Designing a Research Study

Timothy Petersen

Follow this and additional works at: https://digitalrepository.unm.edu/anesthesiology_pubs

Recommended Citation

This Presentation is brought to you for free and open access by the Anesthesiology at UNM Digital Repository. It has been accepted for inclusion in Anesthesiology Research and Scholarship by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Designing a Research Study

What do you mean- “design” a study?

- Studies don’t come in a box
- Many things to consider
- Decisions to make
  - None need be prohibitive or scary
  - But each one matters
  - Some will even seem automatic (yay!)
- “Study design” is just the sum of these
No study is perfect

- So let your goal be “good enough”… and exhale

So what’s involved?

- Settling on your research question/hypothesis
- Choosing an overall approach
- Deciding which data to gather
- And how many subjects you’ll need
- Reducing bias with randomization & blinding
- Then write the protocol
- Keep those pesky rules and expectations in mind
Time invested up front…

- Is time saved / not wasted during:
  - IRB approval process
  - data collection
  - analysis
  - writing
  - peer review

The research question

Begin with a basic idea
Keep an eye out for opportunities

- Here we do things this way, but at my old institution we did things that way
- Somebody’s passing comment or odd question
- Unresolved questions in literature: review article, intro, discussion section, etc
- Disagreements among colleagues: wanna bet?
- Interesting article: tweak it (this is almost always possible!)
- New-ish treatment with inexplicable popularity
- Planned change to a treatment pathway
- They say always/never do XYZ: evidence for Dr. They’s position?

Start informally

- Can I reduce the amount of LA used in this block and still retain effectiveness?
- Which grip is best for novices on their first efforts at mask ventilation?
- Does it matter which brand of block needle I use?
- Does this drug really reduce intraop blood loss?
- What’s the best sedation protocol for this particular set of pediatric imaging patients?
Hit the literature

An example

Starting with outcomes selection
Compare treatments’ effect on postop pain

- Which treatment better controls postoperative pain?

---

Moving past “what’s better?”

- Formalize the comparison
- Consider all salient points of the setting
  - Which providers?
  - Which patient population?
  - What treatments/groups?
  - What outcome(s)?
Outcomes

- What will you measure?
  - One primary outcome
  - A few secondary ones

- Surrogate vs. “real” clinical outcomes (it’s a spectrum)
  - Lab values, etc
  - Complications, survival, pain-free time, etc

Compare treatments’ effect on postop pain

- Time to first request of pain meds
- Time to first report of any sensation
- Time to first report of pain
- Total opioid consumption, within XX time period
- Max pain score in XX time period; resting or dynamic
- Patient satisfaction overall, or specifically with pain control
- Proportion of patients who ever hit, say, ≥8 on pain scale
- Reduction of opioid-related side effects
- Etc.

Why did you pick this one?
Why not just test all of ‘em?

- Problem of multiple comparisons
- Shorthand: at 0.05 significance level, we have a 95% chance of being “right” on a given test
- With two tests, the chance of being right twice (no errors) is just over 90%
- Ten: 60%
- Twenty: 36% (that’s a 64% chance of ≥1 spurious result!)
- So use statistical tests sparingly
- Adjustments are available, but they’re harsh

It’s a balance

- Clinical interest
- Ease of data collection
- Intended knowledge gap to fill

- That’s the whole point of this talk
Clinical significance

- Always keep this in mind

- I can design a study that will show that donuts increase the relative risk of thumb cancer by 3%

- Who cares?
2-tailed vs. 1-tailed

- **2-tailed analyses**
  - Is there any difference between these treatments?
  - Null hypothesis: they are equal
  - The default

- **1-tailed analyses**
  - We have some solid reason to think that A is better than B
  - Is that really the case?
  - Null hypothesis: they are equal, or B is better
  - Being more specific yields a \( p \)-value bonus (\( p/2 \))
  - Less common
Hypothesis/ Research Question

- Should be succinct but specific
  - We hypothesized that the addition of dexamethasone 8 mg to ropivacaine-based sciatic nerve block would result in a delay in patients’ first request for pain medication, as compared to preop IV administration of the same dose.

- Primary outcome
  - Time to first request of pain medication

- Secondary outcomes
  - Total opioid consumption within first 48 hours postop

Selecting the design
Some of the main types (for us)

- When patients are enrolled, and what happens
  - Prospective
  - Retrospective
  - Observational

- Comparison: superiority vs. equivalence vs. noninferiority
  - Are these different/ is one better?
  - Are they the same (within limits)?
  - Is this one at least not worse than that one?

Benefits and Costs

- Prospective
  - Randomization
  - Consent refusals

- Retrospective
  - Ease of data collection
  - Limited to what’s there

- Observational
  - 100% data capture!
  - Can’t manipulate treatment
More on Randomization

- From a scientific perspective, it’s almost always best
  - But maybe not logistically

- Or maybe it’s just not a good fit for your question

- Sometimes you just want to know how often something happens in the real world

- We’ll come back to this

Moving on to the comparison itself...

- Superiority

- Equivalence

- Noninferiority
Superiority trials

- But wait… let’s have a brief tangent

Confidence interval

- A statement of probability
- Usually a 95% CI
  - “The difference between the group means was 6.5 units (95% CI 3-10).”
- If we were to do this study many times, 95% of the resulting CIs would contain the true difference.
- If $p=0.05$, the 95% CI has zero at one end (e.g. 0 – 3 units)
- If $p > 0.05$, it spans 0
- If $p < 0.05$, it does not
- The CI for a 1-tailed test only omits 5% (say) at one end
OK, getting back to it…

Superiority trials

- So common they’re the default
- Do treatments A and B provide different results on this outcome?
- Hypothesis
  - A is different from B
- Null hypothesis
  - A and B are equivalent
- Hope to get a 95% CI that excludes 0
- Can be 2-tailed or 1-tailed
Equivalence trials

- Treatment A is cheaper, easier, etc than treatment B
- Are the clinical outcomes any different?
- Need an *a priori* clinically significant idea of “different”: Δ
- Hypothesis
  - $-\Delta < 95\% \text{ CI for difference} < \Delta$
- Null hypothesis
  - 95% CI contains Δ or $-\Delta$ (or both)
- Hope to get a 95% CI that fits within $\pm\Delta$
- Must be 2-tailed

Noninferiority trials

- Hybrid of superiority and equivalence; imagine a 1-tailed equivalence trial
- Is treatment A *at least not worse* than treatment B?
  - Shorthand: $A - B \geq 0$
- Still need Δ
- Hypothesis
  - $-\Delta < 95\% \text{ CI for difference}$ (which is infinite on this side)
- Null hypothesis
  - 95% CI includes $-\Delta$
95% CI results and trial types

Results: groups’ difference

Reject null hypothesis?

<table>
<thead>
<tr>
<th></th>
<th>Superiority</th>
<th>Equivalence</th>
<th>Noninferiority</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Δ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>0</td>
<td>✓ -ish</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Δ</td>
<td>✓</td>
<td>✓ , but…</td>
<td>✓</td>
</tr>
</tbody>
</table>

Data to gather
So many data...

- How do I select from the universe of data?

Where to start?

- Age, sex, BMI, etc unless there’s a reason not to
- The outcomes of interest (obviously)
- So many confounders….
  - Beware the rabbit hole
  - Show your groups to be similar enough
  - Consider excluding problem people
- Try to keep data collection simple
  - Number of sources of info; time investment
  - Certain data require HIPAA authorization (:: consent)
    - Worth it?
OK, I’ve decided what data to gather

How many times must I do it? And to whom?

What’s a power analysis?

- Usually, an estimate of the needed sample size
- Based on certain knowledge or assumptions
  - Desired power
  - Type I error rate: $\alpha$ (the $p$ value threshold)
  - Expected effect size (for specific outcome!)
  - Expected variation within groups
  - The chosen statistical test
- Always ask about this; journals and IRB expect it
Power analysis

- **Power**
  - Chance of *avoiding* a Type 2 error: i.e. false negative.
  - $1 - \beta$ (where $\beta$ = type 2 error risk)
  - Usually set at 80%; typically higher with high-benefit studies
  - “If there’s anything there, will we see it?”

- **Alpha (significance threshold)**
  - Chance of having a Type I error: i.e. false positive.
  - Usually set at 0.05; lower with high-risk studies
  - “Will our result be reliable?”

Power, continued

- **Effect size**
  - An estimate of the expected difference between groups

- **Expected variation (e.g. standard deviation)**
  - Within-group variation

- **Where to get these?**
  - Literature
  - Pilot study
  - Clinical experience
  - Minimal clinically-significant effect
Sometimes you really don’t know

- What then?
- “Convenience sample”
- Should still justify the chosen sample size
- With 2 of 3, can calculate the third (all else equal):
  - Sample size
  - Power
  - Effect size (maybe as a multiple of standard deviation)

Equivalence vs. superiority: sample size

- Superiority trials are more efficient

- Rule of thumb: allow 4x sample for equivalence trial as in a corresponding superiority trial
The caveat

- With a superiority trial, a negative result (no stat-sig difference) does *not* mean the treatments are equivalent!
  - Unless the 95% CI somehow managed to be within $\pm \Delta$ anyway

Inclusion and exclusion criteria

- **Inclusion**
  - Usually a shorter list
  - Who do you want?
    - Age $\geq 18$, having surgery, planned nerve block, parturients, etc

- **Exclusion**
  - Can be a longer list
  - Who do you *not* want?
    - E.g. LA allergy in a nerve block study, chronic pain, dementia, prisoners, etc

- **Balance “clean” data vs. generalizability**
Arm allocation

Randomize. Usually

Benefits, etc

- Helps mitigate systematic error
  - Learning effects
  - Staff changes
  - Seasonal variation in patient health
  - Weird stuff that nobody thought of
  - Etc.

- When might it be inappropriate?
  - Investigating effect of a nonrandomizable demographic variable
  - Observational or retrospective studies
What to do

- Use a randomization service:
  - random.org
  - randomization.com
- Conceal allocations until the last moment
  - E.g. sealed numbered envelopes
- Blinding
  - Patient, provider to extent possible, assessor
  - Semiblinded data for analyst (e.g. group 1 vs group 2)

Examples of bad “randomization”

- Coin toss by investigator
- A – B – A – B – A – B
  - Etc., such as AAAA… BBBB…
- Visible allocation list
- Allocation bias is almost never deliberate, but it still affects results
Writing the protocol

What does a protocol do?

- It describes the planned study
  - Justification, background
  - Goals
  - Methods
    - Sample
    - Outcomes
    - Logistics
    - Standards for observations
    - Analysis factors
  - It’s the cookbook
Stuff to keep in mind

- Balance of competing constraints
  - Logistics
  - Sample size
  - Consent
  - Randomization
  - Data-collection duration
  - Not a perfect world, and you don’t have infinite money
  - Circumstances vary. One study’s awesome approach may be terrible in another
More stuff

- Anticipate the criticism: what could be done better?
  - Think of some articles you’ve found to be less than convincing
  - What would happen if you made small changes?
    - Stay flexible during planning
    - Err on the side of simplification
  - What would this study look like under a different strategy: observational, retrospective, prospective?
    - Can you still answer your research question?
    - Is another approach better, cheaper, faster, more awesome?

The protocol

- Background
- Hypotheses
- Outcomes primary and secondary
- Sample
  - Inclusion/exclusion criteria
  - Specific or general sample? Intended generalization
  - Power analysis
- Stated standards for observations
  - Obviously needed for subjective data
  - Objective data: specified time points, methods for observation…
Protocol, continued

- Data management
  - How will it be kept? When will identifiers be removed?

- Planned analyses and statistics
  - $p$ threshold
  - Any interim analysis?
  - Be warned: any post-hoc analyses must be clearly labeled in the poster/manuscript
  - We're not discussing statistical techniques today

Keeping important people happy
Regulatory stuff, etc

- IRB
  - CITI, COI training
  - Consent language

- Clinicaltrials.gov
  - Many journals require prospective registration of clinical trials

- CONSORT diagram
  - Keep a count of exclusions/ consent refusals/ loss to followup
Regulatory stuff, etc

- IRB
- Clinicaltrials.gov
- CONSORT diagram
- DSMB?
- FDA?
- Pre-Award?
- VA?

It’s not so scary
Seriously – it’s not

- There is still lots of room for small studies

- “In a given situation, should I do this, or should I do that?”
  - How would you know?
  - Now you’re halfway there

Recommended

- “Bad Science” by Ben Goldacre
- BMJ “How to read a paper” collection online
- “How to Lie with Statistics” by Darrell Huff (classic)
This is the end

My only friend, the end