Graham Hancock

M.P.H. Candidate, Health Policy Track
graham.hancock@vanderbilt.edu

Practicum Site: Vanderbilt University Medical Center - Institute for Clinical and Translational Research

Practicum Site Supervisor: Laura Jones, M.P.H.

Exploring Probabilistic Record Linkage: Identifying and Archiving Methodologies in Pairing De-Identified Datasets

Keywords: probabilistic record linkage, literature review, de-identified data

Introduction: This practicum experience explored the methodology of collecting, sorting, and packaging relevant data from pre-existing databases for use in research projects, as performed by the Vanderbilt Institute for Clinical and Translational Research (VICTR), housed in the Vanderbilt Medical Center. By working with VICTR staff, the student would familiarize themselves with data handling, conducting independent literature reviews, and planning projects to create the data deliverables necessary for researchers to begin their studies, specifically focusing on efforts to pair de-identified datasets that shared patients.

Methods: The student collaborated on a variety of data management projects from establishing a data dictionary for one dataset to defining, categorizing, and organizing different variables across each row of data; the student also developed several literature reviews to create an informative history and context for assigned team projects. They attended team meetings to provide input and garner greater understanding of the project planning process.

Results: The student served as first author or co-authored multiple deliverables (data dictionary, project proposals, literature reviews) with fellow interns. These were created explicitly in response to real projects presented by clientele of VICTR and serve as pieces of the comprehensive product VICTR delivered in achieving each project's goals.

Conclusions: The practicum served as an excellent opportunity to form and temper skill sets related to the foundational phases of research projects and data literacy when working with datasets provided by other research teams. Moving into the future, this experience will help the student understand how to conduct both independent and collaborative research projects through the distinct phases of conceptualization, planning, and execution. Additionally, the student has gained competency in how to handle bodies of raw data and derive useful patterns and observations while also identifying lower quality data that cannot be used for analysis.