

Soo Min You

New York, NY 10032

sy3269@cumc.columbia.edu | +1-917-431-8855

Education

- 09.2024 – 05.2026 **Columbia University, New York, NY**
Master of Science in Biostatistics (expected)
- 09.2013 – 05.2017 **Cornell University, Ithaca, NY**
Bachelor of Science in Chemical Engineering

Research & Project Experience

- 02.2025 – Present **Research Assistant, Department of Biostatistics, Columbia University**
Advisor: Dr. Tian Gu
Project: Evaluation of time series versus general-purpose LLM on health outcome prediction. Develop risk prediction models using embeddings generated from accelerometer data collected from wearable devices and compare performance with time-series foundation models using NHANES data.
- 10.2025 **NYC Preparedness & Recovery Institute Narrative 2 Numbers (N2N): A Mixed Methods Hackathon, New York, NY**
Project Title: 'Accelerating Case Definition and Outbreak Detection Through Community-Level Data Mining'
Developed a proactive disease surveillance pipeline leveraging unstructured, community-level data for rapid disease detection and phenotype definition.
Awarded the Public Health Impact Prize and named runner-up with an AWS prize of \$1,000.
- 06.2015 – 08.2015 **Associate Intern, Korea Railroad Research Institute**
Conducted data collection and analysis for research activities, and supported resolving technical issues and maintaining accurate laboratory records and reports.
- 06.2013 – 07.2013 **Laboratory Intern, Department of Chemical Engineering, Sungkyunkwan University**
Supported laboratory research on solar and hydrogen energy by assisting with photocatalytic water-splitting experiments, collecting results, and organizing data.
- 07.2012 – 08.2012 **Laboratory Intern, Department of Biochemistry, Yonsei University**
Assisted in gene-targeting research by producing and analyzing genetically modified mice and performing DNA extraction, mini prep, and midi prep using standard analytical equipment.

Work Experience

- 04.2021 – 05.2024 **Senior Scientist, Quality Assurance, Samsung Biologics, Korea**
Qualified and released new API manufacturing facilities in Plant 4, accelerating readiness by more than a year.

Led validation lifecycle for a new mRNA manufacturing facility, shortening timeline by 2 months through modified verification strategies.
Cleaning validation: established cleaning limits for drug substances/products based on toxicity, degradation, and TOC recovery, minimizing cross contamination and ensuring quality and safety.
Risk management: reviewed and approved 20 “Deviation” records to capture policy violations 10 “Change Control” records annually to improve compliance and operational efficiency.

07.2017 – 04.2021

Scientist, Validation, Samsung Biologics, Korea

Computerized System Validation: Supported verification of automated systems such as Manufacturing Control System (MCS) and Building Management System (BMS).

Equipment Validation: Reviewed and approved 30+ verification protocols/reports and implemented centralized laboratory system (LabX).

Data Integrity: Deployed backup servers and security controls to support compliance with 21 CFR Part 11 and Annex 11.

Technology Transfer: Assessed allowable TOC residue for 10 newly introduced products.

Selected Courseworks Projects

05.2025

BIST P8106, Data Science II, Columbia University

Project Title: 'Analysis on antibody response to vaccine and factors associated with incidence of flu'

Conducted predictive modeling of severe flu incidence post-vaccination using boosting, SVM, and interpretable methods, developing risk scores and identifying key demographic and clinical predictors.

04.2025

BIST P8158, Latent Variable and Structural Equation Modeling, Columbia University

Project Title: 'Review of mental health practitioners' perceptions of AI analyzed via CFA and SEM'

Critiqued and evaluated a published study on mental health practitioners' perceptions of AI using confirmatory factor analysis (CFA) and structural equation modeling (SEM).

12.2024

BIST P8105, Data Science I, Columbia University

Project Title: 'Analysis of factors on Global Violence'

Conducted a full data science pipeline on a real-world dataset, including data cleaning, exploratory visualization, formal analysis, and dissemination via a Shinyapp website and screencast.

Teaching Experience

05.2022 – 01.2023

Qualified Trainer for Validator, Samsung Biologics, Korea

Provided comprehensive training on thermal data collection equipment validator to internal employees and supported troubleshooting to resolve common operational issues.

08.2018 – 08.2020

Internal Instructor, Samsung Biologics, Korea

Conducted classroom training on validation protocols for operators, providing in-depth knowledge of purpose, scope, testing methods and procedures in compliance with data integrity.

01.2016 – 05.2016 **Peer Tutor, Cornell University**
Peer-tutored freshmen and sophomores in general chemistry.

Extracurricular Activities

07.2025 – 09.2025	Columbia ISSO GLASS Program Completed the Global Leadership Tier 3 Capstone Project through Columbia's International Students and Scholars Office (ISSO), focusing on culturally competent mentorship.
09.2022 – 03.2024	Over the Wall, Samsung Biologics Served as Secretary of a bouldering club, coordinating monthly activities and managing the club budget
07.2016 – 08.2016	Gotoco, Qingdao, China Designed and delivered engaging English as a second language lessons to high school students, enhancing their confidence and academic performance.
12.2015 – 01.2016	Project Teach, Seoul, Korea Provided academic tutoring and mentorship for low-income students.
2014–2016	Cornell Shimtah (Music Club) Performed in annual concerts as a member of a Korean traditional music group.
2013–2017	Cornell 4C (Catholic Community Club) Served as President (2015–2016), coordinating community activities and cultural events.

Certificates & Skills

Software and Programming: R Studio, Python, Microsoft Office, MySQL, LaTeX

Languages:

Fluent in Korean and English
Conversational in French (DELF/DALF B1, May 2020)

Certificates:

Columbia ISSO Global Leadership Tier 1 (06.2025)
Data Integrity Training (Nov 2019), Cleaning Validation (03.2023)
Coursera: Chemical Biology from University of Geneva (04.2023)
Coursera: Python Data Structures from University of Michigan (04.2021)
Coursera: Programming for Everybody from University of Michigan (02.2021)