Evidence-Based Urology
How Does a Randomized Clinical Trial Achieve Its Designed Goals?

Homayoun Sadeghi Bazargani,[1] Sakineh Hajebrahimi[2]

Purpose: To discuss the methodological considerations of a standard and applicable randomized clinical trial (RCT).

Materials and Methods: Using a predefined strategy, we conducted systematic computerized search of the MEDLINE (1966 to 2011) and EMBASE (1980 to 2011) databases to identify all English language educational articles discussing the RCT methodological aspects. Full text versions of identified studies were reviewed in blinded fashion for key methodological and statistical characteristics.

Results: Randomized clinical trials in surgery are the highest level of the primary research evidence in evidence-based medicine. There is increasing demand for implementation of RCTs in urological daily practice.

Conclusion: Randomized clinical trials' report should be absolutely clear, simple, and easy to understand with well-defined internal and external validity. Efforts should be made to design high quality RCTs in urology. There are substantial needs for urologists to their knowledge about RCT.

[1]Rehabilitation & Physical Medicine Research Center, Department of Statistics and Epidemiology, Faculty of Health and Nutrition, Tabriz University of Medical Sciences, Tabriz, Iran
[2]International Evidence Based Urology Working Group, Iranian Center for Evidence Based Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

Corresponding Author:
Sakineh Hajebrahimi, MD
International Evidence Based Urology Working Group, Iranian Center for Evidence Based Medicine, Tabriz University of Medical Sciences, Tabriz, Iran
Tel: +98 411 336 7373
Fax: +98 411 335 7328
E-mail: hajebrahimis@gmail.com

Approximately 1000 years earlier, Avicenna (an Iranian physician) wrote one of his "Canon of Medicine" book's chapters entitled "The recognition of strengths of the medicines characteristics through experimentation". Avicenna had the first known treatise on clinical trials. Almost 830 years later, Fisher did his famous formal randomized clinical trial (RCT).[1]

Randomized clinical trial is a research method, in which the participants are assigned either as intervention or control groups to compare the study results. Randomized clinical trials make the foundation of systematic reviews, evidence-based practice guidelines, and health technology assessment in clinical practice. Therefore, appropriate reporting of RCTs'

results in published articles is very crucial. Consolidated standards of reporting trials (CONSORT) was initially recommended in 1996, which focused on sample size, randomization, allocation concealment, blinding, statistical analysis, primary and secondary (adverse events) outcomes, and overall generality of the evidence.[2]

Although RCTs stay at the top of evidence hierarchy pyramid after systematic reviews, there is a big lack of evidence yet. Of 4856 published articles in four leading urology journals from 1996 to 2004, only 4% were RCTs, of which only 1% was a surgical RCT.[3]

According to CONSORT statement, the quality of RCTs has improved,[3] but still many

fundamental problems may convert the gold standard position of RCTs to bronze situates. Furthermore, the strength of RCTs depends on internal, external, and social validity of the study design. Most of the published evidence of critically analyses of RCTs have shown that results of RCTs might be invalid. Randomization, allocation concealment, blinding, clear statistical method of analysis (intention to treat), and follow-up period are unreported in most of the studies.[3,4]

Our aim was to demonstrate how an RCT achieves its designed goals. This study also discusses limitations of surgical RCTs and suggests some solutions.

Let's start with a short clinical scenario

A 75-year-old man comes to your office with severe lower urinary tract symptoms (LUTS) and recurrent urinary retention. Diagnostic studies confirm outlet obstruction due to the prostate enlargement. Your resident asks about effectiveness of green light laser in treatment of this old man. Therefore, you intend to search for high quality evidence to support the best intervention. Transurethral prostatectomy is considered as the gold standard surgical treatment for benign prostatic hyperplasia.[2]

To our knowledge, a gold standard means an intervention or a test with highest effectiveness and reasonable cost, which has been demonstrated in some valid and relevant RCTs. On the other hand, an RCT is considered as the gold standard study for casualty for an interventional research question. In following sections, we are going to explain what makes an RCT acceptable.

Are the results of the trial valid? (Internal Validity)

What question did the study ask?

To provide reliable evidence regarding a research question in any RCT, it is quite crucial to have relevant reasonable objectives and to prioritize them based on clinical importance and reliability of the expected emerging evidence from each objective. Other than the overall aim of the study, the specific objectives defined in clinical trial studies do not stand at the same level of importance. It is a quite consistent rule in designing middle-phase clinical trials to consider a single primary objective and possibly several secondary ones.

The primary objective of an accurate clinical trial is based on estimating a clinically important outcome as objectively as possible. However, it is critical to know that the researcher should not trade off some important and relevant patient reported outcomes (PROs) for some other less important ones, just due to lower objectivity in PROs measures.[5] Based on a relevant hypothesis for primary objective of the study, a primary efficacy variable is defined. It is also helpful to indicate the relevant endpoints.

The secondary specific objectives are usually defined for outcomes of less clinical importance, less objective measurement, and those less expected to be affected by the intervention or drug.

In designing of an RCT, PICO must be considered: Patients or Population, Intervention, Comparison, and Outcomes. In the aforementioned scenario, Patient (P): Old man with severe LUTS and urinary retention, Intervention (I): Green light laser prostate resection, Comparison (C): Transurethral resection of the prostate (TURP), Outcome (O): symptoms and quality of life improvement.

The sample size estimation is made to fulfill adequate statistical power to test the hypotheses correspondent with primary objective of the clinical trial. However, secondary objectives may passively benefit from power estimation for the primary objective.

Treatment allocation designs

There are several treatment allocation designs in RCTs, including parallel design, cross-over design, and sequential design, etc. Here, we will only discuss the parallel and cross-over designs.

The most popular design in RCTs is the parallel design. Methodologically, design, analysis, and interpretation of an RCT with parallel design appears to be easier than other types.[6] In parallel

design, subjects in two or more groups (trial arms) are followed up in parallel to compare study outcomes. In this design, each participant is assigned to a single treatment strategy.

Most RCTs consist of two arms. One arm is usually investigating a new treatment strategy, which can be called as test treatment or investigational treatment. The investigational treatment strategy may vary to be a new drug, a different dose of a drug, a different form of a drug, a surgical method, a medical device, an educational plan, a rehabilitation program, other types of interventions, or a combination of interventions.

The subjects in second treatment arm may be assigned to an active treatment, which can be a standard treatment or a treatment with previous evidence to have some efficacy in treating the disease or condition of interest. Alternatively, the subjects in the second treatment arm may be assigned to receive placebo or even no treatment. However, due to ethical obligations, active treatment is used in many clinical trials. This is of importance for studies on a serious condition for which at least one active treatment is available and known to have some benefits.[6] There may be also a treatment in common use or a traditional remedy.

If the new treatment strategy is not compared to the active treatment, the question is raised whether to use placebo or no treatment. Placebo seems to be vital if the clinical trial is going to be blinded. The second reason for prioritizing placebo to no treatment is the placebo effect. Wikipedia defines placebo effect as follows: "Sometimes patients given a placebo treatment will have a perceived or actual improvement in a medical condition, a phenomenon commonly called the placebo effect". Although there is some consistency regarding existence of placebo effect in terms of some subjective measurements, this may not be true for many other outcomes.[7-11] Henry K. Beecher was possibly the first scientist who quantified the placebo effect in 1955. Nevertheless, later analysis of the data used by him showed that, contrary to his claim, no evidence of any placebo effect was found in any of the studies cited by him.[12] Another fact to be

considered is the possible risk in using placebos, which may be true in using sham surgery or injectable placebos.[13-18] Transurethral resection of the prostate is a gold standard procedure. In this scenario, a patient who clinically needs a surgery (by consecration of ethical issues) may be put on green light laser of TURP. A sham surgery as a control arm of surgical trials will be acceptable if all ethical issues followed well.

In a cross-over design, each subject receives more than one treatment strategy and the order of receiving each treatment is randomized.[6] No doubt, a washout period should be defined between crossing the treatments, based on presumed effect decay rate for the treatments, to prevent additive or interactive effects of two consecutive treatments on the study outcome. Each subject in a cross-over design serves its own control, which helps in decreasing the noise and confounding. Therefore, smaller sample size is needed for a cross-over RCT compared to a parallel design in similar conditions. On the other hand, if a cross-over design is applied, appropriate statistical methods should be used to properly manage the correlated nature of data. Other than the limitations in defining a washout period, one major disadvantage of cross-over design is that it may not be practically used in acute conditions or if the outcomes occur only once. Furthermore, care should be taken of the diseases like multiple sclerosis having recurrent periods of exacerbations and remissions by nature. Millar introduced another drawback of cross-over design where treatment effects become distorted by and confounded with their order of administration, and proposed its prevention.[19]

Was the assignment of patients to treatments *randomized*?

Randomization is a process through which study subjects are assigned to different trial treatments only by chance. Control of unknown confounders is the popular, but not the sole advantage of randomization.[20] Method of randomization is dependent on sample size, end points, confounding, and prognostic factors.

There are several types of randomization in clinical trials.[21] Two types of randomization

well-known by the researchers are simple and blocked randomization. In simple randomization, subjects are assigned to treatment strategies by chance. The probability for any eligible nominee would be nonzero without any restriction. It is an easy to do method using random tables even if computer programs are not available. The main drawback in such a randomization strategy is the lack of guarantee for number of subjects in each group to be equal or follow a predefined proportion. If the sample size is large enough, this is not a major concern, but in case of a small sample size, this may lead to some problems. Other than design limitations, it may lead to loss of statistical power.

Blocked randomization, another popular method among researchers, makes it possible to assign subjects to either equal or predefined size of blocks and trial arms. Other than what we discussed about equal-sized study groups in simple randomization, blocked randomization has the possibility of a blinded or open label analysis at the end of each block. This may help re-estimate sample size early through the study when the information used for sample size calculation prior to start of the clinical trial appears to be doubtful. Another advantage in blocked randomization is that if, for any reason, the RCT is stopped before achieving the full enrollment, higher power of study may be reached in case of blocked randomization. Nevertheless, one major concern in blocked randomization is allocation predictability in some subjects.[22,23]

One acceptable alternative for simple and blocked randomization in case of small sample size and concern on confounding can be the minimization method. Although minimization may not be defined as a pure randomization, but has proven to yield reliable results.[24-26]

Random assignment of participants to study or control group produces comparable groups at the end of the study and ensures that the mere difference between groups is due to intervention. It means each group is a random sample of eligible study subjects; hence, both are representative of that population. Equalization of the numbers in the groups is not enough in our scenario; one includes number of patients with moderate LUTS and other includes same number of severe LUTS, or younger people in one group versus older ones in the other group. In such situations, stratified recruitment with respect to severity of disease and age must be done.

Were the groups *similar* at the beginning of the trial?

After an appropriate randomization, we need to separate the person who generates allocation from those who accesses eligibility. In other words, allocation should be concealed by using third party schemes, including pharmacy randomization, telephone randomization service, web-based service, or sealed and opaque envelopes. Allocation concealment must be done in patient selection phase, but blinding is in process phase and for intervention.

Were measures *objective* or were the patients and clinicians "*blinded*" to the administered treatment?

Blinding is a key point in many RCTs to reduce information or ascertainment bias. If study subjects are not blinded, knowing which group they are assigned to, may affect their responses to the received intervention. Possibly, knowing that they have been assigned to a group who will receive a new treatment may lead to favorable expectations or anxiety. Blinding those involved in conducting the research, including investigators, physicians, patient enrollers, randomization implementers, health-care providers, and routine data collectors, is also important.[27] Blinding turns vital if the outcome of interest is more subjective while its necessity decreases for more objective outcomes.

Although blinding is a familiar word among clinical researchers, there seems to be some confusion in understanding the terminology of blinding, such as single-, double- and triple-blind, masking, and allocation concealment.[28]

Schulz and colleagues state: *Blinding (masking) indicates that knowledge of the intervention assignments is hidden from participants, trial investigators, or assessors. While, non-blinded (open or open label) denotes trials in which*

everyone involved knows who has received which interventions throughout the trial.[28]

Single-blind usually means that participants are blind to the treatment type and stay blind throughout the study period. In a double-blind trial, study subjects, investigators, or assessors usually remain unaware of the intervention assignments throughout the study.[29] In a triple-blind RCT, the statistician is also blind to the assignment type. Some authors have also used triple-blind terminology instead of double-blind when the assessors and investigators have been separate. In such a scenario, if the statistician is also blinded, authors may get persuaded to use quadruple blinding and some have also dared to define quintuple blinding.[27] We think that more important than the terminology used in reporting clinical trials is to clearly explain how the blinding is done in the trial and whether the blinding process remained perfect or not. As discussed earlier, use of placebos compared to no-treatment strategy has at least advantage of making the blinding possible.

In clinical research fields, there are situations in which drugs cannot be formulated in a way to ensure similar galenical forms in trial arms. For example, a tablet form of a new drug needs to be compared with another form of an active treatment, eg, capsule or topical ointment. This limitation prevents a simple blinding. In such a situation, a technique called as "double-dummy technique" may be used. A placebo is produced similar to the drug in investigational group and is added to the treatment protocol in active control group. Vice versa, a placebo is produced similar to active treatment and will be added to treatment protocol in investigational group. This will help do the blinding, but the number of tablets for instance is increased, reducing compliance of patients.[30]

One last note we would like to add is that blinding itself is not a golden guarantee for the RCT. Thus, a well-designed RCT with relevant methodology should not be disqualified due to lack of blinding. Application of blinding is after allocation and in procedural phase. Therefore, to reduce the emotional effects of the studies, patients have to be blinded to their interventions. However, in most of the surgical trials, blinding

of surgeon is impossible. In this situation, outcome assessor should be an independent and blind investigator. Therefore, the term of double-blind in surgical trials is meant as blind patient and outcome assessor.

What were the results?

It is not uncommon to read an RCT with strong conclusions on efficacy of a new treatment, but using only a Chi-square test performed as a statistical method resulting in a P value less than .05. This is not the sole pitfall in statistical methodology of published clinical trials and many other examples can easily be found in literature. In this study, we only focus on two statistical considerations crucial to RCTs.

How large was the treatment effect?

When a difference in primary outcome of an RCT is observed, the first question will be, "How likely the observed difference is to be by chance?" This can be easily answered using an appropriate statistical test.

Suppose you are comparing the efficacy of two different surgical procedures (A and B) in treating vesicoureteral reflux and find out that of 80 patients in group A, 40 gained successful treatment, while 20 out of 80 patients in group B achieved successful results. The descriptive statistics are indicative of a difference in success rate between the treatments. As we know, this difference is only observed in our sample and we do not know how likely it would occur in reality in a larger population.

This is a random error term which may be understood by P value, but can be derived using a statistical test. The most common statistical test in this situation would be a Chi-square test that gives us a P value equal to .001. This means that we have found an association, which is less likely to be due to chance. We call this "assessing randomness of association". A major flaw is to stop the analysis here and make conclusions only based on these results. The Chi-square test gives us a measure of randomness of association, but we may prefer to have a static for strength of association. Clinicians prefer to choose a

procedure with a clear effectiveness compared to placebo. Sometimes a statistically significant outcome is not clinically important and the results may not be applicable.

What was the measure?

Several measures of strength of association have been developed; three most important of which are discussed here.

A: Relative risk (RR): In epidemiology, RR is the risk of developing a disease relative to exposure. It is a ratio of risk in one group over the risk in another group. In clinical trials, we may consider it as probability of success for investigational treatment over probability of success for the comparison treatment strategy. A relative risk of 1 means there is no difference in risk between the two groups or it means no difference between treatment strategies in clinical trials. Relative risk greater than one suggests higher efficacy of investigational treatment. In preventive clinical trials, like vaccine research, RR is usually expected to be lower than one for the investigational intervention or vaccine.

B: Risk difference (RD): Contrary to RR which is a ratio, RD is an absolute risk measure which is obtained by subtracting risk in one group from the risk in second group. No doubt, many clinicians may be interested in studying the absolute difference in success rates rather than relative success rates when comparing efficacy of treatment strategies.

C: Number needed to treat (NNT): The number of subjects need to be treated ensuring one subject to benefit compared with a control in a clinical trial. For example, if NNT = 3, it means that if three patients get the treatment, one of them will benefit from that treatment compared to control. The larger the NNT, the lower the effectiveness will be. The best NNT is considered

to be 1, where everyone achieves success with investigational treatment and no one with control. In our scenario, if the NNT of green light laser versus TURP is 2, it means the number of patients need to be treated ensuring one subject symptom improvement just because of green light is 2, and if we treat two patients with green light, one is going to have expected outcome.

Contrary to RR and RD, NNT is a measure of effectiveness rather than efficacy, making it more attractive for clinicians and health technology policy makers. Epidemiologists may be more interested in RR, but NNT can be an easy to understand and more beneficial index for clinicians. Number needed to treat is calculated by inversing the RD. As NNT is derived from RD, we recommend the researchers to report RR and NNT in their reports.

It should be taken into account that it is not sufficient to calculate the point estimates of RR, RD, and NNT. We should have an idea how precise the calculated RR, RD, and NNT are. If we have 10 times larger sample size, but with the same response proportions, RR and NNT would also be the same while the results of a larger study can be more reliable. The solution is to estimate some confidence interval (CI) measures of RR and NNT as well. Statistical software packages easily provide you with required statistics. Table demonstrates the calculated statistics for the two examples.

Were all the patients who entered the trial accounted for? Were they analyzed in the groups to which they were randomized?

Non-adherence may be an inevitable part of many clinical trials, especially the effectiveness trials, trials with long-term treatment, and when the treatments used are more likely to have adverse-effects. Follow-up period should be long enough

Two examples of calculating RR, RD, and NNT

	Group	Success	Failure	Total	RR	RD	NNT
Example 1	A	40	40	80	RR = 2	RD = 0.25	NNT = 4
	B	20	60	80	95%CI: 1.3 to 3.1	95%CI: 0.1 to 0.4	95%CI: 2.5 to 10.0
Example 2	A	400	400	800	RR = 2	RD = 0.25	NNT = 4
	B	200	600	800	95%CI: 1.7 to 2.3	95%CI: 0.2 to 0.3	95%CI: 3.3 to 8.0

RR indicates relative risk; RD, risk difference; NNT, number needed to treat; and 95%CI, 95% confidence interval.

in all RCTs; however, there is no ideal definition. On the other hand, all subjects in both groups should be followed up until the end of the study. Therefore, the follow-up period has to be long enough and complete.

Suppose the researcher is going to test the effect of tolterodine on sexual function in women with overactive bladder. The drug is hypothesized to improve sexual function independently or through improving overactive bladder.[31] The patients need to be followed up for several months. Imagine you do an RCT to compare this drug with another new treatment. Despite the researchers wish, the situation may be such that few patients after randomization might discontinue the administered drug or shift to the comparison treatment to which they are not assigned to. Using drugs different from what were allocated during the randomization violates the principle of randomization and may introduce confounding. Several methods are proposed to handle this problem.[32]

    1- "Intent (ion) to treat" analysis: This approach is the most common method to handle such a problem. The non-adherence is ignored and participants are compared through the analysis based on early randomization results. This method well resolves the problem of confounding due to violation of randomization. However, effect size underestimation is the main limitation of this method.[32] Actually, intent to treat analysis can be considered as a measure of effectiveness rather than efficacy.

    2- "As treated analysis": The analysis is based on the actual treatment received by the patient ignoring the randomization. No doubt, confounding variables associated both with adherence and outcome will be a major issue in this method. Measuring such confounders and controlling them, through possibly multivariate analysis, will be a necessity in this regard.

    3- "Per-protocol" analysis: In this method, non-adheres are eliminated from the analysis. This method may introduce confounding effect more than "as treated" analysis.

There may be some instances that non-adherence is due to satisfaction with the treatment and resolution of the main problem in shorter time than expected. Therefore, the patient may stop the treatment and not continue with the study. Using a per-protocol analysis will lead to underestimation of effect size or loss of statistical power of the study. Adalatkhah and colleagues performed a dermatological RCT on moderate acne comparing two drugs. They found that the time to improvement is shorter for new drug and checking multiple measurements showed that some of those who received the new drug felt treated and stopped taking more tablets. A per-protocol analysis or pessimistic permutation of missing data in such a situation may reasonably underestimate the efficacy of new drug. Therefore, a new terminology as "logical intent to treat analysis" to prevent this problem has been presented.[32]

How precise was the estimate of the treatment effect?

Although *P* value could show statistical differences between two groups, but for the size and importance of difference, CI is crucial. The true risk of the outcome in the population is not known and the best we can do is to estimate the true risk based on the sample of patients in the trial. This estimate is called "the point estimate". By looking at CI, we could know how close this estimate is to the true value. If the CI is narrow, then we can be confident that our point estimate is a precise reflection of the population value. The CI also provides us with information about the statistical significance of the result. If the value corresponding to no effect falls outside the 95% CI, then the result is statistically significant at the .05 level. By having a CI even statistically non significant, outcome might be clinically significant. In this situation, we could know how much CI is shifted to true positive rather than true negative.

Will the results help me in caring for my patient? (External Validity/Applicability)

To apply the results of the study to your patient, before making any clinical decision, you have to

answer to following questions:

- Is my patient similar to those in the study that its results can be applied?

- Is the treatment available, accessible, acceptable, and affordable in my setting?

To make a good clinical decision, you have to make sure that new intervention's benefit is superior to its potential harm in your individual patient.

In our scenario, even with acceptable NNT of green light versus TURP, applicability of new treatment should be evaluated.

Take-home message and conclusion

A couple of high quality RCTs are necessary for a clinical decision making in application of a new technique. It means green light laser can be applicable if:

1) Patient oriented characteristics of RCT are similar to your patient.

2) Randomization and concealment are reported.

3) It is blinded.

4) It is controlled by a placebo, sham, or gold standard group.

5) There is a defined, long enough, and complete follow-up period.

6) Patient is oriented to the endpoint.

7) Intention to treat and sub group analyses are done (if applicable).

8) Clinically importance and effects size are demonstrated by absolute risk difference, NNT, and CI.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Sajadi MM, Mansouri D, Sajadi MR. Ibn Sina and the clinical trial. Ann Intern Med. 2009;150:640-3.

2. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996;276:637-9.

3. Scales CD, Jr., Norris RD, Keitz SA, et al. A critical assessment of the quality of reporting of randomized, controlled trials in the urology literature. J Urol. 2007;177:1090-4; discussion 4-5.

4. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. BMJ. 2001;323:42-6.

5. Hanson BP. Designing, conducting and reporting clinical research. A step by step approach. Injury. 2006;37:583-94.

6. Domanski MJ, McKinlay S, Ovid Technologies I. *Successful randomized trials: a handbook for the 21st century*: Lippincott Williams & Wilkins; 2009.p.21-26.

7. Hrobjartsson A, Gotzsche PC. Powerful spin in the conclusion of Wampold et al.'s re-analysis of placebo versus no-treatment trials despite similar results as in original review. J Clin Psychol. 2007;63:373-7.

8. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. J Intern Med. 2004;256:91-100.

9. Hrobjartsson A, Gotzsche PC. Placebo treatment versus no treatment. Cochrane Database Syst Rev. 2003CD003974.

10. Furukawa TA. Review: placebo is better than no treatment for subjective continuous outcomes and for treatment of pain. ACP J Club. 2002;136:20.

11. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med. 2001;344:1594-602.

12. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? J Clin Epidemiol. 1997;50:1311-8.

13. Antal J. [Medical and ethical considerations of sham operation]. Magy Seb. 2007;60:233-8.

14. Clark PA. Sham surgery: to cut or not to cut--that is the ethical dilemma. Am J Bioeth. 2003;3:66-8.

15. Horng S, Miller FG. Ethical framework for the use of sham procedures in clinical trials. Crit Care Med. 2003;31:S126-30.

16. Macklin R. The ethical problems with sham surgery in clinical research. N Engl J Med. 1999;341:992-6.

17. Miller FG. A response to commentators on "Sham surgery: an ethical analysis". Am J Bioeth. 2003;3:W36.

18. Miller FG. Sham surgery: an ethical analysis. Sci Eng Ethics. 2004;10:157-66.

19. Millar K. Clinical trial design: the neglected problem of asymmetrical transfer in cross-over trials. Psychol Med. 1983;13:867-73.

20. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*: Houghton, Mifflin and Company; 2002.p.253.

21. Efird J. Blocked randomization with randomly selected block sizes. Int J Environ Res Public Health. 2011;8:15-20.

22. Berger VW. Do not use blocked randomization. Headache. 2006;46:343; author reply -5.

23. Toorawa R, Adena M, Donovan M, Jones S, Conlon J. Use of simulation to compare the performance of minimization with stratified blocked randomization. Pharm Stat. 2009;8:264-78.

24. Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: a simulation study. Clin Trials. 2011;8:59-69.

25. Han B, Enas NH, McEntegart D. Randomization by minimization for unbalanced treatment allocation. Stat Med. 2009;28:3329-46.

26. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet. 2002;359:696-700.

27. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. Control Clin Trials. 2004;25:143-56.

28. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Ann Intern Med. 2002;136:254-9.

29. Nahler G, Mollet A. *Dictionary of Pharmaceutical Medicine*. 2 ed. New York: Springer; 2009.p.55.

30. Hajebrahimi S, Azaripour A, Sadeghi-Bazargani H. Tolterodine immediate release improves sexual function in women with overactive bladder. J Sex Med. 2008;5:2880-5.

31. Weiss NS. Clinical epidemiology. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern epidemiology*: Lippincott Williams & Wilkins; 2008:519-28.

32. Adalatkhah H, Sadeghi-Bazargani H, Pourfarzi F. Flutamide versus cyproterone acetate / ethinyl estradiol combination in moderate acne: a pilot clinical trial. Clinical, Cosmetic and Investigational Dermatology 2011(In press)