The hare and the tortoise

We all know the fable of the hare and the tortoise. It turns out that, for some reason I cannot understand, someone comes up to make a race with two participants: a tortoise and a hare. Naturally, the favorite of the race is the hare, infinitely faster than the tortoise. It turns out that the hare relies too much and goes to sleep, so when it realizes it can no longer recover the advantage of the tortoise and loses the race against all odds. Moral: Never underestimate others and do not rest on its laurels, or you may spend what happened to the hare.

Sometimes we can think of the clinical trial as a race among participants. This is when the main outcome variable is a time-to-event variable. These variables measure how many participants have the event in question and, more importantly, the time it takes to occur. Time-to-event variables are also called survival variables, although they don’t need to be related to mortality.

Here’s an example. Suppose we want to know the effectiveness of a drug for controlling blood pressure. We give the drug to the intervention group and a placebo to the control group to see how many are better controlled and how long it takes them to improve.

One possibility would be to use risk ratios or *relative risks*. We divide the proportion of patients who improve in the intervention group by the proportion that improve in the control group and get our relative risk. The problem is that we get information about how many more improve in one group than in the other, but we cannot say anything about the temporal aspect. We do not know if they improve sooner or later.

Another possibility is to consider blood pressure control as a dichotomous outcome variable (yes or not) and compute a *logistic regression* model. With this model will get an *odds ratio* that will give us similar information to that obtained by risk ratio, but nothing about the temporal aspect of the occurrence of the event.

The appropriate method to analyze this problem would be to establish the dichotomous measure of blood pressure control, but calculating a model of *proportional hazards regression* or *Cox regression*. This regression model does take into account the time it takes the event to occur.

The Cox regression model calculates the risk of the event in exposed to the intervention compared to unexposed at any given time. To do this, it calculates how much more likely is that the event occurs in the next time interval among subjects who have not suffered the event yet. Taking this measure to the limit, if we proceed to shorten the time interval until
zero, we get to the instantaneous risk, which oscillates with time, but the model calculates an average extrapolation with it. This extrapolation is called hazard ratio (HR).

HR can have values between zero and infinity. The neutral value is one, indicating the same risk in both groups. A value less than one indicates lower risk in the exposed group. Finally, a value greater than one indicates higher risk among exposed, the greater the larger the value of HR.

The HR is not a measure of probability but an odds, so its interpretation is similar to that of the odds ratio, but with the addition that also takes into account the temporal aspect. A common misconception is that the HR reports the temporal duration to the event. For example, an HR = 2 does not mean that the event in exposed develop twice as fast, but that those exposed who have not yet presented the event are twice as likely to present it than the unexposed.

If you want information about the rate at which the event occurs, you have to use other index which is the median time in which the 50% of the participants present the event.

Returning to our tale of the race, the HR would tell us who is most likely to win the race, while the median tell us how much benefit would take the winner to the loser.

And here we leave the subject of the hare, the tortoise and the proportional hazards regression. We have not talked anything about how to represent the results of the Cox regression model. For this a special type of graphics called survival curves or Kaplan-Meier’s are used. But that is another story…

---

**Brown or blond, all bald**

Have you ever wondered why some people go bald, especially men at a certain age? I think it has something to do with hormones. Anyway, it’s something that the affected usually like the least, even though the popular believe that bald are smarter. It seems to me that there is nothing wrong with being bald (it’s much worse to be an asshole) but, of course, I have all my hair on my head.
Following the thread of baldness, let’s suppose we want to know if hair color has anything to do with going bald sooner or later. We set up a non-sense trial with 50 brown-hair and 50 blond-hair participants to study how many go bald and when they do it.

This example serves us to illustrate the different types of variables that we can found in a clinical trial and the different methods that we use to compare each of them.

Some variables are of quantitative continuous type. For instance, the weight of participants, their height, their income, the number of hair per square inch, etc.. Others are qualitative, such as hair color. In this case, we simplify it to a binary variable: brown or blond. Finally, there is a time-to-event type, which show the time it takes participants to present the event in study, in our case, baldness.

However, when comparing differences among these variables between the two groups of the study we have to pick out a method that will be determined by the type of variable that is being considered.

If we deal with a continuous variable such us age or weight between bald and hairy people, or between brown and blond, we’ll use the Student’s t test, provided that our data fit a normal distribution. If that is not the case, the non-parametric test that we would use is the Mann-Whitney’s.

And what if we want to compare several continuous variables at once?. Then we’ll use multiple lineal regression to make comparison among variables.

For qualitative variables the approach is different. To find out if there is a statistically significant dependence between two qualitative variables we have to build a contingency table and use the chi-squared or Fisher’s exact test, depending on our data. When in doubt, we can always use the Fisher’s test. Although it involves a more complex calculation, this is no problem for any of the statistical packages available today.

Another possibility is to calculate a measure of association, such us the relative risk or odds ratio, with its corresponding confidence interval. If the interval do not intersect the line of no-effect (the one), we can consider the association as statistically significant.

But it may happen that we want to compare several qualitative variables at once. In these cases, we’ll use a logistic regression model.

Finally, we’ll discuss the time-to-event variables, a little more complicated to compare. If we deal with a variable such as the time it takes to go bald we have to build a survival or Kaplan-Meier’s curve, which graphically shows what percentage of subjects remain at any moment without
presenting the event (or the percentage that has presented it, according to the way we read it). But it could be that we want to compare the survival curves of brown and blond people to see if there are any differences in the rate at which the groups present the event of going bald. In this case we have to use the log rank test.

This method is based on the comparison between the two curves based on the differences between the observed survival and the expected survival values that we could get if there were no differences between the two groups. Remember that survival refers to the moment to present the event, not necessarily death. With this technique we get a p-value that indicates whether the difference between the two survival curves is statistically significant, but tells us nothing about the magnitude of that difference.

The case of more complex calculation is when we want to compare several variables with a time-to-event-variable. For this multivariate analysis we have to use a proportional hazards regression model (Cox’s regression). This model is more complex than the previous ones but, once again, any statistical software will carry it without difficulty if we feed it with the appropriate data.

And we are going to leave the bald alone for once. We could talk more about time-to-event variables. The Kaplan-Meier’s curve gives us an idea of who is presenting the event over time, but it tells us nothing about the risk of presenting it at any given time. For that we need another indicator named hazard ratio. But that’s another story...