ISN INTERNATIONAL SOCIETY OF NEPHROLOGY | RESEARCH
ACT | CLINICAL TRIALS

CLINICAL TRIALS TOOLKIT

Chapter 5 - How to Build a Statistical Analytical Plan

August 2022

5. HOW TO BUILD A STATISTICAL ANALYTICAL PLAN

By M.D. Ph.D. Augusto Cesar Soares dos Santos Junior (Brazil) and M.D. Brendan Smyth (Australia).

The text below is a transcript of the concepts described in the ISN-ACT Clinical Trials Toolkit. The characters and background story were added to the text to illustrate common research methods used in clinical trials. The names used refer to fictional characters making no reference to actual persons.

5.1 INTRODUCTION

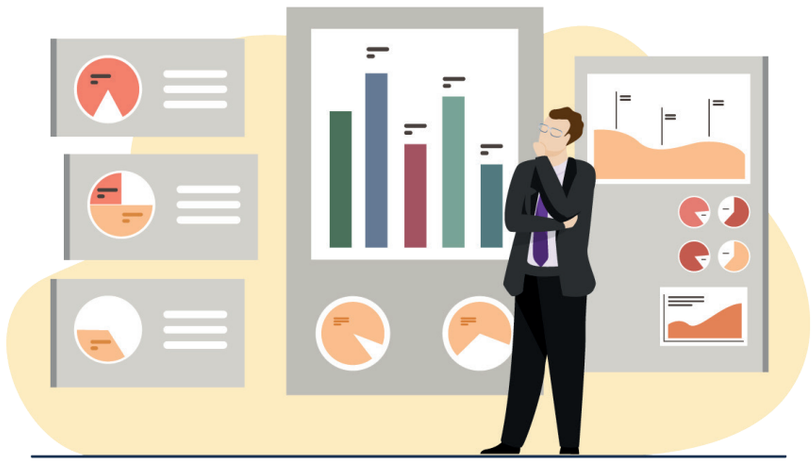
Kidney disease is a public health problem amplified worldwide by inequalities in both emerging and wealthier countries.

To help address challenges related to kidney disease care delivery, the ISN launched the Mentorship Program.

The program facilitates professional relationships to share knowledge and strengthen expertise in the field of nephrology. ISN members create mutual learning partnerships in the ISN global community and receive guidance to further develop their chosen areas. This way, the mentor can enhance their leadership and coaching skills, while the mentee can work on their professional objectives with their mentor.

Dr. Robinson, from Vancouver in Canada, was participating in one such mentorship with Dr. Lucia, a nephrologist based in Tegucigalpa, Honduras. Both nephrologists were interested in projects involving the prevention and treatment of acute kidney injury (AKI), sharing the idea that a properly considered statistical analysis plan is an essential part of the research process.

Having a solid academic reputation and a robust background in statistics, Dr. Robinson was well-qualified to share his knowledge with his mentee, Dr. Lucia. He had already presented a talk outlining the essentials, notably that statistics, a field of mathematics, can be divided into two categories: descriptive and inferential:



Descriptive statistics summarize the characteristics of a data set, answering questions such as: How many people have the

disease? How often did the event occur? What is the spread of test results in the population? Descriptive statistics can point to similarities and differences between groups but on their own are not enough to confirm or refute a hypothesis.

Inferential statistics allow you to test a hypothesis based on the probability theory. It answers the key question: How likely is the difference observed between two groups due to chance alone? In other words, are we just seeing 'noise' in the data, or is something real going on?

5.2 STATISTICAL ANALYSIS PLANS

It was a typical winter's day in the beautiful city of Vancouver. Dr. Robinson, as usual, woke early for his daily walk in Stanley Park. He was preparing to apply for a research grant and had only a few weeks to decide on the best statistical method before presenting his plans at the departmental meeting.

Dr. Lucia was his mentee on this project. Both were aware that, before starting a study, it is good scientific practice to have a Statistical Analysis Plan to ensure transparency and encourage rigorous scientific methods. Researchers are expected to publish their statistical analysis plan beforehand, either summarized as part of a design paper or study protocol or by making it available online.

5.2.1 Primary, secondary, and exploratory analysis

Dr. Lucia was riding her bike to work. She had already had several conversations with Dr. Robinson and was looking forward to working alongside him.

Later that day, Dr. Lucia joined Dr. Robinson online. He was the first to talk about decisions on the primary outcome, which would be decisive in guiding their statistical plan:

“The primary analysis is applied to the study’s primary outcome. It is essential to spell this out from the start. Remember that a P-value of 0.05 simply means a likelihood of one in twenty, meaning that if you do twenty statistical tests, the chances are that one of them will be ‘positive’ with a P-value < 0.05. So, it’s important to specify your primary outcome before the statistical analysis so that readers know you have not simply performed many statistical tests and chosen the one with the significant P-value.”

Without pausing, he added:

“Analyses planned in advance, alongside the primary analysis, are called secondary analyses. Any analysis carried out after looking at the data is always exploratory, meaning that it can only provide suggestions for further research and can’t be used to prove a hypothesis.”

5.2.2 Types of data and variables

Dr. Robinson was an experienced researcher and knew that Dr. Lucia’s comments were fundamental to the statistical analysis plan. However, he wanted to go a step further in the discussion and explain to Dr. Lucia what type of data should be collected:

“An important step before applying any statistical test is identifying the dependent and independent variables. In clinical trials, the dependent variable is the outcome of the study (e.g., mortality rate; change in renal function), and the independent variables are the factors under investigation that may modify the outcome. Independent variables, such as proteinuria and blood pressure, may be used in multivariable analysis. However, multivariable analysis is almost always a secondary analysis in a clinical trial, where the primary analysis is focused on the ‘dependent’ variable of the trial outcome and the ‘independent’ variable of randomized allocation.”

Dr. Lucia took notes as Dr. Robinson spoke; the study design began taking shape in her mind. Dr. Robinson continued with his explanation of key statistical concepts:

"Each variable should also be classified by type, for example, continuous, ordinal, categorical, or dichotomous, with normal (also known as parametric or Gaussian) or non-normal distribution. This helps to determine the right statistical test to use. These variables need to be considered in the planning stages as the type of data will have strengths and limitations in terms of the mathematical tests that can be applied and the interpretation of the results.

For example, if the stage of chronic kidney disease (CKD) is a factor under study, a categorical data analysis must be carried out. However, if the estimated glomerular filtration rate (eGFR) is collected, then a continuous data analysis, which is usually more powerful, can be applied. So, a study design needs to align with the research question and the characteristics of the study variables."

To make sure she had understood the issues, Dr. Lucia summarized,

"So, there can be a trade-off between power versus clinical meaning. A categorical outcome usually has easy-to-understand categories which are meaningful to patients, for instance, a yes/no answer or a stage of a disease. But a continuous outcome can express any value within a range, and a small change may be of uncertain importance. A categorical approach might be simpler to understand than a continuous one, but it can also decrease the power of the study as people with somewhat different results, for example, eGFR of 27 and eGFR of 16, are lumped together in the same category."

Dr. Robinson then turned his attention to the next step: defining the target population:

"Restricting the study population, for instance, targeting only patients most likely to experience worse outcomes, could produce smaller variations between participants (a more homogenous population) and, therefore, easily achieve significant results. Broadening the population, on the other hand, increases the generalizability of the results, but the varied (heterogenous) population can mean that a larger sample size is needed to allow a valid signal to be detected within the noise. These concepts influence not only the choice of statistical plan but also the estimation of the study sample size."

Dr. Lucia was trying to absorb all these elements and once again tried to sum up the concepts:

“It’s essential to decide whether the outcome will be continuous or categorial even before planning data analysis and sample size calculation.”

Dr. Robinson couldn’t agree more. After a long but pleasant discussion, Dr. Robinson planned to spend the rest of his day skiing in the Grouse Mountains. Dr. Lucia was looking forward to a sunset view from El Picacho Park in Tegucigalpa. They were satisfied that by identifying variables and using a study design best suited to the data type and target population, their study would have the best chance of success.

Design and layout: International Society of Nephrology

Editing and proofreading: International Society of Nephrology