JH-CERSI/FDA Workshop
Clinical Trials: Assessing Safety and Efficacy for a Diverse Population

Use of Epidemiologic Studies to Examine Safety in Diverse Populations

Judy A. Staffa, Ph.D, R.Ph.

Director

Division of Epidemiology II

OPE/OSE/CDER/FDA

Use of epidemiologic studies to quantify safety issues and identify risk factors

- Types of studies/data available
- Challenges in identifying certain subgroups in large populations
  - Age
  - Sex (Pregnancy)
  - Race/Ethnicity
  - Comorbidities and other Determinants of Health
  - Genetically defined subgroups

Sources of Drug Safety Information

Types of postmarketing observational studies

- **<u>Prospectively</u> collected data – more flexible/opportunities for custom data collection**
  - Registries
  - Cohorts
  - Case/control surveillance

- **<u>Retrospectively</u> collected data – secondary use – less flexible, but opportunities for linkages/enhancements**
  - Electronic healthcare data
    - Administrative claims
    - Electronic health records

Electronic healthcare data

- **Administrative claims**
  - Collected for reimbursement
  - Diagnoses are coded, usually ICD-9
  - Inpt codes more reliable than outpt codes for diagnoses
  - Pharmacy claims captured by Nat'l Drug Code (NDC)
  - Little clinical detail; maybe access to charts
- **Electronic medical records (EMR)**
  - Collected for clinical care
  - Diagnoses coded in more granular ways
  - Often free text
  - Drugs are those prescribed, not disp.
  - Has clinical detail, but can be tough to extract

Typical examples

- **Administrative claims**
  - CMS (Medicare)
  - Health Core
  - IMS Health Life Link
  - Optimum Insight (UHC, Ingenix)
  - Sentinel
  - Medicaid

- **Electronic medical records (EMR)**
  - GE Centricity
  - CPRD (UK)
  - THIN (UK)

- **Hybrids/Integrated care**
  - Kaiser Permanente
  - Veterans Administration
  - Dept of Defense

Subpopulations of interests – some high level thoughts

Subpopulations of interest: AGE

- **Pediatric**
  - Medicaid – lots of sick/indigent children
  - Commercially insured populations – healthier kids
  - Challenges:
    - Sample size
    - Lack of clinical detail on outcomes of interest – e.g, growth, neurodevelopment, metabolic function
- **Women of childbearing age**
  - Fairly straightforward in most data sources
  - Harder to identify “women of childbearing potential”
- **Elderly (65+ years) – harder than you might think!**

Studying drug safety issues in the elderly

- **Medicare coverage begins in the U.S. at 65 years of age**
  - CMS is primary insurer
    - Part A (Hospitalization)
    - Part B (Outpatient)
    - Part C (Capitated payments – no itemized utilization available)
    - Part D (prescription drug coverage)
  - Many administrative claims data only include "supplemental coverage", so only include claims NOT reimbursed by Medicare
    - Affects ability to ascertain drug-related safety outcomes
    - Linkage to Medicare claims solves problem
  - Some administrative claims systems administer Medicare Part C
    - Can ascertain complete care not visible to CMS
- **Challenge – know what data you are working with!**

Subpopulations of interest: SEX

- **Sex**
  - Can tailor selection of study population to target male or female populations
    - Older males – Veterans Administration

- **Female Sex - Pregnancy**
  - In claims data, easy to identify **DELIVERY** – hard to identify **PREGNANCY**
  - Mother-baby linkages have been developed – link to birth certificates
    - Birth certificates are rich source of additional information
  - Algorithms have been developed – common data models developed
    - MEPREP, DoD, CPRD, Medicaid MAX

Subpopulations of interest: RACE/ETHNICITY

- **Notoriously difficult to study**
  - Inaccuracy – how the information is collected
  - Incompleteness – a problem in most databases
    - CMS/Medicare
      - Patient-reported – validity believed to be high, except for "Hispanic"
    - CPRD
      - Recent study documented improvements in validity and completeness since 2006
    - Sentinel
      - Varies across data partners; not complete for entire data system
- **Challenge – Race/Ethnicity are difficult subgroups to study in currently available postmarketing data resources**

Subpopulations of interest: COMORBIDITIES and other determinants of health

- **Medical conditions in administrative claims data are only identifiable by ICD-9 or ICD-10 codes**
  - Outpatient codes not very reliable
    - Well known strategies to maximize payments
    - "Rule-out" diagnoses are common
  - Inpatient codes more scrutinized
    - Still may need validation
    - Comorbid conditions may be chronic – no hospitalization
  - Success is variable – depending on condition
    - E.g., Cardiovascular risks in diabetic patients taking high doses of olmesartan
- **Other determinants of health (BMI, Smoking) most often not available**

Subpopulations of interest: Genetically defined

- **Genetic data are becoming more available than in the past**
  - Subpopulations with these data available are still relatively small
    - Kaiser Permanente
    - Marshfield Clinic
    - Vanderbilt University
  - Linkages to drug exposure and medical outcome data not yet common
    - Ethical/privacy issues and concerns
  - When genetic subgroup is more prevalent in specific group defined by race/ethnicity – harder to get meaningful sample sizes
    - E.g. Antiepileptic drugs and risk for Stevens-Johnson Syndrome (SJS)
- **Bottom line – prospectively collected data may be only option at the current time**

How well can we study subgroups using observational postmarketing studies?

- **Commonly used data sources for retrospective postmarketing observational safety studies have significant limitations for studying many subpopulations**
  - Some characteristics are easier/more difficult than others to define
  - Attaining appropriate sample size is a challenge for most
  - Important to thoroughly understand data source with regard to these characteristics
    - How are these variables collected?
    - How complete are these variables and related information on the subpopulation in the data source?
- **Frequent reason for requesting prospective data collection**
  - Need to provide detailed rationale for appropriate capture of data