



Review

Fifteen common mistakes encountered in clinical research

Glenn T. Clark DDS, MS^{a,1,*}, Roseann Mulligan DDS, MS^{b,1}

^a *Orofacial Pain and Oral Medicine Center, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, CA 90089-0641, USA*

^b *Community Dentistry Programs and Hospital Affairs, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, CA, USA*

Received 9 August 2010; accepted 23 August 2010

Available online 20 November 2010

Abstract

The baseline standards for minimally acceptable science are improving as the understanding of the scientific method improves. Journals publishing research papers are becoming more and more rigorous. For example, in 2001 a group of authors evaluated the quality of clinical trials in anesthesia published over a 20 year period [Pua et al., *Anesthesiology* 2001;95:1068–73]. The authors divided the time into 3 subgroups and analyzed and compared the quality assessment score from research papers in each group. The authors reported that the scientific quality scores increased significantly in this time, showing more randomization, sample size calculation and blinding of studies. Because every journal strives to have a high scientific impact factor, research quality is critical to this goal. This means novice researchers must study, understand and rigorously avoid the common mistakes described in this review. Failure to do so means the hundreds and hundreds of hours of effort it takes to conduct and write up a clinical trial will be for naught, in that the manuscript will be rejected or worse yet, ignored. All scientists have a responsibility to understand research methods, conduct the best research they can and publish the honest and unbiased results.

© 2010 Japan Prosthodontic Society. Published by Elsevier Ireland. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Contents

1. Failure to carefully examine the literature for similar, prior research	2
2. Failure to critically assess the prior literature	2
3. Failure to specify the inclusion and exclusion criteria for your subjects	2
4. Failure to determine and report the error of your measurement methods.	2
5. Failure to specify the exact statistical assumptions made in the analysis.	3
6. Failure to perform sample size analysis before the study begins	3
7. Failure to implement adequate bias control measures.	3
8. Failure to write and stick to a detailed time line	3
9. Failure to vigorously recruit and retain subjects.	4
10. Failure to have a detailed, written and vetted protocol	4
11. Failure to examine for normality of the data	4
12. Failure to report missing data, dropped subjects and use of an intention to treat analysis	4
13. Failure to perform and report power calculations.	5
14. Failure to point out the weaknesses of your own study.	5
15. Failure to understand and use correct scientific language	5
References	6

This review of the literature will describe the 15 common mistakes that novice researchers often make when planning conducting and writing up a clinical research project. These mistakes are usually made during the design phase; but might also be made during the data collection, analysis or manuscript

* Corresponding author.

E-mail address: gtc@usc.edu (G.T. Clark).

¹ Visiting Professors, School of Dentistry, Showa University, Tokyo, Japan.

preparation phases. In addition, hints on how to improve a research project and publication are suggested.

1. Failure to carefully examine the literature for similar, prior research

All research begins with the idea or question. What young or novice researchers often fail to appreciate is that the questions they take an interest in are likely not to be new, but are actually questions that others have thought of and frequently have made attempts to investigate in the past. The way to avoid this mistake is to assume that the question of interest has already been studied and the first job in the research design process is to exhaustively pursue, find and then catalog what has been published. Of course, the novice researcher may have a new variation of the question, or they may be using a new methodology or examining a new population of patients, but it should always be assumed that the core question in some form is likely to have been addressed previously. It now becomes the novice investigator's job to find that information, and consider the positive and negative outcomes of the prior studies in the new research design development.

HINT 1: When selecting and refining the exact focus of a question it is critically important for the novice to read in detail the discussion section of similar articles, for in that portion of the paper, most researchers speculate on what needs to be accomplished next in that topical area to advance the science.

2. Failure to critically assess the prior literature

Once a wise novice researcher has systematically accumulated and categorized the literature concerning the question of interest, the next step is to carefully examine the research papers related to the question of interest to find out what prior researchers felt could have been improved. One strategy to achieve this is to put together a team or group of research colleagues and select the 10–15 most important articles on a topic for the team to review. Ask each member of the team to present a critical analysis of the literature assigned, presenting both the good and bad points. Developing an individual's critical analysis skills will aid novices greatly in designing studies that minimize error. Not only is it necessary to critically analyze the literature before designing a new research project, but it is necessary to include these critical remarks in the introductory section of the resulting final manuscript in order to justify why the study was needed and what you as a researcher did better than previous researchers.

HINT 2: There is an old adage that says: “those who forget history are doomed to repeat it” and it is applicable to research as well. Investigators who repeat work previously done and do not recognize and build on prior efforts are likely to find their work unpublishable.

3. Failure to specify the inclusion and exclusion criteria for your subjects

A common omission from many research papers is the lack of research subject specifications, namely the inclusion

and exclusion criteria. Listing these criteria helps other researchers understand why current results might differ from other published studies. For example your patient population might be younger or your patient population might be from a different racial group or have a different ratio of males to females than were used in other research studies. In any case, it is necessary to specify as best you can the make-up of your subjects. This includes specific criteria for exclusion if you have any. Once you have the inclusion and exclusion criteria, be sure that you actually follow these criteria in selecting subjects for your study.

HINT 3: If a novice researcher is not sure how to develop a list of inclusion and exclusion criteria for a specific research question, look at prior research and use criteria that other researchers have specified.

4. Failure to determine and report the error of your measurement methods

Very few research reports actually provide more than a single sentence saying their examiners were calibrated. They rarely specify the method of training, the standards of performance and the frequency of re-assessment of their putatively calibrated examiners. All methods need replication and every researcher who is attempting the research project needs to be able to answer the question, “What is the error of your measurement method?” Some researchers refer to prior publications when answering this question but a good researcher knows the exact error of his/her own measurement methods and the inter-examiner variation. To find this error value involves conducting a small test-retest experiment. If however a researcher is using multiple examiners to help collect data, these examiners need to be calibrated to a known standard before being given the go-ahead to begin making measurements. If the research project is a long-term project, i.e. lasting for many months or years, it is critical to have examiners who are calibrated and re-calibrated periodically to an accepted standard of performance. Often extensive, complex and difficult studies fail because of the lack of detail to this small issue. A 2001 article examined the effects of measurement error on therapeutic equivalence trials and reported that measurement errors inappropriately favor the goal of showing treatment equivalence [1]. Essentially, this article reported on how imprecise data makes it difficult to tell if there are any real differences between two methods or two treatments. Such imprecision is a disadvantage if your goal is to evaluate that a new method of treatment is better than the old method; however, if you want to show that the new method or treatment is equivalent to or as good as the old treatment then imprecise data benefits this goal of showing equivalence or non-superiority. Another study in 2008 examined the frequency and characteristics of data entry errors in large clinical databases [2]. These authors reported that error rates ranged from 2.3 to 26.9%, with the errors being not just mistakes in data entry but many non-random, clusters that could potentially affect the study outcome.

HINT 4: A good researcher might even make the calibration process an independent research endeavor that could result in a publication of the process in a scientific journal.

5. Failure to specify the exact statistical assumptions made in the analysis

Since most studies will include statistical analysis of the data, specifying the level of significance (called the alpha level) that is acceptable and the exact statistical tests methods used is common place. However, rarely do you see the authors stating what they used as their beta value (type II error) which indicates their chance of a type II error (usually beta is 0.2 or less). The reciprocal of beta (1 minus beta) is then converted to a percent and reported as the power of a study (usually $\geq 80\%$). Novice researchers often do not state the directionality of the testing that they perform, namely whether they are using a one-tailed or two-tailed analysis. In 2007, an excellent review of the literature was published which cataloged and described 47 specific statistical mistakes that are commonly made in the medical literature [3]. These authors strongly suggested involving a statistical consultant early in a study as a way to prevent some of these common mistakes.

HINT 5: Providing statistical test assumption details gives the reader/reviewer the sense that the authors are attentive to detail and honest in describing the research process and the lack of such detail implies the opposite.

6. Failure to perform sample size analysis before the study begins

Most clinical trials that claim two methods are equivalent (or non-superior) are underpowered, which means they have too few subjects. To avoid this mistake, prior to initiation of a research project, it is important to know how many subjects are needed to achieve the minimum power level desired. There are multiple online and commercial computer based programs that will, with minimum information, provide the user with both the power and the estimated group sample size. To achieve sample size analysis it is necessary to understand the nature of the data that is to be collected, i.e. is the data linear or non-linear. It is also necessary to have a reasonable estimate of what effect the intervention will be, called the effect size. Finally it is essential to understand the variability of data collected. Without knowing the variability of the data, the effect size, and the power that is expected, it is impossible to estimate sample size, but with these data sample size estimation can easily be achieved. In a 2001 paper, the topic of equivalency testing and sample size in dental clinical trials was examined [4]. Specifically these researchers examined studies that compared the efficacy of dentures supported by 2 implants versus dentures supported by 4 implants. Such a study design is called an equivalency study. If the 2 methods are found to be equivalent, then one would logically recommend the use of the simpler and less expensive method. The authors found that underpowering a study makes it easier to find equivalency.

HINT 6: For linear data, if the standard deviation is quite a bit larger (e.g. 2–3 times larger) than the difference between the

2 treatment groups, the sample size required to show significance goes up substantially.

7. Failure to implement adequate bias control measures

The single most important mistake that clinical researchers make is the failure to implement adequate bias control measures. Bias control is what distinguishes good from bad research and measures to control for bias include: randomization of subjects to the areas, interventions and control conditions: measurement and analysis of subjects with the investigators blind to the subject status; and having a credible control condition and verifying at the onset and along the way that the subject is truly blind to the group to which they were assigned. This process is called a blinding status check. Double-blinding of researchers and subjects is desirable in a clinical trial to decrease bias. When blinding is not used or when the subject group status is easily detected, subjects will generally try to fulfill the perceived expectations of the researcher. The issue of expectation fulfillment was first pointed out in a study in Hawthorne, Michigan at an electronics plant [5]. The experimenters varied the intensity of electrical lighting available in the plant to see if there was a cause and effect relationship between work productivity and light intensity. Fortunately they varied the electric lighting in both directions, increasing the intensity and decreasing the intensity. What they discovered is that whenever an experiment was being conducted, work productivity increased; thus the phrase “the Hawthorne Effect” entered our scientific lexicon. This term means that any subject is likely to perform to the investigator’s expectations if they are not blind to their status. In 2001 a study examined the influence of study size on study outcome [6]. Specifically a meta-analysis reviewed 190 randomized trials involving 8 different therapeutic interventions divided the various studies into those with more than 1000 participants and those with less than thousand participants. The results of this analysis were that the smaller sized studies had more positive therapeutic effects than those studies with the larger size. These researchers also reported that the larger studies were systematically less likely to report a positive effect, suggesting bias was easier to occur and have an impact in smaller studies. These researchers also looked at other bias control measures such as randomization and blinding and concluded that inadequate randomization and blinding leads to exaggerated estimates of the intervention’s benefit.

HINT 7: Patient’s are remarkably able to detect to which group they have been assigned even though the blinding measures have been implemented; therefore good studies always perform periodic blinding checks.

8. Failure to write and stick to a detailed time line

A detailed timeline or Gantt chart is an essential feature to include in a protocol of a clinical trial. These charts can be created using a Microsoft Office Excel spreadsheet and every step of the trial should be noted in the timeline. The problem often seen with novice researchers is that they lack experience

and cannot estimate realistically the time needed to achieve a specific task. Nevertheless, a timeline is a critical and important overall feature in clinical studies, and failure to create and follow the timeline is a common mistake that is frequently made in clinical research.

HINT 8: Good researchers make a timeline plan that includes critical benchmarks along the way, they post it on the wall for everyone to see and they stick to it!

9. Failure to vigorously recruit and retain subjects

Clinical research implies that human subjects will be involved in the study. Subjects must be identified and recruited and a plan for this recruitment process needs to be developed and written down. A 2009 study actually compared 3 methods of subject recruitment and reported that direct telephone calls to the patient by the investigator were the most effective method [7]. Failure to have a specific recruitment plan and a method for retaining subjects in the study is a common mistake. Moreover, since subject recruitment is often a major issue in research studies, there should be more than one plan for subject recruitment.

HINT 9: Well designed research often fails because of poor subject recruitment and retention procedures so make this a priority.

10. Failure to have a detailed, written and vetted protocol

Before you begin any research project, especially clinical research, a fully developed protocol is critical. Novice researchers often begin research without completing the protocol. Moreover, in addition to writing the protocol, the researcher needs to present the protocol to a peer group, hopefully a peer group with moderate research experience, with the request that the group provide critical comments and suggestions for improvement. There is an old saying “luck favors the well prepared”. In the field of research, being well prepared means a well thought out, detailed written protocol is available and consulted frequently during the conduct of the clinical research project. Once the second phase of the research project starts, the data analysis phase, it is critical that an appropriate statistical methodology be selected and implemented to effectively analyze the data. Typically an experienced clinical researcher will consult a statistician for advice both before beginning the research and after the data has been collected. In the research phase a statistician is critical in helping to conceptualize the analytical methodology that should be used. Ideally the consultation with the statistician needs to continue as the data is being collected and prior to final analysis of the data. In many ways, the statistician serves as an outside auditor attesting to the diligence and honesty of the research process and analysis. It is not uncommon that the data that was planned to be collected, changes for pragmatic and unexpected reasons. This means the analytical plan may need to be adjusted. Although statistical software programs have improved

immensely in the last 10 years, no software program can make up for inappropriate or inexact design of a research project so consultation with an experienced statistician is almost always a necessity. In 2001, a review paper was written which discussed the topic of optimal clinical research design for chronic pain drug efficacy studies [8]. The authors made a list of suggestions that researchers should consider when they design and conduct such studies, but in their conclusions, they strongly suggested that a biostatistician consultant be used throughout all phases of the clinical trial.

HINT 10: The adage that is applicable here is: “the devil is in the details!” This saying refers to the fact that getting a general understanding and agreement that a project will be conducted is not enough. A researcher must also achieve a thorough understanding and agreement on the specifics of the project, which must be adequately documented or it can easily fail.

11. Failure to examine for normality of the data

In the analytic phase, it is important to examine the data that has been collected to see if it is normally distributed. Normality is a concept that applies to continuous linear data and is not applicable to categorical or non-linear dichotomous data. There are statistical programs that will take a data set and examine whether it meets the standards of normality. Data that is unevenly distributed about the mean can sometimes be transformed into more equally distributed data by using a log or log–log transformation. The advantage of transforming the data is that it allows you to continue using parametric statistical methods, as opposed to using non-parametric statistical analysis methods. In general, parametric statistical analysis is a more sensitive method (i.e. has more statistical power) and is preferred over that used to analyze non-parametric data.

HINT 11: A researcher should always look at the raw data obtained from the study displayed graphically since this demonstrates areas where there are problems with the data. The goal is to see if a histogram of the data demonstrates a bell-shaped curve or some other figure.

12. Failure to report missing data, dropped subjects and use of an intention to treat analysis

Statistical consultants will most likely recommend analytical methods that are consistent with an intention to treat methodology. This methodology deals with dropouts. Often novice researchers exclude dropouts from the analysis, and this can alter the conclusions of the study. Regardless of the method of analysis used, it is critical to report all dropped data, missing data, and subject dropouts in a careful and honest fashion. How the project dealt with lost or dropped data must be included in the methods section of the research report. Clinical trials that involve complicated, difficult or prolonged protocols often suffer from subject dropout. Many researchers will implement inclusion and exclusion criteria that reasonably eliminate the non-compliant patient. For example exclusion criteria might

specify that: “subjects that did not complete the health history questionnaire will be excluded from this study” or “subjects that failed to appear for more than one follow-up visit will be excluded”. Sometimes researchers will see the potential clinical subjects more than once during the pre-enrollment phase to determine their eligibility. This pre-enrollment phase frequently is referred to as the run-in phase. A run-in phase in a clinical study is an advantage in that it is easier to identify subjects who are likely to be non-compliant with the protocol and would best be excluded before enrollment. Clearly such a strategy would result in fewer dropouts, which is highly desirable. Unfortunately, run-in designs with many exclusions, make the results less generalizable to the real world population of subjects. Often such trade-offs are made between practicality, and idealism in design. In 1998 a small study was published describing the advantages and disadvantages of a run-in phase to a research protocol [9]. The authors concluded that run in clinical trials overestimate the benefits and underestimate the risks of treatment.

HINT 12: If you have to choose between excluding subjects and having many drop-outs, always choose excluding.

13. Failure to perform and report power calculations

Novice researchers often fail to perform a power calculation on their study. Such a calculation is critical in studies of equivalency. Small studies with low power often find no significant differences between the treatment interventions, however, if the study was inadequately powered then a type II error is more likely. A type II error is the acceptance of a false negative hypothesis. There are in fact multiple software programs that allow researchers to determine the power of their results. In 2001 an article examined how often underpowered reports of equivalency occurred in the surgical literature [10]. Specifically these authors looked at randomized controlled trials, where the control treatment was an active intervention, usually the standard treatment of the day. In these studies a new treatment was compared to the standard treatment and considered to be equal to the standard treatment if the results were equivalent. These researchers looked at 90 randomized controlled trials in the surgical literature and found that 39% of these reports met the standards for equivalency. The other 61% of the reports were typically underpowered and thus subject to a type II error. In 2001 another paper, examined type II error rates in the orthopedic trauma literature [11]. Similar to the results published in the prior study, 90% of this literature was underpowered with the overall power calculated for the 117 papers reviewed being 25%. The standard acceptable power in a study is 80% and therefore the authors concluded that many type II errors were likely to continue to occur in the orthopedic literature thereby affecting critical future research. Type II errors occur because there are too few subjects, but they also occur because there are too many measurements made on too few subjects. If you measure two groups of subjects twice, it is likely that some of the measurements taken on the second occasion will be different. It is also possible to show that the differences are indeed statistically different, if no downward

adjustments are made to the level of significance to compensate for the fact that there were multiple measurements.

One example of spurious associations being made is in the field of genetic polymorphisms. In 2007 one researcher examined why so many statistically significant associations between diseases in genetic polymorphisms are not replicated in future studies [12]. Specifically this paper looked at 10 single nucleotide polymorphisms or SNPs of the COMT gene that have been associated with various specific diseases. The authors concluded that false positive findings are commonplace and initial associations between genetic SNPs and diseases must be interpreted with high caution, since they are frequently not replicated. In 2006, a group of researchers conducted a meta-analysis on the topic of false positive gene associations, specifically those associated with human lymphocyte disease [13]. These researchers suggested that a median sample size of over 3500 subjects was necessary to avoid false positive results. They went on to state that collaborative studies seem like a logical approach for collecting large data sets like this, since individual researchers often do not have the resources to gather such a large data set themselves. A 2010 paper suggested a statistical standard be developed before initial results are accepted [14]. This paper suggested that a true report probability (TRP) score be developed based on data from multiple studies. The authors suggested that the suggested TRP formula would be straightforward and appropriate and help distinguish spurious results from true results.

HINT 13: Remember that “associations never prove causality.” This is certainly appropriate when trying to link genetic polymorphisms and disease, so replicate, replicate, and replicate.

14. Failure to point out the weaknesses of your own study

In the last phase of a clinical trial, the results are written in a manuscript form and submitted for review. Many novice researchers fail to point out the weaknesses of their own study in the discussion section of their manuscript. This is often reason for rejection of the manuscript.

HINT 14: In general hiding your mistakes or obfuscating them with the hope that no one will notice is not a good policy. Keep in mind that “honesty is the best policy” holds here as well.

15. Failure to understand and use correct scientific language

Finally all researchers, experienced and novices must use the correct scientific language when describing their results. Specifically, a single study never proves that a hypothesis is true; it can only reject the null hypothesis. While most people are not comfortable using such cautionary language, this is the correct scientific language. This understanding begins with studying a good statistical textbook which focuses on clinical research design [15]. Actually very few research manuscripts formally state the null hypothesis in the method section, and then

formally reject or accept the null hypothesis in the discussion section, but when this is done it shows a true understanding of scientific research and the limitations of the scientific method.

HINT 15: If you want to be a good researcher, you must study and understand the nuances of the language associated with the scientific process and only by doing this will you also understand the limitations of this process.

References

- [1] Kim MY, Goldberg JD. The effects of outcome misclassification and measurement error on the design and analysis of therapeutic equivalence trials. *Stat Med* 2000;20:2065–78.
- [2] Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. *AMIA Annu Symp Proc* 2008;6:242–6.
- [3] Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research—a review of common pitfalls. *Swiss Med Wkly* 2007;137:44–9.
- [4] Burns DR, Elswick Jr RK. Equivalence testing with dental clinical trials. *J Dent Res* 2001;80:1513–7.
- [5] Gale EAM. The Hawthorne studies—a fable for our times? *Q J Med* 2004;97:439–49.
- [6] Gluud LL, Thorlund K, Gluud C, Woods L, Harris R, Sterne JA. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982–9.
- [7] Schroy 3rd PC, Glick JT, Robinson P, Lydotes MA, Heeren TC, Prout M, et al. A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: results from a colorectal cancer screening adherence trial. *Clin Trials* 2009;6:597–609.
- [8] Harden RN, Bruehl S. Conducting clinical trials to establish drug efficacy in chronic pain. *Am J Phys Med Rehabil* 2001;80:547–57.
- [9] Pablos-Méndez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. *JAMA* 1998;279:222–5.
- [10] Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg* 2001;136:796–800.
- [11] Lochner HV, Bhandari M, Tornetta 3rd P. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am* 2001;83:1650–5.
- [12] Sullivan PF. Spurious genetic associations. *Biol Psychiatry* 2007;61:1121–6.
- [13] Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol* 2006;164:609–14.
- [14] Weitkunat R, Kaelin E, Vuillaume G, Kallischnigg G. Effectiveness of strategies to increase the validity of findings from association studies: size vs. replication. *BMC Med Res Methodol* 2010;10:47.
- [15] Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*, 3rd ed., Philadelphia: Lippincott, Williams and Wilkins; 2007. p. 51–63.