Paper AB11

Registry/Observational Studies: Study Designs, Data Sources & Key steps in designing such studies

Anupama Sheoran, Bayer, Whippany NJ, USA

ABSTRACT

Epidemiology contributes to the ontogenesis of products and their evaluation in real-life conditions by leading the conceptualization, design and conduct of observational studies characterizing populations, diseases, treatment and drug effects through methodological expertise. There is a lot more interest in observational studies (Real Life Evidence Generation) nowadays as the environment is changing in terms of new regulations, increasing use of observational data, accessibility of healthcare databases and clinical data registries to regulatory authorities & clinical research organizations as well as increasing interest and confidence in real-world data. This paper will address and discuss the following topics:

1. Study designs for Observational studies and data sources
    - Study types – overview and explanation
    - Data sources and databases
2. Key steps in designing Observational studies

INTRODUCTION

Let's begin with the scientific definition of Epidemiology. It is defined as the study & analysis of the distribution (frequency, pattern) and determinants (causes, risk factors) of health & disease frequency in defined human populations.

(Hennekens C. Epidemiology in Medicine, 1987)

In short, epidemiology attempts to determine why certain people are getting ill. Way back in 1854 there was a severe outbreak of cholera in London. This outbreak, which killed 616 people, is best known for the physician John Snow's study of its causes and his hypothesis that contaminated water, not air, was the source of cholera. He basically used a geographical grid to chart deaths from the outbreak and investigated each case to determine who had access to the water pump; he came up with a positive proof that the water from the pump was the source of the epidemic. Dr. John Snow is regarded as one of the founding fathers of modern epidemiology.

The science of epidemiology was first developed to discover and understand possible causes of contagious diseases. Much of the current epidemiological research done at the U.S. Centers for Disease Control still uses theories such as Dr. Snow's to track the sources and causes of many diseases. It has been expanded to include the study of factors associated with non-transmissible diseases like cancer. Epidemiological methods are applied to studies of the clinical use of drugs in populations after the Thalidomide tragedy. A very important lesson was learned from the Thalidomide tragedy in the year 1962. The Thalidomide disaster is one of the darkest episodes in pharmaceutical research history. The drug was marketed as a mild sleeping pill safe even for pregnant women. The drug was used against nausea and to alleviate morning sickness in pregnant women. Shortly after, many infants were born with phocomelia (malformation of the limbs).

The negative effects of thalidomide led to the development of more structured drug regulations and control over drug use and development. By passing the Kefauver-Harris Drug Amendments Act in 1962, legislators tightened restrictions surrounding the surveillance and approval process for drugs to be sold in the U.S., requiring that manufacturers prove they are both safe and effective before they are marketed. This real data from the tragedy kindled global consciousness regarding the need to monitor the safety of medicines.

The changing times has led to the growing strategic use of epidemiology in industry by understanding disease area profiles and assessing drug effects (use, safety, effectiveness). Epidemiology contributes to drug development in many ways:

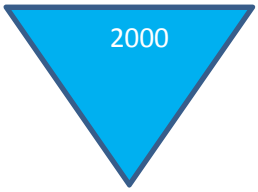
1) By determining risk factors of disease in new therapeutic areas.
2) By validating targets in ongoing pharmacogenomics projects at research stage.
3) During early & late development – by supporting the design of clinical trials.
4) During life cycle management – by conducting drug utilizing studies and data to support the "value story" of a drug.

In the year 1970, healthcare databases became available for epidemiological observational research.

e.g. Canada (Saskatchewan), UK (VAMP-GPRD-CPRD)

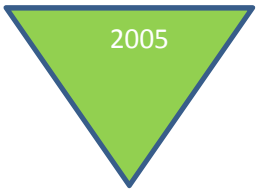
In the 1980's payers were increasingly requiring evidence of health outcomes. The customers started asking for real world data based on field and database research. It was in the 1990's that the regulatory authorities, academia, CRO's and the pharma companies got accessibility of health care databases and by the year 2000, epidemiological expertise was widely used by regulators and pharma industry.

The environment & the industry are changing rapidly. Let's take a look below:



Increasing interest in real world data also by payers and patients

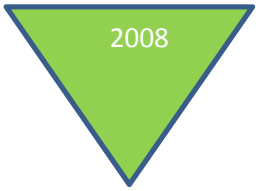
(e.g. NICE 2004)



Expanded use of health care databases with development of

distributed data networks and increasing use of novel analytical tools

(e.g. Sentinel,FDA)

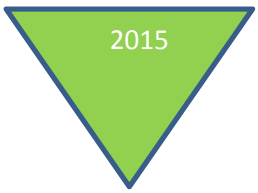


Increasing importance of public private consortiums and partnerships

(e.g. Innovative Medicine's Initiative)

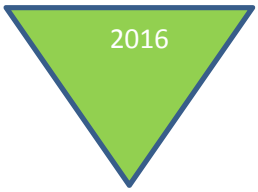
Regulators and payers defining ways to jointly assess the safety and

Effectiveness of drugs (e.g. EunetHTA)



Increasing interest in use of observational data for key decision-making

by regulators and payers. Use of novel, hybrid designs, data sources



Interest in applying real-world data to drug development got a major boost with the passage of the 21st Century Cures Act in 2016.

The main objectives are to take into account the natural history, risks, the health status and health care utilization to analyze the safety & effectiveness as well as the cause of disease. This paves the way to policies, guidelines, public health interventions and even label changes.

Next is evidence generation, producing evidence from the data and ensuring that the evidence is reliable. Later, sharing the evidence to make informed decisions.

E.g. Observational or Registry studies, they include natural history of disease and post-authorization safety and effectiveness studies. By allowing wide patient selection criteria which will include patients with multiple confounding complications, wide age ranges, various socio-economic backgrounds and differing healthcare attitudes, learning how the drug behaves in such complex scenarios provides valuable information. In registry studies sponsor is not dictating how the drug will be used, so there is always the possibility it will be used off label. A review of how the drug is actually used in real-world practice can provide valuable marketing information, hypothesis development for future studies, and indications for use development for future regulatory submissions.

HOW TO DO A REGISTRY/OBSERVATIONAL STUDY?

It is difficult to find guidance on how to do them as there is a lack of consensus standards for such studies.

In the planning phase, identify your stakeholders, the scope of data required, identify the patient outcomes or endpoints, and define the target population. Also get the funding, set up the registry team and finally plan an exit strategy when the study is completed.

In the design phase, details of the study are worked out and a protocol is written. There are only a few options for study designs which are as below:

a) **COHORT DESIGNS** follow over time a group of people who possess a characteristic to see if they develop a particular endpoint or outcomes. It mainly comprises persons with a common baseline characteristic. Cohorts are often defined according to their level of exposure to a drug or intervention.
   Cohort study types:
   - Retrospective cohort study where cases of outcome have already occurred.
   - Prospective cohort study where cases of outcome have not yet occurred.
   - Closed cohort where no members are added during the period of follow-up and no losses other than due to the outcome event.
   - Dynamic (open) cohort where new members can enter and move out during follow-up and loss of follow-up may occur.

   Advantages: The main advantage of the cohort design is that it has a clear timeline separating potential confounders from the exposure and the exposure from the outcome. Cohort studies allow investigators to assess multiple outcomes from given treatments.

   Limitations: If participants need to be recruited and followed over time for the incidence of the outcome, the cohort design quickly becomes inefficient when the incidence of the outcome is low.

b) In **CASE-CONTROL DESIGNS** you gather 'cases' of patients where the events have already happened and who have a particular outcome or who have had a particular adverse event and 'controls' who have not, and then you look backwards to see what proportion had an exposure or characteristic of interest. The control group is chosen from the same population that gave rise to the cases and controls are sampled independently of their status regarding the exposure of interest.

   Advantages: The oversampling of persons with the outcome increases efficiency compared with the full underlying cohort.
   Limitations: The case control study is difficult to conceptualize. Matching doesn't control for confounding in a case-control study as it does in a cohort study.

c) A **CASE-COHORT DESIGN** is a statistical variant of a case-control study. Controls are sampled from a list of people, with each person having an equal probability of being sampled. It is intended to increase efficiency compared with the nested case-control design when selecting participants for whom additional information needs to be collected or when studying more than one outcome.

d) The **CROSS-SECTIONAL DESIGN** is a 'snapshot' of the population status at a specific point in time with respect of disease or exposure variables or both.

POTENTIAL DATA SOURCES

Primary Data: This data is collected by the investigator directly from study participants to address a specific question or hypothesis. This can be collected in various ways such as in-person, telephone, mail or computerized questionnaires.

Secondary Data: Such data is generated during patient care, which is the primary purpose of the data being generated. Depending on their primary purpose, secondary data have various formats e.g. medical/health records, health insurance claims for services provided, lab results, pharmacy dispensing databases, disease and drug registries.

Data from various databases are linked together. Electronic/paper-based healthcare databases provide the primary care and hospital records. Administrative databases provide the insurance claims records. Existing registries for cancer and rare disease, also death registries, pharmacy record databases, census data – all are secondary sources of data.

Examples of Administrative databases (claims databases) -

Saskatchewan database (Canada)

MarketScan and Optum (USA)

GePaRD (Germany)

SNIIRAM (France)

Examples of Physician-based databases (medical records databases) -

CPRD (former GPRD, UK)

THIN database (UK)

Kaiser Permanente (USA)

KEY STEPS TO FOLLOW WHEN DESIGNING AN OBSERVATIONAL STUDY (REAL LIFE EVIDENCE GENERATION)

1) Carefully define the scientific question – Consider up-front potential challenges including sources of bias and ways to prevent/control them.
2) Data sources evaluation – Explore data options and availability of information (exposure, outcomes, covariates etc.). Existing data sources (secondary data source) and new data sources (primary data source).
3) Research approach definition – Look for the most efficient design. Consider all options. Ideally include a comparison group. Also selecting the comparison group is very critical because subjects have a choice as to which intervention they receive.
4) Feasibility assessment – Assess number of people with a specific exposure or outcome, the availability of covariates and follow up period. Validate diagnosis, procedures, and outcomes. Conduct operational feasibility. Determine whether the data source could answer the research questions within the expected timelines and with the required statistical power for the proposed study design.
5) Evidence generation – Study execution according to timelines and milestones. Regular follow-up.
6) Evidence dissemination – Use of evidence for decision-making. Study report, scientific publications.

CONCLUSION

Real world data (RWD) and real world evidence (RWE) are playing an increasing role in health care decisions. Post market safety is monitored using this data and to make regulatory decisions. The 21st Century Cures Act, passed in 2016, places additional focus on the use of these types of data to support regulatory decision making. Companies or medical product developers are using this data and evidence to support clinical trial designs, observational studies to generate innovative & new treatment approaches in conjunction with epidemiological knowledge. There are many reasons for performing such observational studies like to obtain data for reimbursement purposes, the effectiveness of the drug or device in the real world, FDA may require a post-approval study under Section 522 of the Food, Drug and Cosmetic Act, off-label uses (drugs aren't always used the way we think or in the populations we anticipate). This is becoming real because of the use of computers and other tools over the past decades to gather and store huge amounts of health-related data. This data holds potential to allow us to better design and conduct clinical trials in the health care setting to answer many questions previously not achievable. Statistical analysis of registry data is no different than statistical analysis of clinical data. These can easily be as complex and costly as the clinical studies. What makes such studies different is their ability to assess a drug's ability to achieve its intended use in the real world. They are used when alternative solutions don't exist, are outdated, or perhaps unethical. Real world data is transforming the drug development process. Many companies are turning to real-world data. The trend is further fueled by rapid growth in the volume of medical information collected in digital formats and available for analysis. As of 2015, 87 percent of physicians in US and nearly all hospitals were using electronic health records to store patient data. The edge provided by real-world data is already large and will dramatically improve the ability to acquire useful evidence about products. It will allow to get drugs to market more quickly, expand their usage to new indications more quickly and answer critical safety questions to guide the safe and appropriate use of medicines. Companies are bracing themselves to build up their capabilities in this area amidst the new regulatory environment.

REFERENCES

https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

https://en.wikipedia.org/wiki/Thalidomide

http://clinicaldevice.typepad.com/cdg_whitepapers/2011/07/registry-studies-why-and-how.html

https://www.ncbi.nlm.nih.gov/books/NBK49444/

http://pmep.cce.cornell.edu/profiles/extoxnet/TIB/epidemiology.html

https://www.clinicalleader.com/doc/eight-steps-to-more-productive-observational-studies-0001

https://www.amgenscience.com/real-world-data-in-drug-development/

https://www.fda.gov/ScienceResearch/SpecialTopics/RealWorldEvidence/default.htm

CONTACT INFORMATION

Author Name: Anupama Sheoran

Company: Bayer U.S. LLC

Address: 100 Bayer Blvd.

City/Postcode: Whippany NJ 07981 USA

Work Phone: 862-404-7544

Email: anupama.sheoran@bayer.com