Data Analyst
Jim Blair

bit.ly/Persona-download

"I love having a job where I constantly get to solve puzzles and learn about new subjects from the experts."

Bio

After getting a bachelor's in physics, Jim joined the research analytics department at a nearby academic health center. During his 3 years as a data analyst Jim has worked primarily on querying the enterprise data warehouse to find data for clinical researchers. Generating his queries with SQL code, he compiles the retrieved data into reports. He also pulls information for RPPRs, and is beginning to build databases for clinical studies. Jim works on 5 or 6 different projects during any given week, and must be ever mindful of permissions and approvals for sensitive data, as well as time spent on projects for billing purposes. Jim enjoys his work and believes it is a vital first piece of the research puzzle. Starting with understandable, clean, and accurate data is the crux of reproducible science, allowing for complex statistical and mathematical analyses later in the research process. Jim has grown his skills in 'interviewing' the researcher to determine exactly what they are looking for in terms of data. He has also become knowledgeable about where to search for various types of data and when to stop (i.e., the point of diminishing returns).

Education: BS, Physics
Years of experience: 3
Work location: Working remotely is possible but face-to-face time with researchers is a priority

Goals

- To find opportunities to collaborate with biostatisticians, to aid each team's work
- To obtain credit for his work through publication
- To advocate and embody open science practices

Software attitude & use

- Jim embraces the opportunity to learn new technologies and wants to build his skills in Python
- His colleagues are his greatest resource, helping with SQL code, report and form building, and encouragement to experiment with new technologies
- Coding and data tools: SQL and SQL server, R, Tableau, Git
- Research and collaboration: REDCap, the local enterprise data warehouse, Smartsheet, Epic, PowerChart, tools to track clinical studies and their budgets
- General: Microsoft Office Suite, JIRA

Scholarly Outputs

- Co-author on researchers' publications for the study-related databases he built
- Conference presentations and posters
- Documentation on analysis procedures for his department

Pain Points

- The need to reconstruct data from messy sources
- Slow response times from researchers

Motivators

To move beyond data searches and become experienced in more complex analyses

To approach problems from many angles. When an analysis isn't working, Jim finds another approach and strives to figure out what he can do better

To maximize reproducibility in analyses

Wants/Needs

- More time in the day to complete data searches and analyses
- To continue building his skills in a suite of analysis programs
- Wishes original data sources were cleaner
- Greater institutional focus on data management
- A quicker way to know the best communication style to use with each researcher
- A way to share information within his department about effective PI communication

Professional Development

Jim learns the methods of clinical studies and some of the subject matter from researchers as he works to instruct them in analytical methods and jargon

Wants to take a statistics class to understand what biostatisticians do with the data that analysts have collected, which can inform form-building

Self-paced learning through lynda.com

Queries Stack Overflow and user forums

The CTSA Program National Center for Data to Health (CD2H) is supported by the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health (Grant U24TR002306)