Use of Automated Databases for Pharmacoepidemiology Research

Vincent Lo Re, MD, MSCE, FISPE
Department of Medicine (Infectious Diseases)
Center for Pharmacoepidemiology Research and Training
Perelman School of Medicine
University of Pennsylvania

4th MURIA – June 19, 2018

Learning Objectives

• Review databases for pharmacoepidemiology research
  – Registries, claims, medical records
  – Understand their strengths, weaknesses
• Facilitate appropriate database selection
  – Clarify reasons for database selection

Outline

• Overview of automated databases
• Data sources for pharmacoepidemiology:
  – Registries
  – Claims databases
  – Electronic medical record (EMR) databases
  – Hybrid databases
• Appropriate database selection

Automated Databases

• Allow evaluation of health conditions in “real world” settings
• Past 3 decades → ↑ use of electronic data sources containing medical care data
• Efficient, cost-effective way to conduct pharmacoepi research

Automated Database Generation

Ideal Automated Database

- Longitudinal data from all care settings
- Records prescribed, dispensed drugs
- Includes laboratory tests results
- Large representative population
- Linkable to other data sources (via identifiers)
- Confounders of interest available
- Updatable, with access to medical records


“Database studies must be performed within the limitations of a resource not specifically designed to test the research hypothesis”

-Gillian C. Hall, PhD

Automated Databases: Strengths

- Relevant clinical data
- Large, real-world clinical population
- Longitudinal
- Linkable
- Short time-frame from design → results


Potential Limitations of Automated Databases

- Uncertain validity of diagnoses
- Completeness, quality of data
- Instability of population
- Generalizability
- Costs of data


Choosing Among Databases

**Key Point**: The research question dictates selection of the appropriate pharmacoepidemiology database.

- Appropriate study population, size
- Ascertain exposure, outcome
- Relevant confounders measured
- Link with other databases, records

Outline

- Overview of automated databases
- Data sources for pharmacoepidemiology:
  - Registries
  - Claims databases
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

Registry Databases: Overview

- Prospective study of patients with common characteristics
- Developed to evaluate:
  - Natural history of disease
  - Drug effectiveness, safety
  - Quality of life
  - Cost-effectiveness of therapies

Development and Maintenance of Patient Registries

- Collect data on:
  - Demographic characteristics
  - Social history
  - Disease-specific drug treatments
  - Select disease-related outcomes
- Ability to link to other data sources?

Registry Databases: Data Collected

- Large patient numbers
- Usual diagnostic, follow-up procedures
- Contain “real world” therapeutic effectiveness, safety data
- Heterogeneity among sites

Registry Databases: Benefits

- Selection bias (non-sequential patients)
- Variability in data definitions
- Data may not be validated
- Incomplete data on comorbid conditions, outcomes, mortality
- Inability to link with other data sources

Registry Databases: Limitations
Outline

- Overview of automated databases
- Data sources for pharmacoepidemiology:
  - Registries
  - Claims databases
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

Overview of Claims Databases

- Billing for use of healthcare system
- Diagnoses → uncertain validity
  - Hospital: coders; outpatient: clinicians
- Pharmacy claims → dispensed drug (valid)
- Procedure claims → no results
- Concern for lack of completeness
  - No body mass index, tobacco, alcohol data

Examples of Claims Databases

- US government: US Medicaid, Medicare
- US commercial insurance
- Canadian provincial

Where to Find Registries?

https://patientregistry.ahrq.gov

Original Article

Enhancing Cancer Registry Data for Comparative Effectiveness Research (CER) Project: Overview and Methodology

Marc R. Cheu, PhD; Ashley A. Wilkins; RP;彩his Li; Amy Ahn, PhD; Stephen G. Hanft, MD; John E. Wilt, PhD; Mary Kay Field, RN; Asha M. Maheshwary, RN; Maria G. Colosimo, MPH; MSP; Mark E. Ackert; MN; Leslie A. Kane, RN; Betsy W. Shiner, RN; Gloria A. Melito; Tania F. angel; and Thomas W. Hagle, MD, MPH;

Visual and written content is not provided. Further information may be available from the original source.
Outline

- Overview of automated databases
- **Data sources for pharmacoepidemiology:**
  - Registries
  - Claims databases
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

Electronic Medical Record Databases

- Data include:
  - Medical diagnoses
  - Drug prescriptions (not dispensing)
  - Laboratory results
  - Procedures
- Still have concerns for incompleteness
  - Out-of-network care

Clinical Practice Research Datalink & The Health Improvement Network

- United Kingdom medical record databases
- General practitioner: “gatekeeper”
- Available data:
  - Medical diagnoses
  - Outpatient prescriptions
  - Lab results
- Hospital care → Hospital Episode Statistics
  

Outline

- Overview of automated databases
- **Data sources for pharmacoepidemiology:**
  - Registries
  - Claims databases
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

Hybrid Databases

- Administrative AND clinical databases
- Reap benefits of claims and medical record data
- Some may have less diverse populations

Examples of Hybrid Databases

- US health plans:
  - **Individual:** Veterans Affairs, Kaiser Permanente
  - **Group:** Sentinel Distributed Database
- International Epidemiology Databases to Evaluate AIDS (IeDea)
  - International research consortia for HIV data
Example of Hybrid Database #1: VA Health Data

- Largest integrated health care system in US
- Available data:
  - Inpatient/outpatient ICD diagnoses, drugs
  - Procedures, biopsies
  - Laboratory data
- Limitation: emergency $\rightarrow$ nearest hospital
- Linkable (registries, Medicare, Medicaid)

Example #2: International Epidemiology Databases to Evaluate AIDS

- Collects HIV/AIDS data from 7 regions
  - 4 in Africa (Southern, East, West, Central)
  - North America, Asia, Central/South America

- Available data:
  - Medical diagnoses, comorbidities
  - Antiretroviral drugs
  - Laboratory data (e.g., HIV RNA, CD4)

Outline

- Overview of automated databases
- Data sources for pharmacoepidemiology:
  - Registries
  - Claims databases
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

Choosing Among Databases

- Research question dictates database
- Existing “checklists” to guide researchers:
  - ISPE guidelines
  - ISPOR guidelines
- Evaluation based on major domains
Important Questions to Ask

- What is the population covered?
- Are there continuous, consistent data?
  - Exposure, outcomes
  - Confounders of interest
- Is follow-up sufficiently long enough?
- Access to medical records?
- Ability to link to other data sources?

Summary

- All databases have strengths, limitations
- Research question must guide database selection
- Understand accuracy, completeness, appropriateness of data
- Collaborate with expert in data sources