# *DATA SOURCES FOR PHARMACOEPIDEMIOLOGY RESEARCH*

**Dr Joseph O. Fadare**
ASSOCIATE PROFESSOR (CLINICAL PHARMACOLOGY AND THERAPEUTICS)
CONSULTANT PHYSICIAN CLINICAL PHARMACOLOGIST
EKITI SATE UNIVERSITY /EKITI STATE UNIVERSITY TEACHING HOSPITAL
ADO-EKITI , NIGERIA

## DISCLOSURE

- Some of the slides are from educational materials developed for previous training workshops by ISPE.
- Special acknowledgements to Prof. Vincent Lo Re



## OBJECTIVES

- To describe sources of data for pharmacoepidemiology research
- To highlight the strengths and limitations of these data sources



## DEFINITION OF PHARMACOEPIDEMIOLOGY

- Pharmacoepidemiology is the branch of epidemiology that studies the use and effect of medicines in specific populations. It studies the relationships between patients, diseases, and medicines.
- Some examples of applications of pharmacoepidemiology are to:
- Monitor the use and effects of medicines in populations
- Measure the occurrence of diseases
- Study the natural history of diseases
- Measure the characteristics of patients with and without specific diseases

## WHAT IS DATA

- Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation ***(www.merriam-webster.com)***
- Information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer ***(dictionary.cambridge.org)***



## TYPES OF DATA

- Primary data:
  - Original data
  - Can involve all cadres of health care workers
  - Can also be gotten from patients and their relatives
  - Documents used include prescriptions, medical records, dispensing records
  - May be through structured instrument/s
  - Used mainly for drug utilization studies



## TYPES OF DATA (contd)

- Secondary data:
  - Usually administrative and clinical data
  - Health insurance claims databases
  - Re-imbursement data
  - Electronic medical records
  - Aggregate-level data such as sales data (distribution or hospital based)
  - Can be linked with other databases



## TYPES OF DATA USED IN PE STUDIES

- Clinical data
- Field data
- Retrospective observational data
- Registries

## SOURCES OF DATA FOR PE STUDIES

- Registries
- Claims databases
- Electronic medical record (EMR) databases
- Hybrid databases



## ADVANTAGES OF AUTOMATED DATABASES

- Allow evaluation of health conditions in "real world" settings
- Use of electronic data sources containing medical care data of more than 10-30 years
- Cost- effectiveness (time and resources)



## COMPONENTS OF AN IDEAL AUTOMATED DATABASES

- Longitudinal data from all care settings
- Records prescribed, dispensed drugs
- Includes laboratory tests results
- Large representative population
- Linkable to other data sources (via identifiers)
- Confounders of interest available
- Updatable, with access to medical records
- Ideal Automated Data

***Shah BR. Am Heart J 2010;160:8 15.***



## STRENGTHS OF AUTOMATED DATABASES

- Relevant clinical data
- Large, real world clinical population
- Longitudinal and linkable
- Short time frame from design to results

***Suissa S. Nat Clin Pract Rheumatol 2007;3:725 32.***

## LIMITATIONS OF AUTOMATED DATABASES

- Uncertain validity of diagnoses
- Completeness, quality of data
- Instability of population
- Generalizability
- Costs of data



## REGISTRIES

- Prospective study of patients with common characteristics
- Developed to evaluate:
  - Natural history of disease
  - Drug effectiveness, safety
  - Quality of life
  - Cost effectiveness of therapies



## TYPES OF DATA COLLECTED BY REGISTRIES

- Collect data on:
  - Demographic characteristics
  - Social history
  - Disease specific drug treatments
  - Select disease related outcomes
- Ability to link to other data sources?



## REGISTRIES: PROS AND CONS

**PROS**

- Large patient numbers
- Usual diagnostic, follow up procedures
- Contain "real world" therapeutic effectiveness, safety data
- Heterogeneity among sites

**CONS**

- Selection bias (non sequential patients)
- Variability in data definitions
- Data may not be validated
- Incomplete data on comorbid conditions,outcomes, mortality
- Inability to link with other data sources

## EXAMPLES OF REGISTRIES FOR PE STUDIES

- Cancer registries
  - Gastric Cancer Registry
  - Breast Cancer Surveillance Consortium
- Disease registry
  - Children's Health Foundation Pediatric Asthma Registry (link is external)
- Pregnancy registry



## MEDICAL INSURANCE CLAIMS DATABASE

- Billing for use of healthcare system
- Diagnoses cannot be verified
- Coding issues with different hospitals
- Pharmacy claims – dispensed?
- Concern for lack of completeness
- No body mass index, BP, tobacco, alcohol data, etc



## EXAMPLES OF CLAIMS DATABASES

- US Medicaid, Medicare
- Various Medical Schemes in South Africa, Namibia
- HMOs in Nigeria



## ELECTRONIC MEDICAL RECORDS DATABASES

- Generated at the time of visit
- Data include:
  - Medical diagnoses (ICD code)
  - Drug prescriptions (not dispensing)
  - Laboratory results
  - Procedures carried out
- Still have concerns for incompleteness and out of network care

## ELECTRONIC MEDICAL RECORDS CPRD

- Clinical Practice Research Datalink
- Sponsored by the UK MHRA and NIHR
- De-identified patient data collected from over 1000 GP practices in the UK
- See https://www.cprd.com/
- Started in 1987
- Patient count now about 35 million



## EXAMPLE – NIGERIAN HMO DATABASE

| | | | | | | |
|---|---|---|---|---|---|---|
| 66 Yrs | Female | ASTHMA | IV HYDROCORT.,IV AMINO | Ibuprofen | Prednisolone | Salbutamol (Aerosol) |
| 67 Yrs | Male | ASTHMA | Ibuprofen, prednisolone, ve | Amoxycillin | Aminophylline | Hydrocortisone |
| 70 Yrs | Male | ASTHMA | IV HYDROCORTISONE,CA | Paracetamol | Prednisolone | Salbutamol (Aerosol) |
| 69 Yrs | Male | BENIGN PROSTATE HYPERPI | prednisolone, pcm, ventolin. | | Furosemide | |
| 67 Yrs | Male | BENIGN PROSTATE HYPERPI | FRUSEUNIDE,LOSARTAN | Lisinopril | Ampicillin + Cloxacillin | Ascorbic Acid |
| 69 Yrs | Male | BENIGN PROSTATE HYPERTI | 10CC N/S DISTILLED WAT | Amlodipine | Ascorbic Acid | |
| 67 Yrs | Male | BRONCHOLITIS | AMLODIPINE-5MG DLY*5/ | Omeprazole | Artemeter + Lumefantrine | |



## HYBRID OR COMBINED DATABASES

- Administrative AND clinical databases
- Reap benefits of claims and medical record data
- Some may have less diverse populations
- Examples:
  - Veterans Affairs, Kaiser Permanente (USA)
  - International Research Consortia for HIV Data



## EXAMPLES OF HYBRID DATABASES: Veterans Administration Health Data

- Largest integrated health care system in US
- Available data:
  - Inpatient/outpatient ICD diagnoses, drugs
  - Procedures, biopsies
  - Laboratory data
- Linkable (registries, Medicare, Medicaid)

## EXAMPLES OF HYBRID DATABASES: International Epidemiology Databases to Evaluate AIDS

- Collects HIV/AIDS data from 7 regions
- 4 in Africa (Southern, East, West, Central)
- North America, Asia, Central/South America
- Available data:
  - Medical diagnoses, comorbidities
  - Antiretroviral drugs
  - Laboratory data (e.g., HIV RNA, CD4)



## SELECTION OF APPROPRIATE DATABASE FOR RESEARCH

- Research questions?
- The absence of automated databases should not deter discourage us from conducting DUR
- Important questions to ask include:
  - What is the population covered?
  - Are there continuous, consistent data?
  - Exposure, outcomes
  - Confounders of interest
  - Is follow up sufficiently long enough?
  - Access to medical records?
  - Ability to link to other data sources?



## SELECTION OF APPROPRIATE DATABASE FOR RESEARCH

- Research question dictates database
- Available "checklists" to guide researchers:
  - ISPE guidelines
  - ISPOR guidelines



## ETHICAL ISSUES RELATED TO PATIENT DATA USAGE

- Privacy
- Confidentiality
- Security

- There are regulations regulating the use of data in many countries but….
- Usually de-identified data is used but that definition is also debatable…

## CONCLUSION

- Data for PE research can be sourced from different types of databases
- Researchers need to consider the strengths and limitations when making their choice

KEY MESSAGE

- Absence of automated databases should not be an excuse for not conducting DUR



## REFERENCES


- Registry Databases: Overview .Gliklich RE, Dreyer NA. AHRQ publication No. 07 EHC001. http://effectivehealthcare.ahrq.gov/ehc/products/21/12/PatOutExecSumm.pdf
- Suissa S. Nat Clin Pract Rheumatol 2007;3:725 32.
- Shah BR. Am Heart J 2010;160:8 15
- www.iedea.org
- Frank Delisle. Data Sources Available for Pharmacoepidemiology Studies
- European Patients Academy on Therapeutic Innovation. Pharmacoepidemiology