Quantitative Methodology and Analysis Guidebook for Medical Educators

(starting with the two most common research designs that come through the Education Evaluation Core)

By Lindsay Demers, PhD

Director, Education Evaluation Core

Boston University School of Medicine

Caveat:

The overarching goal of this document is to provide clinician and scientist-educators with a methodological and analytic road map for medical education research. The level of detail included here is not sufficient to gain expertise in these areas; rather, think of this as a document you can use for guidance when designing your research and making analytic choices. If you find a topic that is of substantial relevance to your work, please email me for additional resources on the topic, as only the bare minimum of detail is included here.

Further, I have only included here what seems to be most commonly used within the department of medicine. It is likely you will encounter instances that are not covered in this guidebook. If/when this occurs email the Education Evaluation Core for a consultation (if you're within the DoM or part of CELP) or email ldemers@bu.edu for additional resources.

Commonly Used Research Designs in Medical Education

The following are the two most commonly research designs used in quantitative medical education research *in my experience*.

Pre-/Post- (also known as within-subjects design)

The pre-/post design is often used in medical education to look at change within individuals. Typically, a survey or assessment is given prior to some intervention and then given again after participation in the intervention. Often, we look for changes in confidence, comfort, awareness and knowledge before and after participation in an intervention.

This research design is adequate if your scholarly dissemination will be focused primarily on curricular material (such as in MedEdPORTAL or an Education Innovation). However, if you're planning to submit a more traditional research paper, where the data and curricular materials have more equal weight, this design may not be sufficient for higher-tier journals. **It is crucial to look at the journals you plan to publish in to see if they often include pre-/post- designs in their original research articles.** If they don't, you'll want to consider other research designs described later or looking at other journals.

Key Considerations when Using Pre-/Post- Design

- **Be sure you have a way to link the pre- and post- data**
    - To do this anonymously, we often ask participants to create their own unique identifier.
        - Think of questions that you're asked when you forget a password online – things that are not going to change.
            - E.g., First two letters of mother's maiden name, birth month as a number, first two letters of the street you grew up on
        - It helps to explain to participants why we are asking for this information. Otherwise you may get people who don't fill it out seriously and then you will be unable to use their data because you cannot link it.
- **Be sure you're setting yourself up to detect pre-/post- change if it occurs**
    - If you have knowledge items that are too easy, or if you ask a question that many people will strongly agree with at baseline, even if your intervention is amazing, you will not be able to see any change because participants were already at ceiling prior to the intervention.
    - If you can, pilot your questions, or have them reviewed by another content expert.
    - Generally, it is better to include more challenging knowledge questions than easier ones on pre-post knowledge assessments.

Possible Limitations of Using Pre-/Post- Design

- **There is no data showing what would have happened in the absence of the intervention (also known as a counterfactual).** For this reason, you cannot infer causation from your data – meaning that you cannot say unequivocally that your intervention is what caused any change you see.
    - This limitation may particularly be an issue when you're asking about concepts such as confidence or comfort. Unlike a knowledge item, which is harder to "fake", due to social

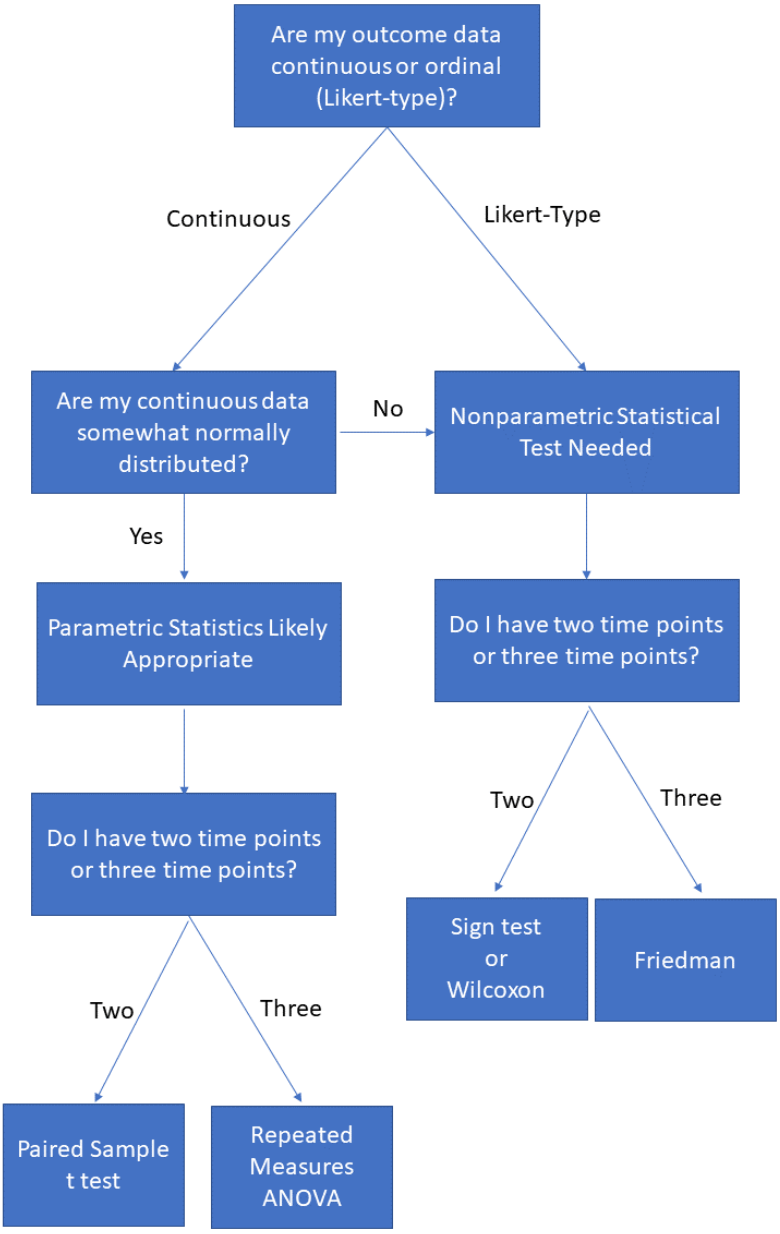
desirability, someone may (consciously or unconsciously) report increased confidence, comfort, etc. after participating in an intervention.
    - For knowledge items, be sure that participants are not explicitly told what, if any, items they got incorrect on the pre- assessment prior to taking the post- assessment. Presumably they will glean this knowledge themselves throughout the intervention and change their answers based on the knowledge they've gained. However, if they are explicitly told what they were wrong about this can make any knowledge change you see less robust.
- **There is a possibility of self-selection bias**, meaning that if your intervention is something that people opt into or volunteer for, they may be fundamentally different from people who choose not to participate. That pre-existing difference between volunteers and non-volunteers may be what is driving the pre-/post- change
    - For example, say you have 20 residents sign up for a two-week long intensive bootcamp on a specific topic and you do a pre- assessment at the beginning of the two weeks and a post- assessment at the end of the two weeks. If you see statistically significant change within participants, it could be explained by the fact that this is a group of residents who are highly motivated to learn about this topic. For that reason, the learning or change you see in the participants may not generalize to those who did not sign up to participate.

Tips for Including a Delayed Post

For many medical education studies that use a pre-/post- design, it can be useful to include a delayed post assessment as well. For example, you cannot ask about behavior change immediately after the intervention because participants will not have had a chance to put anything into practice yet.

- **The length of time between your immediate post and delayed post assessment should be chosen thoughtfully** – how long might it take, realistically, for you to see the changes in behavior, practice, etc. that you're hoping to see?
- **Generally speaking it is harder to get a good survey response rate (i.e.,>80%) on a delayed post survey** because your participants are not in the room (or on Zoom) with you.
    - Be sure you let participants know during the session that you will be sending out a delayed post assessment in X months so that they have it on their radar.
    - If you have any money, consider offering an incentive for participation in the delayed post survey
        - I often use a drawing for a Visa gift card
- **Expect that knowledge gains/comfort/awareness may be attenuated in the delayed post**.
    - If people aren't using the information you provided them with daily, it is likely that there will be some decay in (e.g.) knowledge from immediate post to delayed post. This pattern of results is fairly common. The hope is that they are still outperforming themselves prior to participation in the intervention.
- **If you are interested primarily in changes in behavior or practice**, consider having participants keep a diary or journal of some kind so you can easily track these changes.

Data Analysis Decision Tree for Within-Subjects Designs that include 2 (pre-/post-) or 3 (pre-/immediate post-/delayed post-) timepoints*



*This decision tree is based on the two most commonly used types of data in medical education in my experience (i.e., continuous and ordinal). If your data fall into neither of these you will need additional resources.

Quasi-Experimental Design (a type of between-subjects design)

When you have a comparison group but you do not have random assignment to the intervention vs/ comparison group, that is a quasi-experimental design. It's quasi-experimental because while you do have a group of people not receiving the treatment/intervention as you would in experimental design, it is not true experimental design because the two groups were not created by potential participants being randomly assigned to one of the two groups.

To demonstrate this design in practice, I'm going to use the structure of IM residency. We have approximately 40 IM residents who are split into 4 pods of 10 residents. Assignment to pods is done by the residency office and is not seemingly based on anything systematic. You may say to yourself, well, doesn't this count as random assignment? No, it does not. For it to be true random assignment you as the researcher must do the random assignment yourself. Nonetheless, the fact that assignment to pods is not done in a systematic way, it means that in all likelihood there are no ways in which the pods systematically differ. The fewer systematic differences between your treatment and comparison group[*] the more rigorous your quasi-experimental design.

Examples of Not-So-Rigorous Quasi-Experimental Design

Geriatrician Dr. Ryan Z. Chippendale designs an intervention for PGY2s and PGY3s. She decides to give the intervention to PGY3s and use PGY2s as the comparison group. This methodological choice will make her evaluation less rigorous because the group receiving the intervention also has an entire year of additional training compared to the PGY2s who are providing the counterfactual (what happens in the absence of the intervention). When Dr. Chippendale sees that PGY3s are significantly outperforming PGY2s, is it because her intervention worked or is it simply because the PGY3s have more experience overall compared to the PGY2s? She has no way of knowing! Both options are theoretically as likely to have instilled change between the groups, so you can't rule out one or the other. The more possible explanations there are for any between-group differences you observe, the weaker your evaluation is.

- One way that Dr. Chippendale could improve the rigor of this design is to use the PGY2s as her treatment group and the PGY3s as her comparison. Then, if the PGY2s outperform the PGY3s, we can know that it is not due to additional training. By "stacking the deck" against the intervention group, you gather stronger evidence that your intervention may be having a positive effect.
- Another way this design could be improved upon is to give a baseline knowledge/skill assessment to both PGY2s and PGY3s. If there are no differences between these two groups at baseline, then we will also reduce the likelihood that the group differences we see can be attributed to additional residency training versus participation in the intervention.

Example of a Rigorous Quasi-Experimental Design

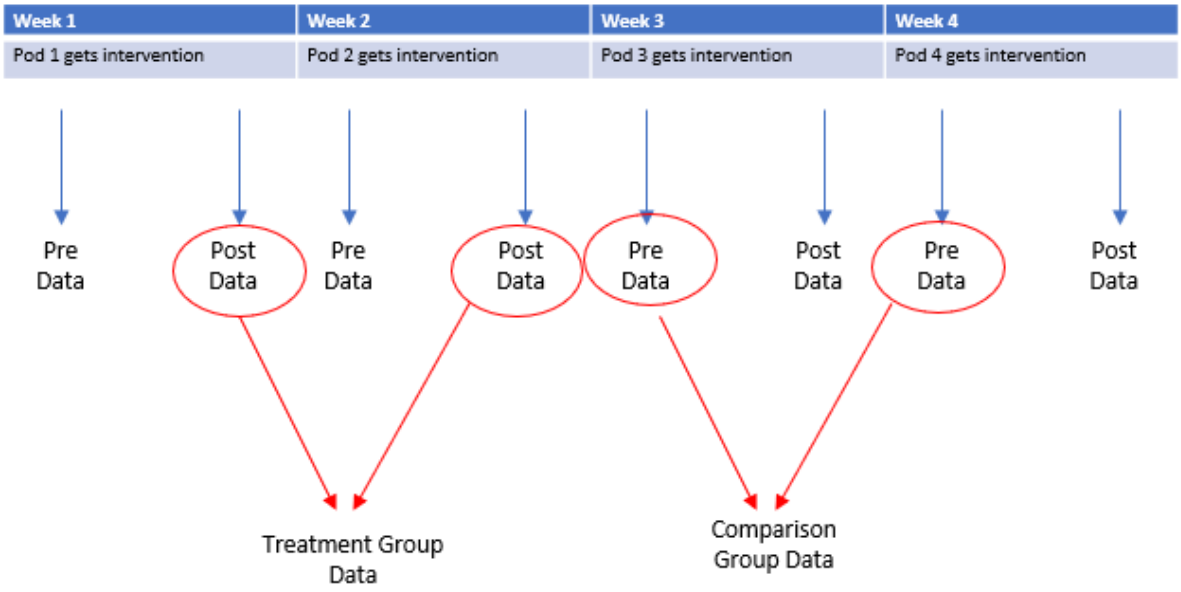
Having learned the error of her methodological ways, in her next evaluation Dr. Chippendale decides to create an intervention specifically for PGY3s in the IM residency. To create a quasi-experimental

[*] Note the use of the term "comparison group" here versus "control group." In quasi-experimental design you do not have a control group because you do not have random assignment.

evaluation, she uses the pre-existing pod structure that residents are sorted into by the residency office as described above. Each Friday for 4 weeks, Dr. Chippendale gives the intervention to one pod of residents. By the end of the 4 weeks all 40 residents have received the intervention. Dr. Chippendale collected pre-/post- data at each session.

Wait, how is this Quasi-Experimental?

In order to take the data she collected and make a quasi-experimental comparison between Pods 1 and 2 and Pods 3 and 4, Dr. Chippendale can compare the post- data of Pods 1 and 2 with the pre- data of Pods 3 and 4.



By strategically comparing the residents' data, Dr. Chippendale has created a fairly rigorous evaluation of her intervention.

Possible Limitations of Quasi-Experimental Design

- **The limitations of a quasi-experimental approach are greatly dependent upon how the treatment and comparison groups were formed**
    - The more likely there is to be a correlation between your grouping characteristic(s) and your outcome of interest, the less likely you are able to eliminate alternative explanations for any group differences you observe.
        - E.g., in the less rigorous example above, there are two likely explanations for why the PGY3s are outperforming the PGY2s, which makes the results less conclusive/rigorous)
    - If you are using a quasi-experimental design, I recommend having a brainstorming session where you try and consider every possible dimension on which your treatment and comparison group could theoretically differ and how, if at all, it may be related to your outcome(s) of interest. The shorter your list, the better! Having a long list after this exercise does not mean you should abandon the approach, however. Rather, it means

you will have a few additional limitations you will need to describe in the discussion section of your paper. On the positive side, the longer your limitations section is, the easier your future directions section is to write 😊

- **If you want to measure quasi-experimental outcomes in learners that are longitudinal, there may be practical and/or ethical considerations**
    - You cannot ethically withhold an intervention from trainees indefinitely. If you want to see differences in behavior after say, 6 months, that would mean you'd need to withhold the treatment from half of the residents for 6 months, which is unethical and also likely impractical.

Tips for Using Quasi-Experimental Design

- **If possible, incentivize participation in your study among the comparison group.** If you are using a design in which the comparison group will not receive the intervention at all, you will be more likely to get data from the people in that group if you can offer an incentive to them for participation. Unlike the people in the treatment group who are receiving something (your intervention), the comparison group has less motivation to respond to your requests for data.
- **Create your treatment and comparison group based on characteristics that are unlikely to be correlated to your outcome measures of interest.** If this doesn't make sense, see previous section again.
- **If you are interested in longitudinal change in your leaners, but you cannot ethically or practically withhold the treatment from the comparison group for as long as you would need to, you should consider a mixed methods approach.**
    - For example, you could make a quantitative comparison between the treatment and comparison groups, but then follow up qualitatively with all participants (assuming everyone got the intervention eventually). By showing that the treatment group is outperforming your comparison group during the window of time prior to the comparison group received the intervention, you are demonstrating the likelihood of effectiveness[*] of your intervention in the short-term. Then, by talking to a sub-sample of those who participated in the intervention (it won't matter if they were part of your treatment or your comparison group for this part) you can get more context-specific and nuanced accounts of how the intervention has subsequently changed participants' behavior, practice, self-efficacy, etc.

[*] Likelihood of effectiveness versus effectiveness because we still don't have a truly experimental approach and cannot unequivocally infer causation from the data.

Data Analysis Decision Tree for Between-Subjects Designs that include 2 or 3 groups (applies to quasi-experimental as well as experimental statistical comparisons)