Evaluating Drug Exposure/Outcomes in Pharmacoepidemiology Databases

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

Education Objectives - 1

- Determine how to define drug exposure in electronic health databases
- Understand limitations of studying prevalent users
- Learn how to select drug comparators

Education Objectives - 2

- Review different types of study outcomes
- Learn methods to validate outcomes within electronic databases
- Review examples of study evaluating validity of health outcomes of interest

Ascertaining Drug Exposure in Electronic Health Databases

Outline

- Selection of drug exposure variables
- Prevalent vs. new user study designs
- Choosing appropriate drug comparator(s)

Outline

- **Selection of drug exposure variables**
- Prevalent vs. new user study designs
- Choosing appropriate drug comparator(s)

Conceptual Considerations for Drug Exposure Measurement

- Link exposure measurement to study question
  - Short term vs. long term
  - Single vs. chronic use
  - Prevalent vs. new users
- Mechanism for exposure/outcome relation
- Consistency, accuracy of exposure measurement
- Changes in exposure status

Methods Available to Measure Drug Exposure

- Sources of drug exposure in databases:
  - Drug prescription data
  - Drug dispensing data
  - Data on payment for medication (i.e., claim)
  - Patient interview, self-report
- Limitations:
  - Drugs prescribed may not be dispensed
  - Drugs dispensed may not be ingested
  - Interviews, self-report may be inaccurate

How to Choose Appropriate Exposure Variable?

- Consider biological mechanism of drug
  - E.g., Drug effects on liver injury after stopping
- Consider limitations of different definitions
- Consider validation
- "What's available" vs. "what's reliable"

Variables Needed to Calculate Cumulative Drug Dose

- Frequency of drug exposure
- Amount, dose of each drug exposure
- Duration of exposure

Outline

- Selection of drug exposure variables
- Prevalent vs. new user study designs
- Choosing appropriate drug comparator(s)

Limitations of Prevalent Users in Pharmacoepidemiology Research

- Prevalent users
  - On drug for some time before follow-up begins
- Limitations:
  - Includes survivors of early period of therapy ("healthy user bias")
  - May miss events early during therapy
  - Covariates for drug use often affected by drug

Ray WA. *Am J Epidemiol* 2003;158:915–20.

New User Design in Pharmacoepidemiology Research

- Identifies and selects new drug initiators
- Follow-up begins at initiation of therapy (t_0)
- May restrict to patients with a minimum period of non-use prior to t_0 (washout)
- Data for patient characteristics are obtained over some time before t_0

Ray WA. *Am J Epidemiol* 2003;158:915–20.

Outline

- Selection of drug exposure variables
- Prevalent vs. new user study designs
- Choosing appropriate drug comparator(s)

Selecting Appropriate Drug Comparator Group(s) - 1

- Should reflect clinically meaningful choices
- Consider study question being addressed
- Patients prescribed a drug are different from those who are not → may relate to outcomes
  - Bias → "confounding by indication"

Selecting Appropriate Drug Comparator Group(s) - 2

- Alternative treatments with similar indication
- Usual or standard care
- Historical comparator
- Comparator from another data source
  - Generalizability?
- No treatment
  - Concern for confounding by indication
  - How to choose time zero?

Operationalizing Drug Comparator Group

- Important considerations:
  - Indication for drug therapy
  - Initiation
  - Dose/intensity of drug comparison
  - Exposure time window

Evaluating Clinical Outcomes in Electronic Health Databases

Outline

- Overview of evaluation of outcomes in electronic data sources
- Steps in validation of clinical outcome
- Example of outcome validation
  - Hepatic decompensation (end-stage liver dz)
  - Other examples

Outline

- Overview of evaluation of outcomes in electronic data sources
- Steps in validation of clinical outcome
- Example of outcome validation
  - Hepatic decompensation (end-stage liver dz)
  - Other examples

Selection of Outcomes in Pharmacoepidemiology Databases

- Types of outcomes to study:
  - Patient-associated morbidity (i.e., disease)
  - Mortality (total or cause-specific)
  - Physiologic parameters
  - Quality of life, lifestyle practices
- Main outcome → most clinically relevant
  - Power/sample size based on primary outcome

Goldberg R. *Am J Med* 2014;127:379-84.

Composite & Secondary Outcomes in Pharmacoepidemiology

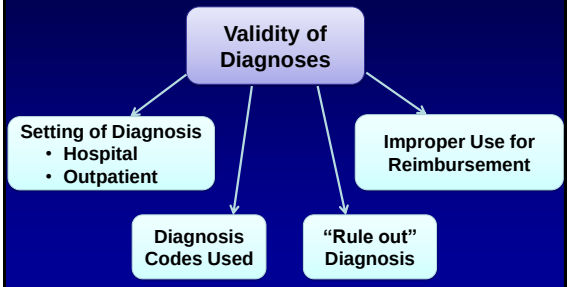
Composite Outcome	Secondary Outcome(s)
Occurrence of any one of multiple endpoints	Additional event(s) of interest
Each component has similar impact on health, changes in same direction with treatment	Typically not specifically powered to assess
Enhances power to detect clinically relevant differences	View results with caution

Goldberg R. *Am J Med* 2014;127:379-84.

Important Aspect of Database Selection: Validity of Outcome(s)

- Study conclusions rest on validity of main outcome(s) evaluated
- Critical element in selection of database is validity (accuracy) of outcome data
  - Diagnosis recorded → have disease
  - No diagnosis recorded → do not have disease

Factors Affecting Validity of Outcomes in Electronic Data



Inaccuracies → Misclassification bias

Validation of Clinical Outcomes in Electronic Data Sources

- Defining, classifying clinical outcomes → crucial in epidemiology
- Cost, logistic hurdles considerable
- Automated, validated algorithms for clinical outcomes are valuable tools

Curb JD. *Ann Epidemiol* 2003;13:S122-8.

What Does Validation Mean?

- Two separate concepts of validation
  - Internal validation: Did clinician record dx?
  - External validation: Was the clinician correct?
- Choice of validation depends on question

Nicholson A. *Pharmacepidemiol Drug Saf* 2011;20:321-4.

Finding False Negatives

- Weakness of validation studies: often do not consider missed cases
  - Patients with condition but no diagnostic code
- Estimate patients diagnosed but uncoded:
  - Sample pts without algorithm, review records
  - Compare rates of diagnosis in database with external source

Nicholson A. *Pharmacepidemiol Drug Saf* 2011;20:321-4.

Outline

- Overview of evaluation of outcomes in electronic data sources
- **Steps in validation of clinical outcome**
- Example of outcome validation
  - Hepatic decompensation (end-stage liver dz)
  - Other examples

Steps in Outcomes Determination

1. Select outcome of interest
2. Formulate definition of outcome
   - Requires review of clinical literature
3. Devise methods to ascertain outcome
   - Diagnosis, procedure, lab, pharmacy data
4. Collect data to confirm outcome
   - Gold standard: medical record
   - Record data on structured forms

Steps in Outcomes Determination

5. Adjudicate the endpoint (validation)
   - Review of forms by clinical experts
6. Determine validity of outcome
   - Positive predictive value
   - Neg. predictive value, sensitivity, specificity
   - Target sample size:
     - Width of 95% CI assuming ↑ PPV (at least 80%)

Evaluation of an Algorithm's Performance Characteristics

	Disease	No Disease
Algorithm+	True+	False+
Algorithm-	False-	True-

$$\text{Sensitivity} = \frac{\text{True+}}{\text{(True+) + (False-)}}$$

$$\text{Specificity} = \frac{\text{True-}}{\text{(True-) + (False+)}}$$

$$\text{PPV} = \frac{\text{True+}}{\text{(True+) + (False+)}}$$

$$\text{NPV} = \frac{\text{True-}}{\text{(True-) + (False-)}}$$

Focus on PPV: If high, confidence that outcomes are true events.

Evaluation of an Algorithm's Performance Characteristics

	Disease	No Disease	
Algorithm+	97	5	102
Algorithm-	3	95	98
	100	100	200

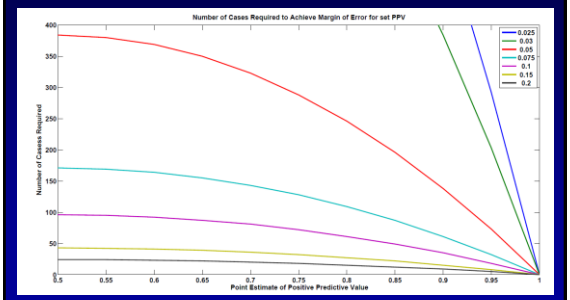
$$\text{Sensitivity} = \frac{\text{True+}}{\text{(True+) + (False-)}} = 97/100 = 0.97$$

$$\text{Specificity} = \frac{\text{True-}}{\text{(True-) + (False+)}} = 95/100 = 0.95$$

$$\text{PPV} = \frac{\text{True+}}{\text{(True+) + (False+)}} = 97/102 = 0.95$$

$$\text{NPV} = \frac{\text{True-}}{\text{(True-) + (False-)}} = 95/98 = 0.97$$

Sample Size Considerations in Validation Studies



Courtesy of Judith Maro, PhD.

Misclassification Bias

- Greater likelihood as PPV decreases
- PPV <80% → misclassification likely
  - Should avoid use of algorithm
- PPV 80-99% → could adjust risk by PPV
- Algorithm PPVs may differ by database
  - Different variables, disease prevalence

Outline

- Overview of evaluation of outcomes in electronic data sources
- Steps in validation of clinical outcome
- **Example of outcome validation**
  - **Hepatic decompensation (end-stage liver dz)**
  - Other examples

Hepatic Decompensation

- Main outcome of chronic liver disease
  - Esp. chronic viral hepatitis
- Few data on hepatic decompensation
  - Lacked methods to ensure validity of events
  - Prevented understanding of:
    - Viral hepatitis, liver disease epidemiology
    - Impact of medications on this outcome

Question: How to identify events validly ?

Specific Aim

- Develop method to screen for, adjudicate hepatic decompensation events
  - Establish case definition of hepatic decompensation
  - Develop method to screen for outcomes
  - Develop method to confirm events

Lo Re V. *Pharmacoepidemiol Drug Saf* 2011;20:689-99.

Study Setting

- 8-Site Veterans Aging Cohort Study (VACS-8):
  - Includes HIV+/- veterans at 8 U.S. VA sites
  - Recruit HIV+, 1:1 age/race/site-matched HIV- patients
  - Administer structured questionnaire yearly
  - Collects data from VA's electronic record system
- Advantages:
  - Large number with chronic liver diseases
  - Ability to screen for outcomes with ICD-9 codes, labs
  - Medical records are electronic

Justice AC. *Med Care* 2006;44 (Supple 2):S13-24.

Study Design / Subjects

- Design: Observational cohort study
- Subjects:
  - All subjects enrolled through 8/15/05 eligible

Steps in Outcomes Determination

1. Select outcome of interest
2. Formulate definition of outcome
   - Requires review of clinical literature
3. Devise methods to ascertain outcome
   - Diagnosis, procedure, lab, pharmacy data
4. Collect data to confirm outcome
   - Gold standard: medical record
   - Record data on structured forms

Steps in Outcomes Determination

1. Select outcome of interest
2. Formulate definition of outcome
   - Requires review of clinical literature
3. Devise methods to ascertain outcome
   - Diagnosis, procedure, lab, pharmacy data
4. Collect data to confirm outcome
   - Gold standard: medical record
   - Record data on structured forms

Step 2. Hepatic Decompensation Outcome Definitions

Diagnosis	Definition
Ascites	1) Reported on abdominal imaging report (Definite) 2) Paracentesis performed (Definite)
Spontaneous Bacterial Peritonitis (SBP)	1) Ascites neutrophil count ≥250 cells/mL (Definite) 2) Bacterial growth from fluid culture (Definite)
Hepatocellular Carcinoma (HCC)	1) Diagnosis from tissue biopsy report (Definite) 2) >2 cm liver mass, 2 imaging studies, w/ cirrhosis (Def) 3) Liver mass on CT/MRI + serum AFP >200 ng/mL (Def)
Variceal Hemorrhage	1) Active bleeding on EGD (Definite) 2) Variceal bleed reported in progress note (Possible)
Encephalopathy	1) Mental confusion documented in note (Definite) 2) Asterixis with ammonia test within 30 d (Possible)

Note: Presence of any one diagnosis in record represented an outcome

Steps in Outcomes Determination

1. Select outcome of interest
2. Formulate definition of outcome
   - Requires review of clinical literature
3. **Devise methods to ascertain outcome**
   - **Diagnosis, procedure, lab, pharmacy data**
4. Collect data to confirm outcome
   - Gold standard: medical record
   - Record data on structured forms

Developing Diagnostic Code Lists to Identify Outcomes

- Important, but challenging, step
  - Same condition described with different codes
- Selection of codes depends on question:
  - Emphasis on sensitivity: select all codes
  - Emphasis on PPV: select specific diagnoses
- Code selection: clinical experts, literature

Step 3. Screening for Hepatic Decompensation

- Patients screened for po… enrollment by:
  - Suggestive ICD-9-CM cod…
    OR
  - Lab abnormalities:
    - Total bilirubin ≥5.0 gm/dL
    - Albumin ≤2.0 gm/dL
    - INR ≥1.7 (no warfarin)

Ascites
SBP
Esophageal variceal bleed
Hepatic encephalopathy
Other (jaundice, hepatorenal)

Hepatologists:
Severe abnormalities in liver synthetic function

- Screening: 1 yr before → 6 mo after enrollment
- Any one code or lab abnormality = Screen +
- Random sample 100 Screen- → confirm absence

Steps in Outcomes Determination

1. Select outcome of interest
2. Formulate definition of outcome
   - Requires review of clinical literature
3. Devise methods to ascertain outcome
   - Diagnosis, procedure, lab, pharmacy data
4. **Collect data to confirm outcome**
   - **Gold standard: medical record**
   - **Record data on structured forms**

Step 4. Medical Record Review to Collect Data

- Subjects who screened + → chart review
- Medical records downloaded from VACS sites
- Text files created for:
  - Progress notes
  - Radiographic studies
  - Surgical pathology reports
  - Laboratory results
- Files imported into electronic database
- Database placed on secure VACS server

Step 4. Medical Record Review to Collect Data

- Data forms collected information from:
  - Abdo ultrasound, CT, MRI: ascites, liver masses
  - Liver biopsy: stage, cirrhosis, cancer
  - Lab data: ammonia, peritoneal fluid
  - Endoscopy: varices (location, bleeding, banded)
  - Notes: encephalopathy, variceal bleed, asterixis
- Abstract: 1 year before → 6 mo after enrolled

Steps in Outcomes Determination

5. Adjudicate the endpoint (validation)
   - Review of forms by clinical experts
6. Determine validity of outcome
   - Positive predictive value
   - Negative predictive value
   - Sensitivity, specificity

Step 5. Outcomes Adjudication

- Data forms scanned in pdf format
- 2 endpoints adjudicators reviewed forms
  - Chronic liver disease specialists
  - Determined:
    - Definite, possible, no event
    - Event date
- Disagreement resolved by 3rd adjudicator

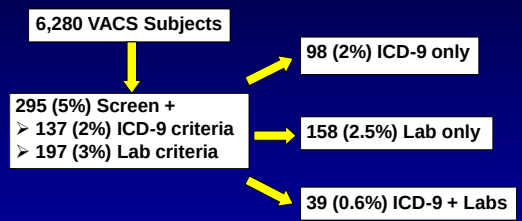
Steps in Outcomes Determination

5. Adjudicate the endpoint (validation)
   - Review of forms by clinical experts
6. Determine validity of outcome
   - Positive predictive value
   - Negative predictive value
   - Sensitivity, specificity

Step 6. Data Analysis

- Positive predictive value of ICD-9 codes/lab abnormalities for decompensation
  - Determine codes / labs with >85% PPV
  - Focus on PPV: if high, clinicians/researchers have confidence that outcomes are true events
- Percent agreement:
  - Concordance between endpoints adjudicators

Results: Subject Selection



Results: Patient Characteristics

Characteristic	All (n=6,280)	Screen– (n=5,985)	Screen+ (n=295)	P-Value
Median age (yrs, IQR)	50 (44 – 55)	50 (44-55)	51 (47-56)	0.001
Hazardous alcohol (no., %)†	2,350 (37%)	2,219 (37%)	131 (44%)	0.01
HIV (no., %)	3,152 (50%)	2,940 (49%)	212 (72%)	<0.001
Hepatitis B (no., %)	372 (6%)	318 (5%)	54 (18%)	<0.001
Hepatitis C (no., %)	2,331 (37%)	2,136 (36%)	195 (66%)	<0.001
HIV + Hepatitis C (no., %)	1,527 (25%)	1,380 (23%)	147 (50%)	<0.001

† Defined by: 1) AUDIT >4 in men or >2 in women and/or ICD-9-CM code for alcohol abuse any time prior to VACS enrollment.

Results: Screen + Patients

Screening Criterion	No. Subjects (%)
ICD-9-CM Code	(n=137)*
Ascites	28 (20%)
Variceal hemorrhage	15 (11%)
Hepatic encephalopathy	9 (7%)
Spontaneous bacterial peritonitis	6 (4%)
Hepatocellular carcinoma	6 (4%)
Other diagnoses suggestive of ESLD	109 (80%)
Laboratory Abnormality	(n=197)*
Albumin ≤2.0 gm/dL	144 (73%)
Total bilirubin ≥5.0 gm/dL	44 (22%)
International normalized ratio ≥1.7	59 (30%)

*Subjects may have had more than one ICD-9 code or lab abnormality recorded

Results: Endpoints Adjudication

- Arbitrator #1: 88 outcomes / 295 (30%)
- Arbitrator #2: 86 outcomes / 295 (29%)

Arbitrator #2	Arbitrator #1: ESLD	Arbitrator #1: No ESLD
ESLD	866	0
No ESLD	2	207

- % agreement = 293/295 = 99%
- Final: 88 outcomes (84 definite; 4 possible)
- Of 100 Screen– pts, no events confirmed

Positive Predictive Values of ICD-9 Diagnosis Codes

ICD-9-CM Code	No. with ICD-9 Code	No. with ESLD	Positive Predictive Value
Ascites	28	24	86%
Spont bacterial peritonitis	6	4	67%
Variceal hemorrhage	15	8	53%
Hepatic encephalopathy	9	1	11%
Hepatocellular carcinoma	6	2	33%
Other possible ESLD diagnoses	109	48	44%
Any ICD-9-CM code	137	57	42%

Lo Re V. *Pharmacoepidemiol Drug Saf* 2011;20:689-99.

Positive Predictive Values of ICD-9 and Lab Abnormalities

ICD-9-CM Codes or Lab Abnormalities	No. with Parameter	No. with ESLD	Positive Predictive Value
Total bilirubin ≥5.0 mg/dL	44	20	45%
Albumin ≤2.0 gm/dL	144	46	32%
INR ≥1.7	59	26	44%
Any laboratory abnormality	197	56	29%
ICD-9 code or lab abnormality	295	88	30%
ICD-9 code + lab abnormality	39	25	64%
1 inpatient, ≥2 outpatient ICD-9 codes for ascites, SBP, variceal bleed	32	29	91%

Lo Re V. *Pharmacoepidemiol Drug Saf* 2011;20:689-99.

Potential Limitations

- Misclassification of outcomes:
  - Minimized likelihood by:
    - Reviewed records of all Screen+ pts
    - Standardized decompensation definitions
    - Classified outcomes: definite, possible, no event
    - Examined events among Screen- pts
- Generalizability: VA health system

Conclusions

- Established liver decompensation definition
- Feasibility of centralized record review
- Developed valid method to identify hepatic decompensation events in VA data
  - Preclude need to review all charts
  - Use in future studies

Outline

- Overview of evaluation of outcomes in electronic data sources
- Steps in validation of clinical outcome
- Example of outcome validation
  - Hepatic decompensation (end-stage liver dz)
  - Other examples

Other Methods to Identify Outcomes in Electronic Databases

Outcome	Source	Algorithm	Definition	PPV
Depression[1]	Medicaid	ICD-9 296.2, 296.3, 298.0, 300.4, 309.0	PHQ-9 score ≥10	66%
Diabetes[2]	Veterans Affairs	ICD-9 250.xx +/- diabetes rx	Self-reported diabetes	93%
Inflammatory bowel disease (IBD)[3]	GPRD	OXMIS 5630C, 5631, 0092ER, 0092LR, 92N	GI consultation, surgery, intestinal biopsy with IBD	92%
Sudden cardiac death/ventricular arrhythmia (VA)[4]	Medicaid	ICD-9 427.1, 427.4, 427.5, 798.1, 798.2	MD-diagnosed cardiac arrest, VA	85%

[1]Kahn LS. *Int J Psych Med* 2008;38:13-29.
[2]Miller DR. *Diabetes Care* 2004:27(S2):B10-21.
[3]Lewis JD. *PDS* 2002;11:211-8.
[4]Hennessy S. *PDS* 2010;19:555-62.

Selection of Analytic Methods Depends on Main Outcome

- Consider types of regression
  - Logistic: cross-sectional, short follow-up
  - Poisson: count of events, incidence rates
  - Cox: time-to-event
- Consider competing risks
  - Event that precludes outcome or alters probability of occurrence

Summary

- Validation of clinical outcomes crucial for electronic data sources
- Suggested steps in validation:
  - Formulate clinical definition of outcome
  - Devise methods to ascertain outcome (codes)
  - Collect data to confirm outcome
  - Adjudication of endpoints
  - Determine validity of electronic outcome

Thought Exercise

- You wish to evaluate the risk of acute liver injury assoc. with oral azole antifungals
  - Concern that ketoconazole may be esp. hepatotoxic
- Questions:
  - What study design would you use?
  - What outcomes should be evaluated?
  - What data source to use to answer the aim?
  - What potential effect modifiers, confounders to collect?