CME

Principles of Data Analysis and Interpretation under Clinical Studies: Prompting Points

*Sada Nand Dwivedi**

Once accurate data collection is completed under a clinical study, to generate quality evidence regarding informed decisions for clinical/public health practice, two important questions arise: how to analyse data appropriately?; and how to interpret the analytical results accurately? For this, one has to follow certain basic principles of data analysis and interpretation. Further, sometimes even analytical methods may be same regardless of involved research questions, hypotheses and objectives of the study, but interpretation is mainly guided by study design. The present write-up has briefly described these principles and their related aspects to make the readers aware about them while dealing with data analysis and interpretation under any clinical study.

Access this article online

Website: www.cijmr.com

DOI: 10.58999/cijmr.v5i02.463

Keywords: Basic Principles, Data Analysis, Analytical Methods, Interpretation of Analytical Results, Study Design, Quality Evidence, Time to Event Analysis, Hierarchical Models

Introduction

There are various steps under research methodology involved under a clinical study.[1,2] One has to sincerely follow involved principles at every step of research methodology to ensure accuracy of collected data, appropriate analysis and interpretation of derived analytical results leading to accurate clinical practice and public health implications. To generate quality evidence regarding informed decisions for clinical/ public health practice, once accurate data collection is completed under a clinical study, two basic questions arise as important steps under research methodology: how to appropriately analyse data; and how to accurately interpret the analytical results? To achieve this, one has to follow certain basic principles of data analysis and interpretation. Also, interpretation of analytical results is mainly guided by study designs[3] even if analytical methods remain same regardless of involved research questions, hypotheses and objectives of the study.[4] While dealing with data analysis and interpretation.[5-7] Under any clinical study, to ensure accuracy in data analysis as well as interpretation, the readers may go through the present write-up which has briefly described related basic principles and other aspects.

International Centre for Health Research, RD Gardi Medical College, Ujjain, Madhya Pradesh, India

Correspondence to: S. N. Dwivedi, International Centre for Health Research, RD Gardi Medical College, Ujjain, Madhya Pradesh, India, Email: dwivedi7@hotmail.com

Submitted: 25/06/2026 Revision: 27/06/2026

Accepted: 06/07/2026 Published: 15/08/2026

Basic Principles of Data Analysis

There are various points one has to be sure while carrying out the analysis. Some of the prompting points are listed below:

Research Question, Hypothesis and Objectives

To begin with one has to be sure that data was collected considering the same research question/ hypothesis/ objective[4] for which answers are expected after completion of its analysis. Otherwise, analysis of the data will remain exploratory with limited indicative observations. In other words, one needs to examine at a glance whether available data has required information to be able to answer the desired question.

Study Design

Awareness about details of used study designs[3] may be mandatory to ensure selection of appropriate analytical methods. For example, data available under matched and unmatched case-control design involve varying analytical methods (e.g., paired and un-paired t-test in case of quantitative variables; McNemar's chi-square test/Chi-squared test in case of categorical variables). Likewise, analytical methods for data collected under cohort study vary as per fixed follow-up time for all

How to cite this article: Dwivedi SN. Principles of Data Analysis and Interpretation under Clinical Studies: Prompting Points. Central India Journal of Medical Research. 2026;5(2):3-10.

study participants and varying follow-up time. Further, intuitively methods for multivariable and multivariate epidemiological modelling often vary as per study designs. However, sometimes analytical methods remain same regardless of study designs. For an instance, in case of a categorical outcome, logistic regression analysis may be used regardless of cross sectional study; unmatched case-control design; and cohort study involving fixed follow-up of all the participants. But, interpretation of the analytical results will be guided by study designs (explained later).

Exposure Variable under the Study

Among various variables in the data set, there is need to have clarity about a planned exposure variable to be mainly examined to be associated with considered outcome. For example, smoking might be highly associated with coronary heart disease (CHD); and breastfeeding may be highly associated with birth-spacing. Among various variables in these two data-sets, smoking and breastfeeding are known as exposure variables respectively. Intuitively, under an intervention study/randomized controlled trial (RCT), considered dose/drug/intervention are known as intervention/exposure variable. For example, while assessing efficacy of optimum radio-active iodine dose to achieve ablation at six months after first dose, among differentiated thyroid cancer patients, radio-active iodine dose will be exposure (i.e., intervention) variable under the study.

Any change in scale of measurement[8] of exposure/intervention variable may result into change in analytical results. To overcome this problem, one should objectively finalise categorization/scale of measurements of not only exposure/intervention variable, but of each and every variable considered in the data analysis. To be more specific, one should try to incorporate more appropriate form of variables in data analysis. For example, in case of smoking/breast-feeding, they need to be involved as time-varying covariates with time dependent effects instead of simply dichotomous, yes or no.

Outcome Variable under the Study

Like exposure variable described above, there is need to have clarity about a planned outcome variable to be mainly assessed to be associated with considered exposure/intervention variable. For example, smoking might be highly associated with coronary heart disease (CHD); and breastfeeding may be highly associated with birth-spacing. Under these examples, CHD and birth-spacing will be known as outcome variables in the related data-sets respectively. Further, under an intervention study/randomized controlled trial (RCT), while assessing efficacy of optimum radio-active iodine dose to achieve ablation at six months after first dose, among differentiated thyroid cancer patients, ablation at six months after first dose of radio-active iodine dose will be outcome under the study.

Like exposure variable, any change in scale of measurement of outcome variable may also result into change in analytical results. To overcome this problem, one should objectively finalise categorization/scale of measurement of outcome variable considered in the data analysis. To be more specific, one should try to incorporate more appropriate form of variables in data analysis. For example, in case of CHD/birth spacing, due to varying time of observations they need to be appropriately considered as time to event (i.e., experiencing CHD or next pregnancy) instead of simply dichotomous, yes or no. On the other hand, while assessing efficacy of optimum radio-active iodine dose to achieve ablation at six months after first dose, among differentiated thyroid cancer patients, considering ablation at six months after first dose of radio-active iodine dose as dichotomous outcome will be appropriate due to involved fixed follow-up time of six months for every patient.

Completeness of the Data

Some of the major issues to be considered to ensure completeness of the data set:

Study population

Understanding of study population facilitates to examine whether covered sample using appropriate sampling methodology[9] under a study is its representative. In other words, the collected data under a study need to have all major characteristics of study population. Otherwise, analytical results may not be generalisable even to its study population. For an instance, findings under hospital based studies may often be not generalisable at community level.

Target sample size

To carry out a conclusive study, regardless of study designs, one needs to cover at least a minimum required sample size decided statistically.[10] Even under pilot/preliminary study, one considers a target sample size following related rules of thumb. Hence, one has to be sure that the available data under a study could at least successfully cover the target sample size as planned in the study-protocol. If not, to begin with, one has to statistically rule out the absence of systematically biased

individuals. In other words, exploratory comparison of two groups (i.e., targeted and covered *vs*. targeted but not covered) in relation to available data on major characteristics may provide important clues regarding probable biases among covered sample size.

Used sampling methodology

Strictly speaking, statistical methods may not be used for analysis of non-randomly[9] collected data. Use of a non-random sampling methodology often distorts representativeness of sample to the study population. Hence, its awareness helps to take precautions while interpreting the analytical results.

Scrutiny of the data and data entry

To maintain quality and completeness of the collected data set, it is recommended to carry out scrutiny of the collected questionnaires about any missing problem and/ or inaccuracy/inconsistency. As such, data collection, scrutiny of the collected data and its data-entry need to run simultaneously. It paves the path of required corrections through generating the error lists (if and when necessary) and handing over to involved investigators.

Missing data

In spite of all efforts to collect data complete in all respects, often there are missing information on outcome variable, and/or exposure variable, and/ or other confounders/ effect modifiers. It might be because of various reasons like incomplete data collection, no response, and malfunction of instruments. In case of presence of this problem among negligible number of study participants (say, up to 5%), one can simply exclude them from the analysis. Otherwise, after exploring the missing patterns (i.e., missing completely at random; missing at random; missing not at random),[11] one can decide to fill missing values using appropriate imputation methods (e.g., conventional imputation methods; through advanced models) before data analysis. However, under clinical studies, efforts should be made to focus on accurate and complete data collection instead of relying on imputation of missing values.

Storage of data

One should always keep back-up files of data for further scrutiny without fail, starting from originally collected data to last version of the finalized data set after due changes in scales of measurements of the variables including outcome variable and exposure/ independent variables on various occasions. Such good practices facilitate in maintaining transparency while handling originally collected quality data and their optimal use in deriving important policy driven clues.

Understanding Data

For appropriate epidemiological analysis, understanding of data set using various steps is unavoidable:

Step-1

To begin with, once raw form is finalized for each and every variable in the data set including exposure/ intervention variable as well as outcome variable, one needs to examine distribution of each variable in the data set before making objective changes in its scale of measurement for further analysis. An independent variable meant for only adjustment (e.g., age) may be retained in its raw form for further analysis.

Step 2

Once objective changes in scales of measurements of various variables are finalized, tabular description of data/descriptive statistics (e.g., frequency; percentage; mean & standard deviation; median and quartile range) for each variable often helps further to decide whether variables in the considered forms are in order or they need further changes in scales of measurement to make them more relevant. Sometimes, even some of the variables may be dropped from further analysis.

Step-3

The association of the outcome variable with each of the independent variables (i.e., exposure/intervention variable as well as with other available variables) needs to be assessed after finalization of scale of measurement of each variable, including outcome variable, and independent variables (i.e., exposure/intervention variable and other available variables). All the relevant variables [e.g., statistically significant; statistically insignificant but clinically significant] need to be retained as a subset of variables for further analysis.

Step-4

As a next step, to identify possible collinearity among independent variables, association between pairs of independent variables needs to be examined. From each pair involving collinearity (i.e., 70% or higher correlation), only one of them needs to be retained in the subset of variables for further analysis. If both are clinically relevant, one more subset of variables may be created for further analysis. Intuitively, the number of subsets of variables for further analysis may vary according to the number of independent variables involving collinearity.

To overcome this problem, a possible alternative approach is to appropriately merge two or more highly correlated variables as one variable. Also, in case of two or more sets of variables, developed epidemiological models may be assessed further to identify an optimal model among them.

Step-5

Other than exposure/intervention variable, the remaining variables in the above-finalized sub-set(s) of independent variables may further be explored regarding their probable confounding and/or effect modification impact[12] on the relationship between exposure/intervention variable and outcome variable. A variable may be only a confounder; only an effect modifier; both a confounder as well as an effect modifier; and neither a confounder nor an effect modifier. Along with observed confounders, as reported elsewhere, additional variables need to be added considering identified effect modifier (s) in the subset of independent variables for analysis through multiplication of each effect modifier with exposure/intervention variable.

Points of Care before Final Data Analysis

To ensure accurate analysis, one has to consider the following points

Point-1

Diagrammatic/graphic description of data/descriptive statistics (e.g., frequency; percentage; mean & standard deviation; median and quartile range) of each variable along with its association with other variables especially outcome helps in understanding of related distributions which might help in further analysis. For example, normal distribution of quantitative variables is an important assumption to carry out certain analysis. Likewise, equality of variance between the groups is another required assumption for many of the comparisons.

Point 2

Considering the finalized form of the outcome variable along with a finalized particular subset of independent variables, including exposure/intervention variable, along with retaining the outcome variable, one has to also retain the original forms of those quantitative independent variables (e.g., age) which are to be considered for adjustments alone. For each of the independent categorical variables consisting of three and more categories, dummy variables have to be created. Otherwise, it may be inappropriately considered as a continuous variable by the statistical packages. For instance, under a study on the association between smoking and CHD, for physical activity consisting of four categories (e.g., no physical activity; mild; moderate and high level physical activity); there will be six dummy variables (i.e., combinations of two) namely high *vs*. moderate; high *vs*. mild; high *vs*. no; moderate *vs*. mild; moderate *vs*. no; mild *vs*. no physical activity.

Point 3

While coding the outcome variable, individuals experiencing an event (e.g., CHD) are coded as 1; and those without an event as zero. Likewise, while coding independent variables, including the exposure variable (e.g., smoking), those exposed are coded as 1, and unexposed as zero (i.e., reference category). In other words, the category of an independent variable with the lowest risk of experiencing the worst outcome has to be considered as the reference category. Otherwise, the analytical results will erroneously reveal a risk factor as a protective factor.

Point 4

To develop an epidemiological/regression model, there are broadly four regression approaches: all possible regression; step-up regression; step-down regression; and step-wise regression, which involve both step-up & step-down regressions. Among them, all possible regression approach provides an optimal model, but it consumes a huge computational time even after using computer-driven statistical programs. To overcome this problem, one preferably needs to use a step-wise regression approach, which is known to provide an optimal model at par with that by all possible regression approaches. Keeping in view probable variation in used nomenclatures for these regression approaches under various statistical programs, one needs to ensure that the regression approach used under a statistical package remains step-wise regression, not otherwise.

Point 5

While developing multivariable regression models, every statistical program has its own default levels of significance regarding inclusion and exclusion of a particular variable in the regression model. It is advisable to decide it more objectively because any change in it may result into due change in developed models. Because of this, strictly speaking, two developed models considering changing levels of significance on same data-set and/or two different data-sets may not be comparable.

Point 5

The researchers often refer to multivariable regression and multivariate regression as synonyms inaccurately. However, to be accurate, modeling of one outcome (e.g., CHD) considering a set of independent variables is known a multivariable regression model; whereas modeling of two or more outcomes simultaneously (e.g., CHD & blood-pressures) considering a set of independent variables is known as a multivariate regression model.

Point 7

Choice of regression method for multivariable regression analysis mainly relies on scale of measurement of the considered outcome. For example

- Quantitative/continuous outcome: Multiple linear regression analysis
- Categorical outcome (Yes/No): Logistic regression analysis
- Time to event (i.e., survival data): Hazard models (i.e., Time to event models)

Point 8

The statistical significance of the analytical results needs to be considered at the planned level of significance in the study protocol. One should not try to convert a non-significant result as significant through inappropriate increase in level of significance. Further, one needs to prefer calculating an interval estimate (i.e., confidence interval at the considered level of confidence), instead of a mere point estimate (i.e., mean; proportion; odds ratio; risk ratio; rate ratio). For example, in case of a planned level of significance of 5%, calculating a 95% [(i.e., 100-5)%] confidence interval (i.e., 95% CI) will be appropriate.

Point 9

For validation of the developed models, one can carry out either internal validation or external validation. If planned, internal validation may be carried out using the Bootstrap method. It is basically through selecting repeated samples of same size through resampling methods with replacement. Estimation of standard error using the sampling distribution of any statistic facilitates a precise confidence interval estimate of the related statistic. For external validation, one has to test the developed model on independent sample data.

Major Analytical Methods

Keeping in view of the objectives with which data were collected and type of the considered outcome, probable analytical methods may be briefly described as follows:

Quantitative Outcome Variable

To describe a quantitative variable, including outcome, mean along with standard deviation need to be reported in case of normal distribution; otherwise median along with the interquartile range becomes unavoidable. To compare the mean of a group with a given hypothetical value, one sample t-test is used in case of normal distribution, the non-parametric equivalent "Wilcoxon test," has to be used. For comparison of means between two independent groups, an unpaired t-test can be used in case of fulfilled assumptions related to normality and equality of variance; otherwise, the non-parametric equivalent Wilcoxon rank sum test/Mann-Whitney test needs to be used. However, to compare two paired groups (e.g., matched groups/pre-post intervention) paired t-test has to be used in case of fulfilled required assumptions; otherwise, a non-parametric equivalent Wilcoxon signed-rank test needs to be used. Likewise, three or more independent groups need to be compared appropriately using one-way analysis of variance or non-parametric equivalent "Kruskal-Wallis test," whereas three or more matched groups may be compared using two-way analysis of variance (i.e., repeated measures ANOVA) or non-parametric equivalent "Friedman test". Inappropriate use of repeated unpaired t-test/Wilcoxon rank sum test and repeated paired t-test/Wilcoxon Sign rank test will result in distortion in the considered level of significance; and also in implications of related results. Further, in case of significant result under analysis of variance, multiple comparison tests are used appropriately to identify pairs of groups/time points with significantly different results.

To quantify association between two quantitative variables, Pearson's correlation coefficient is used in case assumptions related to linearity as well as normality is fulfilled; otherwise non-parametric equivalent "Spearman's correlation coefficient" has to be used. Further, to assess prediction/association of a quantitative outcome variable with another quantitative independent variable, simple linear regression is used in case assumptions related to linearity as well as normality is fulfilled; otherwise one of the non-parametric equivalent regressions needs to be used appropriately. For assessing prediction/association of a quantitative outcome variable with a quantitative/categorical exposure variable after adjustment in relation to various measured or categorical variables, multiple linear regression is used in case assumptions related to linearity as well as normality is fulfilled; otherwise, one of the non-linear or non-

parametric equivalent regressions needs to be used appropriately.

Categorical Outcome (i.e., Two or More Categories)

Such outcomes are also known as binomial/multinomial outcomes. To describe categorical data, absolute frequency along with relative frequency (e.g., %) are used. To compare observed proportion of a specific event in one group to a given proportion in a population, if sample size is sufficient and related assumptions are fulfilled, Chi-squared test is used; otherwise exact statistical hypothesis test (i.e., Binomial test) is used. Likewise, to compare two independent/unpaired groups, Chi-squared test/Fisher's exact test are used. On the other hand, to compare two paired/matched groups, McNemar's chi-squared test is used. Further, to compare three independent/unmatched groups, the chi-squared test is used; whereas to compare three paired/matched groups, the Cochrane Q test is used. Contingency coefficients are computed to quantify association between two categorical variables. Simple logistic regression is used for assessing prediction/association of a categorical outcome involving two categories with an exposure variable. Further, multiple logistic regression is used for assessing prediction/association of a categorical outcome with an exposure variable after adjustment in relation to several measurable/categorical variables. In case of more than two categories in outcome, multinomial logistic regression analysis is used. Further, for a matched case-control study, conditional logistic regression analysis has to be used instead of multiple logistic regression analysis.

Time to Event Outcome

Dival time data. Such outcomes often result due to varying time entry of study-participants and/or their varying follow-ups. To describe survival data in one group, Kaplan Meier Survival curve is used. To compare this curve of two groups (i.e., comparing two independent groups), the log-rank test or the Mantel-Haenszel test is used. To compare two paired/matched groups, conditional proportional hazards regression is used. Likewise, Cox proportional hazards regression is used to compare three or more unmatched groups, whereas conditional proportional hazards regression is used to compare three or more paired/matched groups. As such, Cox proportional hazard regression is used for assessing prediction/association of time-to-event outcome from one/several measured/categorical variables. It may be pointed out here that the considered exposure variable/independent variables may be either of four forms.[13] Fixed covariate with fixed effect; fixed covariate with time-varying effect; time-varying covariate with fixed effect; and time-varying covariate with time-dependent effect. To develop an optimal epidemiological model, one needs to consider more appropriate scales of measurement of the considered independent variables, including the exposure variable and the outcome variable. To achieve further accuracy in the developed model and related analytical results, due to the availability of advanced computational facilities, one needs to also incorporate the structure of the data (e.g., multilevel data) while analyzing it. It is conventionally called multilevel models or hierarchical models.[14]

Use of Statistical Packages in the Analysis

Every statistical package (e.g., Microsoft Excel; SPSS; STATA; R; SYSTAT; BMDP; SAS) may not have computational facilities to carry out a specific analysis. Further, a particular package may not be used to carry out all the required analyses in a study. In other words, desired analyses of one data set may be completed using either one statistical package or more than one statistical package. Further, analytical results under specific analysis of a data set may differ across the available statistical packages for that analysis. For instance, the same regression analyses of the same data through different packages need not be the same due to probable differences in either method of estimation (e.g., ordinary least squares or maximum likelihood estimates); or default/inbuilt considerations (e.g., used default levels of significance regarding inclusion and/or exclusion of a variable in the model). To be more specific, such issues need to be decided more objectively for the analysis and accordingly those packages need to be considered where desired conditions are either in-built or may be altered. Without understanding data and knowledge about the required analytical methods in a study, mere mechanical use of a statistical package may often be deceptive and sometimes disastrous. To overcome this problem, one should ideally be well acquainted with the considerations involved while developing any analytical tool/module in the used statistical package.

Statistical Package: A Note of Caution

As of today, no computer/statistical program can decide basic principles related to appropriate biostatistical analysis of a data set collected with a specific objective. In other words, to carry out a specific analysis of a data set, an appropriate set of statistical methodologies has to

be identified; and the related analytical module needs to be chosen/developed by the analyst under the concerned package. As indicated earlier, required analysis may involve a single statistical package or more than one package. Methods of data analysis generally rely on the research question, type of outcome, and distribution of collected data. It will not be an exaggeration to mention here that even non-applied statisticians, especially youngsters, often get misled while deciding the required analytical tool and methods for analysis of a particular dataset. Further, after appropriate data analysis, even if the method of data analysis is the same, interpretation of the results mainly relies on the study design. If methodologically not comfortable, try to continue with the experienced biostatistician involved, preferably from the planning stage of a study.

Reporting of Analytical Results

One needs to choose relevant parts of the obtained analytical results under a study with a specific research question/objective and considered outcome. For instance, keeping in view of the distribution of data, along with absolute frequency distribution and relevant relative distribution, appropriate descriptive statistics along with respective standard error (preferably 95% CI) need to be reported. Once 95% CI is being reported, reporting *p-values* may not be required. Likewise, while reporting risk measures (e.g., odds ratio in case of case-control study; relative risk/rate ratio in case of cohort study), their respective 95% CI needs to be included without fail.[5-7] There is no point in simply listing regression coefficients along with their standard errors. Under analytical results, one should be strict in describing obtained results, not otherwise.

To make the reporting of study results more lucid and clearer, it is advisable to follow available specific guidelines[15-18] to report studies involving various study designs. For example, an available guideline to describe Case Reports is CARE; Qualitative Research is SRQR/COREQ; Diagnostic/Prognostic Studies are STARD/TRIPOD; Observational Studies are STROBE; Randomized Controlled Trials are CONSORT; and Systematic Review & Meta-Analysis is PRISMA. It may be worthwhile mentioning here that reading available guidelines before planning a particular study may also help in strengthening the related study protocol; smooth completion of the study; appropriate analysis; and interpretation of analytical results.

Interpretation of Results

One needs to be sure whether the study is conclusive or pilot. One should feel free to derive policy implications based on a conclusive study. People often get confused between, comparative cross-sectional study and case control design. To describe analytical results under comparative cross-sectional, preferably association word needs to be used. To describe in terms of risk, one needs to have either a case-control study or a cohort study. However, to talk in terms of causality, one should be sure about fulfillment of the desired recommended criteria. Most important among various points under the criteria is obviously that derived evidence should be drawn through a randomized controlled trial (RCT).[19,20] Further, in case of conflicting results under multiple RCTs on the same topic, one needs to rely on results under systematic review and meta-analysis considering all those studies.[21,22] In summary, inferences including probable policy implications should be drawn in view of overall scientific strength of a study.

In contrast to a conclusive study, regardless of the study designs, one should not be encouraged to derive conclusions from a pilot's study. Under a pilot/preliminary study, one should strictly use suggestive statements while describing the analytical results. It may be necessary to point out here that mere use of statistical methods in analyzing the data may not change a pilot study into a conclusive study. Interpretation should be duly restricted in the light of limitations of the study.

Summary

Keeping in view the above brief description, once accurate data is available under a study, appropriate analytical steps need to be considered to derive accurate and reliable results. Considered change in scales of measurements of any variable, including the exposure variable, other independent variables, and the outcome variable, should be carried out objectively in view of their obvious impact on analytical results and their interpretations. Mere mechanical use of statistical packages may produce disastrous results. Further, only a study following every step under research methodology, especially optimal sample size and power of the study, remains a conclusive study; otherwise a pilot or preliminary study. Interpretation should be drawn in the light of obtained findings and design of the study. Inferences should be done in view of strength of a study.

To be more specific, conclusions may be drawn from a conclusive study only; otherwise it is always better to use suggestive language. Also, while interpreting results, association word needs to be preferred instead of appropriately used risk word in case of case-control and cohort studies. Further, causal inferences may be

drawn only if the study is RCT and related recommended criteria are found fulfilled. In summary, keeping in view of strengths and limitations of the study, it is always needed to write how much evidence a study provides regarding public health and clinical practice.

References

1. Dwivedi SN. What are the major research methodology steps under a clinical study? Central India Journal of Medical Research. 2023; 2(2):4-9.
2. WHO. Health Research Methodology (A Guide for Training in Research Methods). 1992. Oxford University Press, Delhi.
3. Dwivedi SN. Understanding of study designs in clinical research: Major prompting points. Central India Journal of Medical Research. 2024; 3(3):12-15.
4. Dwivedi SN. How to Formulate a Research Question, Hypothesis and Objective for a Clinical Study? Central India Journal of Medical Research. 2023; 2(3):1-5.
5. Armitage P, Berry G, Matthews JNS. Statistics in Medical Research. Fourth Edition. Wiley Inter Science. 2008. ISBN: 9780632052578.
6. KR Sundaram, SN Dwivedi, V Sreenivas. Medical Statistics: Principles & Methods. Wolters and Kluwer (Health), New Delhi. 2015 (Second Edition).
7. Altman Douglas G. Practical Statistics for Medical Research. Chapman & Hall/ CRC. 1991. ISBN 0-412- 27630-5.
8. Dwivedi SN. Understanding of scales of measurements and its consideration in clinical research. Central India Journal of Medical Research. 2023; 2(1):13-16.
9. Dwivedi SN. What are the major sampling methods to obtain a representative sample for a clinical study? Central India Journal of Medical Research. 2024; 3(1):6-11.
10. Dwivedi SN. How to deal with sample size exploration and its reporting in clinical research? Central India Journal of Medical Research. 2022; 1(3):2-6.
11. Chandra, Prem, Dwivedi SN, Bal CS, Kumar Ajay, Pandey Arvind. Innovative role of imputation method/analysis in clinical trial on remnant ablation in differentiated thyroid cancer, Statistics in Transition: Journal of Polish Statistical Association, August, 2007, Vol. 8(2), pp.251-272.
12. Dwivedi SN. Biases including confounding and effect size modification in clinical research and their mitigations. Central India Journal of Medical Research. 2024; 3(2):6-12.
13. Dwivedi SN, Singh R. On assessing the child spacing effect of breastfeeding using Cox-proportional hazards models. Demography India. 2003; 32(2): 160-170.
14. Dwivedi SN, Sundaram KR. Epidemiological models and related simulation results for understanding of contraceptive adoption in India. Int J Epidemiol. 2000 Apr;29(2):300-7. doi: 10.1093/ije/29.2.300. PMID: 10817129.
15. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull World Health Organ. 2007;85(11):867-72.
16. Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., Evans, S. J. W., for the CONSORT Group, (2006). Reporting of Noninferiority and Equivalence Randomized Trials: An Extension of the CONSORT Statement. JAMA 295: 1152-1160
17. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
18. Matthew JP et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29:372:n71. doi: 10.1136/ bmj . n71
19. Bal C, Chandra P, Kumar A, Dwivedi SN. A randomized equivalence trial to determine the optimum dose of iodine-131 for remnant ablation in differentiated thyroid cancer. Nuclear medicine communications. 2012 Oct 1;33(10):1039-47.
20. Dwivedi R, Ramanujam B, Chandra PS, Sapra S, Gulati S, Kalaivani M, Garg A, Bal CS, Tripathi M, Dwivedi SN, Sagar R. Surgery for drug-resistant epilepsy in children. New England Journal of Medicine. 2017 Oct 26;377(17):1639-47.
21. Pathak M, Dwivedi SN, Deo SVS, Thakur B, Sreenivas V, Rath GK. Effectiveness of Added Targeted Therapies to Neoadjuvant Chemotherapy for Breast Cancer: A Systematic Review and Meta-analysis. Clin Breast Cancer. 2019 Dec;19(6):e690-e700. doi: 10.1016/j.clbc.2019.06.001. Epub 2019 Jun 20. PMID: 31337531.
22. Pathak M, Deo SV, Dwivedi SN, Thakur B, Sreenivas V, Rath GK. Regimens of neo-adjuvant chemotherapy in the treatment of breast cancer: A systematic review & network meta-analysis with PRISMA-NMA compliance. Crit Rev Oncol Hematol. 2020 Sep;153:103015. doi: 10.1016/j.critrevonc.2020.103015. Epub 2020 Jun 5. PMID: 32563131.