Review article

J Atheroscler Prev Treat. 2020 Sep-Dec;11(3):87–92
doi.org/10.53590/japt.02.1014

How much do we trust randomised clinical trials, and related guidelines for disease management? A critical appraisal

Demosthenes B. Panagiotakos

Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece, Faculty of Health, University of Canberra, Australia

ABSTRACT

In the "*era*" of evidence-based medicine it is now well appreciated that systematic reviews and meta-analyses of randomized clinical trials are more powerful than other designs in their ability to answer research questions regarding the effectiveness of interventions. In this paper a critical appraisal is made regarding the validity of clinical trials and their role in formulating clinical guidelines.

KEY WORDS: *Randomized Clinical trials, evidence based, guidelines*

In the current *era* of evidence-based medicine (EBM) it is well appreciated that some research designs are more powerful than others in their ability to answer research questions regarding the effectiveness of interventions. Figure 1 illustrates the present belief concerning the hierarchy of studies in terms of their influence on the level of evidence for disease treatment and management. This grading is a framework for ranking evidence of health care interventions and designates the weight should be given to studies when evaluating the same research question. As we can see, on the top are the systematic reviews and meta-analyses of randomised clinical trials (RCT). A clinical trial is a prospective, human study that aims to evaluate the effect and value of one or more intervention(s) against a control, which usually is the current therapeutic approach or an already existed one, and is considered as the "gold-standard" method for evaluating the effectiveness of interventions. If the allocation of the participants to the intervention(s) is made randomly (e.g., "*flip coin*" or, usually, with much more sophisticated approaches) they are called randomised, and it is made to prevent selection bias by distributing the characteristics of the participants - that may influence the outcome – by chance between the groups. By this way, any difference observed in outcome can be attributed only to the intervention. In general, RCT are very good for hypothesis generation and or hypothesis testing.[1] In this paper a critical appraisal is made regarding the validity of clinical trials and their role in formulating clinical guidelines.

Corresponding author:

Prof. Demosthenes B. Panagiotakos, DrMed, FRSPH, FACE
School of Health Science and Education
Department of Nutrition and Dietetics
Harokopio University, Athens, Greece
E-mail: d.b.panagiotakos@usa.net

A Critical Appraisal of Randomised Clinical Trials

As a part of a critical reading of an article presenting and discussing the findings of an RCT, the answers to few questions will help to decide whether the results are trustable and

Submission: 05.07.2020, Acceptance: 05.07.2020

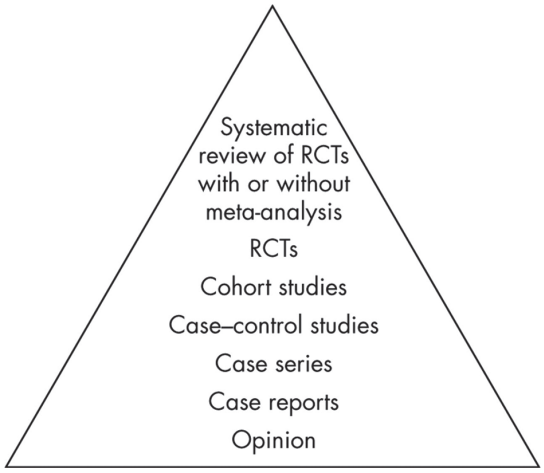

FIGURE 1. The hierarchy of studies regarding their level of evidence, under the context of evidence-based medicine. RCT: randomised clinical trials.

can be applied to "our" patients. Specifically, the validity of the trial's design and methodology should, firstly, be examined. This includes, the design of the study (parallel or cross-over), the randomization and blindness procedures, and the sample size achieved and related power analysis. A second issue that has to be examined is the magnitude and precision of the treatment effect on the a-priori defined endpoint(s), and finally, the applicability of the results to "our" patient or population. The later has to do with the generalization of the RCT findings, which is closely related to the sampling procedures followed on and the characteristics of the tested sample in terms of its representativeness.

To help the reader understand whether can trust the results and conclusions of a RCT, and consequently the guidelines derived based on this evidence, the following questions could be used for a critical appraisal:

- The research hypothesis (question) is clearly stated? (focus)
- The participants were appropriately and randomly allocated to the study's groups? (randomization, avoid selection bias)
- Were participants, and study's personnel blind to the intervention(s) group(s)? (blindness, avoid bias)
- Were all the inclusion and exclusion criteria of the participants clearly defined and rationally supported? (avoid selection / inclusion bias)
- Were participants followed up in the same way? (equity)
- Was the number of participants enough to minimise the role of chance, and to help deriving robust conclusions? (statistical power)
- Are the results presented in a precise way, and the conclusions clearly derived from them?
- Are all important consequences, and side-effects considered, and can the results be applied to our population?

As it was stated above, it is of major importance the research question(s) to be clearly stated. The question should be focused on the problem of interest and should be framed in such a way that even someone who is not specialist in the field would be able to understand them.

Regarding randomisation, simple or more sophisticated (e.g., cluster-, stratified-randomization, etc), it refers to the principle that each participant has an equal probability of being assigned to any given intervention group. By this way, selection bias is eliminated, and a balance of potential known and unknown confounding factors between groups is achieved. While randomisation is performed to eliminate selection bias, it cannot be guaranteed that the study's groups will be similar regarding important participants' characteristics, especially when smaller studies are considered. A way to warrant that the groups will be, as possible as it can be, identical, is to perform stratified block sampling, which, although, may considered as a violation of or the "random" principle, it ensures the equity in participants' characteristics between groups.

However, despite the benefits from randomization, there is always a risk in RCT that a-priori perceptions about the advantages of a specific intervention, - especially when industry funding exists-, might influence the outcome(s), leading to biased results, particularly when subjective outcome measures are used. To control for this source of bias, blinding (or also called masking) of participants (single-) and study's staff (double-) regarding the intervention given, is a solution. Recent methodological studies have shown that blinding of patients and staff prevents bias, since RCT that were not blinded yielded to larger treatment effects estimates as compared to trials in which authors reported double blinding.[2]

Inclusion and exclusion criteria used in the design of a RCT, and a medical study in general, are of crucial importance for the generalization of the results. In particular, the profile of participants entered in a RCT, frames the general population considered, and the population that the study's findings can be applied. In other words, answer to the crucial question, "*does study's findings apply to my patient*?" For example, imagine an RCT performed only in men, or in people < 60 years, or in people without diabetes, etc, it should be clear that the research findings cannot be applied to women, or older people or to diabetic. A related issue that has been discussed several times in the past, is the under-representativeness of women and older people in RCTs, whereas, the interventions tested could also be applied for these population groups, too.

The issue of statistical power, - not only for RCTs, but also for any experimental and observational study - is of major importance, since it reflects the ability of the study to detect an association when such an association truly exists in the (unmeasured) population. The statistical power of an RCT is mainly determined by the known or expected prevalence of the studied outcome in the population, the duration of the study, the magnitude of the expected effect, the design, and the sample size of the study. When the size of an RCT is inadequate (i.e., too "small"), it is very difficult to detect any true differences in the outcome(s) and make robust conclusions. The presentation of study's power, and the description of the related sample size calculations plays an important role in the robustness of the results. However, small-, inadequately sized studies have been frequently, and continue to be, published. Scientists should ignore their results, since there will be a high probability of being observed "by chance". Of the most important key factors for the inclusion of an RCT into a clinical guidelines document is the adequacy, in terms of statistical power, of its sample.

Another key factor is the statistical significance, which refers to the likelihood that the results obtained in a study were not due to chance. Probability of Type I error (also known as *p*- value) and confidence intervals are used in medical research to describe "significance". The choice of a level of "significance" in medical research is subjective; however, for many years, by convention, researchers use a threshold of 0.05, which means that for values less than 0.05, the observed association is a% unlikely to occurred by "chance". An issue that should be discussed now is that, although statistical power favourably increases as the sample size increases, too, *p*-value is inversely associated with sample size (Figure 2). The later means that by increasing the size of a study, the *p*-value is artificially reduced; a fact that may lead to miss-interpretation (or over-interpretation) of study's findings. The past years there is a tendency to present (1-a)% confidence intervals (usually 95%) of the effect size

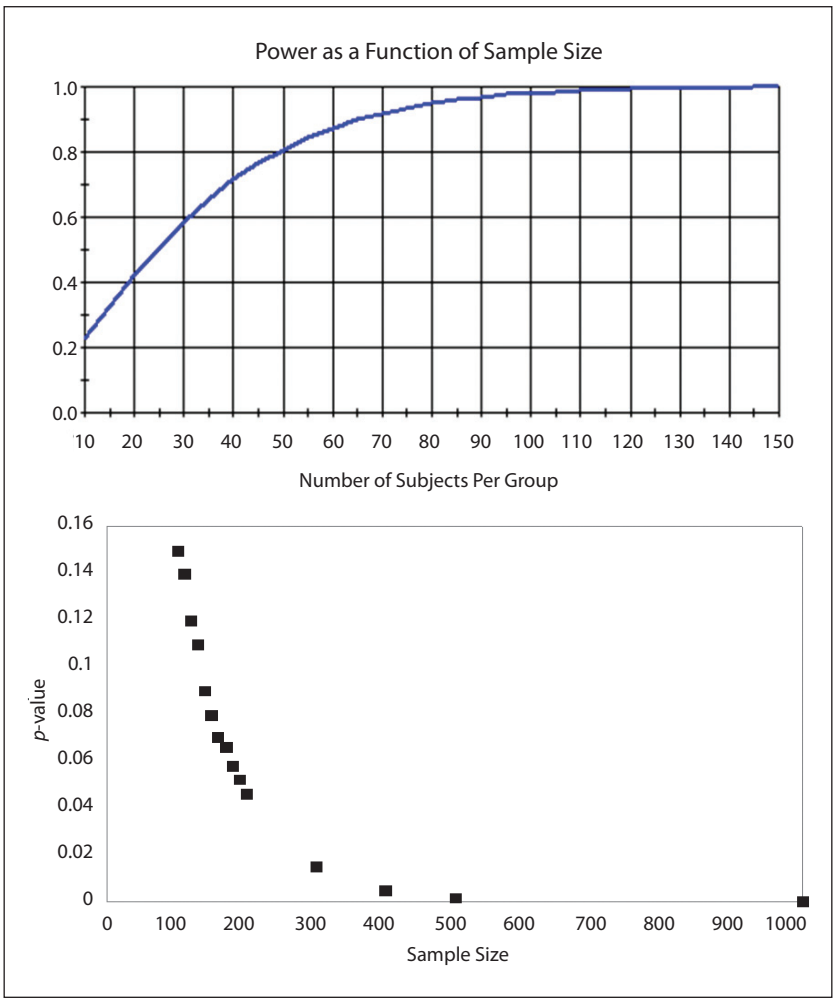

FIGURE 2. Upper plot illustrates the change in statistical power of a specific effect size Δf (e.g., treatmnet difference) as the sample size increases; bottom plot illustrates the reduction in the *p*-value of Δf, as the sample size changes.

measures, instead or in addition to presenting *p*-values. The (1-a)% confidence interval of the intervention's effect size estimate (e.g., mean difference, odds ratio, hazard or risk ratio) is the interval within which by (1-a)% "certainty" (e.g., 95%) the true population effect will lie. It is true that confidence intervals convey more useful information than *p*-values, since they provide information about how precise the measured intervention effect is. The width of the confidence interval designates the precision of the estimate. The wider the interval, the less the precision. A very wide interval may indicate that more data should be collected before anything definite can be decided. However, as in. the case of *p*-value, larger samples lead to narrower intervals.

The past years it has been strongly suggested that to reduce bias, RCT's data should be analysed on an "intention-to-treat" (ITT) basis. It is a fact that the validity of RCT findings largely depends on the randomisation process. However, practice has shown that after the random allocation of the participants to intervention groups, it is almost unavoidable that some of them would not complete the study (dropouts) or change groups, for various reasons. If such participants were excluded from the analysis, the randomization would be violated, and baseline factors in the intervention groups will not anymore be similar. Thus, the findings would not be solely attributed to the intervention but could also be due to confounding. ITT analysis is a strategy which ensures that all participants allocated to the intervention groups will be analysed as they were randomly allocated, regardless whether or not they received the prescribed intervention or completed the study.[4] In few words, ITT analysis ensures clinical reality into research, since it recognises that for various reasons, not all people in the "real" world will receive the intended treatment or complete the study.

Despite all the aforementioned statistical and methodological considerations, in order to decide whether RCT's findings are of value, clinical significance plays also an important role. A "statistically" significant finding may have very little influence in daily clinical practice. Clinical significance reflects the value of the RCT's results to the "everyday" patients and it is important in clinical decision making, regardless of whether the differences were "statistically" significant or not. Judgements about clinical significance of an RCT should be taken into consideration, and particularly, how the intervention benefits (measured e.g., in life-years gained, number-needed to treat, quality of life, etc) and any adverse events, are valued by the patient.[3] Moreover, based on the results of RCTs, clinicians will make decisions not only about the validity of the findings, but also whether they are applicable to their patients. Issues that one needs to consider before choosing to incorporate research evidence into clinical practice, concerns similarities in patient's populations, if side effects outweigh the observed benefits, and the availability and cost of treatments to their local setting. For example, if a specific intervention has found to be effective in hyperlipidemic patients in the USA, it has to be explored whether there is any biological, geographical, behavioural or cultural reason why that particular intervention is also effective in hyperlipidemic patients in Greece. Moreover, it should be considered whether the reported by the RCT side effects may compensate the potential benefits for the patients. In addition, treatment cost and availability are always a barrier in spreading the RCT evidence to the world. There will be no point in prescribing a treatment which cannot either be obtained, or which local hospital or practice is not in a position to support.

The issues discussed regarding the design, analysis and interpretation of an RCT's findings are the main pillars in an attempt to perform a critical appraisal of a RCT. However, another important issue is the role of funding of RCTs. Several papers have dealt with this issue, leading to conflicting results regarding the role and the sources of funding, and particularly from the industry, on trials' outcomes. Although it is not the purpose of this paper, it should be underlined that all sources of research funding should be clearly stated, together with any conflict of interests by the authors.

How much physicians trust guidelines based on RCTs?

Physicians always want their decisions to be based on state-of-the-art evidence. The variations in medical decision making observed in the past years among different countries, regions, and even within hospitals, have led to a more systematic approach to finding the most appropriate strategy for the patient. This approach is the evidence-based medicine (EBM) that is based on systematic reviews of the literature and meta-analyses (when available) of relevant studies, preferably RCTs, and follows a series of steps to gather sufficiently useful information for the unique patient. In the context of EBM, clinical guidelines have become common in the practice of medicine nowadays (Figure 3). Guidelines are usually based on large panels of physicians, health scientists and other professional organizations, that based on the available level of evidence aim to develop a structured roadmap of daily clinical practice. The vast majority of specialty medical societies have published such guidelines, others based on simple if-then-else steps, and others being more complex, following multistep rules that are formalized using algorithms. However, it should be noted that clinical guidelines do not take into account the degree of uncertainty inherent in RCT results, the likelihood of

Classes of recommendations	Definition	Suggested wording to use
Class I	Strong evidence and/or general consensus that a given intervention is beneficial, useful, and effective	Is recommended
Class II	Moderate and conflicting evidence and/or a divergence of opinion about the efficacy of the given intervention	
Class IIa	*Weight of evidence is in favour of efficacy*	Should be considered
Class IIb	*Efficacy is less well established by the current evidence*	May be considered
Class III	Low level of evidence or general consensus regarding the given intervention; the intervention is not effective, and in some cases may be harmful	Is not recommended

FIGURE 3. Classes of recommendations and levels of action usually used in clinical guidelines. Colours from green to orange to red illustrate the level of uncertainty in decision making

intervention's success, and the hazards and benefits of each sequence of action.

However, how much do we trust clinical guidelines? In an editorial paper published in *BMJ* few years ago, Jeanne Lenzer raised the issue "Why we can't trust clinical guidelines".[4] It was supported that despite that it seems difficult to bias clinical guidelines when having so many experts participating under the sponsorship of large professional bodies, there is a number of cases suggesting that this may be common.[5] As reported by Edwin Gale, despite the calls to prohibit or limit conflicts of interests among guideline panelists, most guideline authors have conflicts, making the guidelines less than reliable.[6] For example, in the 2012 released guidelines of the American Heart Association / American Stroke Association notes that " … *every effort has been made to avoid any actual or potential conflicts of interest that may arise as a result of… a business interest of a member of the writing panel …*", according to the conflict of interest disclosures, 13 of the 15 authors had connections with the stroke related industry and 11 had links to companies.[7] Moreover, many other conflicted clinical guidelines have come to light in the past years in a variety of medical disciplines.

Despite all these considerations, the countless advances in medical research, and health services have reduced the level of uncertainty in clinical guidelines when applied in practice. Physicians tend to trust guidelines, even if they may say " … *we stick within the standard of care, because when something goes wrong, want to be able to say: we were just doing what everyone else is doing, even if this was not very good*" In a survey conducted during the 13th Summer School of the Hellenic Atherosclerosis Society (www.atherosclerosis.gr) on July 2020, 648 male and female physicians and other health care scientists were asked about their opinion regarding the validity of published RCT and guidelines for atherosclerotic disease treatment and management. It was observed that 70% reported that they trust published clinical trials, and 75% that they found them useful in their daily clinical practice. However, 72%, reported they have some moderate to strong concerns whether the authors of clinical guidelines are unbiased. Taking all the aforementioned considerations into account, and in order to increase the trust of clinical guidelines, efforts should be directed in increasing the validity, the reproducibility and the transparency of clinical trials. It may be also necessary to consider a wider range of research approaches, including the integration of "real" world studies, that can provide converging evidence on intervention effects.

CONCLUSIONS

An RCT is the most rigorous scientific method for evaluating the effectiveness of health care interventions. However, bias could occur when there are defects in the design and analysis of a trial. It is important for people reading medical reports of RCT to develop skills for critically appraising RCTs, including the ability to assess the validity of trial's methodology, the magnitude and precision of the treatment effect, and the applicability of results into a broader population.

Conflict of interest

None to declare.

ΠΕΡΙΛΗΨΗ

Πόσο εμπιστευόμαστε τις τυχαιοποιημένες κλινικές δοκιμές και τις κατευθυντήριες οδηγίες; Μια κριτική αξιολόγηση

Δημοσθένης Παναγιωτάκος

Τμήμα Επιστήμης Διαιτολογίας-Διατροφής, Σχολή Επιστημών Υγείας και Αγωγής, Χαροκόπειο Πανεπιστήμιο, Αθήνα, Ελλάδα, Σχολή Υγείας, Πανεπιστήμιο Καμπέρα, Αυστραλία

Στην «εποχή» της ιατρικής των ενδείξεων, οι συστηματικές ανασκοπήσεις και οι μετα-αναλύσεις των τυχαιοποιημένων κλινικών δοκιμών είναι από τα πιο ισχυρά εργαλεία στην ικανότητά τους να απαντούν σε ερευνητικά ερωτήματα σχετικά με την αποτελεσματικότητα των παρεμβάσεων. Σε αυτό το άρθρο γίνεται μια κριτική αξιολόγηση σχετικά με την εγκυρότητα των κλινικών δοκιμών και το ρόλο τους στη διαμόρφωση κλινικών οδηγιών.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: *Τυχαιοποιημένες κλινικές δοκιμές, ενδείξεις, κατευθυντήριες οδηγίες*

REFERENCES

1. McGovern DPB. Randomized controlled trials. In: McGovern DPB, Valori RM, Summerskill WSM, editors. Key topics in evidence based medicine. Oxford: BIOS Scientific Publishers; c2001. p. 26-9.
2. Day SJ, Altman DG. Blinding in clinical trials and other studies. BMJ 2000 Aug;321(7259):504.
3. Coggon D. Statistics in clinical practice. London: BMJ Publishing Group; c1995.
4. Lenzer J. Why we can't trust clinical guidelines. BMJ. 2013 Jun;346:f3830
5. Ransohoff DF, Pignone M, Sox HC. How to decide whether a clinical practice guideline is trustworthy. JAMA. 2013 Jan;309(2):139-40.
6. Gale EA. Conflicts of interest in guideline panel members. BMJ. 2011 Oct;343:d5728.
7. Jauch EC, Saver JL, Adams HP Jr, Bruno A, Connors JJ, Demaerschalk BM, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013; 44:870-947.