Statistics in Clinical Trial Design

Here are some of the problems we will talk about. We will discuss them together and find a solution, or find out why it is a foolish question :-)

Phase I: Understanding Toxicity

In early phase clinical trials (phase I) in oncology we want to understand the toxicity of a new drug. In one of the commonly used designs patients are allocated to increasing doses of the drug in groups of 3 at a time. For each patient we record whether the patient experienced toxicity ($y_i = 1$) or not ($y_i = 0$) at the assigned dose. Let Y_j denote the number of toxicities observed at dose j and let N_j denote the number of patients assigned at that dose.
When we first observe $Y_j = 2$ out of $N_j = 6$ patients, we declare this to be the maximum tolerable dose (MTD). The motivation is that this would suggest a prob of toxicity of around $p = 2/6 = 0.33$ which is considered acceptable.

1. **Estimating a success probability:**
   1. Estimate the probabilty of toxicity at the selected MTD? I.e., with 2 toxicities out of 6 patients, what is your best guess for the probability p_j of toxicity at dose j?
   2. At $p_j = 2/6$, what is the probability of observing 2 toxicities out of 6 patients?
   3. Now, consider $p_j = 1/6$ and $p_j = 3/6$. Find again the probability of 2 toxicities out of 6 patients.
   4. How much do you trust the estimated MTD?

Phase II: Understanding Efficacy

Once we understand (or think to!) the safety of a drug or treatment, the next stage in drug development is to carry out a study to find evidence for efficacy.

2. Dose finding: One problem that is considered in phase II studies is to find the right dose to achieve a response. A typical procedure is as follows.

- Fix a grid of doses.
- Allocate patients at each dose.
- Estimate the dose that gives around 95% of max effect

What do you think of that scheme?

3. Backward induction – sequential design: Sometimes we might be able to allocate patients sequentially. That is, decide the allocation for the 2nd set of patients *after* we see the response from the first set.
For example, assume two treatments $d = 0$ (control) and $d = 1$ (experimental therapy) and two outcomes $y = 0$ (no response) and $y = 1$ (response).
And assume the following probabilities

$d = 0$**:** Under control we know everything, The probabilities of response y are

	y	
	0	1
	1/2	1/2

$d = 1$**:** Under experimental therapy we know very little.

	y_1	
	0	1
First patient	2/3	1/3
After one success	1/3	2/3
After one failure	3/4	1/4

"First patient" referes to first patient on $d = 1$. Same for "After one success" and "... failure".
The problem is, initially control ($d = 0$) looks better, and experimental therapy looks a bit worse. But this could flip if we were able to observe one patient on $d = 1$ and we were to observe a success!

- Should we assign the first patient (cohort) to $d = 1$?
- What should we ask first?

Phase III: Comparison with standard of care

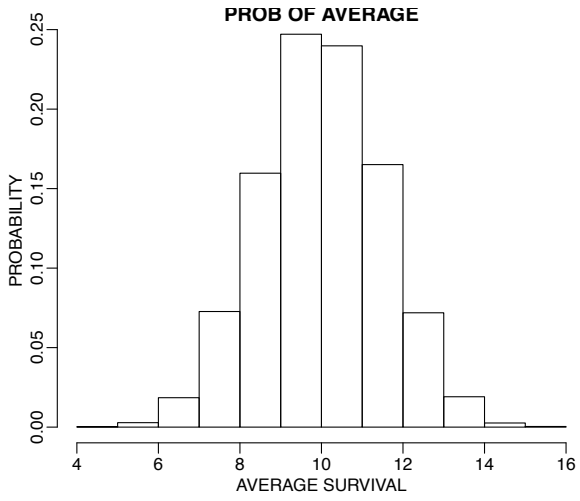
Once we believe that the new treatment could work, we have to carry out a formal comparison and convince the regulatory agency that the new treatment is better, beyond reasonable doubt.

4. Hypothesis testing: There is a very formal framework for carrying out this comparison. The logic is a bit twisted, but it is not our choice! Assume we compare overall survival under treatment (m_1) versus standard of care (m_0). Say $m_0 = 10$ months. Assume we observe average survival of 14 (or 12).
First give names to the two competing hypotheses:

$$H_0 : m_1 = 10 \text{ versus } H_1 : m_1 > 10$$

Now the logic is like in a court room scenario. We have to prove beyond reasonable doubt that H_0 can be rejected. We use probability. We have to show that assuming H_0, the chance of observing our data is very very small, say less than $\alpha = 1\%$.

- Why do we have to allow that 1%? Could we ask for certainty?
- For a (only slightly!) simplified description, assume under H_0, the probability distribution (histogram) for the average survival of 100 patients should look something like this.



Find the probability of an average beyond 14.0 and beyond 12.0.

Here is the rule for carrying out the comparison. It's a ritual :-)

- Find the average survival of the 100 patients.
- If that probability is small beyond reasonable doubt, then we reject H_0. Say "small" is 0.01.
- If it is not, i.e., it's possible to see something as high as that average survival, be happy and don't reject.

5. Randomization: Consider this one. In a big Japanese study with stomach cancer patients we compare *surgery* (standard of care, $d = 0$) vs. *surgery plus adjuvant chemotherapy* (experimental therapy, $d = 1$). The surgeon decides the treatment.

Here is the data, recording overall survival (y) beyond 24 months.

	y		
d	0	1	n
0	126	74	200
1	136	64	200

- What do you think? Which treatment would you recommend?
- Anything we should ask next from the investigators or look for in the data?

6. Multiplicities There have been 16 large phase III Alzheimer trials – all failing. Assume all at $\alpha = 0.05$ (recall the α in the ritual above..)

- What is the probability of this happening if all were flops?
- In the next years there are bound to be more. Assume there are 4 more, for a total of 20 big studies.
  Would you be surprised if some show significant effects?
- Would you recommend a treatment that shows an effect?
- Why is the FDA not allowed to follow your logic? Is it logical?

7. Selection sampling Spitfires (fighter planes of the U.K. Royal Airforce during WW II) often returned from missions with substantial damage from enemy fire. The British high command considered reinforcing the planes where they had been hit most.

- Any thoughts? Do you think that would have been smart?
- Anyone willing to argue for a different recommendation?