Lecture 4
Biostatistics III: Clinical Trials

Michael C Sachs

Design of a Randomized Clinical Trial

Standards of Care

Back to Example

Sample Size

Back to Example

Intend to Treat

Asking for Help

Hypothesis

Give me a hypothesis!

This vaccine will prevent HIV infection

Specific Intervention and control

The vaccine might have multiple doses. Standards of care (SOC) might be required for everyone.

Three injections of MRKAd5 HIV-1 gag/pol/nef + SOC - regular safe sex and STI counseling and testing.

Situational Standards of Care

Different communities have different available standards of care (SOC).

There are heated debates every year - are we ethically obligated to provide the best SOC in the world in every trial as the comparator or are we only obligated to provide the SOC available in the given community where the trial is taking place?

What is the best SOC in the world?

Situational SOC Example

PrEP exists in the US and if someone in the US had a HIV positive partner their doctor would URGE them to use PrEP.

The HVTN has ran and continues to run discordant couple trials in South Africa, where PrEP is not generally available, without providing it. If they did provide it, trials would require many more volunteers and would potentially produce no results...therefore they argue that it is in fact to the benefit of not just humanity, but to South Africans to not provide Prep, even if it is not to the direct benefit of the volunteers.

I personally agree with them....What do you think?

Specific Control type

Can a placebo be used? What is the SOC?

Three placebo vaccinations + SOC - regular safe sex and STI counseling and testing specific to persons life.

Pick an endpoint

what stage of trial is this, remember soft endpoints are fine in early phases...but is this an early phase?

We already have safety and activity information on this vaccine. This is a Phase III trial with a clinical endpoint of HIV infection. But be specific
Time to Incident HIV-1 infections that occurs at least 5 weeks after receiving the last dose of vaccine or placebo.

Translate that into an objective

Be specific!

To determine if this HIV vaccine will reduce incident HIV-1 infections on average when compared to placebo.

Population of interest

What we want and what we can have are different!

All HIV-1-seronegative humans! In reality the best population we can possibly generalize to are HIV-1-seronegative living in South Africa that are healthy enough to receive vaccination.

The sample from Population of interest

We need endpoints and we need volunteers that will stay in the trial

We run in South Africa because unfortunately we need subjects to become infected for the trial to have any power. We also need people willing to be in the trial...it might benefit them, but it might not. We need to balance risk of not observing any HIV infections and risk to the patient.

Enroll volunteers from STI clinics in and around Cape Town.

Sample Size?

There are formulas, there is software, but regardless of what you need to use you need to know

- The end point you plan to use
- Something about the behavior you expect in the control arm
- The treatment effect/ association you want to be able to detect
- **The approximate test you plan to run**
- the power you wish to have to detect the desired difference
- Depending on the endpoint, the variance of the endpoint in the treatment arm

What test to use?

Simple randomized trial with a continuous endpoint

- T-test, Wilcoxon-U rank sum, Z-test ...
- Linear regression, stratified Wilcoxon, adjust for precision variables/covariates

Binary (yes/no) outcome

- Fisher's exact, logistic regression - exact logistic regression
- Unconditional exact test,

Time to event

- logrank, Cox, parametric

What you should use depends on a lot of factors. Often a T-test is a perfectly fine test in a simple randomized trial...is it the 'best' ...? Ask a statistician (shameless plug)

Error and Bias

Randomization, Blinding, Best practice Sampling are to reduce Bias. We also want to reduce the error.

- Power is 1-Type II error - Type II error is not rejecting the null when you should β
- Test Size is Type I error - Type I error is rejecting the null when you shouldn't α

We try to power a trial to have low Type II error but we MUST design a trial so it has controlled Type I error.

Control Type I error

If you pick a standard named analysis, you run 1 of them and that one is the pre-specified analysis you have controlled Type I error.

This is the statistical reason we pre-specify our hypothesis, tests, endpoint etc. to prove that we have control of Type I error and that our p-values mean something.

If you are running more than 1 test, even if it is pre-specified, you may need to adjust the p-value to control the type I error.

Sample Size and Power

ARE EDUCATED GUESSES!

- You are guessing about a lot of things
- Getting null results doesn't automatically mean you got the guess wrong

Sample Size by Simulation

No formulas, No special programs

- Create a data set based on your background information and desired effect size and randomization 1:1 or 2:1 etc.
- Run the test you plan to run in the trial, record if it rejects the null
- Do the above steps many times (100,000 times) and see how often it rejects (Power)

Want to know if the test/software you are using are any good... run this for the null (Type I error)

Sample Size by Formula

There is a simple formula for the $\sim$ sample size of a comparison of two independent means with 1:1 randomization:

$$n = \frac{\delta^2_{\alpha,\beta} * V}{(\mu_1 - \mu_2)^2}$$

- $\delta^2_{\alpha,\beta} = (Z_{1-\alpha/2} + Z_{1-\beta})^2$ For a standard two-side 0.05 test with 90% power this ~ 10.5. $Z_{1-\alpha/2} = Z_{0.975} \approx 1.96$ and $Z_{1-\beta} = Z_{0.9} \approx 1.28$
- V = variance in control group + variance in the treatment group.
- $(\mu_1 - \mu_2)$ is the true difference, not the observed difference you want to be able to detect with power $1 - \beta$

Sample Size by Program

- http://www.stata.com/features/power-and-sample-size/
- PASS
- power.t.test in R (other power calculations in base R)

Test and sample size of Example

- Log rank test accounting for interval censoring for time to infection starting 5 weeks after last vaccination - primary
- Cox model estimated VE by 1-HR
- 3000 subjects to detect 50% VE given expected 8% yearly incident rate in the background to have 99% power
- Sample size was determined by binomial simulation - power should be higher for time to event

This mean the Type II error of the test is $< 1\%$, if you consider p-value < 0.05 significant this single pre-specified logrank test will have 5% type I error. The estimated Cox VE and confidence interval will not match exactly but it will not give a new p-value.

Randomization

How are you going to randomize?

- Hat or Coin flip
- Stratify
- Placebo and/or SOC some other comparator

1:1 stratified by clinic by random block. This means you pick a random even number, then randomly assign to placebo or treatment within that number, within the clinic.

Control Arm and Blinding

Are you going to blind and how?

- to the patients, practitioners, study team, why?

Blinding by placebo, to everyone.

Summary

- Phase III trial to determine if a HIV vaccine reduced the incidence of HIV-1 infection, on average
- We sample from STI clinics in Cape Town because of the high background rate of HIV infection, and we know this limits the population to which we can generalize.
- We randomize and blind to reduce difference in treatment and experiences of subjects.
- We pre-specify the objective, endpoint of incident HIV-1 infections 5 weeks after last dose as well as the testing by Logrank to control the type I error
- We enroll a sample size of 3000 subjects to control the Type II error (increase the likelihood of getting an answer!)

What else do we need?!?!

ITT principle

We worked hard to make and keep those groups we randomized the same so we should analysis them that way.

- If you only analysis those subjects that stayed in the trial and do what they are suppose to this a per-protocol analysis
- ITT means randomized = analyzed.
- Weighting, imputation, to deal with missing data
- Most basic for binary data is worst case scenario
- Composite endpoints where missing data is one of the levels

ITT principle Example

If a vaccine causes unexpected and less then desirable side effects there may be more drop-out and therefore less infections in the vaccine arm.

In a cancer treatment trial, if the treatment works then more SOC subjects may miss visits because they are feeling ill and therefore may have longer times to failure than on treatment. This will cause the treatment to look less helpful than it is.

Ask for Help!

A well designed, run and analyzed clinical trial is a team effort - here is what you need to ask for help

- To get help with sample size - you need a hypothesis, an endpoint, background information and a target effect size
- To get help designing a trial - you need a hypothesis and a population of interest

Clearly money matters, budgets matter, but just like it is unethical to run a trial and get no result it is unethical to spare funds by not getting help just to waste money running a poor trial that doesn't answer any questions. Asking for help before you start the trial will save money in the long run.