

Stat Corner

Biostatistics - I: Introduction, role & applications in medicine

Tariq Zaman, Abbas Raza*

Dermatology department, Allama Iqbal Medical College, Lahore.

* Department of Medicine, King Edward Medical College, Lahore.

Understanding the fundamentals of biostatistics is essential for every clinician in order to plan a scientific study and interpret its results. Unfortunately, this area had been neglected at the level of undergraduate and postgraduate medical education in country. The terms like X^2 test, student t test, p value, sensitivity, specificity etc. are very scary for clinicians and while going through a medical journal, the pages describing the statistical methods are generally skipped. As a part of CME activity for its readers, *JPAD* has started a series of articles covering this dry and boring topic. The authors have simplified the subject in a reader-friendly manner so that it is easy to comprehend and assimilate. (*Editor*).

“The fundamental gospel of statistics is to push back the domain of ignorance, prejudice, rule-of-thumb, arbitrary or premature decisions, tradition, and dogmatism and to increase the domain in which decisions are made and principles are formulated on the basis of analyzed quantitative facts.”....Robert W. Burgess

Statistics is deeply engraved in our daily life e.g. the chances of being successful in a particular assignment or examination, and at a wider scale, success of a particular plan (e.g. healthcare) in a community or winning a war against enemy. All these issues are generally judged qualitatively. Calculating the risks, benefits and probabilities involved in mathematical terms is statistics.

Present series of articles, which will regularly appear in successive issues of the journal, is written to make reader more effective in interpretation of research literature in medicine and be able to design research study and draw appropriate

statistical conclusions. This article is intended to give an introduction to biostatistics, its evolution through time and its role and applications in health sciences.

Historical perspective

From dawn of civilization to about AD 1000, medical treatment was directed against supposed etiologies, e.g, evil spirits, draining bad blood, etc. The first ever mention of statistical probabilities of an outcome is found in Talmud (AD 300). Hindu culture takes the credit of using numerical and decimal system, the foundation of statistics. Arabic mathematician, Khwarizmi, defined rules for adding, subtracting, multiplying and dividing (AD 800). Huygen and John Grant (1660) advanced this knowledge further.

Twentieth century saw an explosion of medical technology and treatment in the fields of surgery, medicine and their subspecialties. It became increasingly necessary to standardize medical practice

Address for Correspondence

Dr. Tariq Zaman,
210 G.T.Road, Baghban pura, Lahore.

on sound scientific basis, and interpret the controversies in the treatment of similar conditions from growing body of evidence. In the mid-twentieth century, “Case Study” and “Case Series” were a common way to prove that a treatment was beneficial. Most research work done was anecdotal than scientific. James Lind, a Royal Navy surgeon, carried out first recorded randomized clinical trial in 1747, applying six different treatments for scurvy cure in sailors. It took 200 years before randomized controlled trials became the standard for evaluation of various therapeutic options.

Sir Ronald Fisher and Austin Bradford are remembered as fathers of modern statistics. They published a series of articles in “Lancet” and “British Medical Journal” (1937-1947), reporting their randomized trials e.g. streptomycin vs. placebo in lung tuberculosis, and stressing upon introduction of biostatistics in medical curricula. Archie Chocrane, a British general practitioner started current movement of systematic statistical reviews of available evidence in medicine, thus laying the foundation of Evidence Based Medicine and Practice.

Statistics and biostatistics

The word “**statistics**” has originated from Latin words *status* (meaning ‘state’) or *statisticus* (meaning ‘state affairs’) and has been used for the information useful for the state, e.g. information about size of population or armed forces, about amount of revenue, reserves or salaries, etc. etc. But now the term ‘statistics’ is used in two different contexts.

Firstly, it is used as a synonym of *data* and refers to systematically arranged numerical facts and figures of any kind,

e.g. statistics regarding economics, business, industry, education, health, diseases, etc. They are collected from records, surveys or experiments.

Secondly, the term statistics is used to refer *a body of knowledge, a discipline or a branch of science*. It is a science of figures. It deals with the principles and methods for collection, processing, analysis, presentation and interpretation of numerical data. It is an inferential science too, and helps us to draw reliable and valid inferences from the available data particularly in the fields of experimentation and research. It is a science of decision-making, and is the only way out to take decisions in the face of uncertainty.

The term “**biostatistics**” is used for the statistics related to the biological sciences like biology, medicine or public health. So biostatistics refers either to the *data* arising out of the biological sciences or to the *science* that deals with the application of statistical methods to the data derived from these fields.

Statistical methods fall into two broad categories, descriptive statistics and inferential statistics.

Descriptive Statistics

This branch of statistics deals with the summarization and description of data. It is the most basic form of statistics that lays the foundation for all statistical knowledge. The methods of descriptive statistics are used to consolidate the data in the form of tables, charts or graphs. They are also applied to compute the numerical quantities that provide information about the central tendency and spread of the data, e.g. mean, median, mode, standard deviation, variance, etc.

Inferential Statistics

It is the branch of statistics that deals with making inferences from the collected data. The methods of inferential statistics are applied to test statistical hypothesis and draw conclusions by computing confidence interval or statistical significance. They are used to estimate population parameters using data obtained from a sample from that population.

Role of statistics in medicine

Statistical analysis is a fundamental component of modern research in all disciplines of science, and medicine is no exception. In medical research the role of statistics starts from the planning phase in deciding study design, selecting sampling technique and formulating research methodology. The statistical tools have their application in data collection and its handling, analysis and interpretation. Statistical treatment of data is required to draw valid conclusions, which have to be evaluated in the light of statistical logic. Statistical methods are also helpful for comprehensible and effective presentation of research data and results. In addition, clear concepts of basic principles of biostatistics are also required to understand the data presented by others.

“A knowledge of statistical methods is not only essential for those who present statistical arguments it is also needed by those on the receiving end.”...R.G.D. Allen

Various statistical tools are extensively used in medical journals for description of data and for drawing conclusions from research findings. Knowledge of statistical methods is necessary to assess the validity of evidence in these articles. Understanding of principles and concepts of biostatistics is an essential prerequisite

to develop critical attitude towards research data and inferences presented at any forum or published in any scientific journal, newspaper or magazine. The clinicians also require some working knowledge of basic statistical techniques while evaluating clinical features, diagnosis, prognosis and management of their patients. In fact, neither medical practice nor research can be properly planned, executed or assessed without understanding the fundamentals of biostatistics.

Application and uses of statistics in health sciences

“A knowledge of statistics is like a knowledge of foreign languages or of algebra; it may prove of use at any time under any circumstances.”...Arthur L. Bowley

The biostatistics has its application and uses in the following broad areas of basic medical sciences, applied medicine and public health:

- In defining the *normal* of various body characteristic and their *limits* (range of normality) in a population, i.e. at what point these measures become ‘abnormal’ or ‘pathological’. For example, normal value and range of blood pressure, heart rate, weight, height, blood counts, serum biochemistry, etc.
- In finding the *difference between means and proportions* of normal variables (body characteristics) in two samples with different parameters or in the same sample at two places or in two different periods, e.g. difference between means of weight or height in various age, sex, racial or ethnic

groups, or difference between heart rate or blood pressure at basal level and after exercise in the same sample.

- In finding the *correlation between two variables*. Whether one variable increases or decreases proportionately with the other variable, and if so by how much. For example, correlation between age and weight, or between height and weight, etc.
- In determining the action, dosage and potency of a new drug in *pharmacological studies*. By statistical techniques the action of a drug is compared in animals or human volunteers with a placebo to determine whether it is due to the drug or by chance. Or the action of the new drug is compared with a standard drug or with various doses.
- In identifying and documenting the *clinical features of diseases*, e.g. signs, symptoms, course, prognosis, complications, etc. and their variability among various groups of patients or their association with various characteristics like age, sex, etc.
- In analyzing the *sensitivity and specificity* of various laboratory tests, clinical procedures or diagnostic criteria.
- In planning and conduct of *clinical trials* in a particular disease for efficacy and safety of a drug, surgical procedure, radiation or physiotherapy, etc.
- In evaluating the role of basic underlying factors in the causation of different ailments in the *epidemiological studies*. For example, determining the role of

contaminated water in diarrhea or dysentery by comparing the proportions of disease attacks in groups of individuals drinking boiled or unboiled water. Similarly, evaluation of the role of smoking in lung cancer or high cholesterol diet in ischemic heart disease, etc.

- In determining the usefulness of various *public health programs*. For example, evaluation of vaccination for a particular infection by comparing the proportions of attacks/deaths in vaccinated and unvaccinated subjects and by analyzing the statistical significance of the difference observed.
- In collecting the *demographic, health and vital statistics*, which are important indicators of health status of the community. They have a pivotal role in formulating health policy and in making future health plans.

In addition, application of statistical principles is necessary in many other areas of medicine and public health, like construction of life tables and survival rate calculation, risk specification, cost analysis, quality of life indices, etc.

Latest trends in statistics

Preceding decade has seen following developments in the clinical application of biostatistics in health care sciences:

Evidence Based Medicine (EBM)

Evidence Based Medicine has been defined as the “the conscientious, explicit and judicious use of best available evidence in making decisions about cases of individual patients”. It was realized in 1980s, that there were large variations in

the treatment strategies offered to patients with similar ailments. For example, rates of prostatectomy (for benign hyperplasia), hysterectomy (for menorrhagia) and cataract surgery varied 2.5, 3.0 and 20 times respectively. Evidence based medicine started with the appreciation of the need for uniformity of treatment for same conditions. This uniformity is based upon statistical analysis of randomized controlled trials on the same subject.

Meta-analysis

Meta-analysis forms the basis of EBM. It involves inclusion of multiple studies according to preset inclusion and exclusion criteria in which the results from all the studies are pooled and analyzed statistically as if it were from one large study. The meta-analysis, also called as “Systematic Literature Review” explains the results quantitatively. In contrast the traditional “Literature Review” examines the relevant literature for general trends and patterns in a particular subject in qualitative terms only.

Guidelines

EBM clarifies one particular facet of disease management (e.g. thrombolysis in acute myocardial infarction). Comprehensive disease management requires treatment according to the best available evidence at each step till recovery. A disease may also not follow the typical pattern. Guidelines are formulated by a panel of experts of the particular disease. Literature is searched for each step of disease management, strength of evidence is judged, and a consensus is reached upon before publication of guidelines.

Role of computer in statistics

Mathematical calculations used in biostatistics have always been feared by

the non-statisticians including health professionals. Statistical analysis of any study is always considered to be the job of statistician. Computers and commercial statistical software packages have not only allayed the fear of mathematical calculations but also offer economy of time. Research worker can now apply the appropriate statistical methods to his research work himself with the help of these softwares. However, a sound knowledge of basic principles of biostatistics still remains mandatory to use these packages.

Conclusions

The medical students, clinicians or research workers should not depend solely on the statistician for statistical analysis. They should learn the basic principles of biostatistics and should apply these methods correctly for unprejudiced assessment and management of the patients, for proper planning of scientific studies, for reinforcement of their own research conclusions and for critical evaluation of the work done by others.

Further reading

1. Bluman AG, ed. *Elementary Statistics – step by step approach*, 5th edn. Philadelphia: McGraw-Hill; 2004.
2. Coogan D, ed. *Statistics in clinical medicine*, 2nd edn. London: BMJ Publications; 1999.
3. Gaddis G, Gaddis M. Introduction to biostatistics; part II, the descriptive statistics. *Ann Emerg Med* 1990; **19**: 309-15.
4. Krishnamurty GB, Kasovia-Schmitt P, Ostroff DJ, eds. *Statistics – An interactive text for health and life sciences*. London: Jones & Bartlett Publishers International; 1995.
5. Mann PS, editor. *Introductory Statistics*, 5th edn. Singapore: John Willy & Sons; 2004.