

Preface

Diagnostic and therapeutic technologies continue to evolve rapidly, and both individual practitioners and clinical teams face increasingly complex decisions. Unfortunately, the current state of medical knowledge does not provide the guidance to make the majority of clinical decisions on the basis of evidence. According to the 2012 Institute of Medicine Committee Report, only 10–20 % of clinical decisions are evidence based. The problem even extends to the creation of clinical practice guidelines (CPGs). Nearly 50 % of recommendations made in specialty society guidelines rely on expert opinion rather than experimental data. Furthermore, the creation process of CPGs is “marred by weak methods and financial conflicts of interest,” rendering current CPGs potentially less trustworthy.

The present research infrastructure is inefficient and frequently produces unreliable results that cannot be replicated. Even randomized controlled trials (RCTs), the traditional gold standards of the research reliability hierarchy, are not without limitations. They can be costly, labor-intensive, slow, and can return results that are seldom generalizable to every patient population. It is impossible for a tightly controlled RCT to capture the full, interactive, and contextual details of the clinical issues that arise in real clinics and inpatient units. Furthermore, many pertinent but unresolved clinical and medical systems issues do not seem to have attracted the interest of the research enterprise, which has come to focus instead on cellular and molecular investigations and single-agent (e.g., a drug or device) effects. For clinicians, the end result is a “data desert” when it comes to making decisions.

Electronic health record (EHR) data are frequently digitally archived and can subsequently be extracted and analyzed. Between 2011 and 2019, the prevalence of EHRs is expected to grow from 34 to 90 % among office-based practices, and the majority of hospitals have replaced or are in the process of replacing paper systems with comprehensive, enterprise EHRs. The power of scale intrinsic to this digital transformation opens the door to a massive amount of currently untapped information. The data, if properly analyzed and meaningfully interpreted, could vastly improve our conception and development of best practices. The possibilities for quality improvement, increased safety, process optimization, and personalization of clinical decisions range from impressive to revolutionary. The National Institutes of

Health (NIH) and other major grant organizations have begun to recognize the power of big data in knowledge creation and are offering grants to support investigators in this area.

This book, written with support from the National Institute for Biomedical Imaging and Bioengineering through grant R01 EB017205-01A1, is meant to serve as an illustrative guide for scientists, engineers, and clinicians that are interested in performing retrospective research using data from EHRs. It is divided into three major parts.

The first part of the book paints the current landscape and describes the body of knowledge that dictates clinical practice guidelines, including the limitations and the challenges. This sets the stage for presenting the motivation behind the secondary analysis of EHR data. The part also describes the data landscape, who the key players are, and which types of databases are useful for which kinds of questions. Finally, the part outlines the political, regulatory and technical challenges faced by clinical informaticians, and provides suggestions on how to navigate through these challenges.

In the second part, the process of parsing a clinical question into a study design and methodology is broken down into five steps. The first step explains how to formulate the right research question, and bring together the appropriate team. The second step outlines strategies for identifying, extracting, Oxford, and preprocessing EHR data to comprehend and address the research question of interest. The third step presents techniques in exploratory analysis and data visualization. In the fourth step, a detailed guide on how to choose the type of analysis that best answers the research question is provided. Finally, the fifth and final step illustrates how to validate results, using cross validation, sensitivity analyses, testing of falsification hypotheses, and other common techniques in the field.

The third, and final part of the book, provides a comprehensive collection of case studies. These case studies highlight various aspects of the research pipeline presented in the second part of the book, and help ground the reader in real world data analyses.

We have written the book so that a reader at different levels may easily start at different parts. For the novice researcher, the book should be read from start to finish. For individuals who are already acquainted with the challenges of clinical informatics, but would like guidance on how to most effectively perform the analysis, the book should be read from the second part onward. Finally, the part on case studies provides project-specific practical considerations on study design and methodology and is recommended for all readers.

The time has come to leverage the data we generate during routine patient care to formulate a more complete lexicon of evidence-based recommendations and support shared decision making with patients. This book will train the next generation of scientists, representing different disciplines, but collaborating to expand the knowledge base that will guide medical practice in the future.

We would like to take this opportunity to thank Professor Roger Mark, whose vision to create a high resolution clinical database that is open to investigators around the world, inspired us to write this textbook.

Secondary Analysis of Electronic Health Records

MIT Critical Data

2016, XXI, 427 p. 108 illus., 100 illus. in color.,

Hardcover

ISBN: 978-3-319-43740-8