Larger Datasets for Rheumatology Research

Kaleb Michaud, PhD
University of Nebraska Medical Center, Omaha, NE
National Data Bank for Rheumatic Diseases, Wichita, KS

Conflict of Interest / Disclosure Statements

Kaleb Michaud receives funds from the Rheumatology Research Foundation.
- UNMC/Mortality due to rheumatoid arthritis and treatments.
- Electronic Mobile Health Tool for Improving Patient Assessment in the Clinic.

Outline

• Introduction to observational data
• Issues & methods
• Data sets
  – Describing
  – Access
  – Examples of research

Obsenotional Study

• “Real-world”
• Informs clinical practice
• Easier -> Larger
• Time: Cross-sectional vs. longitudinal
• Leads to hypothesis-driven trials
• NOT an experiment (RCT)
Efficacy – RCTs

Effectiveness – observational =
  Efficacy
  Adherence
  Sub-group effects
  Safety

Observational Data

- Drug registries
- Disease registries
- Clinic databases & EHRs
- Administrative data
- Some pragmatic trials

Administrative Data

- Observations at times of service
- Largest number of patients (N)
- Well-defined population, exposure, prescription medications & safety measures
- Issues: misclassify diagnosis, poorly defined effectiveness & severity measures, no OTCs

Outline

- Introduction to observational data
- Issues & methods
- Data sets
  - Describing
  - Access
  - Examples of research
Issues

• Confounding
  – Drugs available at different times
  – Health insurance impacts availability

• Confounding by indication
  – Drug exposure not randomly allocated
  – Rx related to disease severity which is linked to effectiveness outcome

Definitions

X = Independent variable or risk factor or exposure
Y = (Dependent) Outcome

Linear Regression

\[ Y = \beta X + \sum \]

2007 2008
25
20
15
10
5
0

0 5 10 15 20 25
More about Regressions

- All regressions are linear, just require transformations
- Y is binary, then use logistic
- Y is continuous, then linear
- X & Y are not interchangeable (order matters!)

Observational Data - methods

- Wide array to reduce biases and confounding
  - Multiple regressions
  - Nested Case Control
  - Propensity scores
  - Instrumental Variables
  - Many more!

Definitions

- Unmeasured confounders

Propensity Scores

- Accounts for selection bias. Good if lots of C.
- 1: Logistic regression: X=βC + ε
- 2: Match each participating (X=1) patient by PS
- 3: Regression w/new sample
- Assumes all confounders are measured
Instrumental Variables

- Becoming more popular as, unlike PS, it does not assume all confounding factors are observed
- Requires adequate instruments
- 2-stage regression

Metformin and the Risk of Cancer

Time-related biases in observational studies

SUNI SODHI, MD, PhD
LAUREN ADKELL, MD, PhD

OBJECTIVE — Time-related biases in observational studies of drug effects have been described extensively to influence therapeutic attribution bias in diabetes, insulin therapy bias, time-window bias, and time-lag bias all leading to greatly overestimate the benefits observed with a drug.

RESEARCH DESIGN AND METHODS — These time-related biases are described and shown to be present in observational studies that have included metformin with important mechanisms at the incidence of and mortality from cancer. As a consequence, metformin increased much attention as a potential cancer agent; these observational studies avoided the conclusions of randomized, controlled trials of metformin in cancer treatment. However, the questionnaire effects observed in these studies are confounded by time-related biases.

RESULTS — We found that 13 observational studies suffered from immortal time bias, whereas other studies did not consider different time-lagging biases when comparing the first-time treatment metformin with second- or third-line treatments. These studies, subject to time-related biases that are available with proper study design and data analysis, led to highly exaggerated effect sizes, with overestimations in cancer risk with metformin ranging from 20% to 94%. These studies that avoided these biases reported no effect of metformin on cancer incidence.

CONCLUSIONS — Although observational studies are important to better understand the effect of changes in study design and analysis in terms of different time-related biases. With respect to metformin, the scientific evidence of its potential beneficial effects on cancer would need to be measured critically before embarking on further long- and short-term studies.

STUDIES OF PREDICTION

Diabetes Care 35:2045-2053, 2012

Figure 1 — Illustration of immortal time bias using a description of patients exposed to metformin and sulfonylurea who died of cancer according to the definition used in the cohort study by Bower et al. (123). The top patient initiated and continued treatment with a sulfonylurea and subsequently added to or added metformin but was classified as a metformin user during the entire follow-up. The time between entry into the cohort and the first metformin prescription that is immortal (clock time) because the subject must survive to receive this first metformin prescription and is misclassified as exposed to metformin when in fact it is exposed to sulfonylurea, leading to immortal time bias.

Figure 2 — Illustration of immortal time bias using a description of patients exposed to metformin and sulfonylurea who died of cancer according to the definition used in the cohort study by Nissen et al. (124). The top patient initiated and continued treatment with a sulfonylurea and subsequently added to metformin but was classified as a metformin user during the entire follow-up. The time between entry into the cohort and the first metformin prescription that is immortal (clock time) because the subject must survive to receive this first metformin prescription and is misclassified as exposed to metformin when in fact it is exposed to sulfonylurea, leading to immortal time bias.
Outline

• Introduction to observational data
• Issues & methods
• Data sets
  – Describing
  – Access
  – Examples of research

Datasets of datasets

• http://www.healthdata.gov/dataset (1977!)
  – http://healthdatapalooza.org

• https://patientregistry.ahrq.gov (138)

• https://clinicaltrials.gov (514 RA open)
“Big” Data Sets – U.S.

- NHIS
- NHS I, II, & III
- NHANES
- HCUP
- MEPS
- NDI
- Framingham Heart Study
- MANY MANY MORE!
- CMS
- VA
- Payors/providers:
  - United Health Care/Optum
  - Alliance
  - Kaiser Permanente
- Rheumatology:
  - RISE
  - NDB
  - CORRONA
  - CARRAnet

National Health Interview Survey (NHIS)

- Multi-purpose health survey
  - National Center for Health Statistics
  - Annual survey of ~106,000
  - In-person household interview
  - Self-reported health information
- Multi-stage clustered sample design
  - Nationally representative of non-institutionalized civilian population
  - Over-sampled Black and Hispanic persons

Fibromyalgia in the NHIS

- Population mean PSD 2.75; NHIS FM mean 16.8
- 1.75% (3.94M) satisfied NHIS criteria
- 1.8% (4.02M) reported being told by a physician or health professional that they had fibromyalgia
- 73.5% reporting fibromyalgia diagnosis did not have fibromyalgia by NHIS criteria

Walitt B et al. ACR 2014 Abstract

National Health and Nutrition Examination Survey (NHANES)

- Multi-purpose health survey
  - National Center for Health Statistics
  - Nationally representative sample
    - ~5000 non-institutionalized US civilians
    - Over-samples persons 60 and older, African Americans, and Hispanics.
  - Collects an array of data including
    - Physical Exams
    - Laboratory Tests
    - Radiographs

CMS

- Centers for Medicare & Medicaid Services
  - US version of single payor
  - Mostly age 65+, Medicaid all ages by state

- [http://www.resdac.org/](http://www.resdac.org/)
  - 5 to 20% of total patients most common
  - Cost: $0-$25K+. Depends on data.

CMS - Rheumatology Research

  - 2006 to 2010, Full US Medicare, RA patients, claims-based algorithm on effectiveness

  - Matched RA clinical registry (BRASS) with Medicare data

  - 2009, 5% Medicare sample, RA w DMARD dispensed, Part B vs. Part D

  And many many more!

Administrative

- Claims-based
- “large US health plan”
- Largest N, larger $

  - "This was a retrospective study using medical claims, pharmacy claims and enrolment information from a large national health plan database."
  - N = 23,619,150!
  - Acknowledgment of Optum, access to United Health Group (~75M)

ACR Registries (RISE)

**Vision:**

*Improved Data to Guide Practice Decisions*

*Elevated Partnerships with Key Stakeholders*

*Enhanced Clinical Research Opportunities*

The ACR has been designated as a Qualified Clinical Data Registry (QCDR) for 2015

RISE collaborative network…
RISE Statistics

- Total Patient Encounters: 1,278,649
- Total RA Encounters: 414,445
- Total Patients: 282,122
- Total RA Patients: 62,809
- Total Providers: 215
- Total Practices: 45

National Data Bank for Rheumatic Diseases (NDB)

- Non-profit, founded in 1998 by Fred Wolfe
- Goal: a generalizable, comprehensive, valid, reliable, believable data bank for research and teaching in rheumatology
- Data Bank: purposeful, organized, systematic data repository driven (usually) by unwritten hypotheses & questions
- Primary diagnoses: RA, OA, FMS & SLE

US NDB Participants (36K)

NDB - unusual

- Patient-direct data, medical record validation
- Observations at 6-month intervals (Jan & Jul)
- Disease registry (& limited drug registry)
- Combines with participating clinic databases
- Observational data biases + participation bias
- 100+ papers
  - Dusad A et al. 2015 A&R
- http://www.ndb.org
RAIN & VARA vs RCTs

Not covered – future

• Analytics from Google/Twitter/Facebook
• EMR download sharing - “Blue Button”
• Personal Health Vault
• Smartphone data
  – Direct, passive, and 2D/3D scanning
• Public cameras...

Larger data: comes in all sizes, shapes, and sources

Questions? arhp@rheumatology.org

Association of Rheumatology Health Professionals

Thank you for participating!