

DOC	JD-HMT-A03
TYPE	Internship position description
TRACK	Healthcare & MedTech
LEVEL	Student / early career

Clinical Data & Biostatistics Intern, Data Review and Study Quality

Study data review, query tracking, descriptive and diagnostic performance analysis, and reproducible reporting for performance studies

KEY TO TERMS USED

QUERY	A written question sent back to a site when data looks wrong, missing, or contradictory.
SENSITIVITY	Of the people who truly have the condition, the share a test correctly flags.
SPECIFICITY	Of the people who truly do not have it, the share a test correctly clears.
DATA LOCK	The point after which a study dataset may no longer be changed.
DENOMINATOR	Exactly which records a percentage was calculated over. Change it and the answer changes.
REPRODUCIBLE	An analysis that reruns from the raw data and produces the same numbers.

INTENDED ROLE

Study data arrives incomplete, inconsistent, and occasionally impossible, and someone has to notice before it reaches an analysis. This role is that check. You will review authorized study datasets for completeness and internal consistency, track queries to closure, and produce descriptive summaries and performance calculations in a reproducible way. The emphasis is on reproducibility and honest reporting, not on advanced statistical modeling.

PRINCIPAL DUTIES

- Review authorized study datasets for missing values, out-of-range results, duplicates, and internal contradictions.
- Write and track data queries, and maintain a query log through to documented resolution.
- Produce descriptive summaries of study populations, enrollment, and completion using an agreed table format.
- Support diagnostic performance calculations such as sensitivity, specificity, and agreement, with denominators defined up front.
- Write analysis scripts that rerun end to end from the source dataset, so that any result can be reproduced on request.
- Document every data handling decision, including exclusions and assumptions, in a dated analysis log.
- Use approved AI tools to assist with code and documentation, then validate outputs against the raw dataset before reporting anything.

EXAMPLE ASSIGNMENTS

Assignments depend on active studies, authorized data access, and the internship timeline. Examples may include:

- A data quality review of an authorized study dataset with a prioritized query list
- A reproducible descriptive summary package with tables, figures, and a rerunnable script
- A performance metric calculation with documented denominators, exclusions, and confidence intervals

YOUR FIRST TWO WEEKS

- Get authorized access to one dataset and write down exactly what each field means.
- Run a first completeness and range check and draft the initial query list.
- Rebuild one existing summary table from the raw data and confirm the numbers match.
- Set up the analysis script to rerun from raw data in one command.

ABOUT HOPKINS MEDTECH

Hopkins MedTech Group LLC is a U.S. company working in medical technology and in vitro diagnostics. Its work covers clinical performance studies, protocol development and site coordination, data review and quality assurance, and regulatory support for United States and European pathways. The group is also exploring AI and digital health through an early-stage research initiative in facial and skin analysis.

REQUIRED QUALIFICATIONS

- Currently pursuing, or recently completed, a degree in statistics, biostatistics, data science, epidemiology, mathematics, life sciences, or a related field, or able to show equivalent self-study or project work.
- Working knowledge of at least one of Python, R, or SAS from coursework, an online course, or a personal project.
- Solid spreadsheet skills and comfort reconciling two sources that are supposed to agree and do not.
- The discipline to document assumptions and exclusions rather than quietly dropping inconvenient rows.
- Discretion with confidential and patient-related data, including strict adherence to authorized access boundaries.
- No prior clinical data management, biostatistics, or professional internship experience is required.

PREFERRED QUALIFICATIONS

Not screening criteria. Coursework, labs, personal projects, and self-study all count as evidence.

- Coursework in biostatistics, study design, epidemiology, or diagnostic test evaluation
- Exposure to version control, notebooks, or any habit that makes an analysis rerunnable by someone else
- Familiarity with concepts such as sensitivity, specificity, predictive value, or inter-rater agreement

WHAT YOU WILL GAIN

- Learn how clinical study data is reviewed, queried, and locked before anyone analyzes it.
- Practice reproducible analysis habits that research and industry teams both demand.
- Understand how diagnostic performance metrics are defined, and how definitions change the answer.
- Leave with a reproducible analysis package you can walk an interviewer through line by line.

DELIVERABLES

- A data quality report with a query log and resolution status
- A reproducible analysis script, with the tables and figures it makes
- An analysis log of assumptions, exclusions, and open questions

SUPERVISION AND EVALUATION

Interns report to a named project lead, meet on a set weekly cadence, and receive written feedback on each deliverable. Work is evaluated on accuracy and traceability of sources, documentation quality, responsiveness to feedback, judgment about what to escalate, and disciplined use of approved AI tools. We expect steady learning and iteration rather than polished perfection on the first day.

WHAT WE DO NOT EXPECT

- Prior internship experience, or a name-brand employer on your CV.
- Knowing the regulations, tools, or vocabulary before you start; the terms box above is where most people begin.
- Certainty. Saying what the evidence does not support is treated as a strength, not a gap.

APPLICATION AND SELECTION

An application is a CV plus a short note naming the track you want and one project from this document you would like to work on, and why. First-round interviews are with the project lead who would supervise the work, not with a screening panel.

DATA HANDLING, PRIVACY, AND RESPONSIBLE AI

Interns work only with data and documents they have been explicitly authorized to access. Patient-identifiable information, study data, client-confidential files, and proprietary materials must never be uploaded to public or unapproved AI services, shared outside the project team, or stored on personal accounts or devices. AI tools may be used only

where the project lead has approved them for that task, and all AI-assisted work must be independently verified before it enters any deliverable. The intern and the project lead remain accountable for the final result.

Hopkins MedTech Group LLC provides equal internship and employment opportunities without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, age, disability, veteran status, or any other legally protected characteristic. Internships are educational in nature; interns do not provide medical advice, make regulatory determinations, or represent the company to any health authority, client, or study site. The AI and digital health initiative described here is early-stage research, not a cleared, approved, or clinically validated product. Assignments, data access, duration, and supervision are confirmed in the final project description.