[Topics in Practice Management]

CHEST

Clinical Databases for Chest Physicians

Andrew M. Courtwright, MD, PhD; and Peter E. Gabriel, MD, MSE

A clinical database is a repository of patient medical and sociodemographic information focused on one or more specific health condition or exposure. Although clinical databases may be used for research purposes, their primary goal is to collect and track patient data for quality improvement, quality assurance, and/or actual clinical management. This article aims to provide an introduction and practical advice on the development of small-scale clinical databases for chest physicians and practice groups. Through example projects, we discuss the pros and cons of available technical platforms, including Microsoft Excel and Access, relational database management systems such as Oracle and PostgreSQL, and Research Electronic Data Capture. We consider approaches to deciding the base unit of data collection, creating consensus around variable definitions, and structuring routine clinical care to complement database aims. We conclude with an overview of regulatory and security considerations for clinical databases.

CHEST 2018; ■(■):■-■

KEY WORDS: clinical database; data security; institutional review board; quality improvement; technical platforms

A clinical database is a repository of patient medical and sociodemographic information focused on one or more specific health conditions or exposures.[1] Clinical databases vary in scope from an individual physician to group practices to health systems to national registries. Although clinical databases may be used for research purposes, their primary goal is to collect and track patient data for quality improvement, quality assurance, and/or actual clinical management.[2] Collection of these data may also allow physicians to capture variations in treatment, understand practice patterns, identify rare events, and improve quality of care.[1] Despite the potential for organizational and individual physician practice improvement, establishing a clinical database may appear challenging and resource intensive, particularly without prior experience. The aim of this article is to provide an introduction and focused, practical advice on the development of small-scale clinical databases for chest physicians and practice groups.

General Guidance

Before project initiation, we recommend that chest physicians interested in creating a clinical database keep in mind a basic guiding principle: the extensiveness of data collection must be balanced with database usability and the resources available for database construction and ongoing data

ABBREVIATIONS: EMR = electronic medical record; IRB = institutional review board; IT = information technology; PHI = protected health information; RDBMS = relational database management systems; REDCap = Research Electronic Data Capture; SQL = structured query language

AFFILIATIONS: From the Division of Pulmonary and Critical Care Medicine (Dr Courtwright), Hospital of the University of Pennsylvania, Philadelphia, PA; and the Department of Radiation Oncology (Dr Gabriel), Hospital of the University of Pennsylvania, Philadelphia, PA.

CORRESPONDENCE TO: Andrew M. Courtwright, MD, PhD, Gates 941, Hospital of the University of Pennsylvania, 3400 Spruce Street, Boston, MA 19104; e-mail: acourtwright@partners.org

DOI: https://doi.org/10.1016/j.chest.2018.01.003

REV 5.5.0 DTD ■ CHEST1506_proof ■ 5 February 2018 ■ 10:11 pm ■ EO: CHEST-17-2388

entry. In our experience, physicians often err on the side of collecting many more variables than necessary, most often because goals for the database have not been clearly defined and there is a desire to collect "everything" that might later become of interest. Particularly when data extraction is difficult, the additional time needed to capture these variables can lead to project stagnation. To achieve the appropriate balance, we recommend identifying overarching goals and targeted questions in advance of project initiation. We also recommend an early self-assessment of capabilities, focused on choosing the appropriate technical platform, describing project participants and roles, and identifying data sources and variables.

Sample Clinical Databases in Chest Medicine

The nature and extent of the clinical database will depend on the project goal or targeted question. Some databases are built to aggregate data on common conditions, exposures, or procedures; some to monitor for safety/harm events; some to track uncommon conditions, exposures, or procedures; and some are designed around quality improvement goals or to capture cost effectiveness. For example, a practice may track all patients with asthma to identify local populations or patients at risk for multiple asthma-related hospitalizations.[3] Or a database may be designed to capture risk factors for 30-day readmission among patients with COPD, given new Medicare reimbursement restrictions in this population.[4]

Alternatively, a practice deciding to introduce a new technology or procedural approach such as transbronchial cryobiopsy might construct a clinical database to capture procedure-related complications and diagnostic yield. If complications are higher than anticipated, this process could allow strategies to mitigate harm events.[5] A clinical database that captured rarer lung diseases seen in a group practice, such as lymphangioleiomyomatosis or smoking-related interstitial lung disease, could be used to track these patients and compare local variations in patient populations with national databases targeting the same disease.[6]

As a final example, growing clinical evidence has led to increased prevalence of lung cancer screening programs. A clinical database could be used to track patients with nodules identified through such a program, both to ensure adequate longitudinal screening follow-up and to capture compliance with society guidelines. Such databases have been also been used to describe the economic impact of hiring advanced practitioners to lead screening programs.[7]

Technical Platforms

In deciding on a technical platform to host a clinical database, physicians should consider the cost; level of technical skill required to develop and maintain the database; hardware requirements (eg, server vs desktop) for deploying the system; security, privacy, and redundancy features; user interface design and ease of data entry; support for concurrent and/or remote access; and data query and analysis capabilities (Table 1). Although private information technology (IT) companies can develop custom databases for a fee, understanding commercially available platforms is essential for the majority of chest physicians who will need to troubleshoot and revise databases as projects evolve. Similarly, web-based platforms such as Google Forms may be an attractive alternative to traditional locally installed platforms, but there are significant privacy concerns that limit the use of most fully web-based approaches.

In our experience, the most commonly used clinical database platform is Microsoft Excel, which is ubiquitous, low cost, and easy to use. Excel requires no specific coding skills for construction, data entry, or data retrieval. In addition, it is available on most office and home computers and requires minimal training for basic use. There are, however, three primary disadvantages to this platform. First, Excel has very limited internal security options for controlling access and maintaining data privacy. Second, Excel does not natively support concurrent multiuser access or remote access: only one user can edit a file at a time from one physical computer. This often leads to a proliferation of copies of Excel data sets, which creates data reliability concerns. Excel files are also often moved and shared through insecure means such as e-mail, USB drives, or web-based file sharing applications, which creates high risk for security and privacy breeches.

Third, and often overlooked, is that Excel allows users to enter data of any type and any form in each cell without restricting variables to predefined formats. For example, a column intended to store the date of diagnosis can contain valid dates, invalid dates (eg, April 31), and free text comments such as "circa 2015 – not sure of exact date." Excel spreadsheets also impose a "flat" data model that make it difficult to capture the one-to-many

REV 5.5.0 DTD ■ CHEST1506_proof ■ 5 February 2018 ■ 10:11 pm ■ EO: CHEST-17-2388