Annals of Clinical and Medical Case Reports

Mini Review

ISSN 2639-8109 | Volume 8

Bio-Statistics Newer Advances, Scope & Challenges in Bio-Medical Research

Murthy NS[1] , Kunnavil R[2] and Shivaraj NS[3]

[1]Former Research Director, DRP and Professor and Research Coordinator, M.S. Ramaiah Medical College & Hospitals (MSRMC), Bangalore, India.

[2]Lecturer in Biostatistics, National Institute of Unani Medicine, Bangalore, India

[3]Assistant professor Statistics, Department of Community Medicine, M.S. Ramaiah Medical College & Hospitals (MSRMC), Bangalore, India

[*]Corresponding author:

Radhika Kunnavil.
Lecturer in Biostatistics, National Institute of Unani Medicine, Bangalore- 560091, India,
Tel: 7829606040,
E-mail: radhik121@gmail.com

Received: 25 Feb 2022
Accepted: 09 Mar 2022
Published: 14 Mar 2022
J Short Name: ACMCR

Citation:

Kunnavil R, Bio-Statistics Newer Advances, Scope & Challenges in Bio-Medical Research. Ann Clin Med Case Rep. 2022; V8(14): 1-7

Keywords:

Biostatistics; Bio-medical research; Scope; Applications; challenges

1. Abstract

Biostatistics also known as biometry which means 'measurement of life' is a branch of applied statistics which deals with collection, compilation, analysis and interpretation of data related to biomedical sciences. It provides a key to better understanding of the medical discipline. Biological data are always subjected to variation and are affected by various environmental, social and genetic factors etc. Biostatistics proves a tool for analysing the data taking into account the variability and to elicit meaningful conclusion in research. In this era of evidence-based medicine. Nowadays, the discipline of biostatistics acts as a basic scientific entity of public health, health services and biomedical research. Biostatistics has evolved as a branch of science in biomedical research that uses a combination of statistics, probability, mathematics and computing to resolve perplexing questions mathematically. Because research questions in biology and medicine are diverse, biostatistics has expanded its domain to include any complex quantitative model to answer research questions. Biostatistics has reach in wide spectrum of biomedical domains. From drug development in a laboratory to development of various disease control and prevention measures at community level and a lot more. The important aspect for biostatistics is that it helps in valid clinical decision making by elucidating and quantifying the contribution of chance. Overtime this field has expanded its wings and now it is widely used across many systems of medicine including Ayurveda, Unani etc. The computer programmes and advent of newer methods has increased its scope and applications. This paper aims to through light on the newer advances in biostatistics. This paper explores the newer methods, scope & few challenges in Bio-Medical Research.

2. Introduction

Bio-statistics has pivotal role in the development of medical and biological sciences as well as in the development of various disease control and prevention measures. Nowadays, the discipline of biostatistics acts as a basic scientific entity of public health, health services and biomedical research. Over the last few decades, bio-statistics have become more quantitative, stochastic, evidenced based with the growth of medical sciences and public health-oriented research. Emerging disciplines such as Machine learning, Clinical Epidemiology, Molecular Biology, Genomics and Pharmacokinetics have all contributed to making medical and health sciences depend more and more on Biostatistics. The present write-up focusses on role and use of bio-statistics in the epidemiology, bio-medical research as well as some to touch upon newer methods, scope & few challenges in Bio-Medical Research.

3. Bio-Statistics as a Stream

Over the last few decades, the development of statistical methods has expanded. Statistics applied to medical research – biostatistics – may now be regarded a subject in its own right, research in medicine and public health has been both a benefactor and a source of new difficulties as a result of this new technique. In reality, biostatistics has evolved into a distinct field of study that solves issues in the biological sciences by combining statistics, probability, mathematics, and computing. Biostatistics has broadened its area to encompass any quantitative, not just statistical, model that may be used to answer these issues, due to the diversity of research questions in biology and medicine. Biostatistics is a field that aims to provide information. Consequently, biostatistics draws quantitative methods from fields including statistics, operations-research, economics, and mathematics in general; and it is applied to research questions in fields such as epidemiology, nutrition, environmental health, and health services research, genomics and population genetics, clinical medicine, and ecology.

The significance of biostatistics and biostatisticians in medical research has long been acknowledged by the biomedical community, and statistics in medicine may now be regarded a successful paradigm for the incorporation of statistics into scientific practice. The relevance of biostatisticians in the biomedical profession may be seen in the fact that they are frequently asked to contribute as advisers on renowned committees and journals. Furthermore, specific statistical publications such as Biostatistics, Biometrics, Biometrika, and many other biostatistics-related journals are held in high esteem [1-2].

4. Types of Research Investigations in Bio-Medical Field

Quality biomedical research is based on a foundation of careful study design. Over the last several decades, newer and innovative concepts and statistical methods for the design and analysis of data in biological studies have been established and are being used. Design of studies such as case-control studies, cohort studies, clinical trials, and survival studies has been the center of development. The application of epidemiologic ideas and techniques to the design, conduct, and analysis of clinical trials is a major development, with comprehensive applications described in the following paragraphs. Observational and experimental research studies are the two categories of scientific research studies in biomedical field. Selection of subjects on whom measurements are made is one of the most essential problems that occur during the formulation of statistical methods of research [3-5].

5. Why is Statistics Necessary in Bio-Medical Field?

Without appropriate inferences, empirical research in any field is incomplete, and biomedical research is no exception. Both the design of diverse biomedical research investigations and the evaluation of outcomes need the use of proper statistical tools. Biostatistics has now become a crucial component in several research domains as a result of expansion of quantitative approaches with in biomedical sciences (bio-chemical, physiological, clinical parameters, or evidence-based medicine). Medicine is a science in which chance plays an important role. Statistics as a science aid in quantifying the role of chance, whereas statistics as an art aids individual clinicians in making accurate diagnostic, prognostic, and therapeutic judgments. It also aids health programme administrators and policymakers in the planning, monitoring, and evaluation of public health efforts. A health indicator can be used to describe one or more aspects of an individual's or population's wellbeing (quality, quantity, and time), and also to define public health problems at a specific moment in time, to indicate changes in levels of a population's or individual's health over time, to define distinctions in population health, and to assess the extent to which a program's objectives are being met. Similarly, validity metrics including sensitivity, specificity, positive and negative predictive value are used to evaluate the quality and usefulness of a diagnostic test or to determine the efficiency of a marker in disease diagnosis [3].

5.1. Epidemiology and Biostatistics

Epidemiology is the branch that studies diseases occurrence and its reasons in different groups of people. Epidemiological data is used to design and evaluate disease prevention initiatives, as well as to guide the treatment of patients who already developed disease. Biostatistics and epidemiology have historically had such a significant link. The early public health experts were doctors basically keen to understand the path wherein ailments eventuate in populations, their causes, as well as their interrelations with various medical and non-medical aspects. These innovators' challenges included not just the study of epidemics and non-communicable diseases such as the connection between smoking and lung cancer, but also the evaluation of therapies. Many had strong analytical reasoning abilities and were well-versed in statistical methods. Then, beginning in the 1930s, epidemiology began focusing upon that study of chronic diseases. The same prospective research strategies that had been so clearly appropriate in the study of infectious diseases became untenable. And it was statisticians, particularly Cornfield and Mantel, who provided a rationale for clarify case-control inference. With concerns about bias related to possible confounding factors, biostatisticians have become more interested in growing on the prerequisites for valid inference. They also began searching at other areas of epidemiological research, including models for evaluating the effects of potential disease risk factors, such as dose response models. These effects are quantified employing probabilistic notions such as the odds ratio or relative risk, which can be estimated appropriately based on the type of study (case-control, cross sectional, or cohort) used for each research project. The large number of statistical methods required in epidemiology has led to the publication of numerous books on statistical applications in epidemiological contexts [6-9].

5.2. Clinical trials and biostatistics

Clinical studies are a crucial component of medical research. Scientific advance can lead to better ways of diagnosing, detecting, and treating diseases and medical conditions as a consequence of these clinical studies. Clinical trials are research studies which use human subjects to evaluate novel therapies or drug combinations, modern surgical or radiotherapy approaches, or new procedures in order to improve illness diagnosis or quality of life for patients. Most hospitals now participate in drug testing, which are only started once laboratory investigations show that a new treatment or technique is safe and also has the potential to be more efficient than existing options. Statistics have become increasingly important in the field of pharmaceutical development in recent years. From planning through conduct and interim analysis to final analysis and reporting, statistics is essential at each and every stage of a clinical trial [10-11]. The statistician is typically in responsible for formulating randomization schedules, advising on sample size, establishing framework for deciding treatment differences, and evaluating response rates. In most instances, the statisticians will also act as a liaison with the Independent Data Monitoring Committee. Several novel and recurring challenges in the drug development process require special attention. Ongoing development of statistical methods for handling subgroups in the design and analysis of clinical trials; alternatives to "intention-to-treat" analysis in the presence of noncompliance in randomized clinical trials; methodologies to address the multiplicities resulting from a variety of sources, methods to assure data integrity etc all of which are inherent in the drug development process. These concerns continue to be a source of contention for statisticians working in the pharmaceutical industry across the world. Furthermore, the engagement of statisticians from all backgrounds continues to enrich the profession and contribute to social health improvements. Biostatisticians' significant methodological contributions to clinical trials research has led to the development of a new journal, Pharmaceutical Statistics, which was just published in 2002 and is already placed in the JCR ranking for Statistics and Probability.

6. Advanced Statistical Areas of interest in Bio-Medical Field

In addition to routine descriptive and inferential statistics, generalised linear models, survival analysis, and Bayesian methods etc have already had a significant impact on the medical statistics in recent years (in diagnostic, epidemiological and clinical trials contexts). Regression analysis or linear discriminant analysis are statistical approaches used in the biomedical profession to predict a dependent variable using additional independent variables/ features or to divide persons into two or more classes of objects or events based on illness status. The approaches outlined above have aided in the reduction of dimensionality as well as the classifying of people as sick or non-diseased [7-13].

7. Modeling-Approach in Epidemiological Research/ Bio-Medical Filed - Generalized Linear Models

Modeling of health and disease process has been a complex phenomenon. Several models have been employed for the analysis and interpretation of data in the biological field. In the forgoing sections, it is proposed to describe three different types of general linear models which have been extensively employed as multivariate procedures in bio-medical field viz. (i) Age-period cohort models, (ii) Logistic regression model and (iii) Survival analysis.

7.1. Importance of time-related analysis

Cancer incidence/mortality rates from population-based registries (which gather data on all cancer cases in specified areas) give information on regional and temporal variation in cancer risk by personal variables including age, sex, and racial or ethnic groupings. "Time," the third element of an epidemiological description, provides information on geographic areas and serves as the foundation for determining how effective cancer-prevention methods are. Changes in cancer pattern over time are of critical importance in cancer control efforts. Age at risk, calendar year, and birth cohort impact are the most often evaluated time-related confounders. The trend analysis helps to understand the question such as how cancer risk has been changing, why and what is likely to happen in future. Cancer trend analysis is important information for the public health and health care planning. Trend analysis reveals information on the disease's aetiology and the significant variance in its prevalence across different geographic areas. Cancer incidence/mortality trends can also be used to forecast future cancer patterns, which can help shape future public health policy. The suitable tools for assessing trends in cancer incidence/mortality data include data modelling by age, birth cohort, and calendar time period. "Age-period-cohort (ACP) models" are the name for these models. These models are based on regression models with a Poisson distribution [14]. Using the data from the Indian Population Based Cancer Registries for the past twenty-five years, the aforesaid modelling approach was used to predict the changes in the incidence of common malignancies. Some malignancies, such as breast, ovarian, corpus, and uterus, were found to be rising at a rate of around 1-2 percent every year, according to the research, and the similar trend was seen in women of a younger age range [15-18].

7.2. Modelling of the data in case of binary outcome event

When the dependent variable (outcome variable) is binary in nature, such as whether an event occurs or not, taking values of unity or zero, the assumption required for fitting a multiple linear regression model of the type $Y= \alpha +\Sigma^{k}_{i=1}\beta X_{i}$ is violated because it is unreasonable to assume that error distribution is normal. Multiple logistic regression analysis (LR) is used as a multivariate approach to discover the independent predictors of the outcome variable instead of multiple linear regression analysis. The key difference between LR and multiple linear regression models is that instead of utilising the dependent variable as is, we utilise a model based on the dependent variable's logit transformation to meet the required assumptions. As a result, in the LR model, we forecast the proportion of subjects (P) who have a specific characteristic, or, alternatively, the probability of having characteristics for any combination of explanatory factors [8,19,20]. A dichotomous outcome variable is linked to a set of "k" known or suspected causes (regression variables), as well as probable confounding and effect modification variables, in this model. Covariates or explanatory variables are a collection of k regression variables or risk factors. The approach of maximum likelihood estimation is used to estimate the unknown parameters in the model, α, and β_i. In most cases, the likelihood inference is preceded by the fitting of a hierarchy of models, each of which contains the last variable. The likelihood ratio test or a test based on Wald statistics are used to test the hypothesis. This modelling approach was used to find independent risk variables linked to illness or a negative outcome event. The logistic regression approach was used in research to look at the impact of maternal and perinatal outcomes in different degrees of anaemia. Mild anaemia had the best maternal and perinatal outcomes, according to the research. Low birth weight kids, induction rates, surgical deliveries, and protracted labour have all been linked to severe anaemia [21].

7.3. Studies on survival analysis

Understanding the link between time and the occurrence of vital and health-related events requires the study of lifetime data. In the biomedical area, time-to-event data is regularly encountered for study. This type of study is known as "survival analysis." The time passed between a subject's enrollment into the research and the occurrence of an event that is related to treatment is the outcome variable in follow-up/survival studies. The outcome variable has been dubbed survival time, and the event of interest (the onset of a disease) has been dubbed failure. In oncology, for example, the focus is usually on the patient's chance of survival after a surgical procedure. Issues of censoring and truncation hamper the analysis of this sort of survival trial.

The analysis of lifetime data is important in understanding the relationship between time and occurrence of vital and health related events. Time-to-event data is frequently encountered for analysis in bio-medical field. Such analysis is called as "survival analysis". In follow-up/ survival studies, the outcome variable is the time elapsed between the entry of a subject into the study and the occurrence of an event is related to treatment. The event of interest (development of a disease, death) has been referred to failure and the outcome variable as the survival time. In oncology, for example, interest typically centers on the patient's time of survival following a surgical intervention. The analysis of this type of survival experiment is complicated by issues of censoring and truncation.

Censoring occurs when we do not fully observe the patient's survival, due to death unrelated to the cancer under study, or disappearance from the study for some reason. The other factor is truncation, which basically occurs when some patients can't be observed for some reasons related to the survival itself. A common example of this is in HIV/AIDS studies of the incubation period (i.e., time from infection to disease). The follow-up starts when the HIV virus is detected and the moment of infection is retrospectively ascertained. Several survival parametric models such as Exponential and Weibull distributions were introduced to model the survival experience/follow-up data analysis of homogeneous populations incorporating the censoring schemes. The distribution of survival times must be known to apply these models. However, when the distribution of survival is not known, the non-parametric method of Kaplan-Meier curve developed in 1959 has been a well-known estimator of the survival function, and it is extensively used in epidemiological and clinical research [20-24].

In order to take into account diversity of situations, which were encountered in practice, Cox in 1972 developed a modelling procedure termed as Cox-proportional hazards model under a very rigorous theoretical backup. The classical proportional hazards model of Cox (1972) is also widely used whenever the goal is to study how covariates affect survival. This model is an important tool in the follow-up/survival studies for modelling the effect of risk factors/prognostic factors when the outcome of interest occurs with time. In the model, the hazard for an individual is a part of the product of a common baseline hazard and a function of set of risk factors. By applying the above modelling procedure, the independent risk factors associated with the development of precancerous lesions of cancer of cervix was evaluated [25]. Similarly, in another study the treatment effectiveness for curing of a gastro intestinal bleeding was evaluated which employed an experimental design [26]. However, when the assumption of proportionality does not satisfy, then a classical approach for the analysis of data of this type is the time-dependent Cox regression model (TDCM). Advantages of Cox's regression model include its easy interpretability and its availability in the majority of statistical packages.

7.4. New issues in survival analysis

When survival is the ultimate result yet intermediate phases are discovered, a generalisation of the survival process occurs. In this case, a series of occurrences is witnessed, resulting in many observations per person. Intermediate stages might be based on categorical time-dependent factors like transplantation, clinical symptoms (e.g., bleeding episodes), or a complication during the disease (e.g., metastases), or biological markers (like CD4T-lymphocyte levels). Multi-state models (MSMs) were accessible in the 1990s, allowing for a better grasp of the disease process and a better comprehension of how the time dependent covariate impacts the illness's evolution. Compared to Cox's regression model, these contemporary models offer significant advantages. They provide a better understanding of the illness process by indicating the risk of moving from one condition to another (transition intensities), as well as a variety of additional data, such as the average time spent in each state and survival rates for each stage. Differences in the course of sickness among subjects in the population can also be explained by covariates on transition intensities. MSMs, in particular, can show how various variables effect different transitions, which is impossible to do with other models like the TDCM. In reality, the risk of mortality in individuals who have undergone different therapy is unlikely to be the same. Furthermore, prognostic markers linked to the risk of mortality may vary based on the treatment received, for example. There is currently a substantial amount of research accessible on the analysis of MSMs [27-31].

8. Some Recurrent and Emerging Issues in Biostatistics

In terms of both the continued improvement of traditional approaches and the introduction of new techniques to meet new issues, modern biostatistics faces a variety of obstacles. We next turn our focus to a number of emergent topics that biostatisticians should investigate further, including bioinformatics, spatial statistics, neural networks, and functional data analysis, as well as big data analysis.

8.1. Statistical methods in bioinformatics

A very rapidly emerging influence on biostatistics is the on-going revolution in molecular biology. Molecular biology is now evolving towards information science, and is energizing as a dynamic new discipline of computational biology, sometimes referred to as bio-informatics. Bio-informatics merges recent advances in molecular biology and genetics with advanced statistics and computer science. The goal is increased understanding of the complex web of interactions linking the individual components of a living cell to the integrated behaviour of the entire organism. The availability of large molecular databases and the decoding of the human genome may allow a scientist to plan an experiment and immediately obtain the relevant data from the available databases. This is an area in which statistical scientists can make very important contributions. Several biostatistics departments (mainly in the U.S.) have already been renamed as “Biostatistics and Bioinformatics” [32-33].

8.2. Spatial statistical methods in health studies

In numerous types of public health and epidemiological studies, the investigation of the geographical distribution of illness incidence and its link to possible risk factors plays an essential role. Geographic epidemiology is the overall term for this field, and there are four major statistical areas of interest: (a) Given "noisy" observed data on illness rates, disease mapping tries to construct a map of the genuine underlying geographical distribution of disease incidence.

(b) Ecological studies look for correlations between sickness incidence and potential risk factors in groups rather than individuals, with groups frequently defined by geographic location. Such studies are helpful in discovering the cause of sickness and may contribute in suggesting future research paths as well as prospective preventative strategies. (c) Disease clustering research focuses on finding geographical locations with a considerably higher risk of disease, or evaluating the evidence of heightened risk near potential sources of hazard. The exploration of control measures when the aetiology of observed clustering has been established, or the targeting of follow-up studies to determine explanations for observed clustering in disease incidence. (d) Environmental assessment and monitoring is concerned with determining the geographical distribution of health-related environmental elements and exposure to them in order to develop appropriate controls or take preventative action. Given the scope and relevance of the issues raised by spatial epidemiology, it's no surprise that there's been a lot of interest in this field in recent years [34-36].

8.3. Neural networks in medicine

Many researchers have been drawn to neural networks (NN) techniques in medicine, and these approaches have been used in a variety of biomedical applications, including diagnostic systems, biochemical analysis, image analysis, and drug discovery. Neural networks, which mimic the behaviour of human neuron networks, have the potential to be beneficial in a wide range of applications. NNs, unlike humans, are not influenced by factors like as weariness, working environment, or emotional state. NNs are frequently employed in diagnostic systems, for the diagnosis of cancer and heart issues, and for the analysis of many types of medical pictures (such as tumour detection in ultrasonograms, classification of chest x-rays, and tissue and vascular classification in magnetic resonance imaging). Many researchers are interested in neural networks (NN) approaches in medicine, and these NNs are being used experimentally to model the human cardiovascular system: diagnosis can be achieved by building a model of an individual's cardiovascular system and comparing it to real-time physiological measurements taken from that patient. NNs are also employed in the research and development of cancer and AIDS medications. In addition to classical and current statistical approaches, neural networks are increasingly being viewed as an extension to generic statistical methodology [37-39].

8.4. Functional data analysis and medicine

Because of technology advancements in recent years, many scientific domains including applied statistics are increasingly measuring and recording continuous (i.e., functional) data. Many current devices, for example, allow biomedical researchers to obtain functional data samples (mainly as curves, though also as images). Because functional data is displayed as a curve, the curve is a good starting point for functional data analysis. Functional data often involves a large number of repeated measurements per subject, and these measurements are typically captured at the same (generally similarly spaced) time intervals and with the same high sampling rate for all participants. The derivatives of these curves, as well as the positions and values of extremes, are occasionally of interest. In endocrinology, for example, investigations of hormone levels after various pharmacological dosages; or in neuroscience, for example, studies to estimate the firing rate of a population of neurons, where the unit of research is each individual neuron's firing curve. Another example is the study of growth curves in which many characteristics of growth, such as height and lung function, are studied [40]. The goals of functional data analysis are generally exploratory in nature, with the goal of representing and displaying data in order to highlight noteworthy qualities that may then be used as input for further research. Other goals might include estimating individual curves from noisy data, characterisation of homogeneity and patterns of variability across curves, and evaluation of the correlations between curve forms and variables. Despite substantial recent advances in functional data analysis, the statistical community has a huge problem in developing new tools to deal with functional data. Aside from the books already mentioned, the special issue on functional data, which is set to appear in the journal Computational Statistics and Data Analysis soon (2007), may be an excellent place to start thinking about new areas of statistical study and prospective applications.

9. New Statistical Methods which are Likely to Play a Key Role in Biomedical Research Over Coming Years

The following new statistical methods are likely to play a key role in biomedical research over coming years: (i) bootstrap (another computer-intensive methods); (ii) Bayesian methods); (iii) generalized additive models;(iv) classification and regression trees (CART); (v) models for longitudinal data (general estimating equations); and (vi) models for hierarchical data, (vii) big data analysis [41]. Modern health research involves increasingly sophisticated statistical tools and computerized systems for data management and analysis. During the past few years' tremendous amount of software has been made available to support statistical computing requirements for biomedical research. Bio-statisticians have to be extremely familiar with various statistical software packages such as STATA, SAS, SPSS,R etc.

The traditional component of biomedical courses will probably focus on areas of mathematical statistics including probability theory, inference, re-sampling methods (e.g. bootstrap), linear regression, analysis of variance, generalized linear models, survival analysis (including multi-state models), nonparametric methods, and data analysis. In addition, new methodologies like spatial statistics, neural networks, smoothing regression methods (such as generalized additive models) and operations research are strongly recommended. The decision technologies, tools and theories of operations research and management Sciences have long been applied to a wide range of issues and problems within health care.

10. Conclusion

Biostatistics is a fundamental scientific field in public health, health services, and biomedical research. With the rise of medical sciences and public health-oriented research over the last few decades, biostatistics has grown more quantitative, stochastic, and evidence-based. Emerging fields including machine learning, clinical epidemiology, molecular biology, genomics, and pharmacokinetics have all led to a growing reliance on biostatistics in medical and health sciences. Medicine is a science in which chance is a significant factor. Statistics as a science may help quantify the influence of chance, but statistics as an art can help individual doctors make appropriate diagnostic, prognostic, and therapeutic decisions. The use of Biostatistical methods to address issues in clinical trials, survival analysis, Data modelling using Generalized linear models, genetics, ecology and machine learning etc are gaining much popularity in the present era of in epidemiological biomedical research.

References

1. Altman DG and Goodman S. Transfer of technology from statistical journals to the biomedical literature: Past trends and future predictions. Journal of American Medical Association. 1994; 272(2): 129-132.
2. Suarez CC, Manteiga WG, Ciencia A. Pensamiento y Cultura CLXXXIII 725 mayo-junio. 2007; 353-361.
3. Hennekens CH, Buring IE. Design strategies in epidemiologic research. In: Mayrent SL (ed). Epidemiology in medicine. Boston: Little, Brown and Co. 1987; 16-20.
4. Mathew A, Murthy NS. A step towards quality medical research. National Medical J India. 1998; 11(6): 283-286.
5. Prentice RL. Statistical methods and challenges in Epidemiology and Biomedical Research, Handbook of statistics, Elsevier, 2008.
6. Rao NSN, Murthy NS. Applied statistical methods in Health Sciences, JAYPEE Brothers Medical Publishers (P) LTD, New-Delhi, 2008; 2nd ED, Pg no-1-246.
7. Armitage P, Statistical methods in medical research. Oxford: Blackwell Scientific Publications. 1983; 167-188.
8. Breslow NE, Day NE. Statistical methods in cancer research. The analysis of case-control studies. International Agency for Research on Cancer, Lyon, 1980.
9. Breslow NE, Day NE. Statistical methods in cancer research. The analysis of cohort studies. International Agency for Research on Cancer, Lyon, 1987.
10. Murthy NS, Mathew A, Sharma JB. Design of research studies in clinical Obstetrics and Gynaecology: Obtetrics & Gynaecology Today. 2001; 7(1): 379-386.
11. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II: Analysis and Examples. British Journal of Cancer. 1977; 35(1): 1-39.
12. Armitage P, Berry G. Statistical Methods in Medical Research (3rd edition). Blackwell. 1994.
13. Houwelingen (van) HC. The future of Biostatistics: Expecting the unexpected". Stat Med. 1997; 16(24): 2773-2784.
14. Clayton DG, Schifflers, E. Models for temporal variations in cancer rates. II: Age-Perid-cohort models.
15. Murthy NS, Burra U, Chaudhry K, Saxena S. Changing trends in incidence of breast cancer-Indian Scenario. Indian Journal of Cancer. 2009; 46(1): 73-74.

16. Murthy NS, Shalini S, Suman G, Pruthvish P, Mathew A. Changing trends in incidence of Ovarian Cancer-Indian Scenario. Asian Pac J Cancer Prev. 2009; 10(6): 1025-1030.

17. Murthy Shalini P, Nandakumar BS, Suman G, Pruthvish S, Mathew A. Estimation of trends in incidence of cancer of corpus uteri - Indian Scenario. European Journal of Cancer Prevention. 2010; 20, 1: 25-32.

18. Shalini CN, Murthy NS, Shalini S, Pruthvish S, Mathew A. Trends in Rectal Cancer Incidence - Indian Scenario. Asian Pacific J Cancer Prev. 2011; 12(8): 2077-2082.

19. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press 1982.

20. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 1958; 53: 457-481.

21. Malhotra M, Sharma JB, Batra S, Sharma S, Murthy NS, Arora R, et al. Maternall and perinatal outcome in varying degrees of anaemia. Inter J Gynae & Obstet. 2002; 79(2): 93-100.

22. Cox DR. Regression models and life tables. Journal of the Royal Statistical Society. 1972; B34: 187-220.

23. Kalbfleisch JD, Prentice RL. Statistical Analysis of Failure Time Data. New York: Wiley, 1980.

24. Mathew A, Pandey M, Murthy NS. Survival analysis, caveates and pitfalls. Eur J Surgical Oncology. 1999; 25: 321-329.

25. Muthry NS, Mathew A, Yoele BB, Happanen N. Cohort data analysis: Is logistic regression or Cox proportional hazard model or Poisson regression model? Ever green problems in Epidemiology (Ed) LeenaTenkanen, Publication 3, School of Public Health, University of Tampere, 1999.

26. Murthy NS, Sehgal A, Satyanarayana L, Das DK, Singh V., Gupta MM, et al. Risk factors related to biological behaviour of precancerous lesions of uterine cervix. Br J Cancer. 1990; 61(5): 732-766.

27. Sarin SK, Lamba GS, Kumar M, Mishra A, Murthy NS. A comparison of endoscopic ligation and propronolol for the primary prevetioin of varical bleeding. New Engl J Medicine. 1999; 340(13): 988-993.

28. Klien JP, Moeshberger ML. Survival analysis techniques for censored Truncated data, Springer Science Business India, 2006.

29. Moeshberger ML, Klien JP. Statistical methods for dependent competing risks, Lifetime Data Anal. 1995; 1(2): 195-204.

30. Klien JP, Moeshberger ML. Independent or dependent competing risks: Does it make difference. Tyler and Francis. 1987.

31. Andersen PK, Borgan O, Gill RD, Keiding N. Statistical Models Based on Counting Processes, Springer, New York, 1993.

32. Hougaard P. Analysis of Multivariate Survival Data, Springer, New York, 2000.

33. Vapnik V. The nature of Statistical learning Theory, Springer-verlag, 1995.

34. Ewens WJ, Grant G. Statistical Methods in Bioinformatics (Statistics for Biology and Health), Springer, New York, 2005.

35. Cressie N. Statistics for Spatial Data. John Wiley & Sons New York, 1993.

36. Spatial Prediction and Kriging, Wiley Series in probability and Statistics for Spatial Data Wiley Serie4s, in Probability and Statistics, 1993.

37. Ripley BD. Pattern Reconnition and Neural Network, Cambridge University Press, 1996.

38. Radhika Kl, George PS, Mathew BS, Mathew A. Comparison of artificial neural network with logistic regression as classification models for prediction of breast cancer patients' outcome. International conference in Epidemiology; 2012.

39. Ripley BD. Statistical aspects of neural networks. In Barndorf-Nielson OE, Jensen JL, Kendall WS (Eds). Networks and Chaos-Statistical and Probabilitic aspects, London, Chaman & Hill Press, 1993.

40. Ramsay J, Silverman B. Applied Functional data Analysis, 2nd Edition, Springer, New York.

41. Kunnavil R, Murthy NS. Healthcare Data Utilization for the Betterment of Mankind - An Overview of Big Data Concept in Healthcare. International Journal of Healthcare Education & Medical Informatics. 2018; 5(2): 14-17.