Statistical Methods In Clinical Trials, Validity Analysis And Evidence Based Medicine & Meta-Analysis

(A) PRINCIPLES & STATISTICAL METHODS IN CLINICAL TRIALS

CLINICAL TRIALS
The drug characteristics are generally assessed through animal experiments or laboratory tests before they are recommended for use. But, however successful these experiments may turn to be, the drug ultimately has to be tried on human beings to assess their efficacy as compared to the existing lines of treatment, their side effects and dosages. Such trials are called CLINICAL TRIALS. In a clinical trial the effect of exposure / intervention on the outcome on a group of subjects is studied:

Exposure / intervention: drug, surgery, diet, exercise or health education
Outcome: recovery, improvement, survival, increase / decrease in the value of the variables etc.

EFFECTIVENESS
Is a measure of the benefit resulting from an intervention for a given health problem under the usual conditions of clinical care for a particular group. Under this evaluation, in addition to measuring the efficacy of an intervention, it also measures its acceptance by those to whom it is offered. Thus, effectiveness answers the question, “does the practice do more good than harm to people to whom it is offered?”

FOUR PHASES OF TRIALS
PHASE - I : Toxicology, Pharmacokinetics, Safety on human volunteers are studied
PHASE - II : To study treatment effect on small number of patients
PHASE - III : Randomized (multi-centric) controlled trials (RCT)
PHASE - IV : Marketing the drugs - strategies and modalities and to study long term side effects

In this section statistical aspects to be taken care of in RCT are discussed:

STEPS IN STATISTICAL ASPECTS OF RCT

1) Estimation of minimum sample size based on the objective(s) of the clinical trial
(Chapter on Sample size estimation and selection of sample from the population applying appropriate sampling method may be referred)

2) Selection of patients
(a) Exclusion & inclusion criteria
(b) Comparability of subjects in the different groups
(c) Control group (placebo / standard treatment)
(d) Method of selection of patients adopting appropriate sampling method simple random method of allocation may be applied. Stratified sampling method of allocation may be applied if there is heterogeneity in the population of patients. For example, if the population consists of all age groups and age is a factor which affects the response to the treatment, stratified sampling method will increase the precision of the estimate of the response parameter.

3) Treatment specifications: dose, frequency, route, duration and other related aspects should be clearly spelt out.

4) Ethical considerations: It is very important that the study is approved in ethical angles by the Ethical Committee in the Institution following the accepted guidelines. Without the ethical clearance of the study the findings of the study are not acceptable

5) Follow-up of patients: Recording of various information systematically in suitably designed proformae is very important in clinical trials. Since the clinical trial is carried out over a period of follow-up depending upon the requirement the records have to be kept systematically and carefully to rule out loss of data

6) Drop out: Drop out of the patients over the study period may be a problem. The drop out rate has to be kept as low as possible for accepting the results scientifically

7) Coding the treatment: Carefully coding the treatment if it is feasible and de-coding it only after the results are
available is very important to maintain confidentiality of the trial and trial results. A third person will have to keep the code of drugs and it will be decoded only by him/her after the results are made available. Of course this will be possible only if the two or more treatment methods look alike (pills or capsules).

8) Data analysis applying appropriate statistical methods: Appropriate statistical methods have to be applied to the data for the scientific validity of the results. Depending upon the type of variable, number of groups, design of the study and the objectives of the study, appropriate statistical methods should be applied.

9) Interpretation of the results validly and meaningfully, mentioning the drawbacks of the trial and cautioning the interpretations.

ALLOCATION OF PATIENTS IN THE DIFFERENT GROUPS

(1) Biased randomization (alternate, odd / even): This method can be biased due to the preference that may be given by the treating doctor due to the severity of the disease. By making double blind allocation this problem can be taken care of to an extent.

(2) Balanced randomization: (random number table from books or computer generated numbers). This is the ideal method of allocation. This can be done either by using the table of random numbers or by generating them through computer software.

Stratification:- If the study population is heterogeneous with respect to, say, age, severity, type etc. which could affect the outcome, stratified allocation would be better. The study population may be stratified according to the factor which makes the population heterogeneous and required samples can be selected from each stratum and allocated randomly to the different treatment groups. This method will make the estimate of the response parameter more precise ( less standard error )

DESIGNS OF CLINICAL TRIALS:- Basically there are two types of design used in clinical trials-Parallel & Cross-over designs

(1) PARALLEL DESIGN: Two or more independent groups with different treatments- Allocation of total patients is done to the independent groups randomly. For example, 40 patients may be randomly distributed to the two treatment groups (Treatment-A & Treatment-B) randomly. This is the most commonly used design.

(2) CROSS-OVER DESIGN: If the number of patients available for the study is limited and there is no carry over effect w.r.t. the treatments after a certain period, Cross-over design may be adopted. This design results in reducing the sample size since the same patients will be used in each group after a certain period of time. For example, out of a total of 20 patients, the first 10 patients may be given treatment-A and the second 10 patients, the treatment-B. After a certain period the first 10 patients will be given treatment-B and the second 10 patients, the treatment-A. This design is recommended provided there is no carry over effect of the drugs after a certain period which has to be ascertained statistically and clinically from earlier documented studies.

A--------Period--------> B
( Set-A )

B--------Period--------> A
( Set-B )

BLINDING:- Blinding in clinical trial is done to avoid bias. Three types of blinding is available.

<table>
<thead>
<tr>
<th>TYPE</th>
<th>PATIENT</th>
<th>INVESTIGATOR</th>
<th>EVALUATOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINGLE</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>DOUBLE</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>TRIPLE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

The most commonly and recommended blinding is the Triple blinding. The patient, the Investigator and the statistician who does the analysis are blinded and gives unbiased results.

SHORT AND LONG TERM CLINICAL TRIALS

In a trial to find out the efficacy of a new drug in comparison to the standard drug in the treatment of common cold or influenza, the outcome is expected within a short time but, in trials on cancer patients, the outcome will take long time to show. Specific practical problems may be faced in such trials.

PRACTICAL PROBLEMS IN LONG TERM CLINICAL TRIALS

1) Necessity of dedicated investigators because of the long period of study
2) Systematically maintained registers
3) Drop outs / withdrawals due to side effects / partial improvement
4) Patient consent & compliance
5) Necessity of change in treatment due to side effects or ethical reasons
6) If multicentric trial, problem of keeping uniformity in the methodology & execution of the trial and data analysis
7) Necessity of interim evaluation
8) In multicentric trials, coping with conflicting results
9) Specific statistical methods to analysis the end point results has to be applied for example survival analysis

STATISTICAL METHODS FOR DATA ANALYSIS

1. Descriptive methods
2. Inference methods

For the details on the methods of Descriptive & Inference analysis, corresponding chapters may be referred in the earlier issues of this Journal. Some specific issues in clinical trials and the methods of analysis for the same are discussed below:

INTERIM ANALYSIS

It is always suggested that a certain number of interim analysis may be planned in the clinical trial. Number of interim analysis may be decided according to the requirement and convenience. Interim analysis helps to monitor the progress of the trial and to see whether all planned activities are going on as per the plan. Any aspect like, selection of patients, criteria for giving treatments and recording of responses requires any modification due to any reason, that may be incorporated in the trial without affecting the design and conduct of the trial. Also, it helps the investigator to find out whether anticipated statistical significance has been achieved in the improvement rates between the treatment methods based on the patients already included in the trial. If anticipated statistical significance has been achieved based on the lesser number of patients, the trial may be stopped for ethical reasons to avoid treating the patients with the lesser effective treatment method. The only modification which has to be made in the analysis is that the level of statistical significance (p-value) has to be changed depending upon the number of interim analyses. If the p-value fixed for statistical significance is 0.05 and the number of interim analysis is 3, the p-value for statistical significance has to be fixed based on O’Brien formula. The p-value in the first interim analysis will be much lower than 0.05, for the second interim analysis. p-value will be slightly higher than the p-value fixed for the first interim analysis and for the final analysis p-value will be fixed as 0.05. For more details of finding out the p-value for each interim analysis, the paper given under Books for further reading may be referred.

INTENTION TO TREAT ANALYSIS

Drop-out, Withdrawal, change of Treatment (within or outside the CT Protocol) — due to serious side effects, general negligence etc. are ideally to be avoided, but, in practice, may not be possible. It affects the balance of Randomization and introduces bias in the Treatment comparisons. To avoid this, analysis may be done as per the original grouping itself. This analysis is called — INTENTION TO TREAT ANALYSIS.

It may not sound logical. It is not recommended for all Clinical Trials. When Treatment A is not effective and for ethical reasons another Treatment may have to be given, for the benefit of the patient, this type of analysis may be recommended. Comparison of Coronary by-pass surgery & Medication for unstable Angina pectoris may be an example. Medication may not be effective in some patients and for ethical reasons and keeping the treatment for the benefit of the patient, coronary by-pass surgery may have to be done in such patients. This will affect the randomization of patients. It is advisable to do the analysis in both ways — i.e., Original grouping & according to the grouping after the change over.

STATISTICAL ANALYSIS FOR ADJUSTMENT FOR THE CONFOUNDING VARIABLES

PROGNOSTIC (CO) FACTORS — CONFOUNDING FACTORS

Effect of the Treatment could be related to many variables: — Sex, Age, Severity of disease, Duration of the disease, Personality Variables, Diet, Smoking habit, Use of Alcohol, Clinical & Laboratory variables (BP, Heart rate, Blood Sugar level) etc. These variables may be called — confounding factors or prognostic co-factors. While comparing the Response variable between the Treatment Groups, the effect of these Co-Factors on the Response Variable has to be studied and adjusted. Consider the following results obtained from a clinical trial:

<table>
<thead>
<tr>
<th>Response</th>
<th>Group</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>(Standard drug)</td>
<td>17(45.9%)</td>
</tr>
<tr>
<td>Not-improved</td>
<td>ND</td>
<td>38</td>
</tr>
</tbody>
</table>

χ² = 1.41 (p = 0.24) — The difference in the improvement rates between the two treatment groups (SD and ND) is statistically not significant. Assume that the investigator knows that age of the patient may affect the response to the treatment. Statistical significance of the difference in the age distribution between the two groups has to be tested before comparing the effect of the two drugs.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Group</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-30</td>
<td>(Standard drug)</td>
<td>20 (36.4%)</td>
</tr>
<tr>
<td>31-60</td>
<td>ND</td>
<td>35</td>
</tr>
</tbody>
</table>

χ² = 7.92 (p = 0.005) — Significant

On comparing the age distribution between the two groups, it was found that there is a statistically significant difference in the age distribution between the two groups. Hence the comparison of the treatment effect should be done separately in the two main age groups, assuming that the improvement rate might be the same statistically, within each age group.

15-30
Response                         Group

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>ND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>10(50.0 %)</td>
<td>10(33.3 %)</td>
</tr>
<tr>
<td>Not-improved</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 0.78 \ (p = 0.38) \] – Not Significant. Though comparatively higher improvement rate was observed in the SD group compared to ND group, the difference was statistically not significant.

31-60

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>ND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved</td>
<td>07(20.0 %)</td>
<td>10(66.7 %)</td>
</tr>
<tr>
<td>Not-improved</td>
<td>28</td>
<td>05</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 8.22 \ (p = 0.004) \] – Significant with 99% Confidence. Statistically significant difference was observed in the improvement rates between the two treatment groups in the age group 31-60 years. The effect of the ND was statistically better than that of SD. Statistical difference was not observed when the analysis was done for the combined age group.

However, if there are many confounding factors, say 5 factors, each having two sub-groups we will have 10 tables for analysis adjusting for only one variable at a time. If all the 5 confounding factors have to be considered together, there will be \( 2^5 = 32 \) tables for which analysis has to be done for 32 tables and it will be difficult for interpretation of the results.

Mantel-Haenzel method of Chi-square may be applied to adjust for a few confounding factors and to get a p-value for statistical significance after adjusting for the confounding factors. However, the best method to take care of many confounding factors and to get a p-value for statistical significance after adjusting for all the confounding factors, Multivariate logistic regression analysis can be done. In this analysis the inter-associations among all the confounding variables and with the type of treatment will be taken care of in the comparison of treatment effects between the treatment groups. For example, comparing the improvement rates in vision between two treatment approaches adjusting for the confounding variables such as age, gender, nutritional status, eye care, duration of watching TV etc. A description of this analysis is beyond the scope of this chapter and hence the books given under references may be referred for the details of this analysis.

(B) STATISTICAL METHODS IN SCREENING AND DIAGNOSTIC TESTS

In Epidemiological studies diagnostic tests play very important role based on clinical observations or on laboratory techniques, by means of which individuals are classified as healthy or having the disease under investigation. This forms a part of the screening programme in Epidemiological studies for early diagnosis of the disease. The test should be as far as possible, simple one, which will be feasible in the field set-up. The suspected cases will be referred for further clinical and laboratory tests for more accurate diagnosis.

One of the principal goals in the practice of medicine or in a clinical trial or in an epidemiological study is to make, as far as possible, as close as possible and as accurate as possible the correct diagnosis based on various clinical and laboratory tests. In this process care should be taken to avoid again as far as possible to declare him/her as a case falsely or to declare him/her as a normal falsely. That is if it cannot be completely avoided, the efforts should be taken to minimise the probability of the wrong classification. The aim should be to minimise the magnitude of uncertainty as low as possible. Since the diagnosis based on the information obtained is a probabilistic computation, the process of making diagnosis is known as diagnostic reasoning or clinical decision making.

The major criteria of the screening test are – Validity of the test, easy applicability, acceptability and less costly.

Validity of a screening test is measured by its ability to correctly categorize persons who have the disease as test positive and those without the disease as test negative. These are measured by the Sensitivity, Specificity and Predictive values of the positives and negatives and accuracy by the test.

Analysis

Type of variable

In most of situations the test under investigation could be of discrete type. For example, test is positive or negative. Some investigations are of continuous type such as cholesterol, hemoglobin, albumin, heart rate etc.

Discreet variable

For easy analysis and interpretation it is better to arrange the data in \( 2 \times 2 \) Contingency table

Gold standard

The method / procedure that is used to define the true state of the person examined, the most accurate diagnosis for confirming the presence of the disease

Example

For evaluating a certain serological test for diagnosis of pulmonary tuberculosis, the Gold standard could be the result of sputum culture for mycobacteria.

Index Test

The test whose validity / discrimination power is to be
investigated in comparison to the gold standard. Without a gold standard a new test cannot be tried upon and recommended for use.

Table-1: 2 x 2 --- Contingency Table

<table>
<thead>
<tr>
<th>TEST</th>
<th>GOLD STANDARD (TRUE)</th>
<th>DISEASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a (TP)</td>
<td>b (FP)</td>
</tr>
<tr>
<td>--</td>
<td>c (FN)</td>
<td>d (TN)</td>
</tr>
</tbody>
</table>

TP = True positives (both test & disease positive) = a
TN = True negatives (both test & disease negative) = d
FP = False positives (test positive, but, disease negative) = b
FN = False negatives (test negative, but, disease positive) = c

Total positives by the GS = a + c
Total negatives by the GS = b + d
Total positives by the Test = a + b
Total negatives by the Test = c + d

Validity of an index test is mainly judged in terms of FIVE parameters: ---
Sensitivity, Specificity, Predictive values of the positives (P+) & the negatives (P-) and Accuracy

Sensitivity = Proportion of the diseased cases rightly detected as diseased by the test = TP / total positives by the GS = a / (a + c)
Specificity = Proportion of the non–diseased rightly detected by the test as non–diseased = TN / total negatives by the GS = d / (b + d)
P+ = TP / total positives by the test = a / (a + b)
P- = TN / total negatives by the test = d / (c + d)

Accuracy = Proportion of those correctly detected as diseased/non–diseased by the test = (a + d) / (a + b + c + d)

Higher the Sensitivity, lower the false negatives will be. Similarly higher the Specificity, lower the false positives will be.

For a good test a careful analysis is required keeping in mind the importance of sensitivity over specificity or vice-versa, keeping the right balance between them and also keeping the predictivity of the test in mind.

If we don’t want to miss a case, sensitivity should be as high as possible. If we don’t want that unnecessarily a normal should not be declared as a case, specificity should be as high as possible. At the same time the test should yield high predictivity of positives & negatives.

Sensitivity is influenced by the severity of the disease. Any test result is more likely to be positive in advanced stages of the disease. Specificity is influenced by the state of health in the non-diseased sample.

Predictive values of the Positives and Negatives by the test are very important in evaluating a screening test. Predictive values of the Positives (P+) measures the Probability that a person actually has the disease given that he/she tests positive. Predictive value of the Negatives (P-) measures the Probability that an individual is truly disease free given that he/she tests negative by the screening test. Predictive values are affected not only by the Validity parameters, but also by the population characteristics to which the test is applied. For rare diseases, the major determinant of P+ is the prevalence rate. In such a situation, the results that are positive will mostly be false positives. Predictivity of a test closely depends upon the prevalence of the disease, even if the sensitivity & specificity remain constant. As the prevalence increases, positive predictivity increases. But, negative predictivity decreases. If the P+ & P– are low (false positives & false negatives are more), the test will have lesser validity.

While Sensitivity and Specificity refer to the Accuracy of the test, P+ and P– refer to the estimation of the Probability of the presence or absence of the disease.

The computation of the Validity parameters are explained for the following data:

Table-2: Results of a simple & easy vision test for visually impaired (with confirmed test done in hospital) in school children are given below:

<table>
<thead>
<tr>
<th>Simple field test</th>
<th>Visually impaired (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Positive</td>
<td>80</td>
</tr>
<tr>
<td>Negative</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Sensitivity of B = 80/100 = 80.0% Specificity of B = 4840/4900 = 98.8% 
P+ of B = 80/140 = 57.1% False positives = 42.9%

Only 1 out of 2 children who had the test positive had Visual impairment.
P_ of B = 4840/4860 = 99.6 % False negatives = 0.4%

Obviously, it would be desirable to have a screening test with high sensitivity and specificity. But, usually that won't be possible and generally there would be a trade-off between these two parameters. The criteria for fixing acceptable values for these two parameters depend upon weighing the consequences leaving cases undetected (false negatives) against erroneously classifying those not having the disease as diseased (false positives). Sensitivity should be increased at the expense of specificity when the penalty associated with missing a case is high such as when the disease is serious and definitive treatment exists. Specificity should be increased relative to sensitivity when the costs of risks associated with further diagnostic tests are substantial, such as breast cancer for which definitive diagnostic evaluation is biopsy.

After the validity of the test has been evaluated, its reliability should be studied. Reliability of a test refers to consistency of results when repeat examinations are performed on the same persons under the same conditions. Four sources of variability can affect the reproducibility of the results of a test-biological variation (inherent in nature), instrument variability, intra-observer variability and inter-observer variability. While the biological variation cannot be manipulated the other types of variation can be minimized by standardization of instruments and techniques and giving appropriate training for measuring and recording the information to the staff who are involved in doing and evaluating the screening tests.

Feasibility of the screening test is determined by many factors such as acceptability by the screenees, cost of effectiveness, time and transport and testing facilities available. While Pap smear test may be easier and acceptable for detecting cancer, sigmoidoscope for detecting colon cancer may be more difficult and non-acceptable and costlier.

**Likelihood –Ratio (LR)**

A simpler expression for measuring the results of Diagnostic tests is by computing the Likelihood Ratios (LR). LR is defined as follows:

\[
\text{LR}^+ = \frac{\text{Probability of the test result in the diseased persons}}{\text{Probability of the test result in the non-diseased persons}}
\]

\[
\text{LR}^- = \frac{\text{Probability of the test is negative in diseased}}{\text{Probability of the test is negative in non-diseased}}
\]

LR is a very useful way to characterize diagnostic information. The clinician / epidemiologist needs to remember only one parameter value instead of two parameters and it is easy to interpret the meaning of this parameter.

**Example : for the data given in Table-2**

\[
\text{LR}^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}} = \frac{0.8}{0.012} = 66.7
\]

\[
\text{LR}^- = \frac{1 - \text{Sensitivity}}{\text{Specificity}} = \frac{0.2}{0.988} = 0.2
\]

The test result would be expected to be positive in 66.7 children with visual impairment as compared to a positive result in a person without visual impairment.

The test result would be expected to be negative in 5 children without visual impairment as compared to a negative result in one child with the visual impairment.

Larger is the value of LR+, the stronger the association between having a positive test result and having the disease and thus better the diagnostic value of the test. Similarly, smaller the value of LR−, stronger is the association between having the negative test result and not having the disease and thus better the diagnostic value of the test. It has been shown that a value of 10 or greater for LR+ and a value of 0.1 or less for LR− can be considered as an indication of a test with high diagnostic value. Likelihood ratios do not vary according to the prevalence of the disease.

**Continuous variables**

Validity parameters can be computed for different cut-off points of the values of the continuous variable and a trade-off between sensitivity and specificity can be done to arrive at an ideal cut-off point. This is done by plotting a curve taking Sensitivity along the Y-axis and (1 – Specificity) along the X-axis. The resulting curve is called ——Receiver operating characteristic curve (ROC curve)

ROC curve is a graphical method for depicting the trade-off between True positive rate and False positive rate. A summary index of overall test performance can be computed as the area under the ROC curve. The greater the area, the better the test performance. The highest possible value for the area under the ROC curve is 1, which is equivalent to a perfect test. The area under the diagonal line corresponds to a test that does not distinguish between persons with or without the disease of interest. The closer an ROC curve is to the upper left hand corner of the graph, the more accurate it is, because the true positive rate is 1 and the false positive rate is Zero. ROC curves are useful graphic methods for comparing two or more diagnostic tests.

**Example: Table 3**

The validity parameters of a test variable for different cut-off values for these two parameters depend upon weighing the consequences leaving cases undetected (false negatives) against erroneously classifying those not having the disease as diseased (false positives). Sensitivity should be increased at the expense of specificity when the penalty associated with missing a case is high such as when the disease is serious and definitive treatment exists. Specificity should be increased relative to sensitivity when the costs of risks associated with further diagnostic tests are substantial, such as breast cancer for which definitive diagnostic evaluation is biopsy.
The cut-off value - 20, with a sensitivity and a specificity of 75%, could be a better cut-off value as compared to all other cut-off values, even though LR+ is much lesser than 3 and LR- is much higher than 0.1.

A summary index of overall test performance can be computed as the area under the ROC curve. The greater the area, the better the test performance. The highest possible value for the area under the ROC curve is 1, which is equivalent to a perfect test. The area under the diagonal line corresponds to a test that does not distinguish between persons with & without the disease. The closer an ROC curve is to the upper left hand corner of the graph, the more accurate it is, because the true positive rate is 1 and the false positive rate is zero. ROC curves are useful graphic methods for comparing two or more diagnostic tests.

STATISTICAL METHODS IN EVIDENCE BASED MEDICINE AND META ANALYSIS IN CLINICAL RESEARCH

In the fastly developing clinical research and decision making process, a new paradigm has emerged in the recent past - “EVIDENCE BASED MEDICINE”.

EVIDENCE BASED MEDICINE is based on: Systematic as well as Unsystematic clinical experience and evidence mainly by literature searching and gathering all available information (evidence) and applying formal scientific & statistical methods in evaluating the clinical literature.

Easy access to the computer and Internet facilities, literature search has become much easier and faster. This is a method which helps the clinicians in making decisions about the care of individual patients using the current best evidence concisely, explicitly & judiciously.

While reviewing the evidences in clinical trials and epidemiological investigations, several important questions need to be asked before taking a decision: similarity of the groups to be compared, at the start of the study, allocation of patients to different groups, whether random or not, validity of the diagnostic tests, drop out rate, the results on the treatment effect, its precision, its clinical importance and applicability etc.

STATISTICAL METHODS IN EVIDENCE BASED MEDICINE

One of the important statistical method commonly used in EBM is the RISK ANALYSIS.

Risk of an event due to a factor can be expressed in two ways – Proportion or Odds

(1) Proportion = \( \frac{a}{a+b} \)

Example:
- a: No of smokers who have lung cancer
- b: No of smokers who didn’t have lung cancer
- \( \frac{a}{a+b} \): Proportion of smokers with lung cancer

(2) Odds = \( \frac{a}{b} \)

Example: Ratio of lung cancer cases to non-cases in smokers.
Similarly, BENEFIT of an event due to a factor can be expressed in two ways

(1) Proportion = \( \frac{a}{a+b} \).

Example:
- a: No of patients who responded to the drug positively
- b: No of patients who didn’t respond to the drug positively
- \( \frac{a}{a+b} \): Total no of patients who received the drug
- \( \frac{a}{a+b} \): Proportion of patients who responded to the drug positively.

(2) Odds = \( \frac{a}{b} \)

Example:
- Ratio of responded to the non-responded in those patients who received the drug.
Combining these two measures - (proportion and odds),
META – ANALYSIS IN CLINICAL RESEARCH

Meta – analysis has been defined as the ‘Statistical analysis of a collection of analytic results for the purpose of integrating the findings’. The last few years have seen rapidly increasing interest in meta – analysis in the medical research literature. The results from a collection of independent randomized studies can be summarized in a systematic and quantitative way using a meta – analysis.

The main objective of such an analysis is to obtain information about treatment effects that cannot be ascertained from any of the studies taken alone. Any individual study may either be too small to detect moderate treatment effects, say on mortality, or too limited to allow generalization to other patient populations. We should like to know, overall, whether a treatment has a beneficial or harmful effect. Reviews of treatment or therapeutic areas are frequently carried out when new compounds are developed. They could perhaps benefit from such a quantitative approach in addition to qualitative and subjective summaries.

A meta – analysis can be viewed as an extreme form of multi – center study. There is a continuum from the true multi – center study, in which all centers follow an identical protocol, to a collection of studies addressing the same general therapeutic question but in results obtained from different independent studies caution has to be taken w.r.t. the design,objectives,inclusion & exclusion criteria of patients,drug delivery etc.This poses problems in getting similar studies for applying Meta-analysis.

OBJECTIVES OF META ANALYSIS

- To enhance power by increasing the size of sample which may not be large enough in individual studies especially in a rare medical condition.
- To restore uncertainty when the reports disagree.
- To improve estimates of effect size.
- To answer questions (not always possible) not posed at the start of individual trials.

These functions are particularly applicable to randomized control trials, because such trials are often too small to detect clinically important differences.

There are a number of reasons why Meta analysis is an important technique in Clinical Trials .It is now recognized that narrative reviews of a set of clinical trials can be misleading, being potentially distorted by the selection of evidence -mainly those providing positive results which might include small and large studies, randomized and non-randomized trials and inadequately analysed trials.

The human mind is not equipped to consider simultaneously a large number of alternatives.Confronted with the results from a large number of studies, giving different results, some significantly favouring one type of treatment, some favouring another treatment and some giving statistically non-significant results, it would be difficult to comprehend them and take a decision. Yet that is exactly the scope of the problem faced by a researcher attempting to integrate the results from a large number of studies.

Meta analysis will help us in reviewing systematically the available evidences,to provide quantitative summaries of the results from each study, to combine these results applying valid statistical analysis and to provide over all interpretation to help the clinicians and pharmacologists to take valid clinical decisions.Combining results will provide more statistical power and precision for detecting treatment effects.

Systematic reviews of articles on and related to the study variables is an important primary task the researcher has to do before attempting meta-analysis.Cochrane Library is the main source of gathering information on various clinical trials .There are many Centres of Cochrane Library at different Geographical sites such as New England Cochrane Centre, Australasian Cochrane Centre, Canadian Cochrane Centre, U.K. Cochrane Centre, French Cochrane Centre and so on.Various information related to several clinical studies have been stored systematically for the benefit of the researchers.

META ANALYSIS MODELS

While combining results from various studies we come across two types of variation viz; Within Study Variation & Between Study Variation. Fixed Effect Model is applied if only Within study variation need to be considered and the statistical analysis normally applied is :Mantel-Haenzel method . Random Effect Model is applied if both within and between study variation has to be considered and the statistical analysis normally applied is: –Dersimonian & Laird method.

If Between Study Variation is substantial relative to Within Study Variation larger studies will get more weightage under Fixed Effect Model than in Random Effect Model. In Random Effect Models,weights given to each Trial are more evenly distributed and small Trials get relatively more weight. Confidence Limits in Fixed Effect Model will be narrower than in Random Effect Model.Discussion on the details of doing the analysis applying these methods is beyond the scope of this article in the Journal.A typical way of presentation on the combined results based on the results from various studies is by drawing the Forest Tree as follows:

Suppose there are 9 studies considered for Meta-analysis and the study parameter is Odds of improvement in the condition
of the patient with drug-A as compared to the improvement in the condition of the patient with drug-B. Each line in the Forest Tree indicates the 95% Confidence limits of Odds ratio which is marked by a dot. The diagram shows that odds of improvement is better with drug-A (left hand side of the Centre line) than drug-B (right hand side of the centre line) in most of the 9 studies except in the 8th study and to some extent in the first and 9th studies. By applying both the Fixed effect & the random effect models, the combined Odds ratio, with 95% Confidence limits, shows that drug-A is better than drug-B w.r.t. the improvement rate.

Books For Further Reading


Note:
This is the last chapter in my presentation on "Biostatistics -Principles and Methods", applied to Medical Research. I hope that the readers are / will be benefitted from the presentation of various applications of Biostatistical methods described in the Eight chapters , in planning their research projects / Thesis reports and in analysing their data applying appropriate statistical methods and thus helping them in conveying their research findings systematically with statistical & scientific validity. The readers may contact me through E-mail for any clarification or any query w.r.t. any aspect covered in the various chapters. Needless to mention I enjoyed writing these chapters with many examples in research in Ophthalmology and in that process I learned a lot in the field of Ophthalmology, especially the important terms and terminologies in this important branch of Medicine. Let me use this opportunity to express my Good Wishes to all the Ophthalmologists, especially the researchers , meaningful, scientifically and statistically valid and useful research for the benefit of the people, especially the patients.

Prof Sundaram was previously the Head of Biostatistics at All India Institute of Medical Sciences. Currently he heads the Department of Biostatistics at Amrita Institute of Medical Sciences, Kochi.