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

CPeRT Center for Pharmacoepidemiology Research and Training

ISPE

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

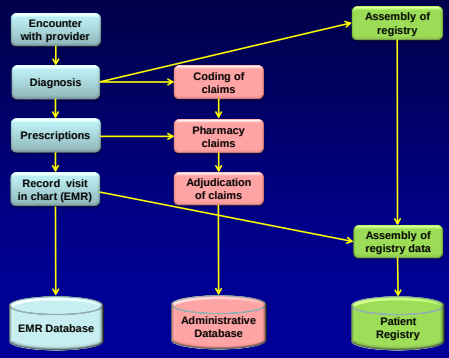
Outline

- **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



Schneeweiss S. *J Clin Epidemiol* 2005;58:323-37.

Ideal Automated Database

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

Shah BR. *Am Heart J* 2010;160:8-15.

"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

Suissa S. *Nat Clin Pract Rheumatol* 2007;3:725-32.

Potential Limitations of Automated Databases

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

Suissa S. *Nat Clin Pract Rheumatol* 2007;3:725-32.

Choosing Among Databases

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

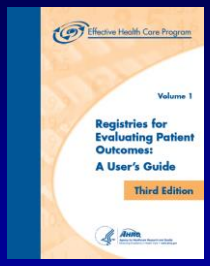
- Appropriate study population, size
- Ascertain exposure, outcome
- Relevant confounders measured
- Link with other databases, records

Outline

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

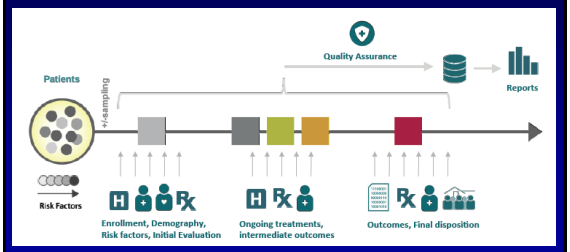
Registry Databases: Overview

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



Gliklich RE, Dreyer NA. AHRQ publication No. 07-EHC001.
http://effectivehealthcare.ahrq.gov/ehc/products/21/12/PatOutExecSumm.pdf

Development and Maintenance of Patient Registries



http://www.pcori.org/assets/11-Gliklich-Slides-Registries.pdf

Registry Databases: Data Collected

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

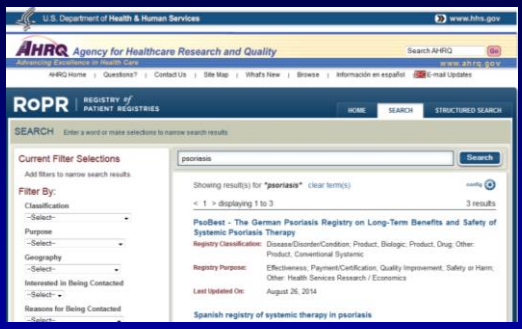
Registry Databases: Benefits

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

Registry Databases: Limitations

- 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

Where to Find Registries?



https://patientregistry.ahrq.gov/

Where to Find Registries?

Original Article

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

Vivien W. Chen, PhD[a]; Christie R. Eheman, PhD[b]; Christopher J. Johnson, MPH[c]; Monique N. Hernandez, PhD[d]; David Rousseau, BS[a]; Timothy S. Styles, MD, MPH[a]; Dee W. West, PhD[e]; Meichin Hsieh, MSPH, CTR[a]; Anne M. Hakeemeyerth, PhD[a]; Maria O. Celaya, MPH, CTR[a]; Randi K. Rycroft, MSPH, CTR[a]; Jennifer M. Wike, MBA, MPH[a]; Melissa Pearson, CTR[a]; Judy Brockhouse, MPH, CTR[a]; Linda G. Mulvihill, RHIT,CTR[a]; Kevin B. Zhang, PhD[a]

Abstract: Following the Institute of Medicine's 2009 report on the national priorities for comparative effectiveness research (CER), funding for support of CER became available in 2009 through the American Recovery and Reinvestment Act. The Centers for Disease Control and Prevention (CDC) received funding to enhance the infrastructure of population-based cancer registries and to expand registry data collection to support CER. The CDC established 10 specialized registries within the National Program of Cancer Registries (NPCR) to enhance data collection for all cancers and to address targeted CER questions, including the clinical use and prognostic value of specific biomarkers. The project also included a special focus on detailed first course of treatment for cancers of the breast, colon, and rectum, as well as chronic myeloid leukemia (CML) diagnosed in 2011. This paper describes the methodology and the work conducted by the CDC and the NPCR specialized registries in collecting data for the 4 special focused cancers, including the selection of additional data variables, development of data collection tools and software modifications, institutional review board approvals, training, collection of detailed first course of treatment, and quality assurance. It also presents the characteristics of the study population and discusses the strengths and limitations of using population-based cancer registries to support CER as well as the potential future role of population-based cancer registries in assessing the quality of patient care and cancer control.

Key words: cancer treatment, CER support, methodology, population-based registry

Journal of Registry Management 2014 Volume 41 Number 3

Outline

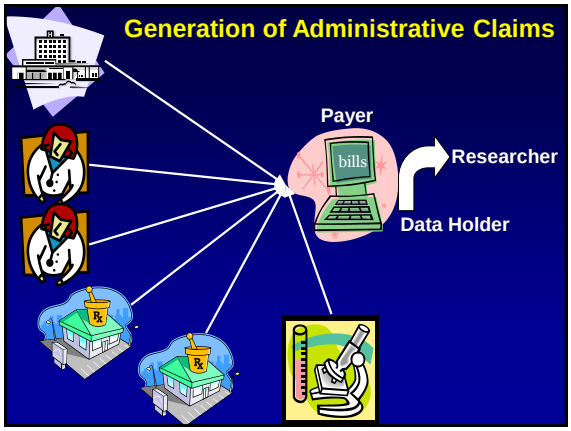
- Overview of automated databases
- **Data sources for pharmacoepidemiology:**
  - Registries
  - **Claims databases**
  - Electronic medical record (EMR) databases
  - Hybrid databases
- Appropriate database selection

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

Schneeweiss S. *J Clin Epidemiol* 2005;58:323-37.

Generation of Administrative Claims



Examples of Claims Databases

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

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

http://www.cprd.com/

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 → nearest hospital
- Linkable (registries, Medicare, Medicaid)

Justice AC. *Med Care* 2006; 44 (Suppl 2): S7-12.

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

Southern Africa
Lesotho
Malawi
Mozambique
South Africa
Zambia
Zimbabwe

IeDEA

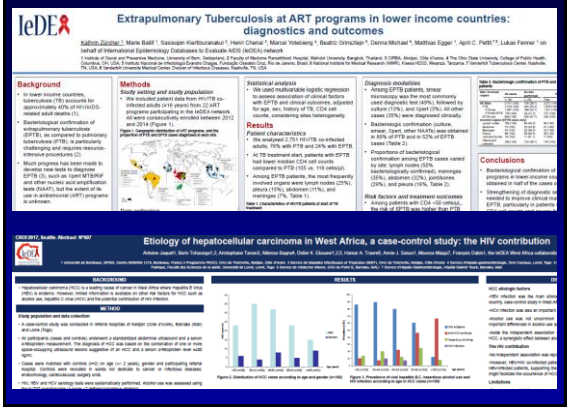
www.iedea.org

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)

IeDEA

www.iedea.org



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

Hall GC. *Pharmacoepidemiol Drug Saf* 2012;21:1-10.
Berger ML. *Value Health* 2009;12:1053-61.

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?

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2012; 21: 1–10
Published online 8 November 2011 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/pds.2229

COMMENTARY

Guidelines for Good Database Selection and use in Pharmacoepidemiology Research†

Gillian C. Hall[1]*, Brian Sauer[2], Alison Bourke[3], Jeffrey S. Brown[4], Matthew W. Reynolds[5] and Robert Lo Casale[6]

[1]Grimsdyke House, London, UK
[2]Salt Lake City VA IDEAS Centre & Division of Epidemiology, The University of Utah, Salt Lake City, UT, USA
[3]CSD Medical Research, London, UK
[4]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA
[5]United BioSource Corporation, Lexington, MA, USA
[6]Department of Epidemiology, Merck & Co. Inc, West Point, PA, USA

International Journal of Epidemiology
© International Epidemiological Association 1996

Vol. 25, No. 2
Printed in Great Britain

A Framework for Evaluation of Secondary Data Sources for Epidemiological Research

HENRIK TOFT SØRENSEN,* SVEND SABROE** AND JØRN OLSEN†

Sørensen H T (Department of Internal Medicine V, Aarhus University Hospital DK-8000 Aarhus C, Denmark) Sabroe S and Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology* 1996; **25**: 435–442

Summary

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

Epidemiologists