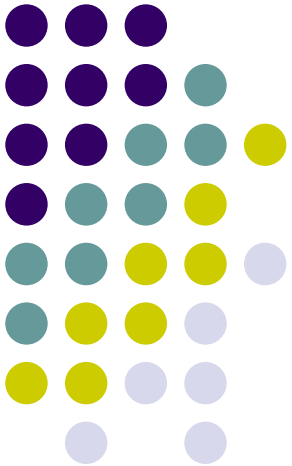
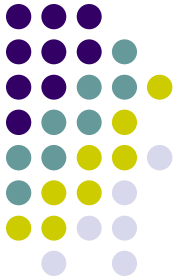
"Where are the data?"
Identifying population data to evaluate risk, use, cost and benefit of medical products

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.

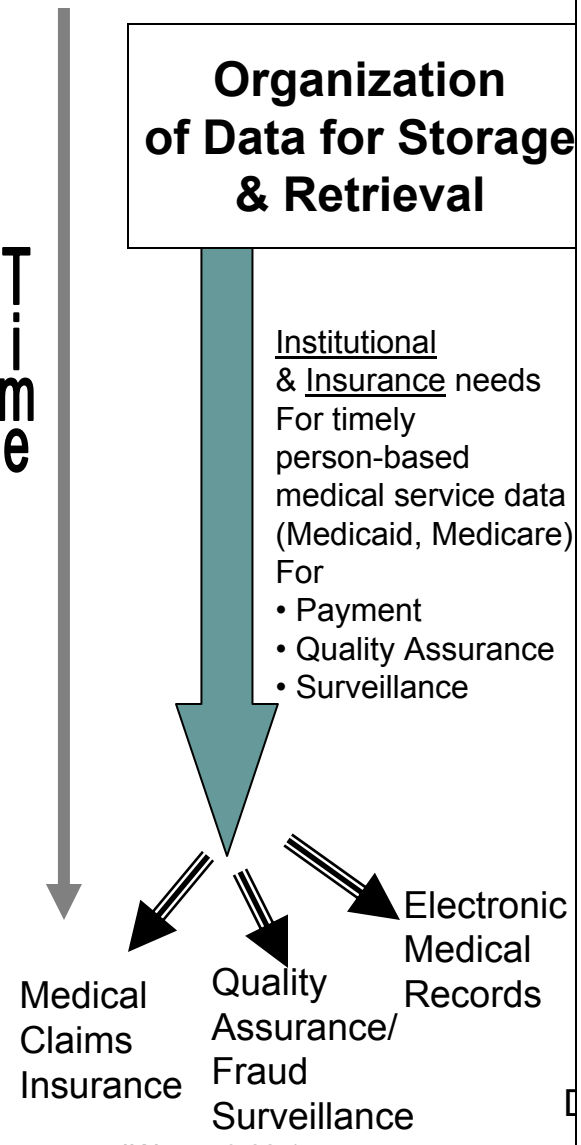
Where are the Data?
Goals

- 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

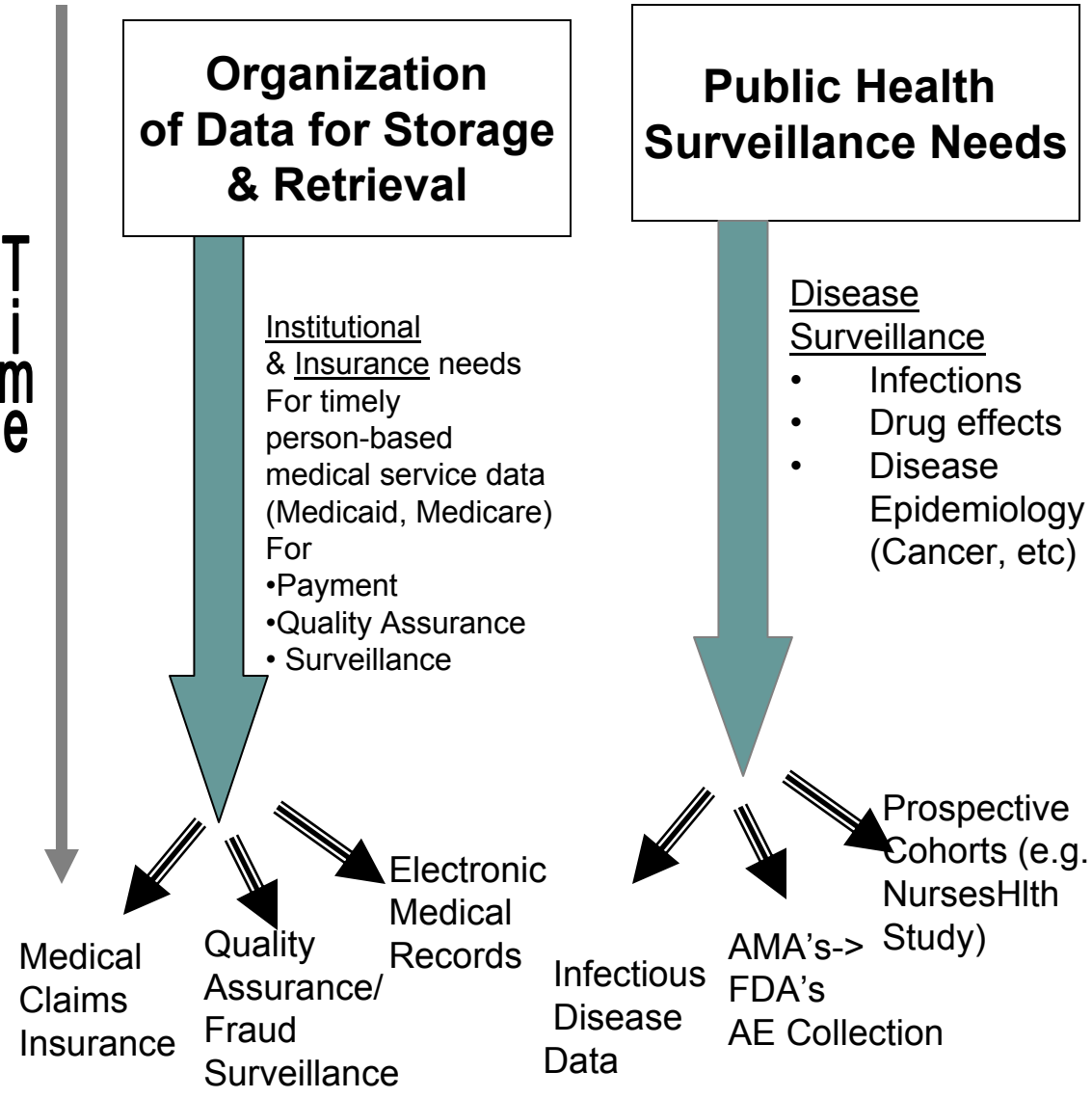
History: Illustrative Examples

- 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
    - Propoxyphene: petition to withdraw from the market in late 1970's and capturing relevant data. Presentation of Matrix of databases to profile the "Darvon Problem" at Int.Clin. Pharmacology meeting in London, 1979.
  - 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)...

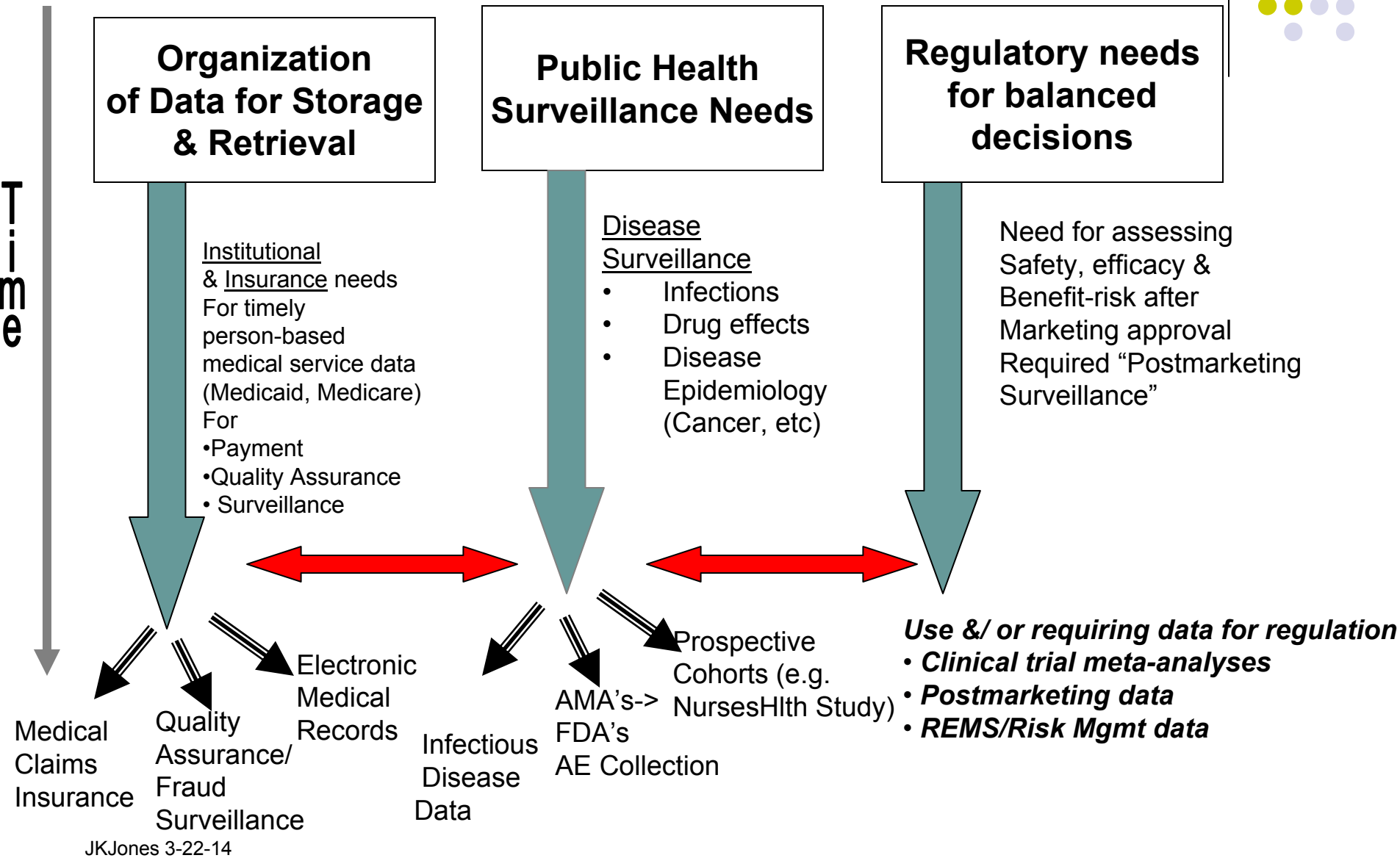
Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes



Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes



Drivers for Collection of Medical & Pharmaceutical Data: Utilization, Effects, Outcomes

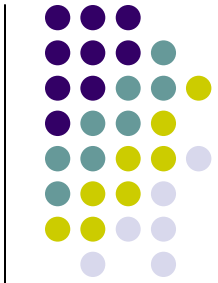


History

- 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,

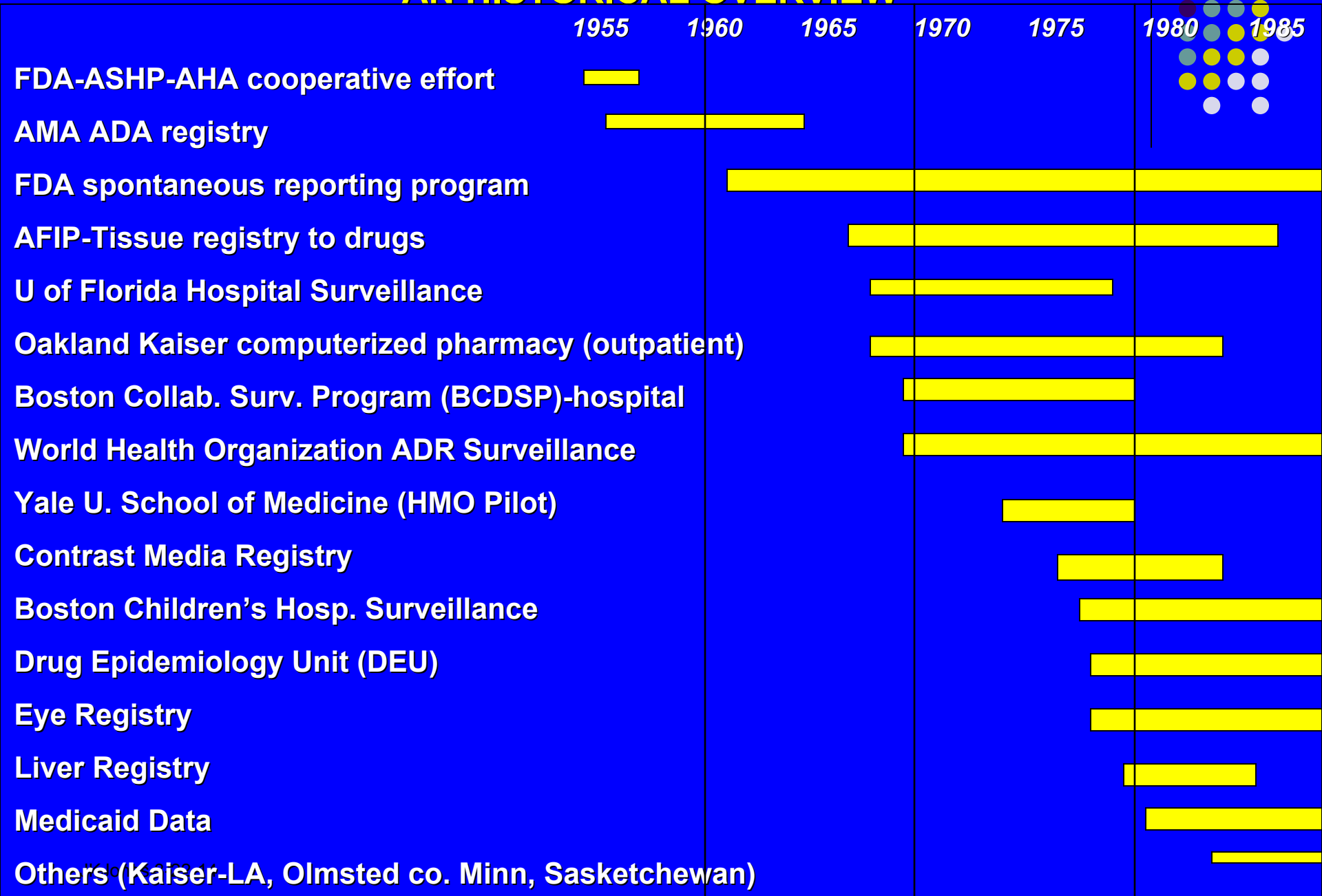
- From mid-1960's to mid-1970's: Medical insurance data:
  - Healthcare legislation requires detailed management of insurance data on medical visits, procedures, drugs & hospitalization as well as providers
    - Example: Medi-Cal
    - Use: to identify specialties as well as patterns of prescribing abuse drugs
- HMO data: Kaiser Oakland developing database on member patients-later in contract with FDA

History

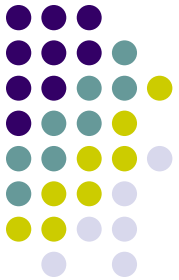
- At FDA
  - Predecessor of AERS database, started in 1969 in part from earlier data from AMA (chloramphenical and other AEs).
  - By 1978, AE database had >200,000 AE reports but reporting only ~12,000 reports/year and access to data via flatfile retrieval (took 24-48 hours).
  - Early 1980's FDA's 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



Data Resources Available to FDA in 1980



Epidemiological Resources/Data

Spontaneous Reports

WHO

Sponsor (~85%)

Medical Community ~10-15%+

BCDSP

Drug Epidemiology Unit

Medicaid 2+ States

Puget Sound HMQ

Boston Children's Hospital

FDA

Liver Registry

Eye Registry

Tissue Reactions Registry (AFIP)

NIDA Drug Abuse Data

IMS Use Data

= FDA Funded

=Part FDA Funded

JKJones 3-22-14

Emergence of Defined Needs for Data

Epidemiology → Pharmacoepidemiology → Pharmacoeconomics

Data Needed

Descriptive → **What is the problem**? → Case Definitions
Non-Quantitative

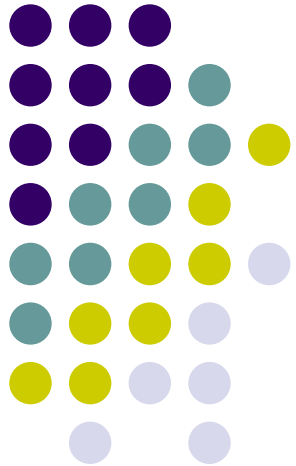
- Spontaneous Reports
- Literature reports
- Case Series

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

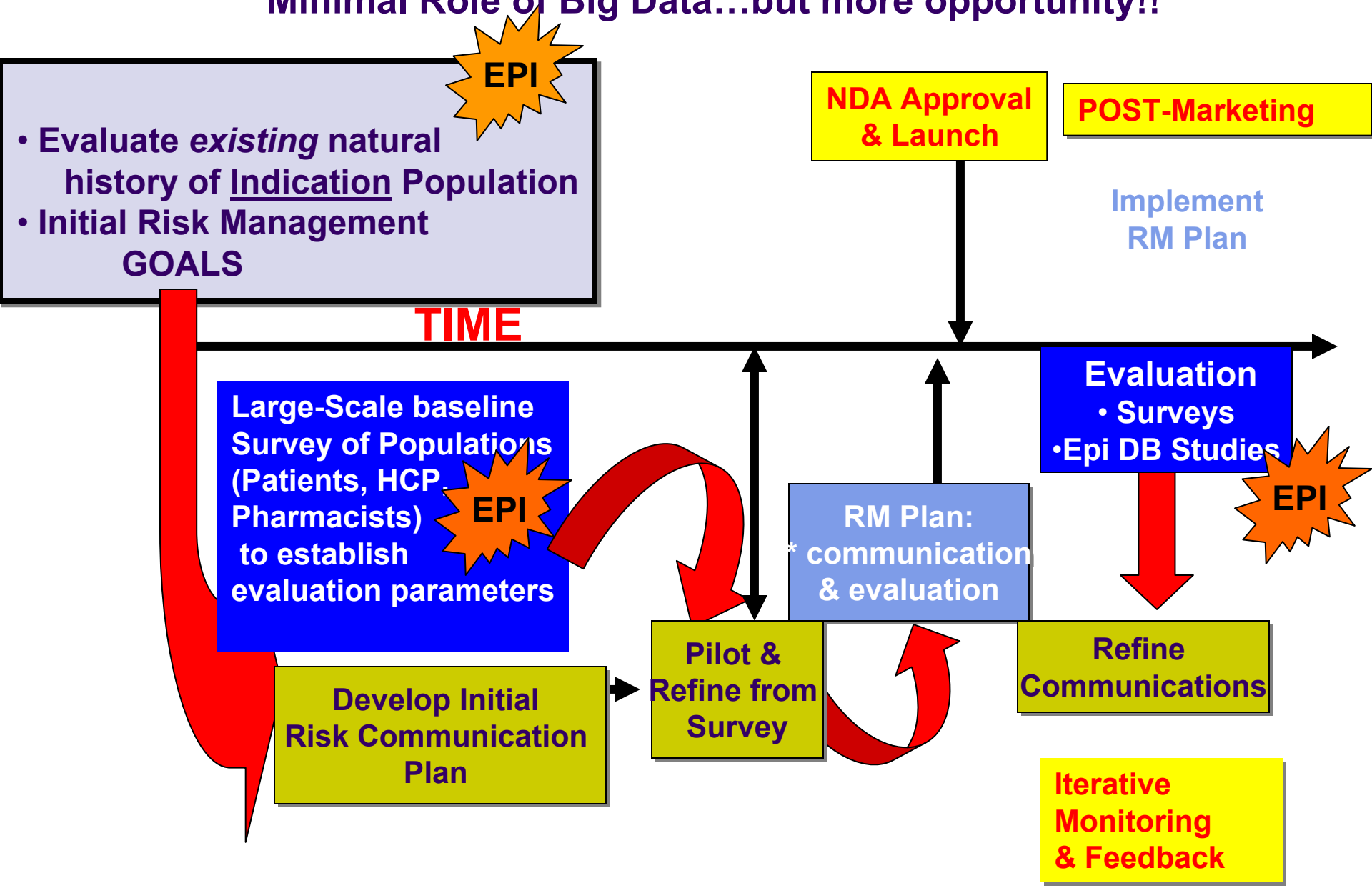
Fast Forward to this decade
When are the Data Used?



Fast Forward to this decade
When are the Data Used?

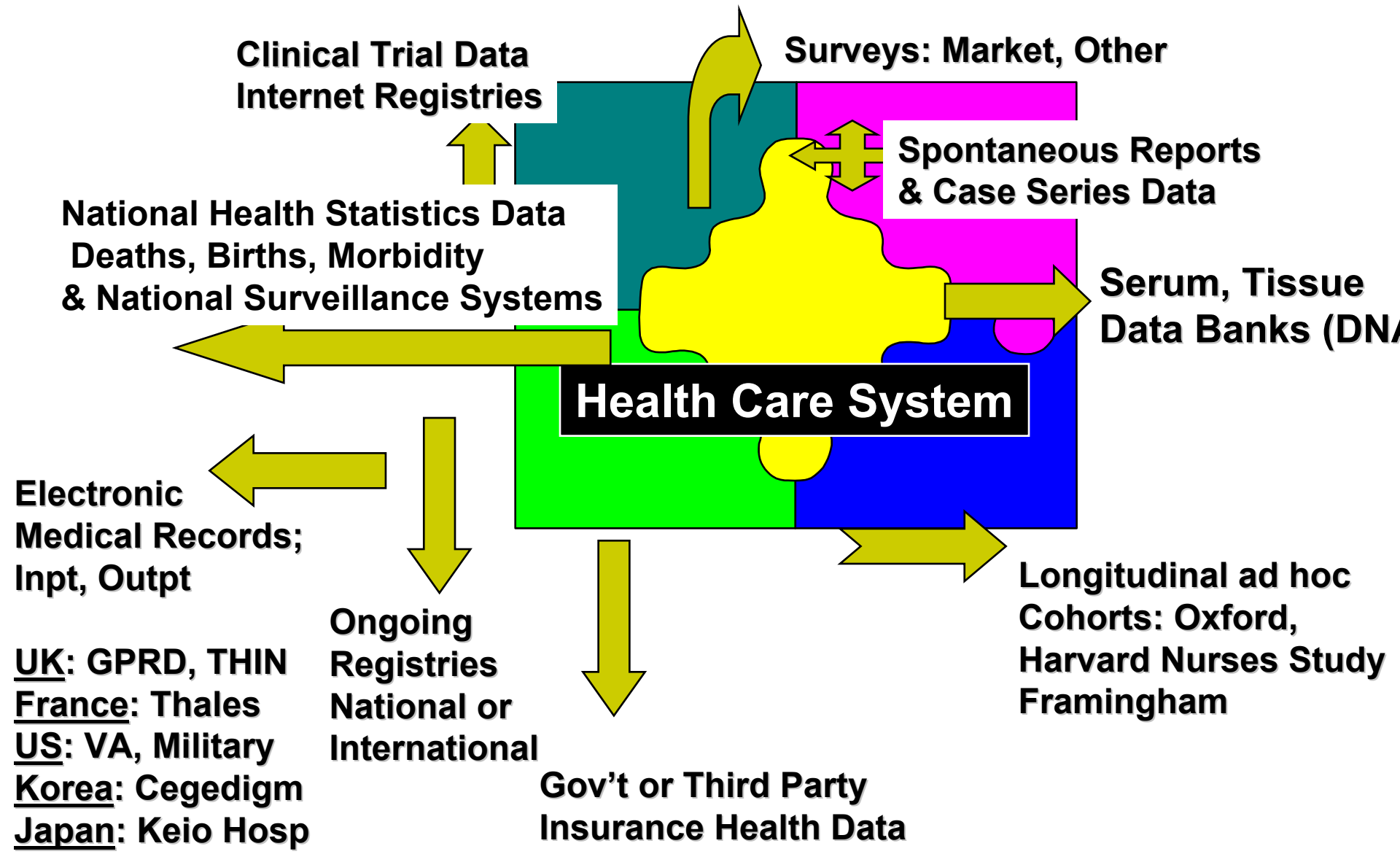
- No longer used *only* for postmarketing safety studies
- Applications from Pre-clinical to Postmarketing are emerging
  - Premarketing and Early clinical development
    - Profile of the Indication population
    - Evaluation of risks and their measurements in REMS after approval
- Evaluation of potential 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
Minimal Role of Big Data…but more opportunity!!



JKJones 3-22-14

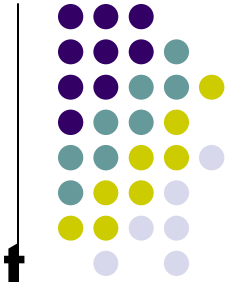
2000+ The Growing Array of Global Health Care Data



JKJones 3-22-14

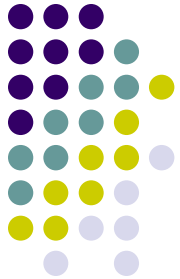
Fast Forward to this decade

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**.

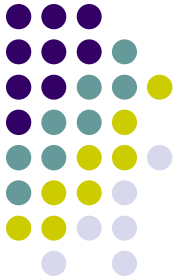
B.R.I.D.G.E. TO DATA®

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)

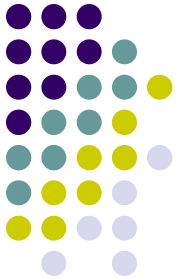
Population Medical Data

Health Care Data: Example of a Medical Claims Profile

Longitudinal

DATE	Diagnosis	Prescription	Procedure	Result	Provider	Reimb Cost
5/6/07	Osteoarthritis	ibuprofen	Arthroscopy		P13456	300
5/6/07		ibuprofen			P14445	10
5/21/07	Diabetes	glyburide			P14445	20
5/21/07		piroxicam			P14445	25
5/21/07			HGB A1C	7.0	P35499	12
5/21/07			ECG		P14465	60
5/21/07	Arthritis				P13456	75
6/15/07	GI Bleed		Hospitalization		H33421	5020
6/16/07	Arthritis H33421 Diabetes Peptic ulcer					
6/26/07		ranitidine			P14445	40
6/26/07		glyburide			P14445	20
6/26/07		sucralfate			P14445	10

What we are learning



- Types of Data
- Linkages
- Variation in Coding Systems
- Needs for conformity & standards

Types of Data

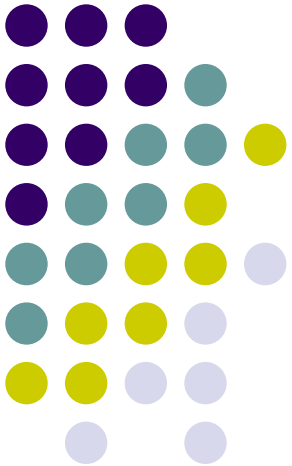
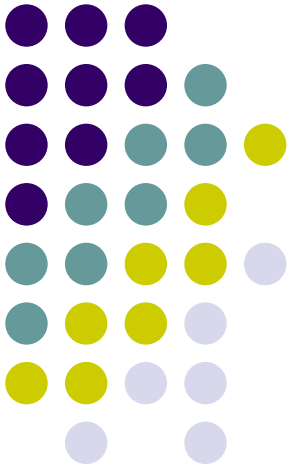


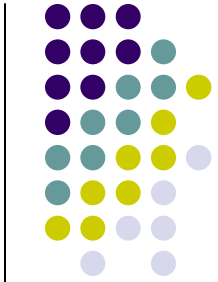
Table 1. Examples of Data Fields Used in Profiles (by Category)

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

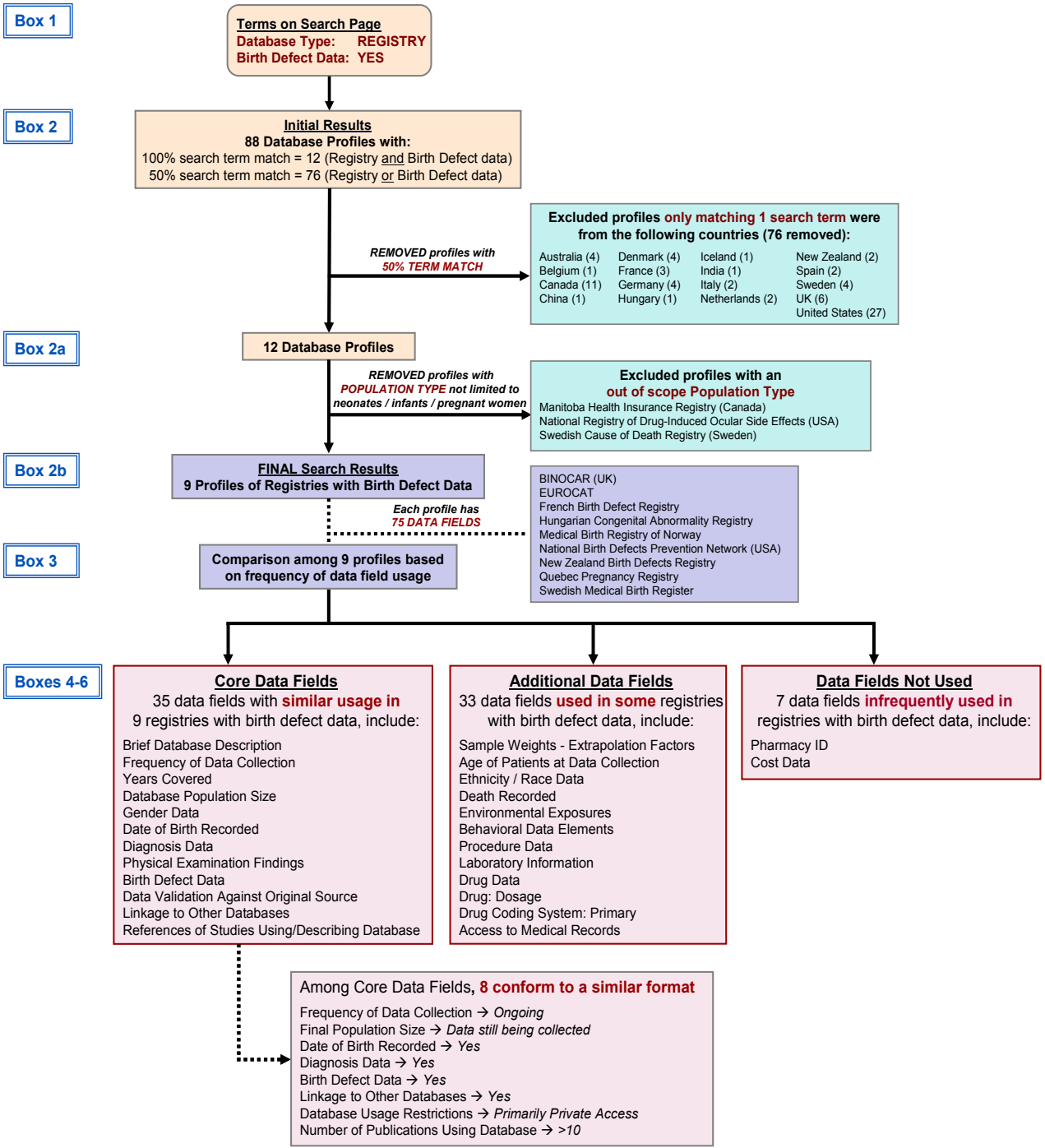
Types of Databases



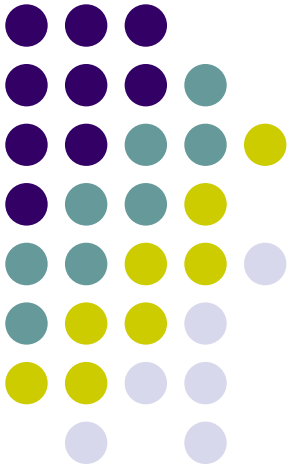
Specific types of Databases



- **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**


Box 1
Terms on Search Page
Database Type: REGISTRY
Birth Defect Data: YES
Box 2
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)
REMOVED profiles with 50% TERM MATCH
Excluded profiles only matching 1 search term were from the following countries (76 removed):
Australia (4)
Belgium (1)
Canada (11)
China (1)
Denmark (4)
France (3)
Germany (4)
Hungary (1)
Iceland (1)
India (1)
Italy (2)
Netherlands (2)
New Zealand (2)
Spain (2)
Sweden (4)
UK (6)
United States (27)
Box 2a
12 Database Profiles
REMOVED profiles with POPULATION TYPE not limited to neonates / infants / pregnant women
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)
Box 2b
FINAL Search Results
9 Profiles of Registries with Birth Defect Data
Each profile has 75 DATA FIELDS
BINOCAR (UK)
EUROCAT
French Birth Defect Registry
Hungarian Congenital Abnormality Registry
Medical Birth Registry of Norway
National Birth Defects Prevention Network (USA)
New Zealand Birth Defects Registry
Quebec Pregnancy Registry
Swedish Medical Birth Register
Box 3
Comparison among 9 profiles based on frequency of data field usage
Boxes 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

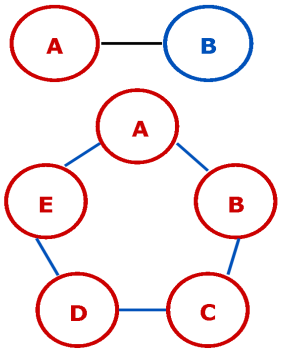
B. Indirect Linkages
- Iceland→Approval→ Death Registry

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')

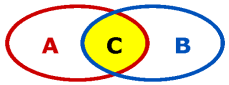
- Norwegian national registers
- Manitoba Population Health Research Data

B. Indirect Linkage (n=18)

(Linking DB 'A' to DB 'B' requires an extra step)

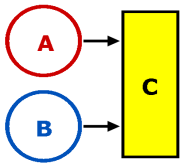
- Icelandic Cancer Registry needs approval prior to linkage of datasets to Cause of Death Registry

C. Formed by Linkage (n=38)



C.1 Combination of Database Subsets (DB 'A' sub-set links to DB 'B' subset to form new DB 'C')

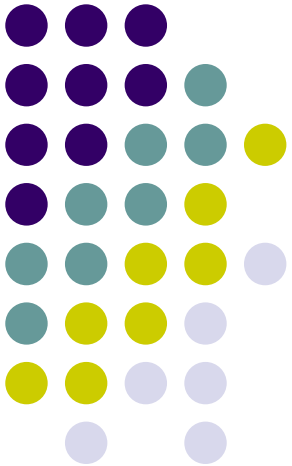
- SEER - Medicare Database (USA) linkage of SEER cancer registries data, and the Medicare enrollment and claims files



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

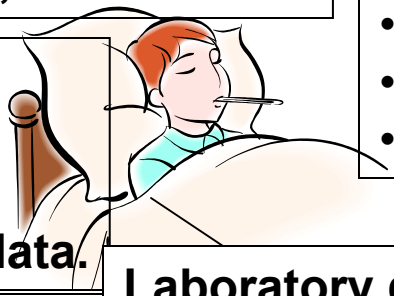


Perspectives on a Patient: The many dimensions of Big Healthcare Data

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

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

Institutional data
- **HCPs**
- **Surgical**
- **Economic, etc.**



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

Laboratory data

Radiology data

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

Pathology data

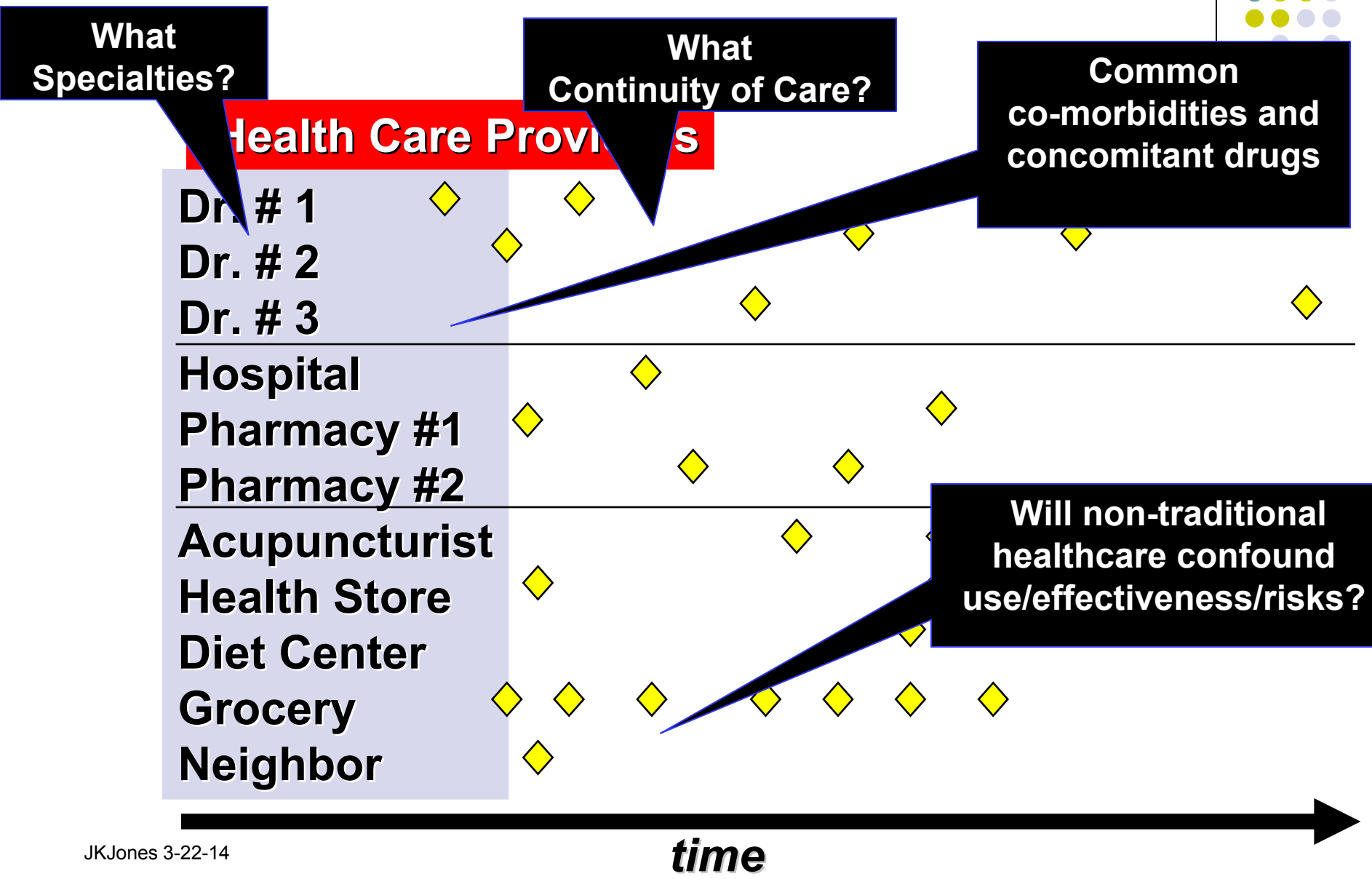
Electron Microscopy data

Genetic data

Microbiome data

Profiling Patient Experience

Actual patient with X disease x 20,000 to profile the environment of future product's use



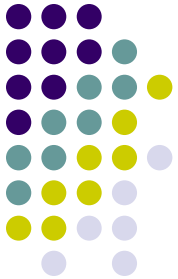
The Future

- 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