Table 2
Justification for Each Covered Topic

No.	Topic	Justification
1	Healthcare laws and regulations	Databases maintain electronic protected health information; thus an understanding of the Privacy and Security Rules are paramount. Examples include Safe Harbor de-identification (45 CFR §164.514(b)(2)), access controls and management, audit controls, data integrity, authentication, encryption, retention, and data backup and recovery.
2	Database development, SDLC, and Prototyping	Because the course is about databases and database management, a detailed understating of development is required. Topics include the Systems Development Life Cycle (SDLC), prototyping, and abstract data modeling. SDLC and prototyping are two commonly used systems development frameworks. The former is for large-scale systems requiring detailed planning (e.g., electronic health record systems), while the latter is for rapid, iterative production of smaller projects (e.g., registries). Because database design falls into both categories, an understanding of the processes and appropriate adoption are necessary. Abstract data modeling (e.g., enterprise data modeling) is customary during initial planning and analysis to gain a high-level understanding of the project and is the foundation for the subsequent topic.
3	Conceptual design, logical design, and normalization	Conceptual design, logical design, and normalization are standard processes in the creation of a database definition; thus these concepts are a necessity. Conceptual design captures graphically (in the form of Entity-Relationship [ER] diagrams) the information to be stored within a database. It records basic entities (e.g., Patients), attributes, and relationships between entities (e.g., between Patients and Encounters), as well as integrity control properties (refer to database modeling in the next section for details). Logical design translates ER diagrams into database implementation specifications. These specifications are refined to remove anomalies (insert, update, and delete anomalies) through normalization. The result is a database definition primed for physical design.
4	Physical design	The last of the design steps, physical design applies platform-specific details to the logical model, for example, the application of database management system (DBMS)-specific data types. Additionally, appropriate file organization and indexing structures are selected to maximize performance. File organization pertains to data storage on disk and thus directly affects performance. An index, if implemented correctly, can decrease query execution times by storing bits of information commonly used (e.g., primary and foreign keys) in small, high-performance data structures generally residing in memory. The designated indexing structure can have a profound effect on performance. Once this process is complete, the database is ready for implementation.
5	SQL	Structured query language (SQL) is the standard interface language for DBMSs and is thus another critical skill for students to learn. The creation, management, and manipulation of data and databases are controlled via various types of SQL commands (refer to Structured Query Language in the subsequent section for details).
6	Administration	Database administration covers a wide swath of activities pertaining to the management, backup, recovery, performance tuning, and security of a database. Entwined with privacy and security, database administration protects the data from unauthorized access (45 CFR §164.312(a)(2)(i) and 164.312(d)), modification, injection, and deletion (45 CFR §164.312(c)(1)), while backup and recovery assist with contingency planning (45 CFR §164.308(a)(7)(ii)(A)-(B))—all vital knowledge and skills for health informatics and health information management professionals. Administration also includes concurrency controls, transactions and conflicts (e.g., lost updates, dirty reads, and incorrect summaries), and locks (e.g., shared and exclusive).
7	Quality and integration	Students must understand that capturing and storing information in an efficient manner does not imply that the data will be of high quality; controls and permissions conceived during the design process and updated throughout the database's lifetime are required. Additionally, knowledge of data cleansing and transformation techniques required for data integration from other sources is necessary, given the nature of healthcare systems and continued advancement in information exchange. Topics include data governance, extraction, quality characteristics, profiling, cleaning, transformation, and loading.
8	Archival and retrieval systems, data warehousing, and ETL	Archival and retrieval systems and data warehouses are key technologies in healthcare. Both store vast quantities of historical data and are used for advanced interrogation. Data are loaded into these systems via the process of extracting, transforming, and loading (ETL) information from multiple independent and heterogeneous data sources. This process requires extreme care (see previous topic) and generally results in de-normalized entities. Students are exposed to concepts in this area such as multidimensional modeling, various ETL types (e.g., consolidated, federated, and propagated), the appropriate use of these systems, and the laws pertaining to data retention.