“Where are the data?”
Identifying population data to evaluate risk, use, cost and benefit of medical products

ASCPT Workshop: Registries and Databases in Clinical Research

Judith K. Jones, MD, PhD.
The Degge Group, Ltd.
March 22, 2014
Disclosure

- Dr. Jones is a pharmaceutical consultant and is supported by varied grants and fees from medical product manufacturers. She is president of a company, The Degge Group, Ltd., that is the recipient of these grants and honoraria.

- The presentation today is not supported by any specific commercial entity, but does relate to a non-profit activity, DGI, LLC. (of which she is Chief Editor) that collects information on healthcare databases.
Where are the Data?

Goals

- History --How it started—
  - One story, from a drug safety & pharmacoepidemiology perspective

- What we are learning & relevance to
  - Clinical pharmacology
  - Pharmacoepidemiology
  - Pharmacoeconomics
  - Health services research

- “Big Data” is very relevant to All of these areas

- The Future
**History: Illustrative Examples**

- Conceptualizing public health information needs
  - **How to get information?**
    - US 1960’s: Chloramphenical associated Aplastic anemia: the US initiation of spontaneous reports in the 1960’s
  
  - **US 1970’s: Looking for different types/sources of data**
  
  - **EU: 1970-80’s “Where are the data?”**
    - The Subacute Myelo-optic Neuropathy (SMON) epidemic of clioquinol-associated blindness: the stimulus for a global search for data. This resulted in the global RAD-AR effort, later the International Medical Benefit Risk project in Geneva (’90s), later became nonprofit B.R.I.D.G.E. TO DATA in US(2000’s to present)...
Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes

Organization of Data for Storage & Retrieval

Institutional & Insurance needs
For timely person-based medical service data (Medicaid, Medicare)
For
• Payment
• Quality Assurance
• Surveillance

Medical Claims Insurance

Quality Assurance/Fraud Surveillance

Electronic Medical Records

AMA’s > FDA’s AE Collection
Need for assessing Safety, efficacy & Benefit-risk after Marketing approval
Required “Postmarketing Surveillance”
Use &/or requiring data for regulation
• Clinical trial meta-analyses
• Postmarketing data
• REMS/Risk Mgmt data
Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes

**Organization of Data for Storage & Retrieval**
- Institutional & Insurance needs
  - For timely person-based medical service data (Medicaid, Medicare)
  - For
    - Payment
    - Quality Assurance
    - Surveillance

**Public Health Surveillance Needs**
- Disease Surveillance
  - Infections
  - Drug effects
  - Disease Epidemiology (Cancer, etc)

- Prospective Cohorts (e.g. NursesHlth Study)
- AMA’s->FDA’s AE Collection

**Medical Claims Insurance**

**Quality Assurance/Fraud Surveillance**

**Electronic Medical Records**

**Infectious Disease Data**

JKJones 3-22-14
Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes

**Organization of Data for Storage & Retrieval**
- Institutional & Insurance needs
  - For timely person-based medical service data (Medicaid, Medicare)
  - For
    - Payment
    - Quality Assurance
    - Surveillance

**Public Health Surveillance Needs**
- Disease Surveillance
  - Infections
  - Drug effects
  - Disease Epidemiology (Cancer, etc)

**Regulatory needs for balanced decisions**
- Need for assessing Safety, efficacy & Benefit-risk after Marketing approval
- Required “Postmarketing Surveillance”

Use &/or requiring data for regulation
- Clinical trial meta-analyses
- Postmarketing data
- REMS/Risk Mgmt data

Medical Claims Insurance
Quality Assurance/Fraud Surveillance
Electronic Medical Records
Infectious Disease Data
AMA’s-> FDA’s AE Collection
Prospective Cohorts (e.g. NursesHlth Study)
History

- Conceptualizing public health information needs, specifically,

  - WHAT Information?
    - Demographics (i.e., WHO?)
    - Diagnoses, Procedures, Treatments (i.e., WHAT?)
    - Institutional locus (Office, Hospital, ER) (i.e., WHERE?)
    - TIMING of the information (i.e., WHEN?)

  - What is its Context?
    - Local or Generalizable to the entire population
    - Longitudinal, cross-sectional (or both?)

  - How Valid is the information? Can it be reproduced?

  - Timing of the information relative to other data, e.g., exposure
History

Examples,

- From mid-1960’s to mid-1970’s: Medical insurance data:
  - Healthcare legislation requires detailed management of insurance data on medical visits, procedures, drugs & hospitalization as well as providers
    - Example: Medi-Cal
    - Use: to identify specialties as well as patterns of prescribing abuse drugs

- HMO data: Kaiser Oakland developing database on member patients-later in contract with FDA
History

- At FDA
  - Predecessor of AERS database, started in 1969 in part from earlier data from AMA (chloramphenical and other AEs).
  
  - By 1978, AE database had >200,000 AE reports but reporting only ~12,000 reports/year and access to data via flatfile retrieval (took 24-48 hours).
  
  - Early 1980’s FDA’s Drugs & Biologic’s Division of Drug Experience collecting data sources, some with known numerator & denominator to complement AE system to evaluate “signals”
    - Medicaid data from Michigan, Minnesota
    - Boston Collaborative Drug Study Program
    - Registries: Liver, skin, Radiologic contrast
## US Efforts in Post Marketing Surveillance

### An Historical Overview

<table>
<thead>
<tr>
<th>Year</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>FDA-ASHP-AHA cooperative effort</td>
</tr>
<tr>
<td>1960</td>
<td>AMA ADA registry</td>
</tr>
<tr>
<td>1965</td>
<td>FDA spontaneous reporting program</td>
</tr>
<tr>
<td>1970</td>
<td>AFIP-Tissue registry to drugs</td>
</tr>
<tr>
<td>1975</td>
<td>U of Florida Hospital Surveillance</td>
</tr>
<tr>
<td>1980</td>
<td>Oakland Kaiser computerized pharmacy (outpatient)</td>
</tr>
<tr>
<td>1985</td>
<td>Boston Collab. Surv. Program (BCDSP)-hospital</td>
</tr>
<tr>
<td></td>
<td>World Health Organization ADR Surveillance</td>
</tr>
<tr>
<td></td>
<td>Yale U. School of Medicine (HMO Pilot)</td>
</tr>
<tr>
<td></td>
<td>Contrast Media Registry</td>
</tr>
<tr>
<td></td>
<td>Boston Children’s Hosp. Surveillance</td>
</tr>
<tr>
<td></td>
<td>Drug Epidemiology Unit (DEU)</td>
</tr>
<tr>
<td></td>
<td>Eye Registry</td>
</tr>
<tr>
<td></td>
<td>Liver Registry</td>
</tr>
<tr>
<td></td>
<td>Medicaid Data</td>
</tr>
<tr>
<td></td>
<td>Others (Kaiser-LA, Olmsted co. Minn, Saskatchewan)</td>
</tr>
</tbody>
</table>
**Emergence of Defined Needs for Data**

Epidemiology → Pharmacoepidemiology → Pharmacoeconomics

**Data Needed**

**Descriptive** → What is the problem? → Case Definitions

**Non-Quantitative**

- Spontaneous Reports
- Literature reports
- Case Series

**Quantitative** → How often does the problem occur? In whom? & Risk Factors.

- **Population-based data**: longitudinal or cross sectional
  - Ad Hoc for longitudinal cohorts, case control
  - Collected for administrative reasons (i.e., insurance claims)

**Data with varying details**: including demographics, timing & site, diagnoses Therapies & procedures, laboratory tests, physical exam, social status, genetic testing, costs billed, costs reimbursed
Fast Forward to this decade
When are the Data Used?
Fast Forward to this decade
When are the Data Used?

- No longer used *only* for postmarketing safety studies

- Applications from Pre-clinical to Postmarketing are emerging
  - Premarketing and Early clinical development
    - Profile of the Indication population
    - Evaluation of risks and their measurements in REMS after approval

- Evaluation of potential risks and REMS design can be piloted in Phase III

- Postmarketing studies using databases for:
  - Further evaluation of potential risks
  - Evaluation of REMS effectiveness:
    - Behavior of stakeholders (physicians, pharmacists, patients)
    - Safety, if event or risk factors common & identifiable in database

- Evaluation in Different populations, countries requires multiple databases
Planning for Big Population Data Needs in Drug Development

GOALS

- Evaluate existing natural history of Indication Population
- Initial Risk Management

EPI

TIME

Evaluate existing natural history of Indication Population
Initial Risk Management

Large-Scale baseline Survey of Populations (Patients, HCP, Pharmacists) to establish evaluation parameters

Develop Initial Risk Communication Plan

NDA Approval & Launch
POST-Marketing
Implement RM Plan

Evaluation
- Surveys
- Epi DB Studies

Refine Communications
Iterative Monitoring & Feedback

RM Plan: communication & evaluation
Pilot & Refine from Survey

JKJones 3-22-14
2000+ The Growing Array of Global Health Care Data

Health Care System

Clinical Trial Data
Internet Registries

National Health Statistics Data
Deaths, Births, Morbidity
& National Surveillance Systems

Electronic Medical Records;
Inpt, Outpt

UK: GPRD, THIN
France: Thales
US: VA, Military
Korea: Cegedigm
Japan: Keio Hosp

Ongoing Registries
National or International

Gov’t or Third Party
Insurance Health Data

Surveys: Market, Other

Spontaneous Reports
& Case Series Data

Serum, Tissue
Data Banks (DNA)

Longitudinal ad hoc
Cohorts: Oxford,
Harvard Nurses Study
Framingham

JKJones 3-22-14
Fast Forward to this decade
Where are the Data?

- Proliferation of many databases and frequent use

- Recognition of the value of population-based data on very large populations,
  - Ex: Sentinel data, Consortia in the EU of multiple database, IMED (former OMOP)

- Epidemiology, Pharmacoepidemiology & Pharmacoeconomic researchers seek multiple diverse global data sources, but need information on their use & limitations

- Electronic data sources expand the breadth of data inquiry
Fast Forward to this decade
Where are the Data?

- Resources to locate data appear in
  - International Society for Pharmacoeconomics & Outcomes Research (ISPOR)
  - International Society for Pharmacoepidemiology & Drug Safety (ISPE)

- EU efforts
  - EnCePP- Identifying qualified databases for EU
  - Innovative Medicines Initiative (IMI) has funded several consortia and databases
  - EUROCAT-consortium of Birth Defect Data

- B.R.I.D.G.E. TO DATA® US based: Subscription Online database of detailed outline of ~230 global population databases
Where are the Data?
The RAD-AR Project - now B.R.I.D.G.E. TO DATA®

- In 1987-8, Ciba Geigy’s global effort to improve risk assessment & response:
  - The Risk Assessment of Drugs - Analysis & Response (RAD-AR) Project.—evolved into the International Medical Benefit Risk Project, a charity in Geneva.

- One aspect was to answer the question: Where are the data? Out of this, what is currently www.bridgetodata.org was launched.

- This project has identified >230 databases in >30 countries
  - Online subscription “database of databases”
  - Profiled using 10 categories and 75 standard subcategories
  - Helps epidemiology, health economic researchers & academics to find optimal databases for
    - Research
    - Teaching tools
    - Template for those developing databases.
KEY DATABASE FEATURES

- Types of Databases include:
  - Longitudinal Claims & EMR data
  - Drug or Disease specific cohorts
  - Registries
  - National Surveys & National Surveillance Systems
  - Spontaneous Reporting Systems

- ~230 Standardized Database Profiles
- 75 Defined data fields
  - Glossary of database & coding terms (international terms included)
- Profiles from 32 Countries
- Continuously updated
Critical Aspects of large datasets -- especially describing healthcare

- What are the functional needs for applying the data to decision-making?
  - For research: timeliness, validity, linkage or integration with contemporaneous dataset
  - For healthcare: timeliness, validity, reliability, ease of understanding (amenable to visualization)
## Population Medical Data

### Health Care Data: Example of a Medical Claims Profile

<table>
<thead>
<tr>
<th>DATE</th>
<th>Diagnosis</th>
<th>Prescription</th>
<th>Procedure</th>
<th>Result</th>
<th>Provider</th>
<th>Reimb Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/6/07</td>
<td>Osteoarthritis</td>
<td>ibuprofen</td>
<td>Arthroscopy</td>
<td></td>
<td>P13456</td>
<td>300</td>
</tr>
<tr>
<td>5/6/07</td>
<td></td>
<td>ibuprofen</td>
<td></td>
<td></td>
<td>P14445</td>
<td>10</td>
</tr>
<tr>
<td>5/21/07</td>
<td>Diabetes</td>
<td>glyburide</td>
<td></td>
<td></td>
<td>P14445</td>
<td>20</td>
</tr>
<tr>
<td>5/21/07</td>
<td></td>
<td>piroxicam</td>
<td></td>
<td></td>
<td>P14445</td>
<td>25</td>
</tr>
<tr>
<td>5/21/07</td>
<td></td>
<td></td>
<td></td>
<td>HGB A1C</td>
<td>7.0</td>
<td>P35499</td>
</tr>
<tr>
<td>5/21/07</td>
<td></td>
<td></td>
<td></td>
<td>ECG</td>
<td></td>
<td>P14465</td>
</tr>
<tr>
<td>5/21/07</td>
<td>Arthritis</td>
<td></td>
<td></td>
<td></td>
<td>P13456</td>
<td>75</td>
</tr>
<tr>
<td>6/15/07</td>
<td>GI Bleed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/16/07</td>
<td>Arthritis H33421</td>
<td></td>
<td>Hospitalization</td>
<td></td>
<td>H33421</td>
<td>5020</td>
</tr>
<tr>
<td></td>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peptic ulcer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/26/07</td>
<td>ranitidine</td>
<td></td>
<td></td>
<td></td>
<td>P14445</td>
<td>40</td>
</tr>
<tr>
<td>6/26/07</td>
<td>glyburide</td>
<td></td>
<td></td>
<td></td>
<td>P14445</td>
<td>20</td>
</tr>
<tr>
<td>6/26/07</td>
<td>sucralfate</td>
<td></td>
<td></td>
<td></td>
<td>P14445</td>
<td>10</td>
</tr>
</tbody>
</table>
What we are learning

- Types of Data
- Linkages
- Variation in Coding Systems
- Needs for conformity & standards
Types of Data
### Table 1. Examples of Data Fields Used in Profiles (by Category)

<table>
<thead>
<tr>
<th>Category</th>
<th>Data Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>Database description, Database source, Years covered, Population type, Date of last update</td>
</tr>
<tr>
<td>Population Dynamics</td>
<td>Population size, Sample weights – Extrapolation factors</td>
</tr>
<tr>
<td>Demographic Data</td>
<td>Age, Gender, Date of birth, Death recorded, Other demographic data</td>
</tr>
<tr>
<td>Physician &amp; Practitioner Info</td>
<td>Physician ID &amp; Specialty, Pharmacy ID</td>
</tr>
<tr>
<td>Diagnoses/Signs &amp; Symptoms</td>
<td>Diagnosis data, Diagnoses coded (coding systems), Max. number of codes, Physical exam findings, Environmental exposures, Behavioral data elements</td>
</tr>
<tr>
<td>Procedures</td>
<td>Procedure data, Procedures coded (coding systems), Laboratory information</td>
</tr>
<tr>
<td>Drug Information</td>
<td>Drug data, Drug dosage, Drug coding system(s), Additional drug information</td>
</tr>
<tr>
<td>Economic Data</td>
<td>Type of cost data (if applicable)</td>
</tr>
<tr>
<td>Validation &amp; Linkage</td>
<td>Data validation, Access to medical records, Linkage to other databases</td>
</tr>
<tr>
<td>Administrative Data</td>
<td>Database contact data, Database usage restrictions, References of studies using/describing the database</td>
</tr>
</tbody>
</table>
Types of Databases
Specific types of Databases

- **Longitudinal Claims & EMR data**
  - Examples: GPRD & THIN in UK, DOD, VA Regenstrief in US
  - Most common, flexible - captures much data
  - Data collected prior to hypothesis so some bias decreased

- **Drug or Disease specific cohorts**
  - Two armed for comparison, or one-armed (old PMS cohorts)

- **Registries**
  - Common method for rare disorders and risk management:
    Birth Defects, rare diseases, ways to observe treated population

- **National Surveys & National Surveillance Systems**
  - Common in many countries for birth defects, cancer, infectious disease

- **Spontaneous Reporting Systems**
Terms on Search Page
Database Type: REGISTRY
Birth Defect Data: YES

Initial Results
88 Database Profiles with:
100% search term match = 12 (Registry and Birth Defect data)
50% search term match = 76 (Registry or Birth Defect data)

12 Database Profiles
REMOVED profiles with 50% TERM MATCH

Excluded profiles only matching 1 search term were from the following countries (76 removed):
Australia (4) Denmark (4) Iceland (1) New Zealand (2) Belgium (1) France (3) India (1) Spain (2) Sweden (4) UK (6) United States (27)

FINAL Search Results
9 Profiles of Registries with Birth Defect Data

Excluded profiles with an out of scope Population Type
Manitoba Health Insurance Registry (Canada)
National Registry of Drug-Induced Ocular Side Effects (USA)
Swedish Cause of Death Registry (Sweden)

Comparison among 9 profiles based on frequency of data field usage

Box 1
Box 2
Box 2a
Box 2b
Box 3
Box 4-6

Core Data Fields
35 data fields with similar usage in 9 registries with birth defect data, include:
Brief Database Description
Frequency of Data Collection
Years Covered
Database Population Size
Gender Data
Date of Birth Recorded
Diagnosis Data
Physical Examination Findings
Birth Defect Data
Data Validation Against Original Source
Linkage to Other Databases
References of Studies Using/Describing Database

Additional Data Fields
33 data fields used in some registries with birth defect data, include:
Sample Weights - Extrapolation Factors
Age of Patients at Data Collection
Ethnicity / Race Data
Death Recorded
Environmental Exposures
Behavioral Data Elements
Procedure Data
Laboratory Information
Drug Data
Drug: Dosage
Drug Coding System: Primary
Access to Medical Records

Data Fields Not Used
7 data fields infrequently used in registries with birth defect data, include:
Pharmacy ID
Cost Data

Among Core Data Fields, 8 conform to a similar format
Frequency of Data Collection → Ongoing
Final Population Size → Data still being collected
Date of Birth Recorded → Yes
Diagnosis Data → Yes
Birth Defect Data → Yes
Linkage to Other Databases → Yes
Database Usage Restrictions → Primarily Private Access
Number of Publications Using Database → >10
Data Linkages
Types of Database Linkages

A1. Direct Linkages
- Insurance DB to Cancer Registry
- MRFIT to Nat Death Index

A2. Multiple Direct Links
- Norwegian Dbs
- Manitoba Hlth

B. Indirect Linkages
- Iceland → Approval → Death Registry

C. Formed by Linkage
  C1. Combination of DB Subsets
    - SEER CA registry to Medicare
  C2. Merged Databases
    - Combined single Registries, i.e., NARCOM (US) for MS.

A. Direct Linkage (n=81)

A.1 Direct Linkage (DB ‘A’ links to DB ‘B’)
- Korean Health Insurance Review Agency (HIRA) Database links to Korea Central Cancer Registry
- Multiple Risk Factor Intervention Trial (MRFIT) links to National Death Index (NDI)

A.2 Multiple Direct Linkage (Network of linkages across DBs ‘A’ through ‘E’)
- Norwegian national registers
- Manitoba Population Health Research Data

B. Indirect Linkage (n=18)

(Blinking DB ‘A’ to DB ‘B’ requires an extra step)
- Icelandic Cancer Registry needs approval prior to linkage of datasets to Cause of Death Registry

C. Formed by Linkage (n=38)

C.1 Combination of Database Subsets (DB ‘A’ subset links to DB ‘B’ subset to form new DB ‘C’)
- SEER - Medicare Database (USA) linkage of SEER cancer registries data, and the Medicare enrollment and claims files

C.2 Merged Databases (DB ‘A’ merges with DB ‘B’ to form new DB ‘C’)
- North American Research Committee on Multiple Sclerosis (NARCOMS) Registry formed by multiple regional MS registries
- AIHW National Diabetes Register (Australia) formed by the National Diabetes Services Scheme database (NDSS) and the Australasian Paediatric Endocrine Group’s (APEG) state and territory databases
Patient-centered Data
Perspectives on a Patient: The many dimensions of Big Healthcare Data

- National, global population data
  - Effective treatments
  - Outcomes
  - Economic, etc.

- Local epidemiologic data
  - Rates
  - Outcomes
  - Economic, etc.

- Disease data
  - Natural History
  - Prognosis
  - Treatment options & data

- Therapeutic data
  - Biopharmaceuticals
  - Surgical, Devices
  - Economic, etc.

- Laboratory data

- Radiology data

- Pathology data

- Electron Microscopy data

- Genetic data

- Microbiome data

- Institutional data
  - HCPs
  - Surgical
  - Economic, etc.
Profiling Patient Experience
Actual patient with X disease x 20,000 to profile the environment of future product’s use

- What Specialties?
- What Continuity of Care?
- Common co-morbidities and concomitant drugs
- Will non-traditional healthcare confound use/effectiveness/risks?
The Future

- More diverse use of DBs for decision-making in both practice & commercial setting (e.g., throughout the cycle of medical product development) and importantly, public health.

- As DB’s proliferate, essential need for standardization of:
  - Coding systems & definitions
  - DB structure
The Future

- Greater awareness & standardization will support:
  - Development of new, more useful DBs for public health and practice uses
  - Support for the practice of multi-country studies and standardization will facilitate meta-analyses
  - Focused product development, planning for risk management and surveying postmarketing for use, risk & benefit
Thank you & Credit to my database team:

- Anokhi Kapasi PhD
- Sharmila Kamani
- Dave Honig
- Varinder Singh
- Bao Nguyen, PharmD, MPH
- Earl Goehring