

Critical Analysis of a Randomized Controlled Trial

Balkrishna D Nimavat¹, Kapil G Zirpe², Sushma K Gurav³

ABSTRACT

In the era of evidence-based medicine, healthcare professionals are bombarded with plenty of trials and articles of which randomized control trial is considered as the epitome of all in terms of level of evidence. It is very crucial to learn the skill of balancing knowledge of randomized control trial and to avoid misinterpretation of trial result in clinical practice. There are various methods and steps to critically appraise the randomized control trial, but those are overly complex to interpret. There should be more simplified and pragmatic approach for analysis of randomized controlled trial. In this article, we like to summarize few of the practical points under 5 headings: "5 'Rs' of critical analysis of randomized control trial" which encompass Right Question, Right Population, Right Study Design, Right Data, and Right Interpretation. This article gives us insight that analysis of randomized control trial should not only based on statistical findings or results but also on systematically reviewing its core question, relevant population selection, robustness of study design, and right interpretation of outcome.

Keywords: Critical analysis, Evidence based medicine, Randomized control trial.

Indian Journal of Critical Care Medicine (2020); 10.5005/jp-journals-10071-23638

INTRODUCTION

"Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital." [Aaron Levenstein]

Being up to date with knowledge is pivotal in world of evidence-based medicine. Sometimes, it is also crucial in terms of medicolegal aspect and to improve best current practice. In view of this background, plenty of articles and trials are emerging out in various journals every day. Among all types of study design, randomized control trial (RCT) is considered as supreme in terms of strength of evidence. Appropriately planned and vigorously conducted RCT is the best study design to see the intervention-related outcome difference, but simultaneously poorly conducted biased RCTs will misguide the reader. It is ideal to read RCTs and optimize clinical practice, but it is critical to understand strong and weak points of those RCTs before being dogmatic about their result or conclusion. There are many methods to appraise the RCTs, but in this article, I tried to simplify the points under 5 headings with mnemonic 5'Rs' that helps to understand things in better way (Flowchart 1).

STEPS FOR CRITICAL ANALYSIS OF RANDOMIZED CONTROL TRIALS

Formulate Right Question/Address Right Question

As Claude Lévi-Strauss said, "The scientist is not a person who gives the right answers, he is one who asks the right questions."

It is crucial to look for right question that possesses characteristic such as innovative, practice changing, knowledge amplifying, and above all having some biological plausibility.

Does Randomized Control Trial Address New/Relevant Question? Does Answer to this Question Lead to More Information that will Help to Improve Current Clinical Practice or Knowledge?

Questions arises from any of topic are mostly of two types: background questions and foreground questions. RCTs are the experimental design that usually target foreground questions that are more specific to establish intervention/drug and their effect/outcome relationship. Foreground research question has four components to get relevant information like Population,

¹Critical Care Unit, Sir HN Reliance Hospital, Ahmedabad, Gujarat, India

^{2,3}Department of Neuro Trauma Unit, Grant Medical Foundation, Pune, Maharashtra, India

Corresponding Author: Balkrishna D Nimavat, Critical Care Unit, Sir HN Reliance Hospital, Ahmedabad, Gujarat, India, Phone: +91 9930093731, e-mail: dr_bk_adrenaline@yahoo.com

How to cite this article: Nimavat BD, Zirpe KG, Gurav SK. Critical Analysis of a Randomized Controlled Trial. *Indian J Crit Care Med* 2020;24(Suppl 4):S215–S222.

Source of support: Nil

Conflict of interest: None

Intervention, Control, and Outcome (PICO format). Whether study question and design are ethical and feasible for relevant population can be decided by FINER criterial.¹

Outcome are the variables that are monitored during study to observe presence/absence of impact of intervention on desired population. Outcome is also labeled as events or end points. Most common clinical end points are mortality, morbidity, and quality of life. It is decisive to choose right end point with their background knowledge and its relevance to formulated question (Fig. 1).²⁻⁴

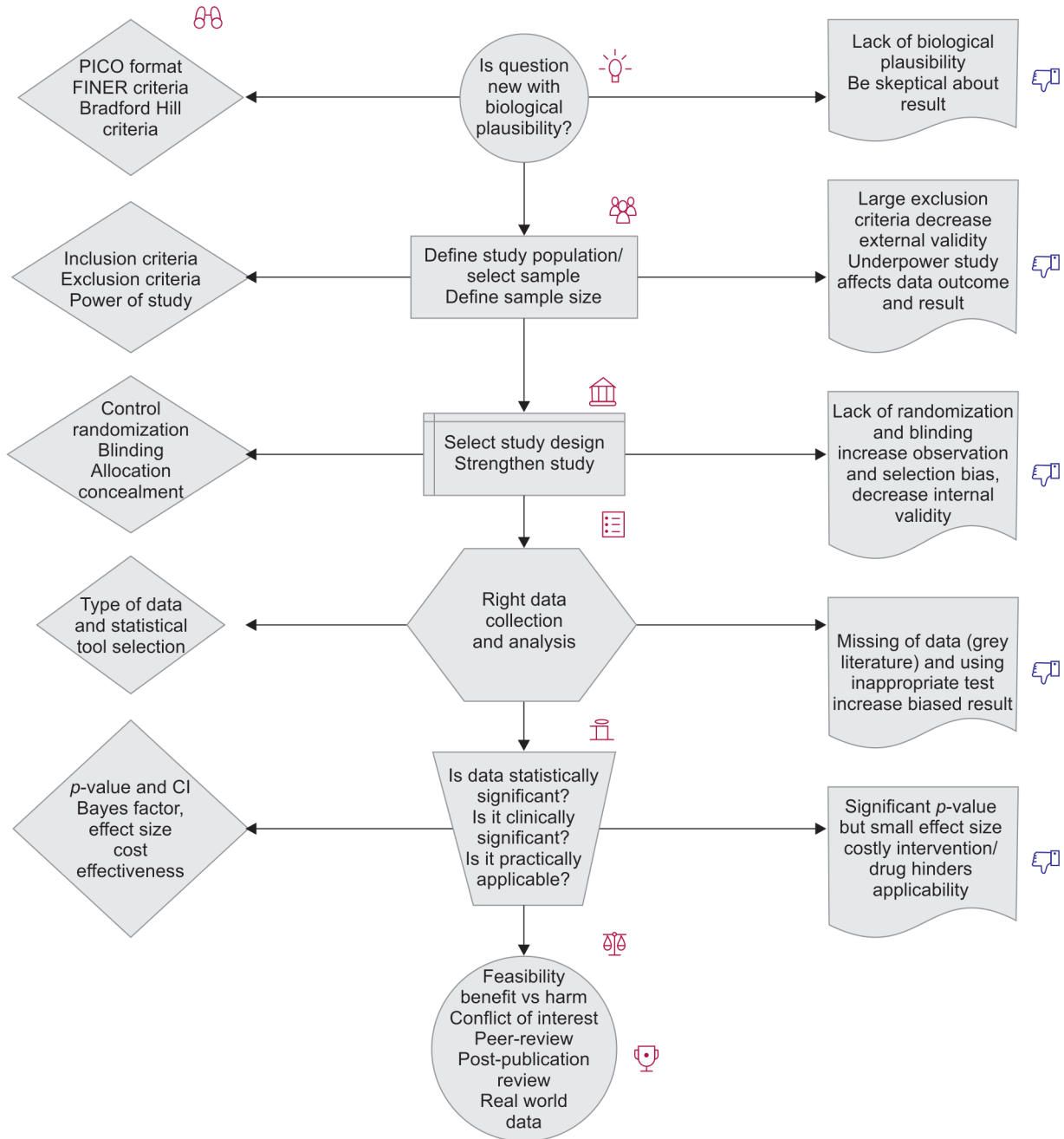
So, it is evident that no single end point is perfect, but end points should be accessed in the context of clinical question, power, and randomization.

Is Cause and Effect Having Biological Plausibility?

Biological plausibility is one of the essential components to establish that correlation means causation. Just mere association or having significant *p*-value without biological plausibility is like beating a dead horse (purely punitive). That means statistically significant data make least sense or should be interpreted with caution if they lack biological plausibility, and data that are unable to give statistical significance but have strong biological plausibility with vigorously conducted study should be evaluated again and discussed before rejection.⁵

To determine whether correlation is equivalent to causation, many criteria and methods are available. One of such criteria is Bradford Hill criteria. It is also important to understand that

Flowchart 1: Presentation of "critical analysis of RCT"



knowledge of biological plausibility is dynamic and evolves with time. It is possible that there is true causation, but biological knowledge at that time is unable to explain it (Table 1).

Right Population

Define Target Population/Does Sample Truly Represent Population?

RCTs are usually conducted on group of people (sample) rather than whole population. It is important for the trial that selected sample truly represents the baseline characteristic of the rest of the population. Inferential leap or generalization from samples to population is also not that simple and most of the time not full proof.

External validity in RCT represents at what extent the study result can be generalized to real-world population. Internal validity gives idea about how vigorous trial is conducted and generates robust data. If RCTs have poor internal validity, result made on that trial cannot be used firmly due to higher chances poor quality data and higher chances of bias for that given sample. Limitation of external validity means trial sample or defined sample is not true representative of rest of population. In a simplified way, if internal validity is questionable, applying it on larger scale is irrelevant, and second if trial having limited external validity (by having large exclusion criteria), applicability of RCT conclusion to rest of the population should be done with caution and less reliable. External validity is improved by changing



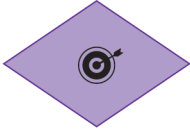


	 Peculiar features	 Pros	 Cons
Primary endpoint	For which trial is powered and randomized Mortality is a commonly used primary endpoint	Easily measured unequivocal patient-centered outcome	Heterogeneity in critical care illness syndrome Heterogenous composite outcome Mortality rate changes with time (as power of study depends on previous mortality rate) Timing of intervention affects outcome (early vs late) Lack of standardization of timing of measurements (like 28/30/60 days mortality)(2)
Secondary endpoint	Not powered enough to help in decision-making detect findings those are less frequent	Additional information about causal mechanism, work as pilot project for exploring novel hypothesis	If multiple secondary endpoints are investigated, then it is likely that one will be positive simply due to random chance (if used, then trial should have only 1–2 secondary endpoints and should be powered at initiation of trial only)
Surrogate endpoint	Rather than direct measurement of effect, reliable and valid indirect measurement of outcome carried out	Faster, easier, sometimes cheaper, selected when event rate is low	They should be validated prior to use surrogate endpoint shows same biological gradient/close response to intervention/exposure that translated to clinical endpoints
Composite endpoint	Two or more events merged into one endpoint to show overall treatment outcome	Reduces flaws associated with the multiple testing increases event rates and increase power of study	Each component of composite endpoint and should be screened for validity, reproducibility appropriateness, accurate measurement. When there is heterogeneity of treatment effect on different component of composite endpoints, interpretation of the results will be difficult (3)(4)

Fig. 1: Types of endpoint and their pros and cons

Table 1: Factors help to formulate sound question^{1,6}

PICO format	Finer criteria	Bradford Hill causality criteria
Population	Feasible	Strength of association (effect size)
Intervention	Interesting	Consistency (reproducibility)
Control	Novel	Specificity
Outcome	Ethical	Temporality (cause before effect)
	Relevant	Biological gradient (dose gradient response)
		Experimental evidence
		Biological plausibility
		Coherence
		Analogy

inclusion and exclusion criteria, while internal validity can be boosted by controlling more variables (reducing confounding), randomization, blinding, improving measurement technique, and by adding control/placebo group.⁷

Size of Target Population/Is Sample Size Adequate?

Another important step is to choose adequate sample size that gives relevant clinical difference that is statistically significant. Sample size estimation should be done prior to trial only and should not be deviate while study is ongoing to prevent statistical error. Study size is affected by multiple factors such as acceptable level of significance (alpha error), power of study, expected effect size, event rate in population (prevalence rate), alternative hypothesis, and standard deviation in population. There are formulas to calculate

sample size, but it is more important for us to understand the relationship of each factor with sample size.⁸

For phenomenon or association where effect size is large, even small size sample will solve the purpose. In traditional concept, we learnt that large sample size is good but that it is not true all the time, as even clinically nonsignificant difference will be highlighted when large sample size is analyzed. For certain disease where prevalence rate is low (rare events), it is not possible to do RCTs (where observational study solve the purpose).

Tool used for sample size estimation is “Power of the study”. Power of study represents how much study population required to avoid type II error for that study. Power of study depends on variable factors such as precision and variance of measurements within any sample, effect size, type I error acceptance level, and type of statistical test we are performing.⁹ Sample size also depends on expected attrition rate/dropout rate/losses to follow-up and funding capacity of trial.

Right Study Design

Experimental design considers better over observational design, as they have better grip on variables, and cause–effect hypothesis can be established. Experimental study design is again divided into preexperimental, quasi-experimental, and true experimental. Quasi-experimental and true experimental design is differentiated by absence and presence of randomization of groups. Randomized control trial is true experimental design, and it delivers higher quality of evidence over other designs as having remarkably high internal validity and presence of randomization. But RCT has its own limitations such as complex study design, costly by nature,

ethical issue with limitations (as intervention/medicine use), time consuming, and difficult to apply on rare disease or conditions.

Strengthen Study Design/Are Measures Taken to Reduce Bias (Selection or Confounding Bias)?

Interventional studies/RCTs are designed to observe the efficacy and safety of new treatment for clinical condition, it is particularly important that outcome does not happen by chance. To reduce confounding factors and bias, variety of strategies such as selection of control, randomization, blinding, and allocation concealment are helpful. Control arm is used for comparison to derive the more reliable effect of intervention. Control are of four types: (1) Historical, (2) Placebo, (3) Active control (where standard treatment used), and (4) Dose–response control (where control have different dose/gradient of intervention compared to interventional arm). Randomization helps to reduce the selection bias and confounding bias. Randomization can be done by computer-generated or random number table from textbook. Randomization techniques are of different types such as simple, block randomization, stratified, and cluster randomization. Reliability of sample randomization gets compromised if used for small sample. Block randomization is better method when there is large sample size, and follow-up period is lengthy. It is also important that block size should not be disclosed to the investigator, and if possible, block size should vary with time and randomly distributed to avoid predictability. Stratified randomization is used when there are specific variables having known influence on outcome. In cluster randomization, rather than individual a group of people are randomized. Blinding is a method to reduce observation bias. Study can be open labeled/unblinded or blinded. Blinding has different types such as participant blinding, observer/investigator blinding, and data analyst blinding. Allocation concealment secures randomization and thus reduces selection bias. The difference between allocation concealment and blinding is that allocation concealment is used while recruitment and blinding after recruitment.¹⁰

Right Data/Is Appropriate Tool/Method Used to Analyze Data?

"We must be careful not to confuse data with the abstractions we use to analyze them." [William James]

Study methodology should include mentioning type of research, collect and analyze data, tool/method used, and rationale of using those tools. After collecting data, the next step is to decide which statistical test should be used. Choosing the right test depends on few parameters: (1) purpose/objectives of study question (whether it is to compare data or establish any correlation between them); (2) how many samples are there (one, two or multiple); (3) type of data (categorical and numerical), (4) type and number of variables? (univariate, bivariate, or multivariate); and (5) Relationship between groups (paired/dependent vs unpaired/independent). Based on these differences, possible combinations arise. Table shows different combination and methodological tests used to analyze data (Table 2).

In RCTs, many times we have seen subgroup analysis or *post hoc* analysis; for a reader, it is very important to understand limitation of those analyzes. Subgroup analyzes are usually considered as a secondary objective, but in era of personalized medicine and targeted therapies, it is well recognized that the treatment effect of a new drug/intervention might not be same among study population. Subgroup analyzes are therefore important to interpret

the results of clinical trials.¹³ Subgroup analyzes is helpful when (1) to evaluate safety profile in particular subgroup, (2) to access consistency of effect on different subgroup, and (3) to detect effect in subgroup in otherwise nonsignificant trial.¹⁴ Subgroup analysis is much criticized by 2 ways: (1) Chances of high false-positive findings as multiple testing and (2) chances of false-negative when inadequate power (because of small sample size). It is exceedingly difficult to come to conclusion based on subgroup analysis and practice it. Still there are few scenarios where clinicians consider validity of subgroup analysis when prior probability of subgroup effect is more (at least more than 20% and preferably >50%), small number of subgroups (≤ 2) are tested, subgroup has same baseline characteristic, and when hypothesis testing of subgroup decided prior only. To reduce false-positive rate in subgroup findings, the clinician can take help of Bayes approach.¹⁵ *Post hoc* analyzes, type of subgroup analysis defined by, 'The act of examining data for findings or responses that were not specified a priori but analyzed after the study has been completed'. If possible, prespecified subgroup analyzes should be done compared to *post hoc* analyzes, as they are more credible.¹³

Right Interpretation (Giving Meaning to Data)

"Everything we hear is an opinion, not a fact. Everything we see is a perspective, not the truth." [Marcus Aurelius]

Is this RCT Result Difference by Chance/Statistically Significant?

Is p-value significant? Purpose of data collection and analysis is to show whether there is difference between two groups or not. Now this difference can be due to chance or true difference. To rule out difference by chance, many tools are used in statistics: *p*-value is one of them. *p*-value is a widely used yet highly misunderstood and misinterpreted index. In Fisher's system, the *p*-value was used as a rough numerical guide for the strength of evidence against the null hypothesis and value of which was arbitrary selected to 0.05. In simplified way, *p*-value <0.05 suggests that one should repeat the experiment and word significance is merely indicating "worthy of attention". So once *p*-value becomes significant, one should do more and more vigorous study rather than end of story.¹⁶

Misperception about p-value: Most common misperception about *p*-value are: (1) Large *p*-value means no difference and (2) smaller *p*-value is always more significant?

(a) "Absence of evidence is not the evidence of absence." If the *p*-value is above the prespecified threshold alpha error (mostly 0.05), we normally conclude that the H_0 is not rejected. But it does not mean that the H_0 is true. The better interpretation is that there is insufficient evidence to reject the H_0 . Similarly, the "not H_0 " could mean there is something wrong with the H_0 and not necessarily that H_a is right.¹⁷ (b) *p*-value is affected by factors like (i) effect size (appropriate index for measuring the effect and size of effect), (ii) size of sample (larger the sample size likely a difference to be detected), and (iii) distribution of data (bigger the standard deviation, lower the *p*-value).¹⁸ It is very important to understand that smaller *p*-values do not always mean significant findings, as larger sample size and smaller effect size can give smaller *p*-value.

Is multiple testing done? Another problem with *p*-value is multiple testing, and few of them/last testing shows *p*-value of <0.05.

Table 2: Factors/questions helps to select statistical tool to analyze data^{11,12}

1. Purpose/objective of study:				
A. Compare data				
2. Number of samples	3. Pair/unpair	4. Type and distribution of data		
		Parametric data (like comparing mean)	Nonparametric data (like comparing median)	
1 sample	–	1 sample <i>t</i> -test (<30: <i>N</i>)	One sample Wilcoxon signed rank test	
		1 sample <i>z</i> -test (≥30: <i>N</i>)		
2 samples	Unpair	Unpaired <i>t</i> -test	Wilcoxon rank sum test or Mann–Whitney <i>U</i> test	
	Pair	Paired <i>t</i> -test	Related samples Wilcoxon signed-rank test	
≥3 samples	Unpair	1-way ANOVA	Kruskal–Wallis <i>H</i> test	
	Pair	Repeated measures ANOVA	Friedman test	
B. Compare proportion				
Independent/unpaired		–	Pearson Chi-square test Fisher exact test	
Dependent/paired		–	McNemar test (2 groups) Cochrane Q test (≥3 groups)	
C. Predictors of outcome variables/correlation between variables (type of regression analysis)				
Number of dependent variables	Type of dependent variable	Number of independent variables	Type of independent variable	Test
One	Continuous	1	Continuous	Simple linear regression
		≥2	Categorical	One-way ANOVA
	Categorical	1	Any type of data	Multiple regression
		≥2	Continuous	Logistic regression
	Rare	Any number	Categorical	Pearson Chi-square or likelihood ratio
			Any type of data	Multiple logistic regression
			Any type	Poisson model
D. Degree of association between variables				
		Parametric method	Non-parametric	
		Pearson correlation coefficient	Spearman rank correlation Coefficient	
D. Analysis of survival data/Time to event analysis				
One sample population			Kaplan–Meier test	
Two sampling populations		One feature/categorical variable	Logrank test	
Two sampling populations		Two features/quantitative variable	Cox’s proportional hazards model, regression analysis	

“If you torture the data enough, nature will always confess.”
[Ronald Coase] One success out of one attempt and one success out of multiple attempts have different meanings in terms of statistics and probability. The underlying mechanism of multiple described is as “File drawer problem”. Multiple testing is more about “intention” and the future likelihood of replicability of the observed finding rather than truth.¹⁷

*Is false discovery rate ruled out?/solution of multiple testing *p*-value:*
Tools used to weed out such bad data that seems good: A simple not perfect solution of multiple testing *p*-value is the Bonferroni adjustment, which is to use $\alpha = 0.05/5 = 0.01$ for 5 (independent) tests as a new threshold and adjust the observed *p*-values by multiplying by 5. Problem with this adjustment is that it not only lowers the chance of detecting false-positive but also reduces

true discoveries. False discovery rate (FDR) is another method that controls the number of false discoveries only in those tests having significant result. Adjusted *p*-values using an optimized FDR approach is known as *q*-value. There are other methods to overcome this phenomenon like O’Brien-Fleming for interim analyzes and empirical Bayes methods.^{17,18}

*Is alternative approach to *p*-value used?/Bayes method:* Limitation of *p*-value is that it does not consider prior probability and alternative hypothesis. The evidence from a given study needs to be combined with that from prior work to generate a conclusion. This purpose is solved by Bayes’ theorem/method. Bayes’ factor is the likelihood ratio of null hypothesis to alternate hypothesis. In simple terms, *p*-value should be compared to strongest Bayes’ factor to see the true evidence against null hypothesis¹⁶ (Table 3).

Table 3: Properties and differences between Bayes' factor and *p*-value^{16,19}

Property	<i>p</i> -value	Bayes' factor
Effect size	No	Yes
Consider alternative hypothesis	No	Yes
Data	Observed + hypothetical	Only observed data
Computation	Easy	Complex
Interval estimation	Confidence interval	Credible interval
Intention of the researcher (result affected by stopping or measurement criteria)	Value affected	Not affected

Is p-value backed up with confidence interval? Confidence interval (CI) describes the range of values calculated from sample observation that likely contains true population value with some degree of uncertainty. CI will help to overcome lacunae of *p*-value by giving more information about significance. It gives idea about size of effect rather than hypothesis testing. Width of CI will give idea about precision/reliability of estimate. CI gives insight about direction and strength of effect and thus clinical relevance rather than just statistical one. *p*-value is affected by type I error while CI is not.^{20,21} Size of CI depends on sample size and standard deviation of study group. If sample size is large (leads to more confidence), it will give narrow CI. If the dispersion is wide, then certainty of conclusion is less and wider CI. Confidence interval is also affected by level of confidence that is selected by user and does not depend on sample characteristics. Most selected level is 95%, but different levels like 90% or 99% can be considered.^{20,21} Another usefulness of CI is in equivalence/superior/non-inferior type of studies, where CI is used as intergroup comparison tool and not the *p*-value.²¹

Is data robust? The Fragility index (FI) measures the robustness of the results of a clinical trial. In simple words, if the FI is high, statistical reproducibility of the study is high. The FI is the minimum number of patients whose status would have to change from a non-event (not experiencing the primary end point) to an event (experiencing the primary end point) to make the study lose statistical significance. For instance, an FI score of 1 means that only one patient would have to not experience the primary end point to make the trial result nonsignificant. In other words, it is a measure of how many events the statistical significance of a clinical trial result depends on. A smaller FI score indicates a more fragile, less statistically robust clinical trial result. Like other statistical tools, FI is also not free from limitations: (1) Only appropriate for RCTs; (2) Appropriate for dichotomous outcome; (3) Not appropriate for time-to event binary outcomes; (4) No specific FI value that defines an RCT outcome as robust and no FI score cut off value considered acceptable; (5) Use of FI scores to assess secondary outcome measures in studies may be limited; (6) Not reliable/difficult to interpret when number of subjects who drop out for unknown reasons is large; and (7) FI strongly related to *p*-value. In view of above-mentioned flaws, FI should not be used as isolated tool to measure strength of effect. Trials with lower scores are more fragile (which is usually in association with the smaller number of events, smaller sample size, and resulting lower study power), and trials with a higher FI

score are less fragile (which is usually associated with larger number of events, larger sample size, and resulting higher study power).²²⁻²⁴

Is this Statistically Significant Difference/Clinically Significant?

Another more common misinterpretation is 'statistically significant is equivalent to clinical significant'. Statistically significant means there is true difference in the data but whether that difference is clinically significant or not depends on many factors such as size of effect (minimum important difference), any harms (risk-benefit), cost-effectiveness/feasibility, and conflict of interest/funding.²⁵

"The primary product of a research inquiry is one or more measures of effect size, not p-values." [Jacob Cohen]

p-value gives idea about whether effect exists or not but does not give idea about size of effect. It is particularly important to mention both effect size and *p*-value in the study. Both parameters are not alternative to each other but rather they are complementary. Unlike significance tests, effect size is independent of sample size.²⁶ Effect size indices can be calculated depending on the type of comparison under study (Table 4).

Interpretation of effect size depends on the assumptions that both group ("control" and "experimental") values are normally distributed and have same standard deviations. Relative risk and odds ratio should be interpreted in the context of absolute risk and confidence interval. Use of an effect size with a confidence interval will deliver the same information as a test of statistical significance, but it gives weightage on the significance of the effect rather than the sample size.

Minimum important difference: Most important and difficult point in clinical significance is to decide what difference is clinically important. There are 3 ways to decide MID: Anchor-based, distribution-based, and expert panel approach.²⁵

Is Randomized Control Trial Result Applicable/Practice Changing?

When any of new intervention or therapy launched, its acceptance and success not only depend on clinical efficacy but also on the costs associated with it. Randomized trials focus on clinical end points such as organ failure, respiratory or renal support, mortality, and morbidity, while contemporary clinical trials include economic outcomes. Therapy with good clinical outcome and low cost is considered as dominant strategy, and in such cases, there is no need of any deep analysis. But problem arises when there is one novel therapy showing some better clinical outcome but having higher cost. In such cases, the most important thing is whether improvement in outcome is worth the higher cost. So, cost-effectiveness helps in balancing cost with efficacy/outcome and comparing available alternative therapies.²⁹

Is any conflict of interest financial/non-financial? A conflict of interest (COI) happens when contradictory interest emerges out to on a topic/activity by an individual/institution. When conflict of interest exists, validity of RCT should be in question, independent of the behavior of the investigator. Conflict of interest can happen at different level/tier like at the level of investigator, ethics committee (EC), or at regulator level. Conflict of interest can happen with sponsors like pharmaceutical companies, contract research organization, or at multiple levels. Nowadays, most of the trials are blinded, so, it is exceedingly difficult for investigator to manipulate the data and thus the result. But it is possible to alter data unintentionally or knowingly at the level of data analysis by

Table 4: Common effect size indices^{26–28}

<i>Index</i>	<i>Description</i>	<i>Effect size</i>	<i>Comment</i>
Between groups			
Cohen's d	Widely used in meta-analysis Uses mean value and standard deviation of both groups.	Small/trivial 0.2 Medium 0.5 Large 0.8 Very large 1.3	It is useful in deciding sample size based on required effect size/power of the study. For continuous data.
Odds ratio (OR)	Ratio of 2 odds In case-control study effect size will be shown by OR	Small 1.5 Medium 2 Large 3	For binary outcome. RR/OR of 1 means the risk is comparable in both groups.
Relative risk (RR)	Ratio of 2 probabilities	Small 2 Medium 3 Large 4	
Number needed to treat (NNT)	It is reciprocal to absolute risk reduction (ARR). It is the number of subjects expect to treat with intervention/drug A to have one more success compared to that of intervention/drug B.		NNT can be used for binary outcome. Does not consider magnitude of baseline mortality rate It should be interpreted with its comparison arm and depending on context. NNT again labelled as NNT-B or NNT-H based on benefit or harm done by intervention
Measure of association			
Pearson's r correlation	Measures linear correlation between two variables X and Y	Small ± 0.2 Medium ± 0.5 Large ± 0.8	Used for strength of association between 2 variables analysis Value range -1 (perfectly negative correlation) to 1 (perfectly positive correlation)

data management team. It is important to check at this level, as most investigators would not even know if results were altered by data analyst. In simple way, conflict of interest can be divided into non-financial type and financial type. Other classifications are negative conflict of interest and positive conflict of interest. More common is that we concern about positive conflict of interest, but negative conflict of interest is also worth observing. Negative COI happens when any investigator/sponsor willfully rejects/gives injustice to potential useful therapy or intervention, just for his own rivalry or benefit.³⁰

It is also very important to know that conflict of interest is not always bad thing, and sometimes it just happens because of nature of question/core problem not because of individual or sponsor.^{30,31} Most common and best approach to handle conflict of interest is by public reporting of relevant conflicts.

Is bias present in randomized control trial? Bias is defined as systematic error in the results of individual studies or their synthesis. Cochrane Risk of Bias Tool for randomized trials mentioned that bias can happen at 6 different levels/domains: generation of allocation sequence, concealment of allocation sequence, blinding of participants (single blinding) and doctors (double blinding), blinding of data analyst (triple blinding), attrition bias, and publication bias. It is worth noticing that financial conflict of interest is not part of this but, it can be motive behind it.³¹

Is randomized control trial peer reviewed or not? Another important thing about article publication and reliability is whether peer

review done or not. Peer-review is the assessment of article by qualified people before publication. Peer-review helps to improve the quality of article by adding suggestion, and second it rejects the unacceptable poor-quality articles. Most of the reputed journals made their own policy about peer-review. Peer-review is not free of bias. Sometimes, quality of this process depends on selected qualified faculty and their preference on article. Like peer-review, post publication review is also especially important and should not be ignored, as it is criticized/analyzed by hundreds of experts.^{32,33}

CONCLUSION

In nutshell, critical analysis of RCT is all about balancing the strong and weak points of trial based on analyzing main domains such as right question, right population, right study design, right data, and right interpretation. It is also important to note that these demarcations are immensely simplified, and they are interconnected by many paths.

REFERENCES

1. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. *Indian J Sex Transm Dis* 2010;31(1):47–50. DOI: 10.4103/0253-7184.69003.
2. Veldhoen RA, Howes D, Maslove DM. Is mortality a useful primary end point for critical care trials? *Chest*, Elsevier Inc 2020;158(1):206–211.
3. cki. "Endpoints used for Relative Effectiveness Assessment of pharmaceuticals: Clinical Endpoints"-February 2013 GUIDELINE Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints.

4. fda. Clinical Trial Endpoints.
5. Farland LV, Correia KF, Wise LA, Williams PL, Ginsburg ES, Missmer SA. p-values and reproductive health: what can clinical researchers learn from the American Statistical Association? *Hum Reprod* 2016;31(11):2406–2410. DOI: 10.1093/humrep/dew192.
6. Höfler M. Emerging themes in epidemiology the Bradford hill considerations on causality: a counterfactual perspective. 2005;2:11. Available from: <http://www.ete-online.com/content/2/1/11>.
7. Patino CM, Ferreira JC. Internal and external validity: can you apply research study results to your patients? *J Brasileiro de Pneumologia Sociedade Brasileira de Pneumologia e Tisiologia* 2018;44(3):183. DOI: 10.1590/s1806-37562018000000164.
8. Bhalerao S, Kadam P. Sample size calculation. *Int J Ayurveda Res* 2010;1(1):55. DOI: 10.4103/0974-7788.59946.
9. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J, BMJ Publishing Group* 2003;20(5):453–458.
10. Sil A, Kumar P, Kumar R, Das NK. Selection of control, randomization, blinding, and allocation concealment. *Indian Dermatol Online J [Internet]* 2020;10(5):601–605. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31544090>.
11. Khusainova RM, Shilova ZVCO. Selection of appropriate statistical methods for research results processing. *Int Electron J Math Educ* 2016;11(1):303–315.
12. Mishra P, Pandey C, Singh U, Keshri A, Sabaretnam M. Selection of appropriate statistical methods for data analysis. *Ann Card Anaesth [Internet]* 2019;22(3):297. DOI: 10.4103/aca.ACA_248_18 Available from: <http://www.annals.in/text.asp?2019/22/3/297/262097>.
13. Tanniou J, Van Der Tweel I, Teerenstra S, Roes KCB. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. *BMC Med Res Methodol* 2016;16(1):20. DOI: 10.1186/s12874-016-0122-6.
14. Srinivas TR, Ho B, Kang J, Kaplan B. Post hoc analyses: after the facts. *Transplantation, Lippincott Williams and Wilkins* 2015;99(1):17–20. DOI: 10.1097/TP.0000000000000581.
15. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015;351:h5651. DOI: 10.1136/bmj.h5651.
16. Goodman S. A dirty dozen: twelve P-value misconceptions. *Semin Hematol* 2008;45(3):135–140. DOI: 10.1053/j.seminhematol.2008.04.003.
17. Kim J, Bang H. Three common misuses of P values [internet]. *Dental Hypotheses*, vol. 7. Medknow Publications; 2016. pp. 73–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27695640>.
18. Dahiru T. P-value, a true test of statistical significance? a cautionary note. *Ann Ibadan Postgrad Med* 2011;6(1):21. DOI: 10.4314/aipm.v6i1.64038.
19. Jarosz AF, Wiley J. What are the odds? A practical guide to computing and reporting bayes factors. *J Probl Solving [Internet]* 2014;7(1):2–9. Available from: <https://docs.lib.purdue.edu/jps/vol7/iss1/2>.
20. Prel J-B, du, Hommel G, Röhrig B, Blettner M. Confidence interval or P-value? part 4 of a series on evaluation of scientific publications. *Dtsch Aertzblatt Online [Internet]* 2009;106(19):335–339. Available from: <https://www.aerzteblatt.de/10.3238/arztebl.2009.0335>.
21. Hazra A. Using the confidence interval confidently. *J Thorac Dis* 2017;9(10):4125–4130. DOI: 10.21037/jtd.2017.09.14.
22. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surgery, American Medical Association* 2019;154(1):74–79. DOI: 10.1001/jamasurg.2018.4318.
23. Andrade C. The use and limitations of the fragility index in the interpretation of clinical trial findings. *J Clin Psychiatry* 2020;81(2):20f13334. DOI: 10.4088/JCP.20f13334.
24. Carter RE, McKie PM, Storlie CB, Kern PE, The Fragility Index: a P-value in sheep's clothing? [cited 2020 May 14]; Available from: <https://academic.oup.com/eurheartj/article-abstract/38/5/346/2422087>.
25. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Aust Crit Care [Internet]* 2010;23(2):93–97. DOI: 10.1016/j.aucc.2010.03.001 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20347326>.
26. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. *J Grad Med Educ* 2012;4(3):279–282. DOI: 10.4300/JGME-D-12-00156.1.
27. Altman DG. Confidence intervals for the number needed to treat. *BMJ, BMJ Publishing Group* 1998;317(7168):1309–1312. DOI: 10.1136/bmj.317.7168.1309.
28. McGough JJ, Faraone SV. Estimating the size of treatment effects: moving beyond P values. *Psychiatry, Matrix Medical Communications* 2009;6:21–29.
29. Hlatky MA, Owens DK, Sanders GD. Cost-effectiveness as an outcome in randomized clinical trials. *Clinical Trials [Internet]* 2006;3(6):543–551. DOI: 10.1177/1740774506073105 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17170039>.
30. Ghooi R. Conflict of interest in clinical research. *Perspect Clin Res [Internet]* 2015;6(1):10. DOI: 10.4103/2229-3485.148794 Available from: <http://www.picronline.org/text.asp?2015/6/1/10/148794>.
31. Savović J, Akl EA, Hróbjartsson A. Financial conflicts of interest in clinical research. *Intens Care Med, Springer Verlag* 2018;44:1767–1769.
32. Kelly J, Sadeghieh T, Adeli K. Peer review in scientific publications: benefits, critiques, & a survival guide. *EJIFCC [Internet]* 2014;25(3):227–243. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27683470>.
33. Keserlioglu K, Kilicoglu H, ter Riet G. Impact of peer review on discussion of study limitations and strength of claims in randomized trial reports: a before and after study. *Res Integr Peer Rev [Internet]* 2019;4(1):19. DOI: 10.1186/s41073-019-0078-2 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31534784>.