A basic approach to clinical trials

By David Ortego

Randomized controlled trials are some, if not the best method to obtain evidence based results that can be later applied in daily clinical practice. But when I read one for the first time, I easily got lost and did not understand many things. After reading the article, I was afraid of coming to a conclusion that was not the correct one. This is important since depending on who's the reader, it could affect the patient's outcome. In order to avoid this, I will give a brief introduction to the reader of the blog to some basic concepts that are not as intuitive as one could think but are key for understanding the clinical trials, these concepts can later be expanded by further research on books, articles, databases, etc. The examples that I will be using are very basic and the result numbers that are mentioned are completely invented. I am sure many other details are also relevant when considering clinical trials and treatments, but for the sake of understanding the concept I'll keep them very simple. For the better understanding of the examples, I would like to note that the zolmitriptan group can sometimes be called the treatment group and the ibuprofen group, the control group.

Let's start from the beginning:

So let's imagine we find a clinical trial whose title looks interesting, for example, the use of Zolmitriptan (a drug usually used to treat migraine) is better for treating migraine than the non-steroid anti-inflammatory drug (NSAID) Ibuprofen. One of the first things that you see is the hypothesis, but while on the title they say that Zolmitriptan is better than Ibuprofen, on the hypothesis they are saying exactly the opposite, what's going on?

This is where the concept of null hypothesis comes in. A null hypothesis is when you make a statement that you ultimately want to reject and then accept an alternative hypothesis, which is the one that in reality you want to demonstrate. To continue with this example, the null hypothesis could be Ho = Ibuprofen is better in reducing the amount of headache events than zolmitriptan on patients with migraine, but this is not what they want to demonstrate , they want to demonstrate the opposite, so the alternative hypothesis would be H1= Zolmitriptan is better in reducing the amount of headache events in patients with migraine.

Understanding the results:

Now that we understand what they are trying to demonstrate, the researchers will do an experiment to demonstrate this hypothesis. Depending on the trial, it could be explained in a different way on the methods, for example, they could start talking about what's the type of trial they are making (multicenter, randomized, etc), the different committees that will supervise the trial, they mention what's the population that they included in the study and which was excluded and how was the treatment established as well as the end points or outcomes. I won't be focusing on this part of the clinical trial; I will briefly focus more on the statistical section of the articles. So after this they talk about statistics. In this section, we will see the concepts that are the ones I'll be talking about on this blog, which are sample size, p value, false positive result (type 1 error), confidence intervals, statistical power and false negative result (type 2 error) .

When we are reading the article, one critical thing to take into account in order to get a result that is statistically significant and be said to be close to the truth is sample size. If the clinical trial has a small sample size, this could give imprecise information and maybe a negative result, but it does not mean that, continuing with the example of ibuprofen and zolmitriptan, ibuprofen is better than zolmitriptan for migraine, it means that they did not have enough people, enough evidence to demonstrate it.

So what's with that famous p? Why is it everywhere?

After briefly understanding why sample size is important, if we continue reading the clinical trial, we see something that is usually repeated when showing a result, which is the p value. P value is a result which could help us determine whether the data that they have collected is statistically significant or not. It is related to α, which is the risk of having a false positive result. But what's a false positive? It means that they accept the alternative hypothesis (H1= Zolmitriptan is better than ibuprofen in reducing the amount of headache events in patients with migraine) when in reality it is false, so we think we are getting the result we want when in reality we are not.

In order to make sure that they don't get a false positive, they want α to be very small, it is usually given a value of 0,05 or 5%, but it could be given an even smaller number. If they give α this value it means that the probability of getting a false positive is 5%. I mentioned that p value is related to α, in order to accept that a result is statistically significant, one way of expressing it is by saying that p value is less than 0,05 , which means that the probability of getting their result , considering that Ho is true is less than 5% , since the chance of getting their result when they say that Ho (= Ibuprofen is better than zolmitriptan in reducing the amount of headache events on patients with migraine) is so small, it must mean that its false, so they then have to accept the alternative hypothesis, which is that H1= Zolmitriptan is better than ibuprofen in reducing the amount of headache events in patients with migraine, which is the result they are looking for.

The problem with p value is that it is a very unstable parameter, one could make an experiment and get a p value of 0,04 , which means that we would reject the null hypothesis Ho but then repeat it and get a p value of 0,051 , in this case we would not be able to reject the null hypothesis. If the p value is very close to their threshold, α= 0,05, then we can look at the sample size on the article and if it is too small, the positive result that they got on the article could have been by chance and therefore we should interpret it with skepticism.

What are those numbers between brackets and a CI next to it?

If we continue reading the article about ibuprofen we could also see another parameter that is next to the p value, that is called confidence interval or CI. For example, we read the following: the amount of headaches that the control group (the group with ibuprofen) had during the trial is 10 with a 95% CI (7-12) p value of 0,01. A confidence interval is an interval where, with a

certain amount of probability, in this case 95%, it includes the true value of the population. So in the example it means that , the true value of headaches that the general population could have by using ibuprofen for the treatment of migraine with a 95% of probability is one between 7 and 12. Now let's look at the group with zolmitriptan, again, these numbers are completely invented and are only used to explain the concepts, the number of headache events that they got is 2 with a 95% CI (1-5) p value of 0,0001 , this means that the number of headaches a population that is using zolmitriptan for the treatment of migraine could suffer, with a 95% of probability, is between 1 and 5. If we read these results, we could assume that they are statistically significant.

Let's put another example, let's imagine that instead of those values, the researchers found out that the amount of headaches that the control group had during the trial is 10 with a 95% CI (7-12) p value of 0,3 and on the group with zolmitriptan, is 5 with a 95% CI (5 -10) p value of 0,2 , in this situation , the CI of the zolmitriptan includes the null hypothesis value H*o* , which is 10 , this means that the number of headache events that the population could have with zolmitriptan , with a 95% probability is between 4 and 10 , so there is a chance that people with zolmitriptan could have the same amount of headaches than with ibuprofen, so is there really a difference between both groups?, this means that we cannot reject the null hypothesis, because there's a probability of getting the same amount of headaches with both drugs.

This concept can be summed up as if a 95% CI does not contain the value of the null hypothesis H*o ,* then the result must be statistically significant, with $p<0,05$, which is the first example that I explained. On the contrary, if the 95% CI does contain the value of the null hypothesis, then the result must not be statistically significant, with $p>0,05$, which is the second example.

After understanding CI, we can now focus on p value and how narrow is the CI, since not only is it important to get a statistically significant result but also a clinically relevant one. Following the migraine example, let's imagine this study lasted 30 days, it is not the same to be statistically significant with 5 headache events with 95% CI (4-7) p value of 0,0001 on the zolmitriptan group and 8 headache events 95% CI of (8-11) with a p value of 0,0003 on the ibuprofen group than to be statistically significant and have a 1 headache event with a 95% CI of (0-1) on zolmitriptan with a p value of 0,0001 and 8 headache events with 95% CI (8-9) and a p value of 0,0003 on ibuprofen. On the first example one could argue if the difference is clinically relevant or not, on the other hand, looking at the second example, the amount of headaches due to migraine that a patient had on 30 days with zolmitriptan was significant and clinically relevant, going from 8-9 migraine headaches to only 1 or even 0 when using zolmitriptan is very important for the patient, making this second drug a way better treatment option in comparison to ibuprofen. These differences are due to sample size, the smaller the sample size, the bigger the differences can be, and the bigger the CI but as you increase the sample size, you will be able to see smaller differences in the treatment group and a narrower CI. Depending on if the researchers want to measure big changes or very small changes, the sample size can be bigger or smaller.

To finish with CI, we can say that it is a more reliable method of telling whether the result that we read on the articles is statistically significant or not, rather than simply looking at the p value.

If there is a type 1 error, there has to be a type 2 error, right?

Finally, I would like to comment on another concept that can be seen on clinical trials which is statistical power, or simply power. What do they mean in the articles when they mention this? To understand power, I think it is also important to talk about another type of error, which is a false negative, also known as type 2 error or β, this error occurs when the null hypothesis H_o is false, but you can not reject it because you did not have enough evidence to do so. Explained with a formula, β = 1-power, so Power=1-β.

Traditionally, beta is given a value of 0,20 or 20%, so on the articles, a standard acceptable power is a value 80% or more. But what does it mean? It means the probability of rejecting a false null hypothesis, H_o . This means, with the values given a moment ago about power, if the researcher's study has 80% power, that the result they get has a 20% chance of having a false negative result or, seen from a different point of view, it has an 80% probability of rejecting the false null hypothesis.

This might be better understood with an example, let's say we keep reading the article of ibuprofen and zolmitriptan, on the statistical analysis section they mention that their study has an 80% power and then they get that the number of headache events with the zolmitriptan drug was 1 with a 95% CI of (0-1) with a p value of 0,0001 and 8 headache events with 95% CI (8-11) with a p value of 0,0003 on ibuprofen , discarding the fact that power could have been different than the one they previously calculated, with these results the researchers can with a probability of 80% reject the null hypothesis and accept the alternative one, which is that zolmitriptan is better in reducing the headache events on patients with migraine.

Statistical power, or simply power is directly proportional to sample size, effect size and data scatter, the higher the power, the bigger the sample size needs to be, the bigger your effect size, which is the parameter that the researchers are looking for, in this case is the amount of headaches on patients with migraine, and the less scattered your data will be.

Is the sample size random? How do they know how many people they need?

Before starting a study and after coming up with a hypothesis, the researchers need to decide how many people are going to need to be able to carry out their experiment, so depending on the power that they want they will need more or less people. Nowadays, the sample size can be calculated with programs (some can be found on the internet and are free to use, like the one called GRANMO) by including which power you want, which α you want, which β and what's the effect size you are looking for.

It is important to take into account that power can change, the researchers can previously calculate the power they want, which will be associated to the amount of people they will need to recruit, conduct the experiment and then get a result that is different than the one they

previously calculated, one can later ask, what went wrong? If they don't get the result they want it might have been because the real experiment lacked power. Let's say they expected to get, with 80% power, only 1 headache event with zolmitriptan but then when doing the experiment and gathering the data, they get 4 headache events , a p value of 0,3 and a CI that includes the H*o*. This could indicate that the study lacked power. They could not reject the null hypothesis, if we later calculate the power, although it is not needed because with the result it is clear it didn't have to, we could see that in fact it was not 80%, but less, a reason for it might have been that they did not have enough sample size. Was the H*o* true? Not really, it just means that they did not have enough evidence to reject it, considering that, based on zolmitriptan´s mechanism of action, we know it works against migraine, the researchers probably got a false negative and by repeating the experiment with a bigger sample size, they could get more evidence in order to be able to reject the H*o*.

I am now ready to read a clinical trial!

Although there is a lot more to know about a clinical trial, like understanding the methodology, which includes many other concepts like bias, randomisation, population distribution, having a negative result on a trial, etc, hopefully, after going through these basic concepts, reading a clinical trial will not be as intimidating and as challenging as it was in the beginning of this blog. All the best for the people that are getting started, like me, in the world of clinical pharmacology and statistics.

Useful links:

This Blog is inspired by the article: ohnston, S.C., Amarenco, P., Albers, G.W., Denison, H., Easton, J.D., Evans, S.R., Held, P., Jonasson, J., Minematsu, K., Molina, C.A., Wang, Y. & Wong, K.S.L. (2016). Ticagrelor versus Aspirin in Acute Stroke or Transient Ischemic Attack. *The New England Journal of Medicine*, 375(1): 35-43. doi: 10.1056/NEJMoa1603060.

Free calculator for sample size calculation: https://www.imim.es/ofertadeserveis/software-public/granmo/

Introductory video to clinical trials: https://www.youtube.com/watch?v=bctaWQTYHJc

Statistical test, hypothesis and p value video , this channel has good and short videos about statistics and clinical trials : https://www.youtube.com/watch?v=L2IlcoEb5jk

Easy and simple way to understand null hypothesis and p-value:
https://www.youtube.com/watch?v=eyknGvncKLw

References:

Flechner L, Tseng TY. Understanding results: P-values, confidence intervals, and number need to treat. Indian J Urol. 2011 Oct;27(4):532-5. doi: 10.4103/0970-1591.91447. PMID: 22279324; PMCID: PMC3263226.

Singh AK, Kelley K, Agarwal R. Interpreting results of clinical trials: a conceptual framework. Clin J Am Soc Nephrol. 2008 Sep;3(5):1246-52. doi: 10.2215/CJN.03580807. PMID: 18728215; PMCID: PMC4571159.

Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021 Feb 15;31(1):010502. doi: 10.11613/BM.2021.010502. Epub 2020 Dec 15. PMID: 33380887; PMCID: PMC7745163.