



# Research Week 2022

## Biases in Data from Electronic Health Records for Research: a Literature Review

Frances Hsu, MS<sup>1</sup>

<sup>1</sup>Oregon Health & Science University; hsufr@ohsu.edu  
Department of Medical Informatics and Clinical Epidemiology

### Keywords

EHR, bias, clinical informatics

### Abstract

#### *Objective*

This literature review identifies and organizes potential biases when using data from electronic health records (EHRs) for secondary research.

#### *Methods*

A search in the MEDLINE database using MeSH terms 'EHR' combined with 'bias.' No date limit was imposed but articles were limited to publications in English language. Articles reporting general sources of bias in EHR data and specific example(s) of bias were included. Bias concepts were grouped following the research process so they can be applied at each step.

#### *Results*

Misclassification (information) bias, missing data bias, and bias due to healthcare process were identified in addition to traditional biases found in observational studies (e.g., selection bias, confounding). Misclassification bias occurs from documentation structure and data extraction from EHR. The healthcare process potentially contributes additional sources of missing data bias in EHR data, such as providers' training that selectively seeks and documents certain data but not others.

#### *Conclusion*

This review provides a framework for researchers to evaluate potential biases in their study when using data from EHR and/or registry. Identification of bias(es) would allow researchers to make statistical adjustment or note limitations of their studies, while highlighting issues arising from data collection and access to health care.