Biostatistics for Grants: Pre-Clinical

Robert A. Parker

My Background

- My background is mainly in clinical research
- These slides are based on talks for clinical investigators, but have been somewhat changed for animal / laboratory experiments
- I have worked on some pre-clinical studies (mainly animal studies) but make no claim to be an expert in them

Why Does Biostatistics Matter? (1)

- Rigorous design of studies:
  - what is the question?
  - why does this question matter?
  - how will doing this study answer this question?
  - appropriate basic design
  - appropriate endpoints
  - appropriate measurement techniques
  - appropriate population
  - sample size

Why Does Biostatistics Matter? (2)

- Appropriate analysis
  - valid approach
  - efficient: makes best use of the existing data
  - done correctly

Why Does Biostatistics Matter? (3)

- Appropriate conclusions:
  - valid conclusions based on the results of the study
  - interpretation incorporates uncertainty in the results

Bad analyses can be redone

Bad designs cannot be fixed

Focus on the study design

What Biostatistics Is **NOT** (1)

- Biostatistics focuses on design and statistics
  - ensuring valid results
  - separating signal from noise
- Biostatistics does **NOT** focus on whether results are clinically important

What Biostatistics Is **NOT** (2)

- Ideally you want to be able to say that the difference is ***important***
- ***Important*** is different from ***statistically significant***
- In fact, medical journals routinely use jargon phrases to distinguish the two concepts
  - clinically important
  - statistically significant

What Biostatistics Is **NOT** (3)

"If even one more baby survives, that is clinically important"

Jay Shenai, MD

circa 1990

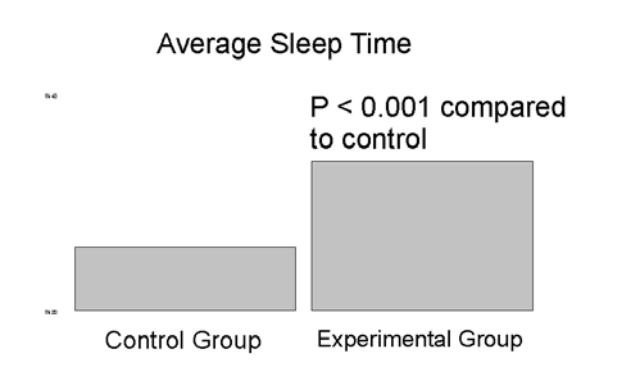
What Biostatistics Is **NOT** (4)

Statement is true ... but no study could ever demonstrate such an effect

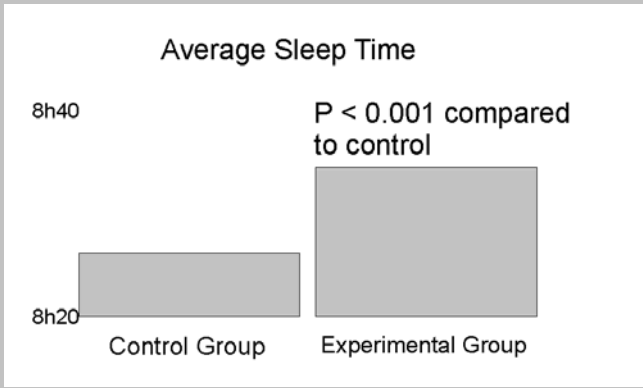
What Biostatistics Is **NOT** (5)

Other studies are able to demonstrate things that are statistically significant, but may not be clinically important

What Biostatistics Is **NOT** (6)



What Biostatistics Is **NOT** (7)



Statistical significance does not imply clinical importance.

Clinical importance does not imply statistical significance.

Outline

- Interventional Study Designs
- Randomization and Masking
- Hypothesis Testing
- Statistical Tests
- P-values
- Material expected for a grant

Outline

- **Interventional Study Designs**
- Randomization and Masking
- Hypothesis Testing
- Statistical Tests
- P-values
- Material expected for a grant

Single-Arm Designs: Do NOT Use Them

Comparative Designs

Comparison Group (1)

Most studies use a comparative design

If you are doing a comparative study,

you need a comparison group

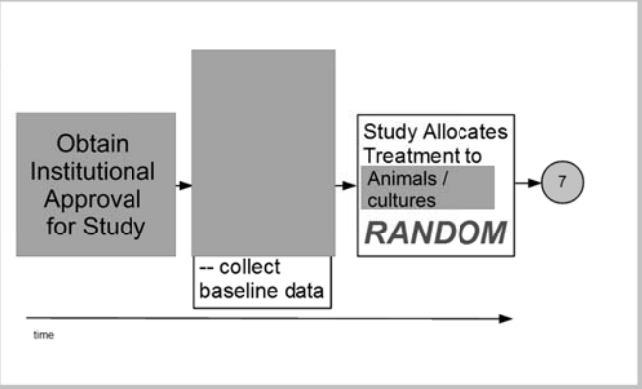
Comparison Group (2)

Three standard “comparators”

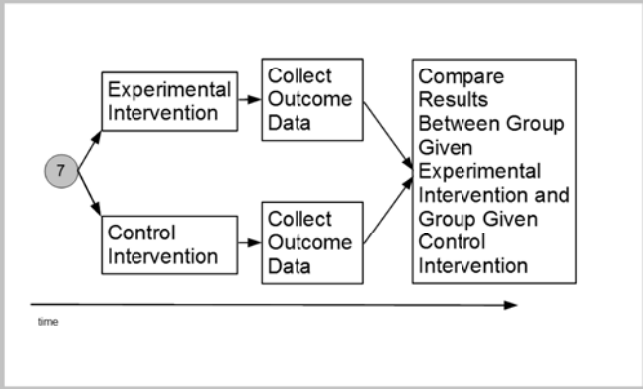
- placebo
- standard of care
- active control

Parallel-Group Designs

Parallel-Group Study (1)



Parallel-Group Study (2)



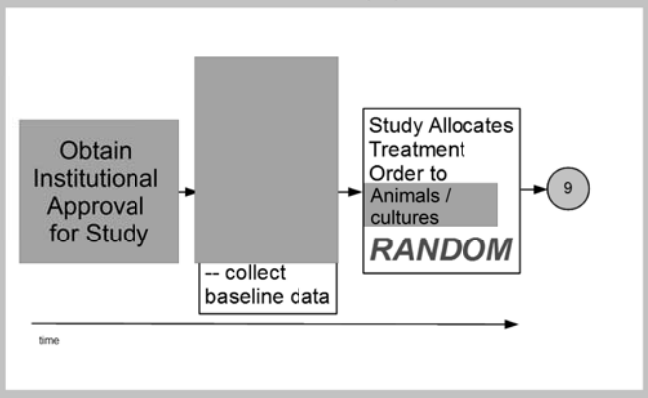
Parallel-Group Study (3)

- The standard parallel-group study
  - can have more than two groups, for example a control group and multiple dose levels
- Study assigns the intervention
  - assignment should be ***randomized***

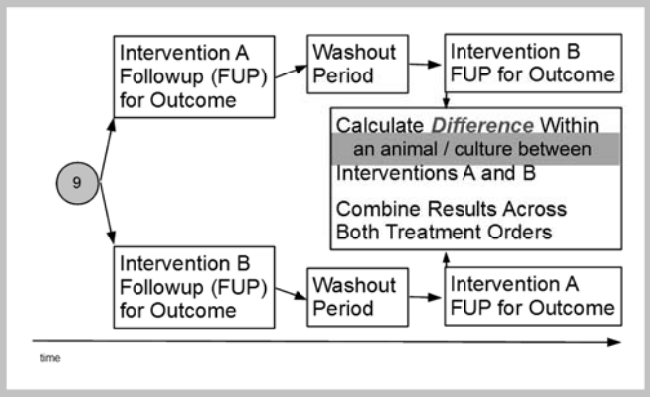
By far the most common design used

Within-Subject Designs: NEED COMPELLING RATIONALE TO DO THIS

Cross-Over Studies (1)



Cross-Over Studies (2)



Bias: The Threat to Validity

Bias (1)

***Bias*: the results observed reflect other factors in addition to (or even instead of) the effect of the treatment**

Bias (2)

- there are multiple potential sources of bias
- it is impossible to completely eliminate the possibility of bias
- it is possible to minimize some of the major biases with careful planning
- the accusation that a bias ***may*** exist is often sufficient to cause the validity of a study to be generally questioned

Bias (3)

Assessor Bias

- the assessor's knowledge of which treatment the animal is receiving ***may*** affect the way the assessor assesses outcome
- such a bias would directly affect the validity of the conclusions of the study

Bias (4)

Laboratory Bias

- the technicians knowledge of which treatment the specimen comes from ***may*** affect the way the test is done / interpreted / read
- such a bias would directly affect the validity of the conclusions of the study

Outline

- Interventional Study Designs
- **Randomization and Masking**
- Hypothesis Testing
- Statistical Tests
- P-values
- Material expected for a grant

Treatment Allocation (Randomization)

Treatment Allocation (1)

Treatment allocation **must** be random

Treatment Allocation (2)

- Assigning animals alternately to treatment group is **not** random assignment
- Assigning the first half of the population to one group is **not** random assignment

Treatment Allocation (3)

Randomization implies that treatment allocation is assigned ***to each individual animal after*** the animal is included into the study

- minimizes the possibility that there is some systematic difference between the two groups -- whether done knowingly or unknowingly

Treatment Allocation (4)

- Random allocation does not imply that there are equal numbers in each group
- Equal numbers in each group (1:1 [experimental:control] randomization) maximizes power for total number of animals under usual assumptions

Treatment Allocation (5)

- Randomization does not imply that one cannot "group" animals in various ways
- Two common methods of grouping are:
  - blocking
  - stratification

Treatment Allocation (6)
Blocking

- animals are grouped into ***blocks*** by time of treatment (“batches”)
- animals within each block are randomized separately
- ***Blocking*** ensures that animals are reasonably well balanced across treatment over time
  - eliminates confounding because of batch variability

Treatment Allocation (7)
Stratification

- Stratification is a method to ensure balance on a specific factor (or factors) which ***are*** used in the randomization
- Stratification does ***not*** ensure balance between other factors which ***are not*** included in the randomization
- Common stratification factors would be gender and age
- Randomization done separately in each stratum

Masking
(also called blinding)

Masking (1)

Almost all of the ***potential*** biases can be minimized if ***everyone*** involved in the study is masked to the actual treatment the patient is receiving

Masking (2)

Masking (also called ***blinding***) is intended to avoid biases due to knowledge of treatment

Masking (3)
Hierarchy of Masking

- ***open label***: no masking **NOT RELEVANT**
- ***single blind***: patient (usually; occasionally may be assessor) masked to treatment
- ***double blind***: patient and assessors (who often are also the health care providers and data collectors) masked to treatment
- ***complete masking***: everyone involved in study masked to treatment

Masking (4)

- the real decision is whether the person doing the experiment (the “assessor”) is masked or not
- if you know the treatment, then you can bias the results (consciously or unconsciously)
- since you can always mask treatment to an animal, all your studies should be double-blind if you can

Outline

- Interventional Study Designs
- Randomization and Masking
- **Hypothesis Testing**
- Statistical Tests
- P-values
- Material expected for a grant

Hypothesis Testing (1)

- A formal method to make ***statistical inferences*** from data
- It answers the question whether the results are ***statistically significant*** or not

Statistical significance does not imply clinical importance.

Clinical importance does not imply statistical significance.

Hypothesis Testing (2)

Components:

- Null hypothesis
- Alternative hypothesis
- P-value
- Type I and II errors
- Power

Hypothesis Testing (3)
Null Hypothesis

- what you want to **disprove**
- generally, the idea that there is no difference between two treatments
  - applies to studies in which you are trying to show that the experimental treatment is better than the standard treatment (or placebo)
- a "straw man"
  - there is always a difference if you can measure something accurately enough

Hypothesis Testing (4)
Alternative Hypothesis

- what you hope to show
- generally, that there is a difference between two treatments
- although the alternative hypothesis may be general (treatment is better than control), to calculate statistical properties for the alternative hypothesis a ***specific value*** for this difference is needed
  - this difference is ***specified*** by the investigator

Hypothesis Testing (5)
P-value

- the probability that the data (or something more extreme) could have been observed ***if*** the ***null hypothesis*** were true
- P-value calculations are not valid if there is ***systematic bias*** in the results

Hypothesis Testing (6)

Types of Errors (1)

Verdict	Defendant	
	Not Guilty	Guilty
Case Not Proven	Correct	Type II
Guilty	Type I	Correct (Power)

Hypothesis Testing (7)

Types of Errors (2)

Decision	Truth	
	Null	Alternative
Fail to Reject Null Hypothesis	Correct $1-\alpha$	Type II β
Reject Null Hypothesis	Type I α	Correct (Power) $1-\beta$

Hypothesis Testing (8)

Types of Errors (3)

Type I error

- ***rejecting the null hypothesis*** even though the null hypothesis is true
  - convicting an innocent person
- the probability of this is denoted by α and is ***pre-specified*** by the investigator
  - conventionally α = 0.05, two-sided
  - if the P-value is less than or equal to α, the results are called "statistically significant"
- considered more serious error, so it is Type I

Aside

- It is much more informative to use the P-value than to only say something is statistically significant or not
- Both P=0.049 and P=0.001 are statistically significant (for α=0.05), but the evidence is much stronger for a signal when P=0.001
- The cutoff α=0.05 (or any other level) is arbitrary

Hypothesis Testing (9)

Types of Errors (4)

Type II error

- ***failing to reject the null hypothesis*** even though the ***specific value*** of the alternative is true
  - not convicting a guilty person
- the probability of this is denoted by β
- considered less serious error, so it is Type II

Hypothesis Testing (10)

Types of Errors (5)

- increasing Type I error (α level) lowers Type II error
  - more likely to make a Type I error, and reject the null hypothesis even though it is true
  - less likely to make a Type II error, and fail to reject the null hypothesis when the alternative hypothesis is true
  - same as convicting on less evidence in a trial: more innocent people convicted, fewer guilty people are let off
- opposite is also true

Hypothesis Testing (11)

Power

- the chance of ***rejecting the null hypothesis*** if the ***specific value*** of the alternative hypothesis is true
  - convicting a guilty person
- the chance of making the ***right decision*** when the ***specific value*** of the alternative is true
- the probability of this is equal to 1-β

Outline

- Interventional Study Designs
- Randomization and Masking
- Hypothesis Testing
- **Statistical Tests**
- Assumptions of P-values
- Material expected for a grant

Statistical Tests (1)

- Fundamental idea:
- Calculate “effect size” (= difference from expected result)
- Calculate “noise” (standard error)
- Compare the ratio of

  effect size / expected noise

  to the theoretical distribution: large ratio implies “statistically significant”

Statistical Tests (2)

Is treatment associated with complete recovery?

	Complete	Partial
ExptTrmt	10	90
Control	5	45

No evidence in this data:

- 10% in the ExptTrmt group completely recovered
- 10% in the Control group completely recovered

Statistical Tests (3)

Is treatment associated with complete recovery?

	Complete	Partial
ExptTrmt	10	90
Control	4	46

Slight evidence in this data:

- 10% in the ExptTrmt group completely recovered
- 8% in the Control group completely recovered

Statistical Tests (4)

Observed data

	Complete	Partial	%Complete
ExptTrmt	10	90	10%
Control	4	46	8%

Expected data

	Complete	Partial
ExptTrmt	9.33	90.67
Control	4.67	45.33

Effect size is 0.67 (plus or minus) (=10-9.33)

Statistical Tests (5)

Observed data

	Complete	Partial	Total
ExptTrmt	10	90	100
Control	4	46	50
Total	14	136	150

Calculation: expected in the ExptTrmt group with complete recovery:

150 (total) x (14/150) x (100/150) = 9.33

Statistical Tests (6)

Observed data

	Complete	Partial	Total
ExptTrmt	10	90	100
Control	4	46	50
Total	14	136	150

Calculation: expected in the Control group with partial recovery:

150 (total) x (136/150) x (50/150) = 45.33

Statistical Tests (7)

Expected data

	Complete	Partial
ExptTrmt	9.33	90.67
Control	4.67	45.33

Effect size is 0.67 (plus or minus)

Noise: 1/sqrt[1/9.33+ 1/90.67 + 1/4.67 + 1/45.33] = 1.68

Ratio = 0.67/1.68 = 0.40, P-value = 0.7

Statistical Tests (8)

Observed data

	Complete	Partial	%Complete
ExptTrmt	200	1800	10%
Control	80	920	8%

Expected data

	Complete	Partial
ExptTrmt	186.67	1813.33
Control	93.33	906.67

Effect size is 13.33 (plus or minus)

Statistical Tests (9)

Expected data

	Complete	Partial
ExptTrmt	186.67	1813.33
Control	93.33	906.67

Effect size is 13.33 (plus or minus)

Noise: 1/sqrt[1/186.67+ 1/1813.33 + 1/93.33 + 1/906.67] = 7.51

Ratio = 13.33/7.51 = 1.78, P-value = 0.08

Statistical Tests (10)

- We have what most would consider to be a clinically important difference in complete recovery (10% vs 8% would be considered important -- it is 25% more people with complete recovery)
- We have a very large study (3000 people)
- But it is still just noise -- not statistically significant!

Statistical Tests (11)

- There are lots and lots and lots of different tests for statistical significance
- Some of the most common are listed on the next few slides

Statistical Tests (12)

- Continuous Data: Within a Group
  - **(Paired) T-test** or **Wilcoxon Signed Rank Test**
- Continuous Data: Between Groups
  - **T-test** (2 groups) or **Analysis of Variance** (> 2 groups)
  - **Wilcoxon Rank Sum Test** (2 groups) or **Kruskal-Wallis Test** (> 2 groups)
- Continuous Data: Relationship Among Variables
  - **Pearson correlation** or **Spearman's Rho**

Statistical Tests (13)

- Categorical data
  - **Fisher's Exact Test** (and extensions) or **Chi-square Tests** (and extensions)
  - **McNemar's Test** (matched data)
- Survival Data (censored data in general)
  - **Log-rank test**

Outline

- Interventional Study Designs
- Randomization and Masking
- Hypothesis Testing
- Statistical Tests
- **P-values**
- Material expected for a grant

P-values (1)

- the probability that the data (or something more extreme) could have been observed ***if*** the ***null hypothesis*** were true
- P-value calculations are not valid if there is ***systematic bias*** in the results

P-values (2)

- there are always additional assumptions when calculating a P-value
- the most basic assumption is that the assignment to group is done at random
  - this implies that there should not be a bias between groups in the way the groups were formed

P-values (3)

- another basic assumption is that the observations are ***independent***
- this means that the result for one measurement is not affected by another measurement
- for matched data (like measurements before and after the intervention) -- the pair is the unit that is independent
- this independence assumption is critical to the interpretation of any set of data

P-values (4)

There are other standard assumptions involved for most significance tests:

- particular underlying distribution (usually the "normal" distribution which is the bell-shaped curve)
  - this also implies symmetry
- common variability ("noise") between groups
  - special tests exist which take these differences into account

P-values (5)

- P-values are based on assumptions
- if the assumptions are not true, then the P-values are not valid

Outline

- Interventional Study Designs
- Randomization and Masking
- Hypothesis Testing
- Statistical Tests
- P-values
- **Material expected for a grant**

Grants (1)

Background on your projects

- 16 abstracts received
- about 10 appear to involve in vivo experiments
  - animals
- some involve human tissue, but appear to be bench work with it
- 2 seem to me to be engineering projects

Grants (2)

All grants need

- clarity
  - what you are doing
  - why you are doing it
  - why it matters
  - how results are judged

Grants (3)

- goal is to let the reviewer like the grant
- **you** must make it easy for the reviewer to like the grant
- so you must make it easy to understand
- this means you need coherence / consistency
  - background / pilot data support aims
  - terms used consistently
  - aims need to hang together
  - **WRITTEN CLEARLY**

Grants (4)

So

- have someone else read it (cold and fast) who is not an expert in your field (not your mentor, not the person next to you in the lab)
- assume that they are RIGHT when they say something doesn't make sense
- fix it / change it / clarify it

because the reviewers will have the same problems

Grants (5)

- If there are potential weaknesses in the grant, you should raise them and discuss why they are (a) really not that bad and (b) unavoidable to pursue this line of research
  - must be convincingly discussed: if it really is that bad why are you wasting your time writing this grant?
  - makes you seem more competent
  - reviewer does not get a chance to score points by finding a "gotcha" in your grant

Grants (6)

What does this mean about the
statistical material
in your application?

Grants (7)
Design (1)

Design of the experiments must be clear (1)

- what you are doing
  - experimental groups / conditions
  - controls
  - how many of them you are doing
- how you are doing it
  - this needs to be the "big picture" view of the experiment

Grants (8)
Design (2)

Design of the experiments must be clear (2)

- how you are assessing the results
  - how do you draw conclusions from the results of your experiment
    - it may be self-evident to you, but it won't be to the reviewers
  - why your design ensures that if you get the results you hope for, you can make conclusions from it
    - it may be self-evident to you, but it won't be to the reviewers

Grants (9)
Design (3)

Design of the experiments must be clear (3)

- why the experiment you propose will answer the question you are asking
  - this should be clear from the background / pilot data but you should tie it together for the reviewer
  - if you can not describe this in a short paragraph, you need to think about what you are doing, since you have focused on the details, not on why the experiment matters

Grants (10)

- **design is the fundamental statistical issue**
- if the design is not right nothing can save the project

Bad analyses can be redone

Bad designs cannot be fixed

Focus on the study design

Grants (11)
Statistical Material (1)

Statistical Material

- sample size
- data analysis
- data management
  - might be needed for some human projects, but not relevant to these types of projects

Grants (12)
Statistical Material (2)

My normal writeup in a grant would be organized like this:

Statistical considerations

Sample size ...

Data analysis ...

Grants (13)
Statistical Material (3)

Sample size (1)

- replicates: need to be clear whether and how many you are using
- animals: need to be clear what the number used in each group will be
- should provide some rationale for this number
- the standard approach is based on a power analysis (expected for animals)

Grants (14)
Statistical Material (4)

Sample size (2): elements of a power statement

- effect size
- variability -- usually reported as standard deviation
- statistical characteristics:
  - alpha (two-sided)
  - power
- which gives you the magic number N

Grants (15)
Statistical Material (5)

Sample size (3)

- need to account for potential problems with the experiments
  - animals dying before you can make the measurement (if death is not the endpoint)
  - animals not getting the syndrome (if inducing the syndrome chemically)

Grants (16)
Statistical Material (6)

Sample size (4)

I personally have no problem with a power statement like this:

- As this line of investigation is worth pursuing only with clear evidence of a dramatic effect (at least twice the standard deviation observed among the animals), we will need 6 animals / group to have > 85% power to detect a difference in <whatever> between the <experimental name> and the control group (alpha=0.05, two-sided). We anticipate that 20-25% of animals will <whatever>, so we will begin with 8 animals / group.

Aside

The magic number 6 assumes a two-group comparison using a t-test.

Grants (17)
Statistical Material (7)

Sample size (5)

- some statisticians would be happier with specific numbers for the effect and the standard deviation
- standard deviation could be justified based on your pilot results if they are suitable

Grants (18)
Statistical Material (8)

Data analysis (1)

- if possible, mention plotting your data
  - mention a couple of specific plots as examples
- looking at data is always the way to begin
- be simple and straight forward
  - use techniques that you can do yourself
- be brief
  - the more details you give the more chance you have of making a mistake

Grants (19)
Statistical Material (9)

Data analysis (2)

- do not get too fancy
  - this raises concerns, unless you have a statistician on the grant
- if you attempt to fake it, it will probably be obvious to the reviewer
  - never ever bluff

Grants (20)
Statistical Material (10)

- Continuous Data: Within a Group
  - **(Paired) T-test** or **Wilcoxon Signed Rank Test**
- Continuous Data: Between Groups
  - **T-test** (2 groups) or **Analysis of Variance** (> 2 groups)
  - **Wilcoxon Rank Sum Test** (2 groups) or **Kruskal-Wallis Test** (> 2 groups)
- Continuous Data: Relationship Among Variables
  - **Pearson correlation** or **Spearman's Rho**

Grants (21)
Statistical Material (11)

- Categorical data
  - **Fisher's Exact Test** (and extensions) or **Chi-square Tests** (and extensions)
  - **McNemar's Test** (matched data)
- Survival Data (censored data in general)
  - **Log-rank test**

What You Need to Remember

Remember (1)

Hypothesis testing is a formal method to make *statistical inferences* about the results

- identifies "signals" from "noise"
- there are two different ways of making an error in hypothesis testing: deciding there is a signal when there is not one, or deciding that there is no signal when there is one
- based on the P-value

Remember (2)

P-values

- are based on the ratio: “effect” / “noise” assuming that the null hypothesis were true
- assumes that there are no biases in the results
- makes many other assumptions as well
- even more assumptions involved when the results are based on modeling

Statistical significance does not imply clinical importance.

Clinical importance does not imply statistical significance.

Remember (3)

- ask someone to read the grant who has not read it before
- take their comments seriously
- give yourself enough time to
  - get over the annoyance that someone raises issues
  - fix the grant before submitting it

Remember (4)

Grant must be clear

- what you are doing
- why you are doing it
- why what you are doing answers the question

Remember (5)

- if using animals (or people): the number needs to be clear, and have a reasonable justification
- if doing data analysis, it should be straightforward unless you are working with a statistician (or are yourself an expert)

Focus on the study design