Statistical Basic Steps to Be Considered on Planning a Research

Cristina Pires Camargo, MD∗
Rolf Gemperli, MD, PhD†

The requirements for publishing an article are increasing due to the higher competition level. Some methodological aspects are essential to produce a good-quality article. The researcher must consider some basic items while planning a research protocol: the research question, the endpoints, the study design, and statistical analysis (SA) that includes sample size calculation. All of them should be carefully planned and described in detail. Once established the study design, the researcher can find some guidelines that define specific checklists for each kind of study design.2 Some of them are very popular, such as Consolidated Standards of Reporting Trials (CONSORT) for randomized controlled trials (RCTs) (http://www.consort-statement.org/), STrengthening the Reporting of OBServational studies in Epidemiology (STROBE) for observational trials (https://www.strobe-statement.org/), and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for systematic reviews and meta-analysis (http://www.prisma-statement.org/). To give reliability to the research protocol by making it auditable, the researcher should register the protocol at dedicated platforms, for example, “clinical trials.gov” for RCT (https://clinicaltrials.gov/) and Prospective Register of Systematic Reviews (PROSPERO) for systematic reviews and meta-analysis (http://www.crd.york.ac.uk/prospero/register_new_review.asp).

SA deals with scientific data and their variability during the research.2,5 SA suggests that whether the presence or the absence of a difference between groups is random or real (types I and II errors). SA is also helpful in determining the occurrence of bias, which is related to the internal validity of the research. The good quality of the research (internal validity) is mandatory for translating the findings from the study subjects to the general population (external validity/generalizability).3,4,5

Due to the clinical, ethical, and financial implications of research in human subjects, we discussed some basic statistical steps to be considered while planning a surgical research.5

RESEARCH QUESTION

Establishing the research question is the most essential step of a project planning. A vague or inaccurate research question increases the odds of failure of the research in its different steps. An adequate, precise, clinically sound research question makes easier for the researcher to choose both the best endpoint/outcomes and the valid tools for the endpoint analysis.5

When selecting the research question, the researcher is usually moved by the clinical relevance, but feasibility, ethics, and innovation should be considered as well.3

It seems obvious, but the research protocol should answer only 1 main question. This is the primary endpoint. Although a research may have several secondary endpoints, the sample size should be calculated to give mathematical meaning to answer the main question (please see it below in Sample Size).

For statistical purposes, when the researcher defines the question, the inferential statistics begin to be established. It means that if the experiment or observation compares different groups, the statistical plan demands a comparative test. Eventually, the researcher wants to establish a relationship, for example, vitamin C supplementation and collagen production in a wound-healing process, and possibly infer how strong or weak this relationship occurs. In this situation, correlation statistic tests will be necessary. Another scenario is when the researcher wants to predict one or more outcomes, for example, body mass index greater than 35 predicts a higher adverse event rate of abdominoplasty. Logistic or linear regression is the more appropriate statistical approach for this scenario (Table 1).

But, no matters if the research will deal with a comparison, association, or prediction, the research question will always test the hypothesis: null hypothesis (H0)—no difference between the groups of comparison; alternative hypothesis—there is difference between the groups of comparison.

SAMPLE AND SAMPLE SIZE

Most of the times sampling in health science is non-probabilistic, meaning that the researcher will sample available subjects from the site(s).3,5

Disclosure: The authors have no financial interest to declare in relation to the content of this article.
The researcher should try to select a population of patients that best reflects the expression of the disease in the general population, encompassing all the spectrum of the disease (mild, moderate, and severe cases). If this is not the case, the researcher should be aware of this limitation of the study with obvious implications for its generalizability, making this fact transparent to the reader. In addition to the choice of a representative sample population, the researcher will have to calculate the sample size needed to prove or reject the research hypothesis.

There are several mathematical formulas for sample size calculation. To apply them adequately, the researcher must define the following:
- Number of groups or arms
- Independent and dependent variables nature (numeric, categorical, nominal)
- Is it a comparison or association or predictive study?
- Effect size
- Study design
- Alpha and power

As mentioned above, the sample size calculation is calculated to test the primary endpoint. It means that the sample size is usually insufficient (underpowered) to compare secondary endpoints. For this reason, the secondary endpoints are equally known as exploratory questions. The function of exploratory questions or secondary endpoints is to guide the researcher toward the next steps in the investigational research line.

STUDY DESIGN

The study design selection is conditioned to the research question. When drugs, surgical techniques, or any kind of intervention are being compared, the adopted study design is the RCT. In the RCT, the usual hypothesis is that an intervention is superior to the standard therapy or placebo. This is the superiority trial. On the other hand, the researcher may choose to adopt an equivalence or noninferiority design for the RCT. This choice will directly affect the sample size and SA. Intuitively, a superiority RCT requires a large sample size than a noninferiority trial. Moreover, there are other nuances to be defined by the researcher concerning the study design: whether cross-over will be allowed between the 2 or more groups/arms, the number of groups/arms. These nuances will also influence the sample size calculation and the selection of the more adequate statistical test.

But if the research questions relate to the identification of an etiologic factor, prevalence, or incidence, an observational trial will suffice. There are basically 3 types of observational studies: cohort, case-control, and transversal studies. In cohort and case-control studies, the investigated etiologic factor and the disease are measured in different moments, either prospectively or retrospectively. In transversal or cross-sectional studies, the variables are measured at the same time. Studies that evaluate diagnostic tests are a typical example of transversal studies. The “novel” diagnostic test and the “gold standard” one are performed and compared in short time frame. All study designs have their particular advantages and limitations (Table 2).

Once defined the study design, the researcher will have to clearly determine the following:
- The outcomes, also known as dependent variables
- The intervention or exposition, also known as independent variables
- Independent and dependent variable nature (numeric, categorical, nominal) and possible distribution (normal, Poisson, Cumulative distribution) and variance (equal, unequal)
- The parameters that will describe the variables, for example, frequency, mean, and median
- The measures of association that will be used to describe the results of the comparison of the variable parameters, for example, odds ratio, relative risk, and number needed to treat

VARIABLES

In scientific research, the variables have different natures. They can be numerical (continuous, discrete), categorical, or nominal (Table 3). The choice of the variable

| Study Design | Advantages | Limitations |
|--------------|------------|-------------|
| Transversal  | Frequent events | No time relation |
|              | Low cost | Low evidence |
|              | No time demanding | |
| Case-control | Rare events | Difficult to define control group |
|              | Low cost | Low evidence |
|              | No time demanding | |
| Cohort       | Several factors can be evaluated | Time consuming |
|              | Frequent and rare events | Expensive |
| Randomized clinical trial | Can determine cause and effect | Selection bias |
|              | Randomization balance known and unknown variables | Time consuming |
| Case series  | No time consuming | Expensive |
|              | Very low evidence | Biased (selection, performing, detection) |
will be determined by clinical relevance.\textsuperscript{5,9} For example, numerical variables are preferable such as the number of cigarettes per day, yet, for the clinical purpose, the best approach is to transform it into a categorical variable (mild, moderate, and severe smokers).\textsuperscript{5}

The best outcome is the clinically most relevant one. However in some clinical situations, an intermediate outcome is adopted because the best one is less frequent which would require a larger sample size and more time. The intermediate outcome is also known as surrogate marker. The selection of a surrogate marker is challenging because it has to reflect the behavior of the clinically relevant outcome.\textsuperscript{10,11} In the setting of a study of a drug for cholesterol reduction, the clinically relevant outcome is the frequency of ischemic heart episodes. A possible surrogate marker could be the reduction of cholesterolemia to normal levels. Another example in plastic surgery is surrogate outcomes to assess wound healing: β-catenin, c-myc, wound fluid, matrix metalloproteinase, and interleukins.\textsuperscript{12}

**DEFINE THE MEASUREMENTS TOOLS**

In this section, it is interesting to discuss some concepts related to measurement.

For statistics, the best variables are the numeric ones because numbers allow calculus (eg, blood pressure, hematocrit, and wound area).

Regarding numeric variables, the researcher must pay attention to the precision and accuracy of the data. Precision is related to random errors or variability in the outcome measure. Accuracy is associated with the systematic error, a measurement of bias. To improve the research quality, the researcher had to consider both precision and accuracy.\textsuperscript{13}

However, it is not uncommon that in plastic surgery, the primary outcomes are not expressed as numeric variables, but as categorical variables such as patient satisfaction or quality of life. These tools (usually questionnaires) are subjective, and for statistical purposes, it is not the ideal variable to deal with. One possible way to address this problem is transforming nominal (completely satisfied to complete unsatisfied) data to categories (+3, +2, +1, 0, −1, −2, −3). This ordinal transformation allows statistical modeling. However, the meaning of this “number” is subjective. To illustrate it: the distance between one cm and the next cm is always the same, but the distance between +4 and +3 is subjective. For this reason, categorical variables are less powered than numeric variables.\textsuperscript{3,8}

It must be underlined that the researcher should adopt scales and scores already validated in the literature (eg, SF-36, BREAST-Q). Otherwise, the external validation (see above) will be compromised, reducing the chances of publication. Moreover, author permission is required to use some questionnaires. Finally, if the original version of the questionnaire is in a foreign language, it will require adaptation and validation to the language of the site where the study will be performed.\textsuperscript{14}

**ALPHA AND POWER**

The determination of 5\% in α value is a default, statisticians assumed as the cut-off value 5\% (p= 0.005) to reject the null hypothesis.

This arbitrary value is assumed to avoid type I error (alternative hypothesis is accepted, but it is not true). This type of error is more important in clinical context.\textsuperscript{15} Additionally, the researcher can assume a power of 80\%–90\% (1-β). This value reflects the chance not to find results when it is true, type II error Table 4.

**STATISTICAL ANALYSIS**

The parameters of variables of 2 different groups can be “compared,” “associated,” or can be tested to “predict” a specific outcome (see above).

If the research question is comparative, for example, compare 2 different techniques for mammoplasty, and the nature of the variable is continuous with a normal distribution equal variance between the groups, the more adequate statistical test is Student t (paired or unpaired). If the project compares more than 2 groups, the SA will be different. To help the more suitable SA, the authors showed a simplified flowchart according to the nature of the variables and study design\textsuperscript{3,16} (Figs. 1–3).

**MISSING DATA**

Most of the experiments lose patients in the follow-up visits. A reasonable dropout rate must be lower than 20\%, and the statistician can consider this lost in the sample size calculation. Moreover, for several reasons, some data can be lost. To avoid these facts, the researcher needs to plan dropout rates and how to deal with missing data at random.\textsuperscript{16}

According to CONSORT\textsuperscript{a} and Food and Drug Administration,\textsuperscript{17} one of the acceptable methods to missing data imputation is last observation carry forward.\textsuperscript{2,16} An alternative to that strategy is to consider the worst outcome for the missing patients from the intervention group and the best outcome for the missing patients from the control group.

---

**Table 3. Variable Nature and Descriptive Statistical Analysis**

| Variable Nature | Descriptive Statistics—Central Tendency | Descriptive Statistics—Dispersion | Examples |
|-----------------|----------------------------------------|-----------------------------------|----------|
| Numerical (continuous, discrete) | Mean | SD | Blood pressure, area, blood volume |
| Categorical | Median | Interquartile interval | Score and scales |
| Nominal | Mode | NA | Socioeconomic status, marital status, gender, binary |

---

**Table 4. According to the Hypothesis and Results, This Diagram Described the Possible Results for a Study**

| H0 Is True | H0 Is False |
|------------|-------------|
| Reject H0 | Type I error (false-positive) | Correct study result |
| Accept H0 | Correct study result | Type II error (false-negative) |
By adopting this conservative approach, we increase the chances of type II error (to accept the H0 when a real difference exists between the intervention and the control group) but reduce the chances of type I error (to reject the H0 when a real difference does not exist between the intervention and control group).3

INTENTION-TO-TREAT AND PER PROTOCOL ANALYSES

Some experiments face problems with adherence and missing data. One way to deal with it is performing the intention-to-treat analysis or full analysis. In this strategy, all randomized patients are included in the denominator; no matters if the patient complete the protocol or not. The advantage is to avoid efficacy overestimation. In others words, the effectiveness will be tested.

The drawbacks of this conservative strategy (ITT analysis) are type II error, deal with heterogeneous population, and misleading results due to a high dropout rate.18

On the other hand, if the researcher only includes the patients who completed the study, this is known as per protocol analysis. The advantages are to evaluate only the patients who received the treatment; it is related to efficacy. The disadvantage is a potential treatment overestimation, that is, a type I error.19

US Food and Drug Administration guidelines advise to perform both intention-to-treat analysis and per protocol analysis. If both analyses show similar results, the confidence in the study results will increase.20

In conclusion, to improve the publishing odds, we discussed some basic statistical aspects to be considered in a scientific project planning. One of the critical issues is the research question. If the authors build a clear, concise question, the chances of developing a good-quality study will increase. According to this question, the authors will select the adequate endpoint, measurements tools, sample size, and SA.

REFERENCES

1. Novack GD. The Importance of A Priori Statistical Planning in Controlled Clinical Trials. Am J Ophthalmol. 2015 Jul;160(1):4–5. e1. DOI 10.1016/j.ajo.2015.03.018.
2. Available at http://www.equator-network.org/Cited. Accessed November 25, 2017.
3. Portney LG, Watkins MP. Foundations of Clinical Research. Applications to Practice. 3rd ed. New Jersey: Upper Saddle River; 2009.
4. Ashton CM, Wray NP, Jarman AF, et al. Ethics and methods in surgical trials. 2009:35:579–583.
5. Carter RL, Schaeffer RL, Marks RG. The role of consulting units in statistics departments. Am Stat. 1986;40:260–264.
6. Jia B, Lynn HS. A sample size planning approach that considers both statistical significance and clinical significance. *Trials*. 2015;12:213.

7. Macefield RC, Boulind CE, Blazeby JM. Selecting and measuring optimal outcomes for randomised controlled trials in surgery. *Langenbecks Arch Surg* 2014;399:263–272.

8. Schulz KE, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

9. Pagano M, Gauvreau K. Principles of Biostatistics. 2nd Edition. Broks/ Cole, Cengage Learning. 2000, 7–95. ISBN-13:978-0-534-22902-3.

10. Gosho M, Nagashima K, Sato Y. Study designs and statistical analyses for biomarker research. *Sensors (Basel)*. 2012;12:8966–8986.

11. Adams-Huet, B., & Ahn, C. (2009). Bridging Clinical Investigators and Statisticians: Writing the Statistical Methodology for a Research Proposal. *Journal of Investigative Medicine: The Official Publication of the American Federation for Clinical Research*, 57(8), 818–824. http://doi.org/10.231/JIM.0b013e3181c2996c.

12. Lindley LE, Stojadinovic O, Pastar I, et al. Biology and biomarkers for wound healing. *Plast Reconstr Surg* 2016;138(3 Suppl):188–288.

13. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65:2276–2284.

14. Swanson E. Validity, reliability, and the questionable role of psychometrics in plastic surgery. *Plast Reconstr Surg Glob Open*. 2014;2:e161.

15. Drummond GB, Vowler SL. Statistical reporting guidelines. Type I: families, planning and errors. *Exp Physiol*. 2013;98:3–6.

16. Kaur M, Sprague S, Ignacy T, et al. Practical tips for surgical research: how to optimize participant retention and complete follow-up in surgical research. *Can J Surg*. 2014;57:420–427.

17. Little RJ, D’Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367:1355–1360.

18. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2:109–112.

19. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377:1391–1398.

20. Guideline for the Format and Content of the Clinical and Statistical Sections of Applications. Rockville, MD: Center for Drug Evaluation and Research, Food and Drug Administration, Department of Health and Human Services; 1988:443–4330.