Electronic Health Records in Behavioral Research

John Loft
RTI International

Diana Greene

Follow this and additional works at: https://repository.upenn.edu/admindata_conferences_presentations_2018

https://repository.upenn.edu/admindata_conferences_presentations_2018/26

DOI https://doi.org/10.23889/ijpds.v3i5.1052

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/admindata_conferences_presentations_2018/26
For more information, please contact repository@pobox.upenn.edu.
Electronic Health Records in Behavioral Research

Abstract
Medical records are a type of administrative record with rich potential for research of behavioral health and health policy. Developments in electronic health records (EHR) can increase access to data contained in medical records but also present some unusual challenges for research. This presentation summarizes recent literature describing the use of EHR in research and identifies issues for consideration in the preparation of research design and protocols for data collection and preparation. The discussion is presented in a framework for evaluation of data quality and fitness for use.

Comments
DOI https://doi.org/10.23889/ijpds.v3i5.1052

This presentation is available at ScholarlyCommons: https://repository.upenn.edu/admindata_conferences_presentations_2018/26
Electronic Health Records in Behavioral Research

Assessing Data Quality

John D. Loft and Diana Greene

ADRF Network Research Conference on Innovations in Administrative Data for Social Sciences
November 13th and 14th
Washington, DC

Improving Data Quality and Standards
Overview

- Background
  - Review recent literature
  - Administrative records in behavioral research
  - Electronic health records as administrative records
  - Comments related to data quality
Administrative records include any data source that is kept by a business, government agency, or other organization for the purposes of their work or function.

- Medical records—clinical, laboratory results, prescriptions, billing
- School records—transcripts, grades, attendance
- Financial or tax records—tax status, ownership
- Voting records—counts of registered voters, party affiliation

Survey Research Laboratory, 2017
Develop sample frame
- Records about foster care placement to study foster parents by county
- Public records of tax status for nonprofit organizations in a state

Analysis
- Estimate accuracy of data across data sets—medical records used to assess accuracy of self-reports in surveys
- Estimate bias—election results to assess self reports in polls

Complement other data with useful substantive information
- Complement clinical records medical events with laboratory results or billing records

Survey Research Laboratory, 2017
Background—Challenges to Using Administrative Data

Access
- Some records are public (e.g., tax status of an organization)—many are protected (medical, education, tax, voting)

Extracting information from administrative records
- Administrative records are formatted and stored in ways that facilitate their intended use and utility to organization—not always ways that are useful or accessible to researchers.
- Additional coding often necessary to format data for research purposes

Linking administrative with other data sources about individuals
- If common identifiers are not available, some form of probabilistic match is necessary

Survey Research Laboratory, 2017
Uses other than direct health care delivery:

- Analysis
- Research (clinical and behavioral)
- Quality and safety measurement
- Public health
- Payment
- Provider certification or accreditation
- Marketing
- Business applications and commercial activities

Safran, *et al.*, 2007
Background—Total Survey Error Model

Total Survey Error

- Sampling Error
  - Sampling scheme
  - Sample size
  - Estimator choice

- Nonsampling Error
  - Specification
  - Nonresponse
  - Frame
  - Measurement
  - Data processing

Mean Squared Error (MSE)

\[ \text{MSE} = \text{Bias}^2 + \text{Variance} \]

Biemer, 2010
Concerns with Research Based on EHRs

- EHR data not collected for research purposes
- Variable data quality
- Coverage issues
- Interoperability across data systems
- Consistency within data systems
EHRs are completed for clinical purposes (diagnosis and treatment) and billing purposes

- Often completed by multiple providers (doctors, physician assistants, nurse practitioners, nurses) using different conventions
- Especially problematic in text-based fields, but also true in fixed-response fields
- Training in and supervision of data entry conventions is variable across providers and by individuals within providers
- Consistent, validated methods for assessing data quality and completeness have not yet been adopted

Cahn et al., 2010; Cowie et al., 2017; Skripacak et al., 2014; Wasserman, 2011; Weiskopf & Weng, 2013
Clinical data related to the patient’s physical condition are available, but recording can be inconsistent across recorders within and across provider settings.

Mental health indicators are poorly documented in EHRs (Anderson et al., 2012; Valuck et al., 2015).

Socio-demographic variables of interest to behavioral researchers (marital/occupational status, insurance coverage) are not available.

Anderson et al., 2012; Cahn et al., 2010; Cowie et al., 2017; Dungey et al. 2016; Skripcak et al., 2014; Wasserman, 2011; Weiskopf & Weng, 2013; Wright, et al., 2015; Valuck et al., 2015; Wright, et al., 2015
Many of the papers we reviewed were limited in scope to a single or only a few providers using the same EHR or data repository.

Coverage relative to a population is a concern in regions where patients receive care from different providers or hospitals operating in different EHR systems that are not linked.

Completeness of records related to severity/complexity of patients condition as well as other patient characteristics.

Also related to individual providers, work load, organizational features.

Romo et al., 2016; Rusnov et al., 2014; Singer et al., 2016; Wright et al., 2015.
Interoperability is “…the ability of two or more systems or components to exchange information and to use information that has been exchanged” (Institute of Electrical and Electronics Engineers, 1990, quoted in Benson, 2010

While a great deal of progress has been made in developing standards for interoperability (HL7, SNOMED, FHIR), it is a moving target

Additional standardization may need to be imposed by researchers for specific research objectives

Benson, 2010; Benson & Grieve (2016), Skripcak, et al., 2014;
Concern about interoperability across data systems has a parallel concern about consistency within data systems used by providers.

Subsystems used to record clinical data may have different standards/convention than subsystems used to record laboratory results or prescriptions or billing codes.

Terry, et al. 2010
Summary—Potential errors in EHRs

Total Error

- Sampling error (if selection of records is based on probability design)
  - Sampling scheme
  - Sample size
  - Estimator choice

- Nonsampling error
  - Specification error
  - Incomplete data
  - Frame
  - Measurement
    - Inconsistent coding across and within records systems
    - Data entry
  - Data processing
delivering the promise of science for global good

John Loft
jloft@rti.org
312.456.5241