Some examples of statistical issues in the efficacy trial

Just 3 or many

A sample of three issues

- There are many issues, but it seems that these three provide examples of the types of decisions that need to be made and will have longer term effects:
  - Randomization ratio.
  - Interim analysis for early decision
  - Role of baseline cofounders

The randomization ratio should be 1 to 1

- Often developers and/or regulators request 2 to 1 or other ratios, presumably to increase the Safety data base.
- Recall that the power of the statistical test for efficacy is based on the total number of cases regardless of the number of subjects enrolled
- The expected number of cases, therefore, depends on a simple formula,
    - Total(r>1) = Total(r=1) + VE(difference in TRT sample size)
- Larger sample sizes, longer waits, greater expense.

Considerations about interim analyses and how the decisions affect the conduct of the study.

These are estimated sample sizes for various desired power, expected VE, Lower Bounds on VE, and attack rates. Let's concentrate on the two shaded lines

Sample sizes for a Lassa efficacy trial (with a 15% increase to cover loss to follow up)						
Power = 90%						
				Attack rates		
VE	LB	Number of events	Critical value	.03%	.04%	.05%
70%	0	32	10	9650	7238	5790
	20	57	17	17189	12892	10314
	30	76	22	22919	17189	13752
90%	0	12	2	4278	3209	2567
	20	17	3	6061	4546	3637
	30	19	3	6773	5080	4064
Power = 80%						
70%	0	25	7	7539	5654	4524
	20	40	11	12062	9046	7238
	30	61	17	18395	13796	11039
90%	0	12	2	4278	3209	2567
	20	14	2	4991	3743	2995
	30	15	2	5348	4011	3208

The interim plan

VE	LB	Number of events	Critical value
70%	0	32	10
	20	57	17
	30	76	22

Suppose we want to have a vaccine with at least VE = 70% and we want to have enough cases that the LB of the 95% CI Is at lest 30%. We would like to do an interim analysis at 32 cases and accept a LB of 0. We need to use an interim technique so we will assure ourselves we are not wasting resources going on. We will use O'Brien-Fleming

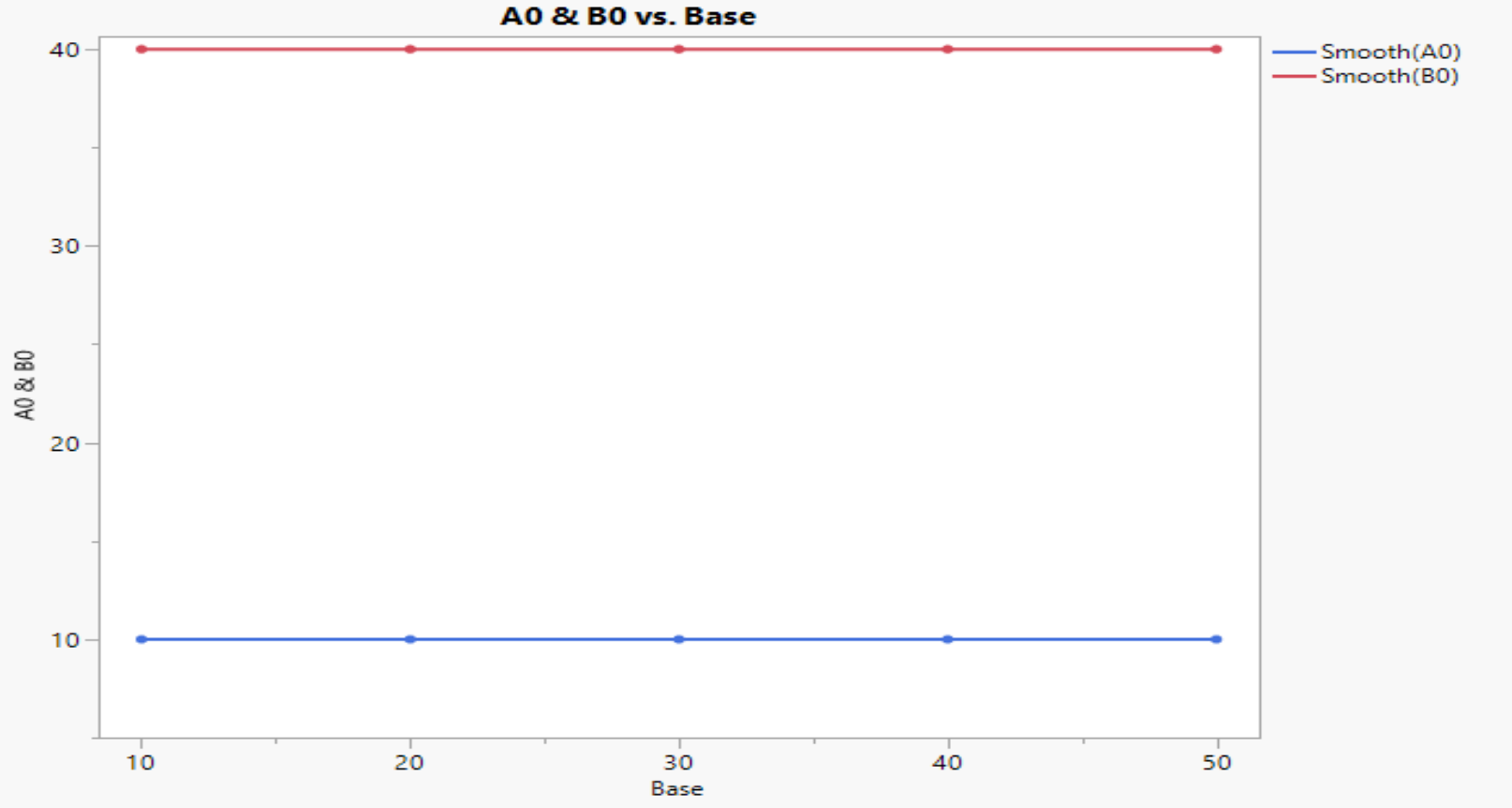
The fact that the final analysis is the same is an oddity since the exact test we use has discrete steps and the O'Brien-Fleming technique is conservative on the first analysis and therefore affect the final less than might be expected.

VE	Which analysis	LB	Number of events	Critical value for success	Critical value for futility
70%	Interim	0	39	<=11	>=17
	Final	30	76	22	NA

The effect of Baseline Cofounders

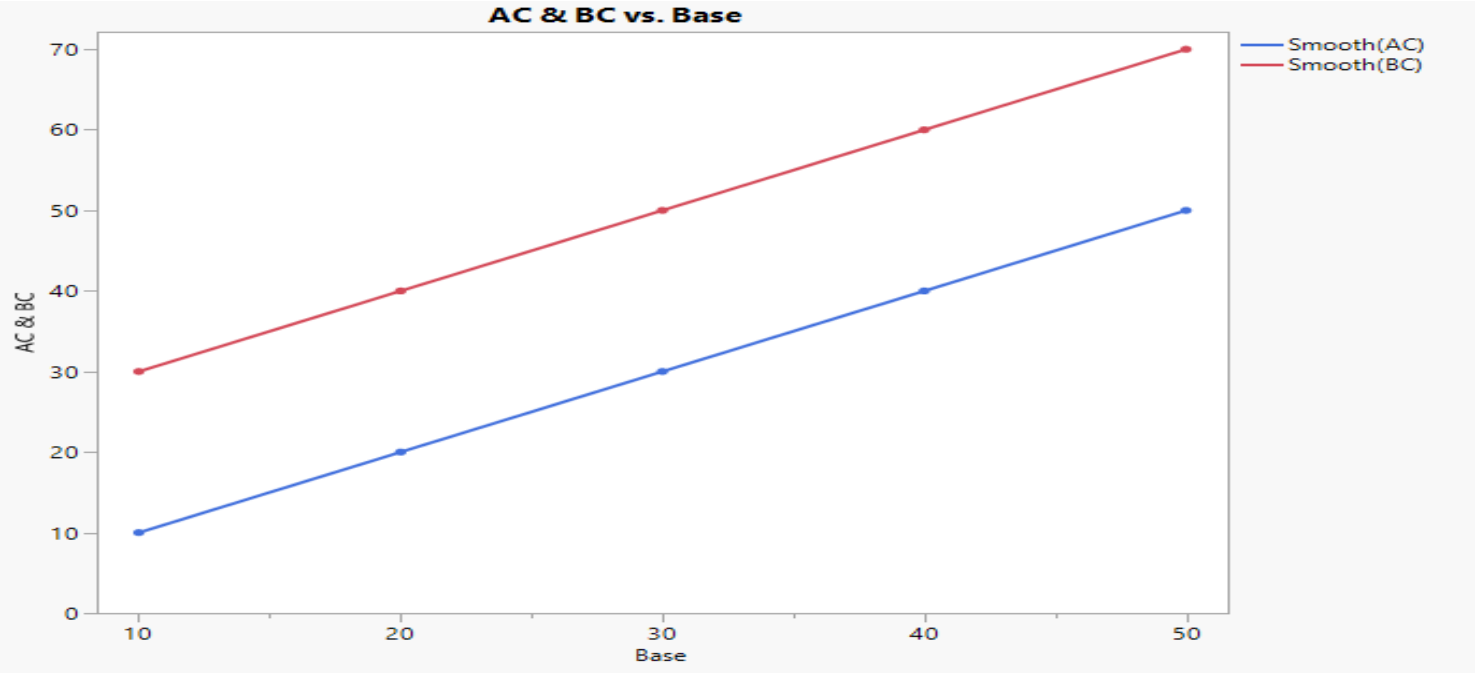
- Variables measured at baseline are not randomly assigned but can affect the response to randomized treatments.
- Depending on the actual effect the appropriate analyses can be complicated.
- The NIH-FDA Web Site BEST refers to these as possible Predictive Biomarkers. Others use the term Measures of Therapeutic Effect.
- Some examples
  - Age in Flu vaccines
  - Baseline serostatus (titer) in Dengue.
  - Probably others

Baseline has no effect



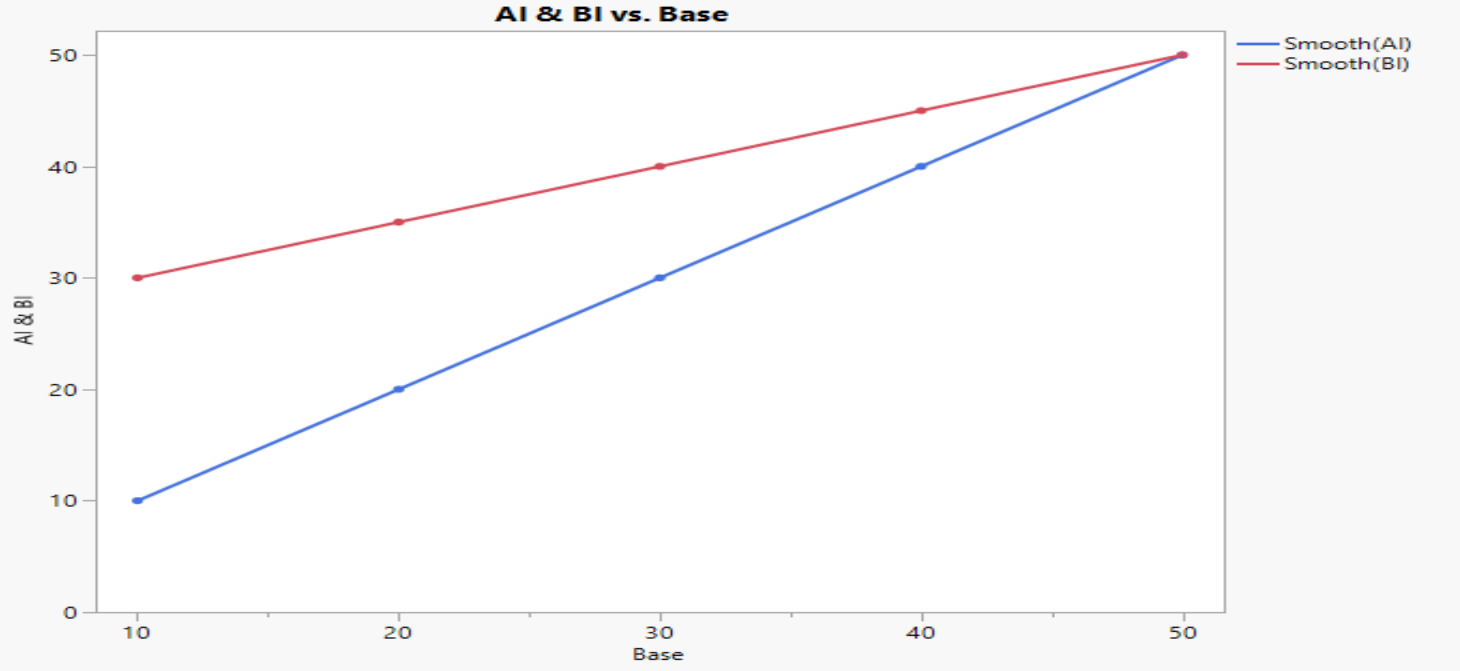
Over the range of the Baseline the responses remain the same. We do not have to consider the Baseline further

Baseline can be used in an ANCOVA, reducing variability and increasing precision.



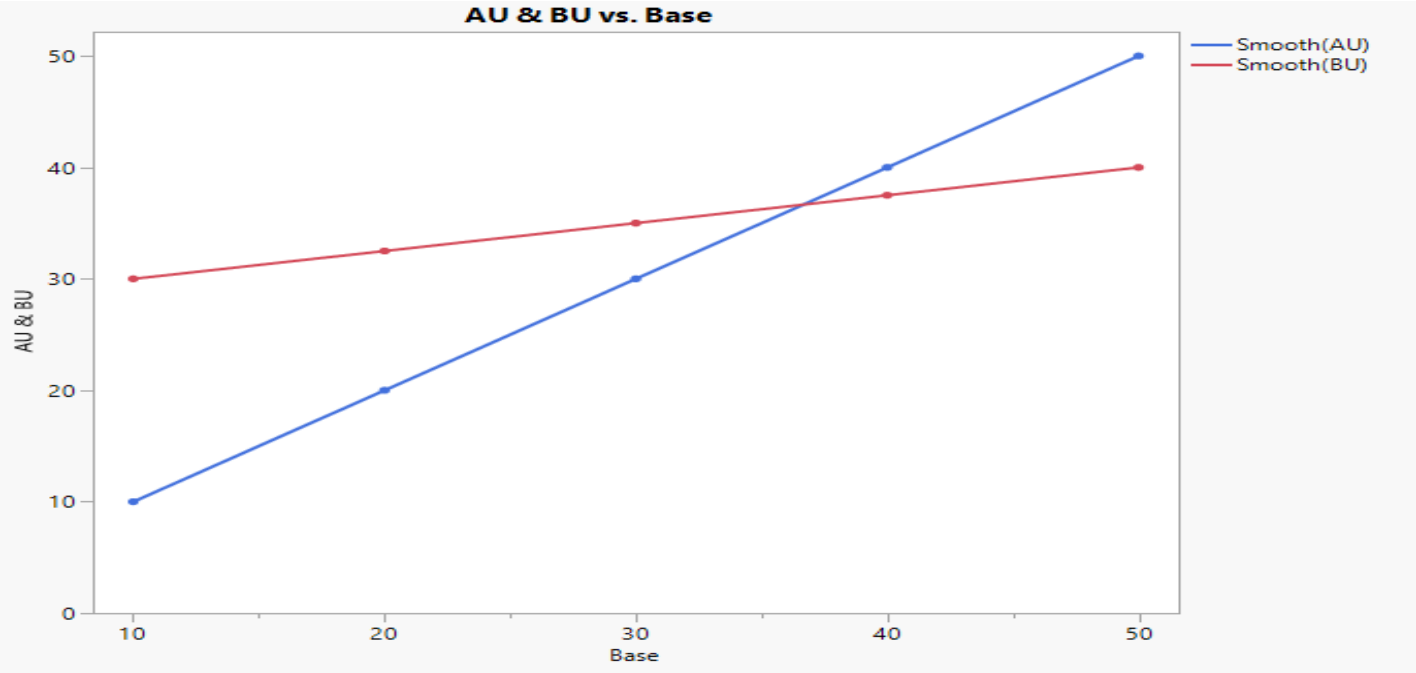
The classical justification for an ANCOVA is attained. The Baseline has an effect, but it is the same on both treatments so that the *difference* is not a function of the Baseline

Baseline complicates the interpretation of results. There are ranges of the Baseline for which the vaccine may have no or little effect



In this case the Baseline affects response differentially and may modify appropriate use.

Baseline may be causing a disaster. There are ranges of the baseline that seem to cause the treatment to have harm



For low values of the Baseline, Treatment A is better, for high values, B is better. We need to be careful about the model and about the estimate of error used in tests and CIs.

Some additional comments about Baseline variables

- The graphs depict a continuous response such as a titer, but the results and cautions are the same for a graph of VE versus a Baseline variable (e.g., age in Flu and titer in Dengue).
- Often when the situation depicted in the last two occurs, simplistic solutions are offered. These include stratifying on the baseline values and defining the endpoint to be change from baseline or fold increase from baseline. These do not solve the problem.
- The issue is that the response (response titer, VE, Safety measure, etc.) is a function of the Baseline and that function must be estimated to inform policy.
- For Lassa we will likely be enrolling subjects with a wide range of baseline pertinent values.