“KMunicate” project briefing

As an extended group exercise, we are going to be data visualisation consultants for a clinical trial researcher, Tim Morris of the UK's Medical Research Council Clinical Trials Unit.

We'll meet Tim at a first Skype call on Tuesday afternoon, where he can answer questions about his study ("KMunicate"), what visualisations he did for it, what he is happy or unhappy about in the publication (BMJ Open paper, attached), and what he wants to achieve from this consultation.

At a second call on Thursday afternoon, we need to pitch one or two options back to him, being very clear about why we think they are good, and what we think their weaknesses might be.

Goal of KMunicate

This was a project undertaken to test out potential improvements to a very common format of chart in medical statistics, called the Kaplan-Meier plot (KM plot). Essentially it was an informal survey, where people were shown the variants and asked to rate them. Although this study compared different visualisations -- which is interesting to us in itself -- our goal is to advise on visual presentation of the survey findings.

What is a Kaplan-Meier plot?

In many biomedical studies, we are interested in the time to an event or "endpoint", such as a heart attack. This is broadly called survival analysis, although the endpoint need not be fatal, and in fact could be desirable. In a clinical trial, participants are randomly allocated to different treatments, so we generally just have to compare the survival between two (or more groups).

Time is plotted from left to right. Each group starts at the top left corner and moves right. As participants have the endpoint, the % "surviving" decreases and the curve drops. For a undesirable endpoint, we would prefer the treatment that gives a higher curve, even if everyone gets the endpoint eventually.

In the classic form of survival analysis, once someone has had the endpoint, they are no longer at risk of it and they are no longer counted in the study. Usually, there are other reasons why participants cease to be at risk; they might ask to leave the study, or have another complication that means it's no longer safe to carry on treating them, in which case they are removed from the pool of people at risk. Statisticians call this "censoring".

KM plots show discrete changes in the numbers of people in three categories: those at risk of the endpoint, those who have already had the endpoint, and those who were censored before they had the endpoint. Because these are discrete integers, the KM plot curves drop at event times and can look like steps.

Typically, uncertainty in the location of the KM curves is calculated as a confidence interval at each timepoint and can be shown as a range around the curve.

What did KMunicate try to improve?
The study proposed four variants on the classic KM plot that provided more information on the numbers of people (at risk, who had the endpoint, or who were censored). There were also two variants showing more about uncertainty.

What was asked of survey participants?

They were shown five options for the numbers of people (classic, extended risk table, lines beneath, areas beneath, areas behind), and three options for the uncertainty (classic, shaded areas, fading lines).

Then, there were two main questions. In each variant, they had to compare them with the classic option, and rate them as "much more useful", "somewhat more useful", "a bit more useful", "no preference", or "less useful".

Finally -- and this is the part we are mainly concerned with -- they were asked to identify their first, second and third preference. Not everyone gave complete information for this. There were 1274 participants in total. 1271 gave ranks for the number variants, and 1255 for the uncertainty variants.

However, in the number variants, only 1228 (96%) allocated 1st and 2nd rank, and only 1023 (80%) went on to allocate 3rd rank. Note also that nobody was asked to allocate 4th or 5th rank, and because of this alone, we might expect there to be twice as many unranked answers as there were participants (200%), but this is actually 224%. How can we re-allocate the excess 24% or otherwise show that we don't know 20% of participants' 3rd rank options.

In the uncertainty variants, 1117 (88%) allocated 1st and 2nd rank, and 898 (70%) allocated all three ranks. In this set of three options, we might assume that the third rank is identified if the first and second are.

Conclusions of the study

The "extended risk table" option was the clear favourite among the numbers options, but second rank was not clear, even if we take the reported ranks at face value and ignore missing data. "Lines beneath" and "areas behind" were unpopular; even if all the missing second place ranks belonged to one of them, they still would not rise above the others.

"Confidence intervals" was by far the favourite in the "uncertainty" options, with the classic reference design in second place, and "faded lines" very unpopular.