# Video 6 - Non-experimental studies

Steve Simon

# Learning objectives

- 1.To distinguish different types of quantitative nonexperimental approaches
- 2.To discuss strengths and weaknesses of qualitative research

Here are the objectives for this week.

There will be a separate recording on single subject designs for those who are curious about it.

# Required reading

- 1.Chapters 7
- 2.Canvas Berkley-Patton et al., article
- 3.Canvas Eton et al., article.

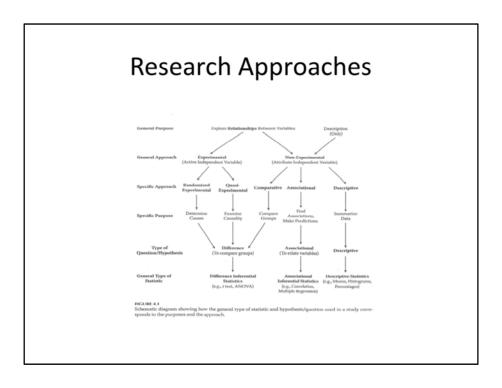
## Optional reading

Wolfe et al. "Effects of HIV-related stigma among an early sample of patients receiving antiretroviral therapy in Botswana"

Rowan et al. "Engagement-in-care during the first 5 years after HIV diagnosis: Data from a cohort of newly HIV-diagnosed individuals in a large US city"

DiLorenzo et al. "Comparison of older and younger individuals with multiple sclerosis: A preliminary investigation"

This is what you should have read already. If you haven't done the reading yet, pause this video and read this material. You'll get more out of the video if you do so.



Going back to the figure that we have used. Now we will talk about comparative, associational, and descriptive studies.

Now you are talking about an attribute independent variable, one that you cannot manipulate or control. You can talk about causality easily for a randomized design but these three designs do not allow you to say anything about causality.

Criteria	Descriptive	Associational	Comparative	Quasi- experimental	Randomized experimental
Random assignment of participants to groups by investigator	No groups	No (only one group)	No	No	Yes
Independent variable is active	No Independent variable	No (attribute)	No (attribute)	Yes	Yes
Independent variable is controlled by the investigator	No	No	No	Sometimes	Usually
Number of levels of the Independent variable	No Independent variable	Usually 5 or more ordered levels	Usually 2–4 levels	Usually 2–4 levels	Usually 2-4 levels
Relationships between variables or comparison of groups	No .	Yes (relationship)	Yes (comparison)	Yes (comparison)	Yes (comparison)

A comparison of the five basic quantitative research approaches.

Figure 7.1 identifies the five types of designs (descriptive, associational, comparative, quasi-experimental, and randomized experimental).

These three non-experimental approaches differ the levels of the independent variable. A large number of ordered levels (equivalent to a continuous independent variable) falls in the associational research type.

DiLorenzo compared older patients to younger patients.

Haberer article includes both quantitative and qualitative research. Interactive voice response (IVR) versus short message service (SMS) as a way of collecting adherence. It is an HIV study where adherence is very important. Mobile technology has a lot of penetration in places like Africa. Parents that could read were randomized, but parents who could not read were assigned to IVR.

They did qualitative interviews with all participants. The first theme was poor understanding of the responses. A second theme was challenges in training. The participants did not feel able to speak up and say that they don't understand.

## Research Approaches

- Comparative research approach
  - Purpose?
  - How is this approach similar to randomized experimental and quasi-experimental?
  - C an this approach tell you anything about cause and effect?

Look at the very end of the introduction to get at the purpose of the research. Review the participant.

Random selection is a better representation of the population. (External validity?)

In any prospective study, you are only getting information for those people who are willing to be part of the study.

- Quantitative Comparative
  - Comparing groups 2 or more
  - Main IV has only a few levels
  - · Statistical analysis -
    - T ypically t-test or analysis of variance
  - Example DiLorenzo et al., 2004 article
- Usually research that is reported will contain elements of more than one design
  - Example Haberer et al., 2010 article

Often your research has the elements of more than one design.

## **Research Approaches**

- Associational research approach
  - Purpose?
  - Characteristics of the variables measured?
  - Can this approach tell you anything about cause and effect?

The compliance study is a good example of when you often end up with mixed methods to get more out of

The purpose is to find associations among variables. The variables collected are typically continuous. This type of approach does not tell you about cause and effect directly. The association might allow you to predict. Continuous really means many levels. The analyses are typically correlation based or multiple regression.

An article by Rowan illustrates this well. Clinical record review and state HIV data. Variations in the engagement in care continuum. Impact of outmigration and death. When you don't have data is it because they are no longer engaged in care, or maybe they have moved on to a different provider. Has care improved over the last five years?

What variables are associated with getting treatment to reduce viral loads. This paper also addresses how to handle people who you do not have complete follow-up on.

Example of smoking cessation usually assumes that anyone who disappears is categorized as a treatment failure.

- Quantitative Associational
  - IV is often continuous / has many levels of an ordered variable (typically 5 or more )
  - Looking at the association between the IV(s) and the DV
  - · Statistical analysis -
    - Correlation-based
    - Multiple regression when there are multiple Ivs
  - Example Rowan et al. 2014 article

# **Research Approaches**

- Descriptive research approach
  - Purpose?
  - When usually used?
  - Characteristics of the variables measured?
  - Can this approach tell you anything about cause and effect?

Usually used for an exploratory. It could be characterization of surveys or polls. It might be performance reports.

- Quantitative Descriptive
  - Descriptive statistics
    - Averages
    - Percentages
    - Histogram
    - Frequency distribution
  - Some publication of strictly descriptive studies
  - More often descriptive component of a study
    - Describe the sample
  - Example Wolfe et al, 2006 study

The Wolfe study is a good example. This study is in Botswana. Effects of stigma and its effect on health related behaviors. Structured questions followed by open ended questions. They are reporting percentages.

- Qualitative 5 main approaches
  - · Phenomenological
  - · Grounded theory
  - Ethnographic
  - · Case study
  - Narrative

Phenomenological: try to understand the meaning that people place on events in their life. Try to understand from the perspective of the individual.

Grounded theory: the goal is to generate theory from data collection: process, actions, interactions. The inferences that you take away need to be grounded in the data you collect. Try to avoid any prior expectations. Some people even recommend limiting the literature review because it might box you in.

You won't know your sample size until you start collecting your data. As you are coding this information, you will be developing categories. These are provisional, there may be revisions as you go along and so the coding process is dynamic and iterative. As coding develops, you may go back to earlier data. Seek disconfirming examples. This is a check of your validity. You can also have audits by an independent reviewer. You are trying to reflect the reality of human experience.

Ethnographic. Group of people who share the same culture.

Case study. A single case or group of related cases.

Narrative. Develop stories.

These approaches are all inductive. You are trying to develop a framework of understanding. Don't go in with a heavy set of expectations.

- Qualitative Phenomenological
  - A llow researcher to "... understand the meaning participants place onto "... events, phenomenon, and activities."
  - Goal "... explain the essence of experiences lived by the participants."
  - Method usually interviews

- Qualitative Grounded Theory
  - Goal "... generate theory from data collected from participants."
  - "... focus on the process, actions, and interactions experienced by ... participants."
  - Method usually interviews

- Grounded theory (Glaser & Strauss, 1967)
  - Inferences firmly "grounded" in the data
  - Prior theoretical expectations avoided
  - Sampling proceeds parallel to data collection & analysis
  - Categories regarded as provisional, subject to revision
  - Research maintains skepticism, seeks disconfirming examples
  - Ultimate purpose construct coherent theory from the data

- Qualitative Ethnographic
  - Goal "... describe a group of individuals who share the same culture."
- Qualitative Case-study
  - Goal "...develop deep understanding of a case or cases."
- Qualitative Narrative
  - Goal "... identify and report stories from the participants."

- Inductive process -
  - Start with the specific (raw data / transcript)
  - Develop a theoretical framework from the data
  - Conceptual categories emerge from the data

- Start the study with a research question
  - This may provide a potential starting point for analyzing text
  - If you do this, DO NOT let it prevent you from seeing what is in the text
- Steps
  - From raw data <U+F0E8> theory building
  - "Iterative" process
    - e.g., as you are developing categories, check them back against the raw data – do they continue to reflect the original data?

Keep alert and look for things that come out in the material that you were not anticipating. As your categories, go back and compare them against the raw data. Make sure that you allow the data to guide you. Often coding from transcripts. But it is also good to have access to the original recordings. Sometimes there is a doubt about what some was saying (joking or sarcasm). Look for tone of voice and non-verbal cues.

Content analysis is a generic term, but you can look at things like ad messages in magazines that appeal to young adults.

Thematic analysis is looking for patterns.

- Analysis process
  - Given research question, may have some idea of a general conceptual structure; serve only as a starting point
  - · Usually work from transcripts
  - If possible, refer back to audio recordings when needed
  - · Qualitative form of analysis
    - Content analysis analysis of the content of communication
    - Thematic analysis identifying patterns or themes in the data

Break up the text into "bits" that represent specific thoughts. Open coding lots of themes. Axial coding looks at the relationships between themes, builds a superstructure of ideas.

- Category Formation assign sections (data bits) to analytical categories/themes "coding"
  - Summarizing the data by identifying similarities and differences / commonalities and contrasts
  - · Categories:
    - More manageable units of information
    - "open coding"
    - Descriptive Concrete
    - Interpretive Abstract
    - Relational
  - Cross-reference categories back to text
    - Makes it easier to revise coding as needed

Software allows you to go back and forth between themes and the original text. Cross-referencing is important. Find examples easily. It helps provide validity checks. Does someone agree with your categorization.

Categories need to be faithful to the data. This is internal validity. But you need to place this in a broader understanding.

You need to be careful about trying to quantify this too much. Keep in mind that just because someone talks about something a lot does not mean that this is the most important thing to them. The most important thing might be something that people are uncomfortable talking about.

- Category Formation
  - Faithful to the data "internal" aspect
  - Meaningful to other categories "external" aspect
  - Balance
    - Generality vs Specificity
    - Inclusivity vs Exclusivity
  - Number of times idea mentioned does NOT necessarily reflect importance

- Labeling / Revising Categories
  - Member-generated " first-order " account
  - Observer-generated " second-order " account
  - · Labels are provisional; may be revised
  - Iteration move back and forth between raw data and labeled categories
  - "Zoom in" vs "Wide angle" view of the coding
  - Revision may end up dividing or combining categories
  - · May establish criteria for category coding

Phrasing of the person versus observations generated by the researcher. Labels should be considered as provisional as you work with more complete information. It is an iterative process. Go back to original data. You're looking at the specific of what's been told versus a broader understanding.

Establishing the criteria may depend on how many people are doing the coding. If you don't have a pre-existing structure, the coding may evolve and may involve 100% review (both raters review everything and resolve discrepancy). Then a third person does an audit.

- Saturation the point at which you are not gaining any new insight, no new categories being identified, no new relationships being defined
  - · Don't need any additional analysis
  - Don't need any additional data collection (if collection & analysis done in parallel)
  - Determinant nature of the information being collection – NOT amount of information
  - Keeping notes of the coding/analysis process
  - Help identify categories/relationships
  - · Help if "stuck"
  - Document analysis process

Saturation is a way to assess your sample size. Have you gotten a representative amount of data.

Really important to keep notes of the coding and analysis process, so you can check your work. Allows you to do qualirty checks on the process.

- Role of Judgment
  - Balancing act -
    - Level of creativity by coder to identify categories/relationships
    - Must reflect the informants thoughts
    - Audit of the coding by an independent person can check for the match between the coding and the source information
  - Low-inference descriptors
  - Look for "negative cases"

You develop a framework, but that framework must always represent what the subjects have said.

Low inference descriptors are those things involving just the basic words or text.

- Quantifying Information
  - Pros & Cons
  - When quantifying might be of value
  - Simple frequency of occurrence does NOT necessarily reflect importance

This was mentioned earlier.

#### **Eton Article**

- Long-term Goal "... build a general, multidomain, patient-reported measure of burden of treatment with wide applicability across diseases and treatments"
- Semi-structured qualitative interviews
  - "identify issues ... illustrative of burden of treatment ..."
  - "inform derivation of a general, patient-reported measure of burden of treatment flexible enough for application across any disease or treatment regimen."

The Eton article has a great description of the coding process. The long term goal develop a burden of treatment with wide applicability across a wide range of treatments.

Identify issues associated with burden of treatment. Derivation of patient report that is flexible enough to apply across any disease.

#### **Eton Article**

- Development of interview
- Content analysis process
  - Multiple coders
  - Identify key themes and subthemes
  - Discussion and consensus to develop coding scheme
  - Repeated check and update of coding process
  - Thematic content saturation
- Resulting major themes and subthemes

Checked for a logical flow from broad to specific. Framework from first five interviews. Reviewed after ten more interviews. Saturation reached after 25 interviews, but they had 7 more that they included anyway.

Described patient population.

Figure 1 shows the themes that emerged.

The third theme was factors that exacebate burdens had subthemes.

## Berkley-Patton Article

- Goal Use qualitative method to explore components of the Information-Motivation-Behavioral Skills model
  - Information + Motivation Behavioral Skills Adherence
- Focus Groups
- Content analysis "... identify themes on expectations and information, personal and social motivation, and behavioral skills related to ART adherence."

Another example on Canvas is an article by Jannette Berkley-Patton et al. Done using focus groups and content analysis.

Information and expectations with subthemes. Specific text included as examples of themes.

# Berkley-Patton Article

- Information and expectations
- Motivation for ART adherence
- Behavioral skills related to ART adherence
- Themes for each of these areas?
- Value obtained from this kind of study?

- Community Based Participatory Research (CBPR)
- Participatory Action Research (PAR)
  - Takes place outside the normal research environment
  - Designed to meet the needs of the community in which the research is conducted
  - Value for the participants that goes beyond the value of the collected information

- Existing research data or data collected from other locations ...
  - May not reflect conditions (e.g., disease incidence) in the community of interest
  - May not reflect underlying risk factors
  - May not indicate how an intervention will work in a specific community

Community Based Participatory Research.

This is increasingly used and make a big contribution. Extent to which the population that you are interested in are very much a part of the process of designing the research. It takes place outside the research process. It hurts internal validity but it is far more relevant, improving external validity. The value goes way beyond. Develop community capacity, improves sustainability, builds skills. More emphasis on these models using PCORI. Most of this work involves the community from the beginning.

- CBPR/PAR research may be more generalizable than research conducted in a clinical setting
  - · Sample characteristics
- Practice-Based Research Networks (PBRNs)
  - Physicians from community settings work together to study research questions of mutual interest

- How the research context influences the research process ...
  - · What issues are routinely encountered?
  - To whom is the research important?
  - · Ripple effect of the research process
  - · Level of engagement of the target population

Other research may not reflect what the community is interested in. Research in a carefully constrained environment may not be sustainable.

Collaboration among practitioners where they can combine efforts. Also valuable for networks of pharmacists.

Issues important to the population and community. There can be a ripple effect. When community members get engaged, they can learn skills that can help in a different context, and inform them about resources that they can use in other areas. The community gains from the whole process.

- Conducting CBPR/PAR
  - · Start simple
  - Think of local comparative advantage
  - Network
  - Collaborate
    - Top-down
    - Bottom-up

Simple. What is important and what is needed. Add value to the community. Identify people who can be resources. This is very much a collaboration. It can be top-down, where you make yourself available to the community. A bottom-up is where you offer your expertise to help them develop a project.

# Nonexperimental - CBPR / PAR

- CBPR/PAR especially appropriate for research involving health disparities
  - Multiple individual and community-level determinants
- Mechanism to translate basic health psychology conceptual models into interventions

There are some many community level issues associated with health disparities.

# Nonexperimental - CBPR / PAR

- Contributions by researchers
  - Know existing research evidence
  - · Can provide support/suggestions
    - Organization
    - Information-gathering
    - Action
  - Help community get in touch with resources

You biring in your knowledge of existing resources, and provide support and suggestions. Advice about resources that they can tap.

## Nonexperimental - CBPR / PAR

- Contributions by community
  - Unique insight into local context
  - Information concerning local needs and priorities
  - Provide "real world" feedback/reaction
  - Provide guidance for adapting programs/interventions to the community
- Community involvement community capacity sustainability

They understand needs and real world feedback. Does this fit in to individual people's needs.

- Definition using existing database/dataset to reexamine variables and answer questions
  - Research question different from what was originally addressed
- Differences could be ...
  - · Variables that are analyzed
  - · Relationships that are explored
  - · Different subsets of cases
  - · Different analysis techniques

Let's have a common definition. You are re-purposing the data.

Some secondary datasets are available strictly for this purpose. These are often national databases collected user taxpayer money and made available to anyone who wants to use them.

Sometimes we collect more data than what we strictly need for our purposes.

Example: One of the interventions used motivational interviewing. Autonomous regulation scale. Some additional information came out that that there was a third subscale after the data was collected. So Mary was approached to see if her data would support this third subscale.

Identifying people who have common data sets that can be combined.

NIH supports secondary analysis of data. Example of RFA on secondary analysis. Publicly available data sets for aging related analysis" document.

- Could be done when able to combine different/compatible data sets
- Advantages
  - Cost
  - Working with large dataset
  - Time efficiency
  - May be a good first step in a research area

The data are already collected, so you don't have to wait around.

Secondary analyses are great for exploratory studies. Initial analysis was changes in rate of adherence. Sixteen different adherence studies contributed by fourteen different groups.

The combination of disparate data sets is one of the hallmarks of big data (the "variety" V of the three V's of big data).

- Disadvantages
  - · Lack of control over data collection process
    - Quality control
    - Specifics of variables and wording
    - Missing data or data entry errors
  - Consider possible sources of error
    - Make judgment on possible impact on research question
    - What can be done to address problems

One of the big ones is the lack of control over the data collection process. You can't change how the data was collected without access to a time machine. You lose control over the variables that were collected, the wording of the questionnaires, the quality checks.

Data quality issues can be an issue at times. Why conduct rigorous quality checks now for data that will only be used by someone else in the future?

Example of depression: some of the data sources might use CESD, BDI, SCID.

- Two types of secondary data sets
  - Individual
  - Aggregate

Data sets that have individual level records: there is separate records an indvidual unit (usually a person, but it could be an event like an emergency department visit).

- Individual separate information available for each subject/unit
  - Can use to measure associations between characteristics among individuals/units
  - Source
    - Previous research study
    - Large regional and national data sets
      - » Surveillance, Epidemiology, and End Results (SEER)
      - » National Death Index

The book gives several resources, and there is a document on Canvas that lists some of the secondary data sets that DBHI has worked with.

- Aggregate information available only for groups of subjects
- Can study associations among groups by comparing them on risk factors or variables of interest (ecologic studies)
- Advantage availability of data

Aggregated data avoids some privacy concerns.

- Aggregate (cont ' d)
- Disadvantage -
  - · Associations are especially susceptible to confounding
  - Associations seen in aggregate don't always hold at the individual level (ecologic fallacy)
- Most appropriately used
  - · Test the plausibility of a new hypothesis, or
  - Generate new hypotheses

An example of the ecologic fallacy. Sales of cigarettes are greater in states that have higher suicide rates. You don't know whether an individual who smokes is more likely to commit suicide.

- Chicken or the egg!
  - Identify existing database and determine what can be done with it – that you are interested in!
    - Get familiar with the data set available variables
    - Consider possible new questions that can be addressed
    - Go to the literature
      - » How has database been used previously?
      - » What has already been answered?
      - » Develop a theoretical framework for your research

You can have a question and find what secondary data sets might help answer that question. Or you can explore a data set that you find interesting and see if there are interesting research questions that this data set might help with.

For the latter, do a literature search where others have used this data set and see what questions they were able to answer.

- Table 13.1

Steps in Finding Research Questions to Fit an Existing Database

- 1. Choose a database.
- 2.Become thoroughly familiar with the database. Make a flow sheet of all variables and how they were measured.
- 3. Identify pairs or groups of variables whose association may be of interest.
- 4. Review the literature and consult experts to determine if these research questions would be

- Chicken or the egg!
  - Search for an existing database that will address the research question you are interested in
    - Start with the research question you are interested in
    - Have a list of variables that are relevant
    - Have an idea of what databases are available
    - Find out if there are special qualifications for working with the database

- Table 13.2

Steps in Finding Databases to Fit a Specified Research Question

- 1. Choose a research question and review the literature thoroughly.
- 2.List combinations of predictor and outcome variables whose relationship might help answer the research question.
- 3. Identify databases that might include the variables of interest.
- 4. Become familiar with each of these databases and

Gaining access can mean different things. Some datasets, you can directly download (no cost). Others are free, but you have to prepare a research proposal and get approval before you can use that data. Others require a small payment. Some data sets require you to sign an agreement first.

- Sources where to find out about them
  - Journal articles / conference presentations
  - · Data libraries
  - U.S. government
    - Census Bureau
    - National Center for Health Statistics (NCHS)
    - Framingham Heart Study
    - Multiple Risk Factor Intervention Trial (MRFIT)
    - NIH Data Sharing Plan requirement

There are lots of resources available in the genetics community that are not covered here.

- Sources where to find out about them
  - · Community-based data sets
    - Evaluate patterns of utilization and clinical outcomes of medical treatment
    - Examples
      - » Administrative and clinical database
      - » Registries
    - Useful for
      - » Examining rare outcomes
      - » Assessing real-world utilization and effectiveness of an intervention
      - » Study effectiveness rather than efficacy
    - Importance of replication

Truman is setting up a system, i2b2, that allows you to study research cohorts. Health Facts.

Kansas City Quality Improvement Consortium. Mid America Research Council.

Data from CMS, Department of Veteran Affairs.

Missouri Information for Community Assessment (MICA). Helps local health agencies and their partners.

Useful for assessing the magnitude of the problem.

[[Contrast national with community level data sets.]]

### **Ancillary Studies**

- Adds new measurements to an existing/ongoing study
  - Advantages
    - Similar to advantages of secondary analysis
    - Inexpensive & efficient
    - Can be added to any type of study
  - Disadvantages
    - Not as valuable as if measures were part of original study
      - » Not a complete set of data
    - Have no control over what the other measures are
    - Need to get the buy-in of the primary investigators

Find a study that is already collecting data and ask them to add some measures that you are interested in.

Example, there are many ways to collect data on adherence. Add some extra measures after the fact to explore interrelationships or to see if newer measures are more sensitive.

## **Ancillary Studies**

- Not unusual for additional measure to be added to longitudinal studies
- How to get started
  - Identify ongoing studies of interest
    - Either predictor or outcome measures of interest
    - NIH RePORTER (Research Portfolio Online Reporting Tool)
    - ClinicalTrials.gov
    - Pharmaceutical companies
    - Contacting experts / researchers
  - Get cooperation of investigators

NCCAM example.

### **Ancillary Studies**

- How to get started (cont 'd)
  - Work out the logistics (measures, where to insert it, is there a cost associated with adding the measure?)
  - Get all appropriate approvals (e.g., IRB to document change in data that are being collected)
  - Work out what data from the main study you will have access to and how you will get the data you need
  - Discuss and agree on issues such as manuscript publication and authorship

An ancillary study requires IRB approval. You need to discuss with the coordinators of the study that you are piggy baking on

- Examples of available secondary analysis data sets

...

- MEPS (Medical Expenditure Panel Survey)
- YRBSS (Youth Risk Behavior Surveillance System)
- NSQIP (National Surgical Quality Improvement Proram)

• ...

MEPS and YRBSS are individual based. NSQIP are event based.

http://www.meps.ahrq.gov

MEPS is nationally representative maintained by the Agency for Health Care Research and Quality. MEPS is largely cross-sectional, but has panels that can address longitudinal questions.

The YRBSS focuses on six health risk behaviors among youths (unintentional injuries and violence, tobacco use, alcohol and other drug use, sexual behaviors. dietary behaviors, and physical activity).

Review the list of secondary datasets prepared by DBHI.

#### Assignment #4

- Turn in the results of a literature search on the topic you are planning to focus on for your research proposal. This is NOT expected to be a literature review; it should provide evidence that you are finding literature that is relevant to your topic. This assignment serves as the start of your References section, using the reference/citation style you plan to use in your proposal.

I don't expect this to be your final literature review but it will help form the literature review. Looking for progress in finding relevant references. No need to turn in a reference section. Use a reference or citation style. Pick a style that is used in your field. Just be consistent. Show me evidence of the literature that is relevant.

### Assignment

- 1. Turn in the results of a literature search on the topic you are planning to focus on for your research proposal. What you turn in should provide evidence of the literature you have found that is relevant to the topic you've chosen for your research proposal. This assignment serves as a start for your References section.
- 2. Prepare for next week's session

# Discussion questions

- 1. What kind of research topics could be addressed using non-experimental designs?
- 2. What kind of research topics could be addressed using qualitative methods?
- 3. What are the ethical issues that need to be addressed when doing these types of studies?

