yiju, Z., **Sunan, G.**, ... Yu, Z. (2023). Machine learning prediction of exposure to acrylamide based on modelling of association between dietary exposure and internal biomarkers. ***FOOD CHEM TOXICOL****,170:* 113498.
- **Sunan, G.,** Tongxu L., Heming D., ... Yu, W. (2023). Serum metabolites mediate causal relationships between intestinal flora and non-cognitive and cognitive features: Mutilvariable mendelian randomization. (***Under Review***)
- **Sunan, G.**, Yimei T., Xuzhi W., ... Yu, Z. (2023). Urinary metabolic analysis under 3-Chloro-1,2-propanediol di-ester and Eicosapentaenoic acid co-exposure: based on non-targeted metabolomics and molecular docking. (***Under Review***)
- Yimei T. [†], **Sunan, G. [†]**, Xuzhi W., ... Yu, Z. (2023). Internal and external exposure associations of 3-monochloropropane-1,2-diol and glycidol using machine learning. (***Under Review***)
- **Sunan, G.**, Mingyi, Y., Zisheng, L., ... Li, L. (2022). Soy protein/chitosan-based microsphere as stable biocompatible vehicles of oleanolic acid: an emerging alternative enabling the quality maintenance of minimally processed produce. ***FOOD HYDROCOLLOID****, 124*(Part B), 107325.

PROJECT EXPERIENCE

Aging-related Disease and Objective Physical Activity — 12/2023-Now

Research Assistant (Johns Hopkins University, *Supervised by Dr. Jennifer Schrack & Dr. Ciprian Crainiceanu*)

- Explored the effects from domain-specific cognitive trajectories on objective physical activity pattern.
- Tested the association results between different step algorithms and aging-related diseases.

Evidence for Serum Metabolites Mediate Causal Relationships Between Gut Microbiota and Cognitive Functions: Based on mendelian randomization (MR) and model averaging — 03/2023-10/2023

Research Assistant (Renmin University, *Supervised by Dr. Yu Wang*)

- Identified instrument variables of gut microbiota (n=211), metabolites (n=483) and cognitive features from GWAS.
- Performed bidirectional two-sample MR methods to obtained corrected causal association; Carried out sensitivity and outlier tests to test assumption and detect pleiotropy influence.
- Calculated the indirect and direct mediation effect by Two-step methods; Evaluated causal strength of metabolites by posterior probability through Bayesian model averaging MR (BMA-MR).

Effects of Accelerated Biology Aging on Depressive Symptom (DS) under Causal Frame — 09/2022-04/2023

Research Assistant (Renmin University, *Supervised by Dr. Yu Wang*)

- Biological age was derived ensemble learning and adjusted for chronological age. Adopted mixed-effects logistic regression to obtain correlation results; Used DAG and backdoor criteria to identify confounding factors; Evaluated and controlled covariates by fast large-scale almost matching exactly approach (FLAME) and PSM.
- Recognized risk subgroups by generalized mixed-effects model tree to support precision health.

UOTM: Uncertainty in Time Series Model Selection Methods (R package) 09/2022-02/2023

Research Assistant (Zhejiang University, *Supervised by Dr. Yu Wang*)

- Quantify the uncertainty of model selection in an autoregressive moving average model; Estimated measure based on Bootstrap, during which Monte-Carlo simulation and real data were used to realize simulation and verification

New Aging Indicators based on EHR data and Ensemble Learning 02/2022-09/2022

Team Project (Zhejiang University, *Supervised by Dr. Ming Chen*)

- Compared interpolation effects of AutoEncoder, KNN, MICE, etc. in EHR data (70000+ entries with 100+ features) under simulated MCAR and MNAR.
- Performed variable selection to identify age-related features; Developed STK-BA based on Stacking with GAM, GBDT, CNN, etc. with meta-model (GAM) to improve internal stability; Adopted Poisson and logistic model to highlight benefits in illustrating diseases and lifestyle effects.

Biomarker-Based Predictive Models for Typical Food Contaminants 12/2021-08/2022

Research Assistant (Zhejiang University, *Supervised by Dr. Yu Zhang & Dr. Jingjing Jiao*)

- Measured concentrations of vivo biomarkers of acrylamide and 3-MCPD from urine samples using UHPLC-MS/MS.
- Compared models by MLR, SVM, Random Forest, XGBoost, LightGBM and CatBoost algorithms with internal exposure data to predict dietary exposure to acrylamide and assessed their performance using R and Python.

Metabolic Profile Analysis of Contaminant Exposure and Dietary Interventions Based on Non-Target Metabolomes and Computational Biology 10/2021-06/2022

Undergraduate Thesis (Zhejiang University, *Supervised by Dr. Yu Zhang & Dr. Jingjing Jiao*)

- Compared metabolic profile between single and co-exposure; Discovered the differential metabolites by OPLS-DA and volcano plots; Conducted enrichment analysis to explore affected metabolic pathways.
- Proposed metabolic fingerprints to analyze exposure differences; identified metabolic enzymes that linked dysregulated metabolites based on KEGG and PDB databases; Explored the action mechanism by CB-Dock 2 and AutoDock Vina.

Brain Targeted Drug Delivery Based on Computer Aided Virtual Screening 02/2019-08/2020

Research Assistant (Zhejiang University, *Supervised by Dr. Baiyi Lu*)

- Compared the applicability of force field simulation methods; Screened 40000+ molecules by using affinity, clustering and similarity indicators; Visualized the 2D and 3D interactions on Pymol.
- Constructed modified Nano-carriers, and assisted to conduct in vivo imaging and cell uptake experiments

WORKING EXPERIENCE

Investment Analyst Intern (Healthcare) 06/2023-08/2023

K2VC (K2 Venture Partners), Beijing

- Deeply conducted 10+ technology analysis, covering in molecular biology, vitro diagnostics, medical imaging, etc.
- Using Python, crawler and ChatGPT to optimize automatic information extraction, judgment, and text integration

Data Analysis Intern 09/2021-03/2022

CDC in Zhejiang Province, Hangzhou

- Introduced multi-omics analysis and identified glucolipid metabolism disorders related markers by dose-response relationship from 24000+ potential variables; Responsible for visualization and model verification.

Entrepreneurial Team Co-founders 05/2020-12/2021

AnGene Care Co., Ltd, Hangzhou

- Promoted portable medical testing devices; Responsible for incubation, business plan and enterprise cooperation.
- Participated in national and provincial start-up competitions three times (2 Gold and 1 Silver Gold), took charge of road shows, and contacted angel investors; Our teams were recommended by 6 academic and business mentors

SKILLS

- **Programming Skills**: R, Python, SQL, SAS, Linux, SPSS, SLURM, Databricks
- **Visualization Skills:** Photoshop, InDesign, LaTex, Markdown, Prism, Origin, Microsoft Office, OmniGraffle
- **Version Control & IDE:** GitHub, Anaconda, Visual Studio, PyCharm
- **Language:** English, Mandarin (Native)