Bojiang Zhang

张泊江 · Clinical Data Engineer — Oncology · CDISC SDTM/ADaM · EDC

Yokohama, Japan · bojiang.zhang.jp@gmail.com · LinkedIn · bojiang.org · github.com/hakupao

PROFESSIONAL PROFILE

Oncology clinical data professional with **4 years** across a clinical CRO and a medical-data provider, specialising in **CDISC SDTM/ADaM** mapping, **EDC build** and end-to-end clinical data management for cancer studies. Independently delivered data for **7 oncology studies (10,000+ patients)** sourced from the National Cancer Center Japan. Works in **Chinese (native), Japanese (JLPT N2) and English**, which suits Japan/APAC oncology research. Combines clinical-data expertise with hands-on programming (SQL/Python/VBA) and AI-assisted tooling to raise data quality and shorten delivery timelines.

CORE SKILLS

- **Clinical Data Standards:** CDISC SDTM, ADaM, CDASH, Define.xml, Controlled Terminology, aCRF; Pinnacle 21; MedDRA, WHODrug
- **EDC & Data Management:** EDC build / test / maintenance, eCRF design, edit-check programming, UAT, database & specification design, data cleaning & validation, SOP authoring
- **Programming:** SQL, SAS, Python, VBA (working familiarity: Java, JavaScript, Rust)
- **Databases:** PostgreSQL, SQL Server, MySQL, MS Access (also SQLite, MongoDB)
- **Oncology Domain:** colorectal, gastric, melanoma, lung, liver & pancreatic cancers; Overall Survival (OS) / Progression-Free Survival (PFS) analysis; baseline, treatment, adverse-event (AE) & laboratory tables
- **Regulatory:** PMDA data requirements
- **Languages:** Chinese (native), Japanese (JLPT N2), English (professional working)

PROFESSIONAL EXPERIENCE

Clinical Data Specialist — Oncology, iTms K.K. Yokohama/Tokyo, Japan · Sep 2023 – Present

- Independently handled study data end to end (specification, mapping, validation and outputs) for **7 oncology studies (600–4,300 patients each; 10,000+ total)** using data from the National Cancer Center Japan, to **CDISC SDTM/ADaM** standards.
- Designed study databases and data specifications from protocols to CDISC standards; **mapped and transformed EDC outputs into SDTM/ADaM** datasets, and programmed edit-check validation rules agreed with hospital clinicians.
- Programmed oncology analysis outputs to sponsor specifications: baseline characteristics, treatment, **adverse-event (AE)** and laboratory tables; age, stage and gene-mutation distributions; **Overall Survival (OS)** and **Progression-Free Survival (PFS)** analyses. Used SQL, VBA, Access and Excel.
- **Standardised** the data-handling workflow into reusable checklists and SOPs, replacing undocumented, experience-based practice. This **cut delivery time by ~30%** and shortened the protocol-to-deliverable cycle to as little as **2 months**.
- Built internal data-processing tools for non-technical colleagues and an **AI-assisted SDTM knowledge base** for the team, improving processing speed and accuracy.
- **Onboarded and trained 5+ new data staff** using handbooks I wrote. Worked in Japanese and Chinese with Japan-based (and some Taiwan-based) oncology research, making sure datasets met PMDA requirements.

EDC Developer — Clinical Trials, EPS Corporation (CRO) Japan · Sep 2022 – Aug 2023

- **Built and maintained oncology-trial EDC databases** (e-Catch) across multiple tumour types: colorectal, gastric, melanoma, lung, liver and pancreatic cancers.
- Designed **eCRFs**, programmed **edit-check** validation logic, and ran **UAT** and version maintenance for live studies.
- Delivered user training for EDC system users.

Test / QA Engineer — Optical & Inspection Equipment, DAC-Engineering K.K. Japan · Dec 2017 – Aug 2022

- Calibrated and tested precision optical and inspection equipment, building a foundation in rigorous QA, test protocols and measurement-data integrity.

OPEN-SOURCE PROJECTS

- **SDTM Pedia**: converted the SDTMIG v3.4, SDTM v2.0 and CDISC terminology source files into **293 structured Markdown files** (63 domains, 1,005 codelists, 37,939 CDISC terms) that LLM tools can use directly, with no vector database.
- **SDTM Mapping System**: config-driven ETL in Python/pandas/MySQL. An Excel mapping workbook drives a **7-step pipeline** from raw study exports to CDISC SDTM datasets and an M5 submission package.
- **MedAudit Diff Watcher**: local-first CSV audit tool with folder watching, row/field diffs and an SQLite audit trail.
- More at bojiang.org · github.com/hakupao

EDUCATION

Bachelor of Engineering — Telecommunications Engineering, Shenyang Jianzhu University — China · 2013 – 2017

School of Information & Control Engineering