

Asian Journal of Research in Biological and Pharmaceutical Sciences

Journal home page: www.ajrbps.com



REVIEW ON EPIDEMIOLOGICAL STUDY DESIGNS

V.J. Divya^{*1}, A. Vikneswari¹, Sini Joseph¹, Rajesh Singh Bista¹

^{*1}Department of Pharmacy Practice, Bharathi College of Pharmacy, Mandya, Karnataka, India.

ABSTRACT

The article gives an overview of traditional classification of epidemiological study designs as well as study designs based on modern epidemiology. Epidemiology is the study of the distribution and determinants of disease frequency in human populations and the application of this study to control health problems. The choice is often between validity, i.e. obtaining the most accurate answer, and feasibility, i.e. obtaining an answer. When the individual is the unit of analysis and the disease outcome under study is dichotomous, then epidemiological study designs can best be classified according to two criteria: (i) the type of outcome under study (incidence or prevalence) and (ii) whether there is sampling on the basis of the outcome. Once this two-dimensional classification system has been adopted, then there are only four basic study designs (i) incidence studies; (ii) incidence case - control studies; (iii) prevalence studies; and (iv) prevalence case - control studies (Rothman *et al*). Continuous outcome measures using longitudinal and cross sectional studies have been mentioned.

KEYWORDS

Epidemiology, Study Design and Two-Dimensional Classification.

Author for Correspondence:

V.J. Divya,
Department of Pharmacy Practice,
Bharathi College of Pharmacy,
Mandya, Karnataka, India.

Email: siyac26@gmail.com

INTRODUCTION

Epidemiology is the study of the distribution and determinants of disease frequency in human populations and the application of this study to control health problems^{1,2}. *Classical* epidemiology is primarily concerned with the statistical relationships between disease agents, both infectious and non-infectious; *ecological* epidemiology studies describe (often mathematically) the ecological interactions between populations of hosts and infectious agents. Other sub-categories, e.g. molecular epidemiology, clinical epidemiology or environmental epidemiology, relate to the techniques and domains in which the quantitative tools are being applied.

Often the techniques and approaches will be different, but the two ubiquitous components are a population-based approach and quantification³. These epidemiological studies are frequently called observational studies because the researcher observes what is happening or has happened without intervening in the natural progression of disease events³.

Choice of study design

Having settled on a study hypothesis and/or the required measure of disease occurrence, the subsequent decision is which type of study is appropriate. The decision will be based not only on methodological but also on practical considerations. For example, the most appropriate study may be too expensive or take too long to provide an answer. In such circumstances a compromise will require to be made - to undertake a study which can be conducted within the budget and time available and which delivers information which is suitable for answering a hypothesis or provides a useful measure of disease occurrence⁴. The choice is often between validity, i.e. obtaining the most accurate answer, and feasibility, i.e. obtaining an answer.

There are a number of broad considerations

1. Ecologic and migrant studies are primarily used to generate hypotheses about the aetiology of disease. If appropriate information is routinely collected, they can be conducted quickly and at low cost.
2. Cross-sectional studies generally are able to determine only associations between risk factor and disease. They can also be the method through which other types of study are conducted.
3. The cohort approach allows identification of multiple disease outcomes from a single exposure, whereas the case-control approach allows identification of multiple exposures associated with a single disease entity.
4. The lack of quality control of data from a retrospective cohort study, particularly on exposure status, would support a prospective approach. Similarly, data may be sufficient for the primary exposure of interest, but may

be lacking on possible confounders that need to be considered.

5. The prospective cohort approach, in theory, also permits setting up systems to notify change in exposure status during the follow-up period, an option that may be lacking in a retrospectively derived cohort with only 'point' data on exposure.
6. Prospective cohort studies suffer from the problems of potential and unknown loss-to-follow-up rates: it is increasingly to track down individuals after a time interval. Assessment of disease status may then be impossible from within the study.
7. Cohort studies are substantially more expensive than the smaller case control approach. The rarer the disease the more impracticable the cohort approach becomes. Studies that involve population screening to derive either current or future cases are more expensive than those that can utilize an existing morbidity recording system, such as a population-based cancer register.
8. Time is relevant in so far as public health questions that require an immediate answer, for example regarding risks from current occupational exposure, might not be able to wait for the 10 years it might take for a prospective study to reach an answer.
9. The availability of data may dictate the choice available⁴.

Design and Analysis of Observational Studies

In observational studies a series of steps are undertaken (Frankena and Thrusfield, 1997)⁵.

1. The objectives of the study are defined
2. The target population is described
3. The sampling method is selected and sample size calculated
4. Disease and exposure factors are measured in the sample
5. Bias (selection, misclassification, information or recall bias and confounding) is evaluated
6. Data is validated
7. Data is analysed
8. Findings are reported

If care is taken in the planning, implementation and analysis of observational studies, risk factors can be identified to allow preventive measures to be instigated⁵.

Each type of observational study is useful under different circumstances. The following Table No.1 provides a guide to the advantages and disadvantages of each type of study and may help in understanding the decisions researchers make in designing an epidemiological study.

The Four Basic Study Designs

When the individual is the unit of analysis and the disease outcome under study is dichotomous, then epidemiological study designs can best be classified according to two criteria: (i) the type of outcome under study (incidence or prevalence) and (ii) whether there is sampling on the basis of the outcome. This classification system has previously been proposed by neil pierce (2012)⁶ Greenland and Morgenstern (1988)⁷ and Morgenstern and Thomas (1993)⁸, all of whom followed previous authors^{9,10} in rejecting directionality (i.e. prospective/retrospective or from exposure to outcome vs from outcome to exposure) as a key feature for distinguishing study designs. Once this two-dimensional classification system has been adopted, then there are only four basic study designs (i) incidence studies; (ii) incidence case - control studies; (iii) prevalence studies; and (iv) prevalence case - control studies (Rothman et al)¹¹. It should first be emphasized that all epidemiological studies are (or should be) based on a particular population (the 'source population') followed over a particular period of time (the 'risk period'). Within this framework, the most fundamental distinction is between studies of disease 'incidence' and studies of disease 'prevalence'. Once this distinction has been drawn, then the different epidemiological study designs differ primarily in the manner in which information is drawn from the source population and risk period⁶.

Incidence studies

Incidence studies ideally measure exposures, confounders and outcome times of all population members. Incidence studies also include studies

where the source population has been defined but a cohort has not been formally enumerated by the investigator, e.g. 'descriptive' studies of national death rates. Furthermore, there is no fundamental distinction between incidence studies based on a broad population (e.g. all workers at a particular factory or all persons living in a particular geographical area) and incidence studies involving sampling on the basis of exposure, since the latter procedure merely redefines the study population (cohort). Three measures of disease occurrence are commonly used in incidence studies. Perhaps the most common measure is the person - time 'incidence rate'; a second measure is the 'incidence proportion' (average risk), which is the proportion of study subjects who experience the outcome of interest at any time during the follow-up period. A third possible measure is the 'incidence odds', which is the ratio of the number of subjects who experience the outcome to the number of subjects who do not experience the outcome. These three measures of disease occurrence all involve the same numerator: the number of incident cases of disease. They differ in whether their denominators represent person - time at risk, persons at risk or survivors. Corresponding to these three measures of disease occurrence, the three ratio measures of effect used in incidence studies are the 'rate ratio', 'risk ratio' and 'odds ratio'⁶.

Incidence case - control studies

Incidence studies are usually the preferred approach to studying the causes of disease, because they use all of the available information on the source population over the risk period. However, they are often very expensive in terms of time and resources, and the equivalent results may be achieved more efficiently by using an incidence case - control study design. In incidence case - control studies, the relative risk measure is the 'odds ratio'. The effect measure that the odds ratio (OR) obtained from this case - control study will estimate depends on the manner in which controls are selected. Once again, there are three main options that define three subtypes of incidence case - control studies. One option is to select controls at random from those who

do not experience the outcome during the follow-up period, i.e. the 'survivors' (those who did not develop the outcome at any time during the follow-up period). In this instance, a sample of controls chosen by 'cumulative sampling' (or exclusive sampling) will estimate the exposure odds of the survivors, and the OR obtained in the case - control study will therefore estimate the incidence OR in the base population. Early descriptions of the case - control approach were usually of this type.

These descriptions emphasized that the OR was approximately equal to the risk ratio when the disease was rare. It was later recognized that controls can be sampled at random from the entire 'source population' (those at risk at the beginning of follow-up) rather than just from the survivors (those at risk at the end of follow-up). This approach, which has been reinvented several times since it was first proposed by Thomas, has more recently been termed 'case - cohort sampling' (or inclusive sampling).

Prevalence studies

Incidence studies are usually the preferred approach to studying the causes of disease, but they often involve lengthy periods of follow-up and large resources¹². Also, for some diseases (e.g. asthma and diabetes), incidence may be difficult to measure without very intensive follow-up. Thus, it is often more practical to study the 'prevalence' of disease at a particular point in time. This approach has one major potential shortcoming, since disease prevalence may differ between two groups because of differences in age-specific disease incidence, disease duration or other population parameters; thus, it is much more difficult to assess causation (i.e. whether an exposure increases disease incidence) in prevalence studies.

Nevertheless, for many common diseases, studying prevalence is often the only practical option and may be an important first step in the research process; furthermore, prevalence may be of interest in itself, e.g. because it measures the population burden of disease. For example, motor neurone disease and multiple sclerosis have similar incidence and mortality rates, but multiple sclerosis represents a greater burden of morbidity for the health services,

because survival for motor neurone disease is so short.

Prevalence case - control studies

Just as an incidence case - control study can be used to obtain the same findings as a full cohort study, a prevalence case - control study can be used to obtain the same findings as a full prevalence study in a more efficient manner. In particular, if obtaining exposure information is difficult or costly, then it may be more efficient to conduct a prevalence case - control study by obtaining exposure information on some or all of the prevalent cases and a sample of controls selected from the non-cases⁶.

Continuous outcome measures

Cross-sectional studies

In the presentation of prevalence studies above, the health outcome under study was a 'state' (e.g. having or not having hypertension). Studies could involve observing the incidence of the 'event' of acquiring the disease state (e.g. the incidence of being diagnosed with hypertension), or the prevalence of the disease state (e.g. the prevalence of hypertension). More generally, the health state under study may have multiple categories (e.g. non-hypertensive, mild hypertension, moderate hypertension and severe hypertension) or may be represented by a continuous measurement (e.g. blood pressure). Since these measurements are taken at a particular point in time, such studies are often referred to as 'cross-sectional studies'. Prevalence studies are a subgroup of cross-sectional studies in which the disease outcome is dichotomous⁶.

Longitudinal studies

Longitudinal studies (cohort studies) involve repeated observation of study participants over time. They represent the most comprehensive approach since they use all of the available information on the source population over the risk period. Incidence studies are a subgroup of longitudinal study in which the outcome measure is dichotomous. More generally, longitudinal studies may involve repeated assessment of categorical or continuous outcome measures over time (e.g. a series of linked cross-sectional studies in the same population). A simple longitudinal study may involve comparing the

disease outcome measure or more usually changes in the measure, over time, between exposed and non-exposed groups. For example, rather than comparing the incidence of hypertension (as in an incidence study) or the prevalence at a particular time (as in a prevalence study), or the mean blood pressure at a particular point in time (as in a cross-sectional study), a longitudinal study might involve measuring baseline blood pressure in exposed and non-exposed persons and then comparing changes in blood pressure (i.e. the change from the baseline measure) over time in the two groups. One special type of longitudinal study is that of 'time series' comparisons in which variations in exposure levels and symptom levels are assessed over time with each individual serving as their own comparison⁶.

DISCUSSION

There is no definitive approach in classifying types of epidemiological studies, and different classification schemes may be useful for different purposes. Each type of design represents a different

way of harvesting the necessary information. The selection of one design over another depends on the research question and takes into account validity, efficiency, and ethical concerns. The figures (Figure No.1 and 2) presented here involve 'ideal types' that are followed in Practice. The modern epidemiological classification presented here is not familiar, which are not followed in practice frequently. Thus, undoubtedly some readers will find the scheme presented here simplistic. Nonetheless, this 4-fold classification of study types has several advantages over other classification schemes. First, it captures the important distinction between incidence and prevalence studies; in doing so it clarifies the distinctive feature of cross-sectional (prevalence) studies, namely that they involve prevalence data rather than incidence data. Secondly, it captures the important distinction between studies that involve collecting data on all members of a population and studies that involve sampling on outcome (this is the widely accepted distinction between cohort and case - control studies).

Table No.1: Advantages and Disadvantages of four types of observational studies

| S.No | Type of study | Definition | Advantages | Disadvantages |
|------|---|---|--|--|
| 1 | Cross-sectional | Examines relationship between exposure and outcome prevalence in a defined population at a single point in time | Less time-consuming than case-control or cohort studies. Inexpensive, Good, quick picture of prevalence of exposure and prevalence of outcome | Difficult to determine temporal relationship between exposure and outcome (lacks time element) May have excess prevalence from long duration cases (such as cases that last longer than usual but may not be serious) |
| 2 | Case-control | Examines multiple exposures in relation to an outcome; subjects are defined as cases and controls, and exposure histories are compared | Relatively inexpensive Less time-consuming than cohort studies, Can evaluate effects of multiple exposures, Efficient for rare outcomes or outcomes with long induction or latency periods | Subject to recall bias (based on subjects' memory and reports) Inefficient for rare exposures (such as Difficult to establish clear chronology of exposure and outcome |
| 3 | Cohort (specifically prospective) | Examines multiple health effects of an exposure; subjects are defined according to their exposure levels and followed over time for outcome occurrence | Can evaluate multiple effects of a single exposure. More efficient for rare exposures and outcomes with long induction and latency periods Can directly measure incidence Clear chronological relationship between exposure and outcome | <ul style="list-style-type: none"> • Expensive • Time-consuming Inefficient for rare outcomes with long induction or latency periods |
| 4 | Ecological | Examines relationship between exposure and outcome with population-level rather than individual-level data (usually defines groups by place, time, or both) | Inexpensive Less time-consuming Simple and easy to understand Examines community-, group-, or national-level data and trends | Subject to the <i>ecological fallacy</i> , which infers association at the population level whereas one may not exist at the individual level Difficult to detect complicated exposure-outcome relationships |

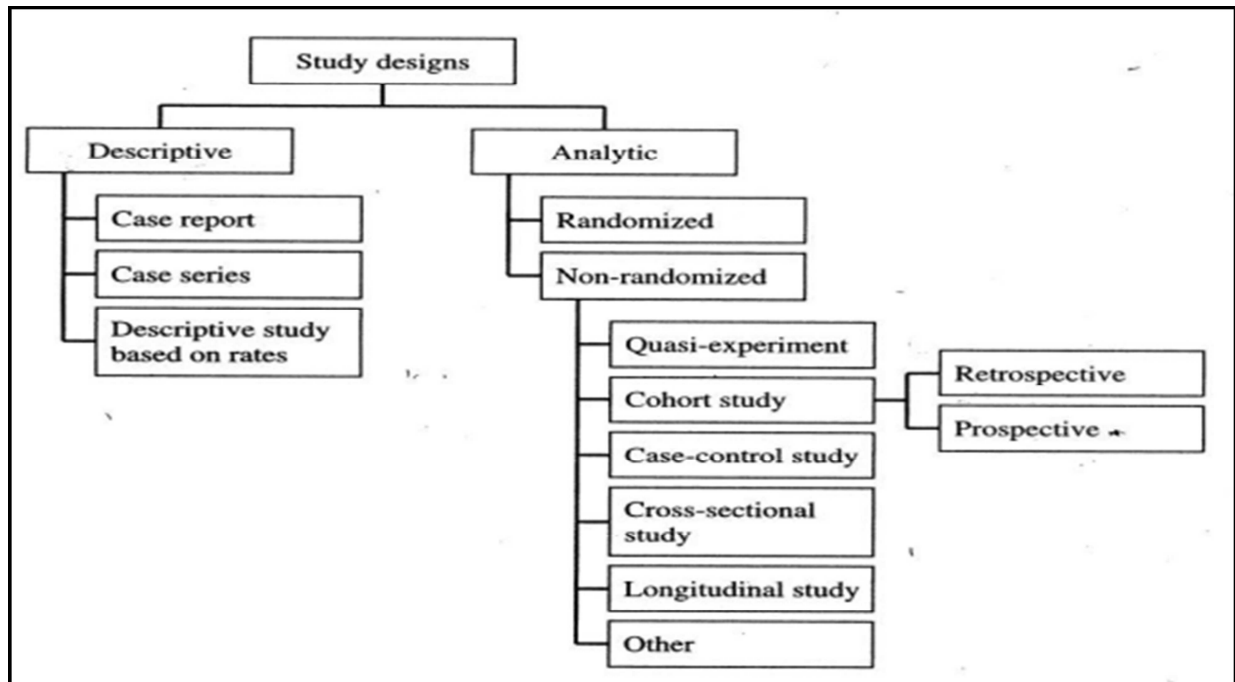


Figure No.1: Major Epidemiologic Study Design

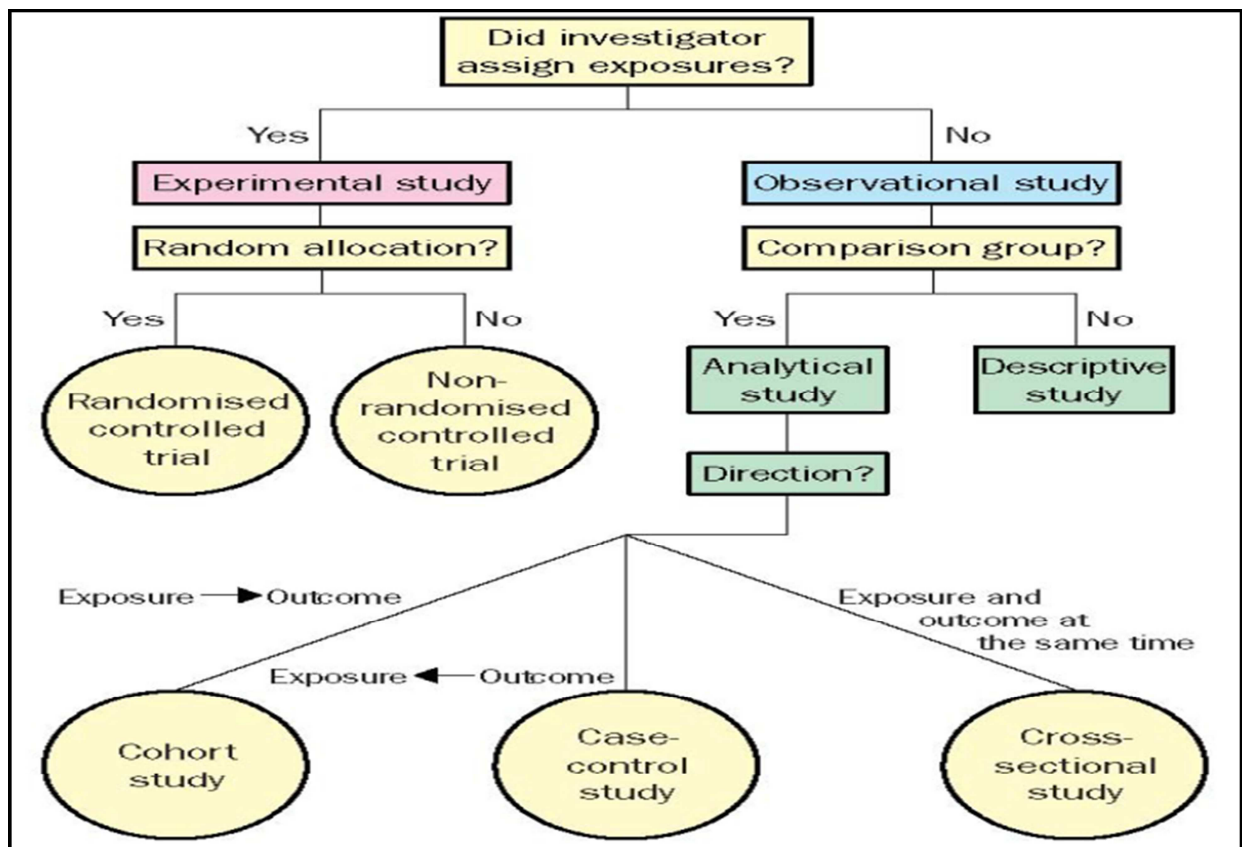


Figure No.2: Algorithm for classification of types of clinical research

CONCLUSION

The article gives an overview of traditional classification of epidemiological study designs as well as study designs based on modern epidemiology.

ACKNOWLEDGEMENT

All authors are would like to thanks Bharathi College of Pharmacy, Bharathinagara, Mandya, Karnataka, India for continuous support and encouragement throughout this review work.

CONFLICT OF INTEREST

None declared.

BIBLIOGRAPHY

1. Mac Mahon B, Trichopoulos D. Epidemiology Principles and Methods, Boston, MA: Little, Brown and Company, 2nd edition, 1996.
2. Last J M. A Dictionary of Epidemiology, New York, NY: Oxford University Press, 3rd edition, 1995.
3. Stuart reid, Ian robertson. Epidemiology Workshop for Equine Research Workers, University of Sydney, 1998.
4. Alan J. Silman, Gary J. Macfarlane. Epidemiological studies-a practical guide, Cambridge university press, 2nd edition, 2002, 31 and 41.
5. Frankena K and Thrusfield M V. Basics of observational studies. In Application of Quantitative Methods in Veterinary Epidemiology Noordhuizen JPTM, Frankena K, van der Hoofd C M and Graat E A M Wageningen Pers, Wageningen, 1997.
6. Neil pearce. Classification of epidemiological study designs, *Int J Epidemiol*, 41, 2012, 393-397.
7. Greenland S, Morgenstern H. Classification schemes for epidemiologic research designs, *J Clin Epidemiol*, 41, 1988, 715-16.
8. Morgenstern H, Thomas D. Principles of study design in environmental epidemiology, *Environ Health Perspect*, 101, 1993, 23-38.
9. Miettinen O S. Theoretical Epidemiology, New York: John Wiley and Sons, Inc., 1985.
10. Rothman K J. Modern Epidemiology, Boston: Little, Brown, 1986.
11. Rothman K J, Greenland S, Lash T L. Modern Epidemiology, 3rd edition, Philadelphia: Lippincott Williams and Wilkins, 2008.
12. Pearce N. Effect measures in prevalence studies, *Environ Health Perspect*, 112, 2004, 1047-50.

Please cite this article in press as: V.J. Divya et al. Review on Epidemiological Study Designs, *Asian Journal of Research in Biological and Pharmaceutical Sciences*, 3(2), 2015, 52-58.