Week 3: Refining your research topic

Andrew Stokes

February 5, 2019

Review

- Automation improves efficiency, transparency and reduces likelihood of mistakes
- In practice, means writing a script in R that takes your raw data, pre-processes it, performs analyses and generates the final output for the research report.
- Significant upfront investment of time and effort, but with long-term time savings

Today

Today is focused on strategies for refining your research topic into a set of empirically testable objectives or hypotheses. We will discuss:

- ▶ Conducting a comprehensive and systematic literature review
- ▶ Critically evaluating studies
- ▶ Visualizing the research question using a problem diagram
- ▶ Formulating research objectives and hypotheses
- ▶ Identifying variables to be measured in the data collection

Getting to know the literature: a critical first step

- Helps you identify what is cutting edge and where you might be able to push the science forward
- Ideas will occurr to you all the time: the better you know the literature, the better you can identify which of your ideas are novel
- You'll be aware of the controversies, dilemmas and paradoxes in your field that beg for further exploration.
- You'll start to notice the gaps in the literature and the limitations of current approaches.

Getting to know the literature: a critical first step

Knowing the literature also allows you take advantage of past insights and all the hard work of your precedessors. Chapter 8 in Writing for Social Scientists (available on Blackboard) has a great analogy on this point. Here is an excerpt:
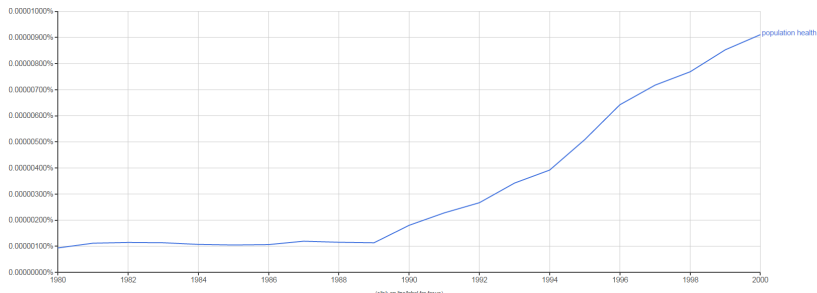
> Imagine that you are doing a woodworking project, perhaps making a table. You have designed it and cut out some of the parts. Fortunately, you needn't make all the parts yourself. Some are standard sizes and shapes - lengths of two by four, for instance - available at any lumberyard. Some have already been designed and made by other people - drawer pulls and turned legs. All you have to do is fit them into the places you left for them, knowing that they were available (p.142)

Not an easy task

But familiarizing yourself with the literature and staying current is not easy. The volume of studies being published has increased exponentially making the task of staying current harder with each passing year. What accounts for this?

- ▶ It's getting easier to work with data, thanks to increased computing power
- ▶ Most data are now readily downlowdable onto a personal computer
- ▶ Statistical procedures that might have taken weeks to process, can now be done in a matter of minutes
- ▶ Proliferation of data sources and the era of big data
- ▶ Expansion of the scientific workforce

Frequency of use of the phrase 'Population Health' in books[1]

Conducting a literature review

▶ We will discuss the major tools for searching the literature, including the standard ones and others that have emerged recently
▶ The goal is to track down all the relevant literature and to do so efficienctly.
▶ Sometimes it will also be useful to use a strategy that is reproducible by others by documenting the search terms you used for your literature review

PubMed: a powerful tool

- ▶ PubMed is an extremely powerful tool for conducting literature reviews
- ▶ Relevant to researchers in the medical and public health sciences, containing some 21 million journal article
- ▶ If you know how to search it effectively it is a gold mine

PubMed: key points

- Use the AND or OR operators to connect search terms (Boolean operators)
- Impose limits to narrow your search
- Look for terms in specific sections of papers, such as those that appear in the title or abstract
- Use MeSH terms (medical subject headings)
- Use the related articles link after you have identified relevant articles.
- Keep a copy of the PubMed helpfile handy if you get stuck.

PubMed Demo[2]

Let's go through several exercises as a class

Google Scholar

Google Scholar is another useful took for literature review and has some unique features.

▶ Excellent for tracking down studies in the grey literature (work that appears outside the peer-revieweliterature, including reports, theses and dissertations and conference abstracts). These studies are important because they can tell you a lot about what's happening now, whereas materials in the peer-reviewed literature often reflect the state of science one or more years in the past (since there is lag between an article's submission and its publication).

Google Scholar

- ▶ Quickly get a list of all the articles published subsequent to the original article that cited it. Once you have identified an important paper, this can be useful for identifying the most recent research done on the topic.
- ▶ Can also use Google Scholar is follow researchers in your field of interest and receive notifications when they have published new research

Google Scholar Demo

https://scholar.google.com/

Other strategies for searching the literature

Study reference cited lists. To identify further articles of interest, you can see what other researchers cited in their papers. If you do this enough times, you will start to see which papers are the most influential in the field, because citations to them will keep popping up. Even better, you can usually go back to the text of the article itself and learn more about the paper and in what context it was cited.

Social media

Social media tools can help too. A lot of academic researchers have embraced social media for disseminating new research and creating repositories of their past studies. Some tools to be aware of include Twitter, ResearchGate and Academia. Learning about the latest interesting research is as simple as getting an account and following people whose work you have read and found inspiring.

Follow journals

Pick a few journals that you like and subscribe to their mailing lists. You can have journals such as *The Lancet*, *PLoS Medicine*, and others deliver content write to your inbox on a weekly or monthly basis. You'll get the table of contents for the most recent issue of the journal right when it is released, without having to remember the publication dates of various journals.

How to read and evaluate what you find

So now you have a large stack of papers. Which ones do you prioritize and how do you read and critically evaluate them?

Prioritizing

- ▶ Start with the articles from high impact journals (journal impact factors are available on Google Scholar Metrics, we will come back to this below)
- ▶ Give priority to articles that have been published more recently. The seminal earlier papers will become clear with time, because you'll keep seeing references to them.

Reading a research paper

- Start with the abstract, then tables and figures (especially the figures)
- Read the Discussion section. This section is a particularly useful part of the paper because it is typically where the authors discuss the literature, the strengths and limitations of their approach and future directions. These can yield helpful insights for refining your own research topic.

Critically evaluating the literature

- ▶ Know that each and every study has limitations, sometimes minor but other times not minor.
- ▶ Finding major problems can be rewarding as it gives you a potential avenue for further work

What to think about while reading a paper

- What angle are the researchers coming from? Is there a vested interest? Who paid for the research?

Conflicts of interest: a case study

What to think about while reading a paper continued

- Is the article peer-reviewed? What journal was it published in? There are a million journals these day and most research can get published somewhere. So its important to note the source of the article and consider the reputation of the journal.

Using Google Scholar Metrics to explore the journal impact factor

What to think about while reading a paper continued

- ▶ If the study was based on a survey, what was the sample size and response rate? Are the data generalizable?
- ▶ How were the data collected? In-person interviews, by phone, by mail?
- ▶ How are key variables measured? Is there potential for bias or measurement error and if so how might it affect the results?
- ▶ Were all potential confounding factors adjusted for?
- ▶ Did the study include a sensitivity analysis and if so did it support the main findings?
- ▶ How wide are the confidence intervals around the main results?
- ▶ Are the reported findings plausible?
- ▶ Are the conclusions supported by the findings?

Many studies don't hold up when replicated

Listen: Planet Money's *The Experiment Experiment*:
http://goo.gl/ALR8Ht

Preparing an annotated bibliography

An annotated bibliography is a list of citations, each followed by a brief annotation. Common elements of an annotated bibliography include:

▶ A concise summary of the article (e.g. major objectives, hypotheses, data, methodology, key independent and dependent variables, results)

▶ A critical evaluation of the article (e.g. potential biases, strengths and limitations of data or methods, assumptions, correspondence between results and conclusions, consistency with other studies

▶ Discussion of relevance of article to your research topic

A useful analogy[3]

Think of your paper as part of a conversation with people interested in the same things you are; the annotated bibliography allows you to tell readers what to check out, what might be worth checking out in some situations, and what might not be worth spending the time on. It's kind of like providing a list of good movies for your classmates to watch and then going over the list with them, telling them why this movie is better than that one. . . You want to give your audience enough information to understand basically what the movies are about and to make an informed decision about where to spend their money based on their interests.

---

[3]From UNC Writing Center: http://goo.gl/hUnCWG

Organizing your papers

Just as automation is valuable for processing and analyzing data, it is also useful when it comes to referencing. In this class, we will use Mendeley, a free and user friendly bibligraphic management tool

Major features of Mendeley

- Free (up to 2 GB cloud storage)
- PC and Mac compatable
- Format references in Word or Open Office
- Web account provides online access
- Captures citation data from PDFs you import
- Search and annotate PDFs
- Share and collaborate

Watch: Mendeley Video

Watch Mendeley introductory video here: https://goo.gl/bLxxZX

Mendeley Demo

- Interactive demo

Additional Resources

▶ Mendeley videos and tutorials:
https://community.mendeley.com/guides/videos
▶ UC Berkeley Tutorial:
http://guides.lib.berkeley.edu/mendeley/getstarted

Systematic analysis of the research question[4]

- ▶ When you are at the point where you have an idea of what you want to do but it is still rough around the edges, it can be useful to *visualize* the research question using a diagram.
- ▶ This can be particularly useful in applied research when there are a variety of stakeholders whose views should be incorporated.
- ▶ The goal of the exercise is to further specify and describe the research question.

---

[4]Notes that follow draw on the IDRC book. See Modules 4 (analysis and statement of problem), 6 (formulation of research objectives) and 8 (variables) for additional guidance.

What's the problem?

A research question can often be stated as a problem. The IDRC text offers three criteria for problems that make ideal research topics:

- ▶ Perceived difference between what exists and the ideal (what is wrong?)
- ▶ Reasons for difference unclear (why is it wrong?)
- ▶ Often more than one answer to the problem exists (what is the best approach to fix it?)

The Problem Diagram

You could then diagram the problem as follows (for more on this see the IDRC book):

▶ Write down the problem in the center
▶ Brainstorm possible causes or contributing factors
▶ Identify further contributiong factors (you can think of these as distal vs. proximate).
▶ Organize the factors into a typology and produce a final diagram.

Example

- ▶ Let's say you are interested in reasons some people adopt health technologies in low income settings while others do not (improved stoves or insecticide-treated bednets, for example).
- ▶ You could re-state the research question as a problem: *very low uptake of improved stoves in low-income countries.*

Perceived difference between what exists and the ideal

The current situation is that few people have adopted improved stoves. Ideally, uptake would be much greater, since the health benefits can be substantial.

Reasons for difference unclear

The explanation for why there is low uptake of improved stoves is not obvious.

Often more than one answer to the problem exists

The solution could be to

- ▶ make fuel more readily available
- ▶ design stoves that are more consistent with culture and cooking practices
- ▶ make the stoves more convenient ot use
- ▶ Lower the price.

In other words, there are many different potential solutions and it's not clear which avenues should be pursued.

Problem Diagram Exercise

Let's try our hand at a problem diagram as a class, focusing on the case of low-uptake of improved stoves. You can find my version here

Objectives and hypotheses

Formulating your research in terms of objectives or hypotheses can help you hone in on what exactly you want to accomplish. This is important for several reasons.

- ▶ It can help to ensure that you only collect the data you need and nothing more
- ▶ It can help provide a logical structure and organization to your study
- ▶ Finally, it easier to communicate what you plan to do to others

The general objective

Research objectives often begin with a general objective followed by several specific ones. The general objective should state what you want to get out of the research in general terms. It should give the what, where and why:

- ▶ What will you do?
- ▶ Where will you do it?
- ▶ Why is the research being carried out?

Specific objectives

- The specific objectives break the general objective into several discrete components.
- The problem diagram identified categories of issues and contributiong factors. Use it as a guide to formulating a set of specific objectives that systematically address each of the categories and the key factors related to the problem.
- Write the specific objectives in a logical sequence and phrase them in succinct operational terms using action verbs.
- The final specific objective often states how the results will be used.

Specific objectives continued

You can often start with three basic specific objectives, adding further ones as needed. Address these to:

- quantifying the problem
- associating the problem with other factors
- providing recommendations

An example is posted on the course website.

Variables

Another important step for focusing the study is to think about the variables you'll need to measure. Using the problem diagram as a reference, you should consider

- ▶ variables needed for each objective or hypothesis
- ▶ which are independent vs. depedendent
- ▶ how they can be measured
- ▶ which need further research before they can be defined

In this class, we will use a worksheet to organize our thinking.

Types of variables

A basic distinction is between the dependent (outcome) variable and independent (predictor) variables. Independent variables can be further distinguished based on their role in the analysis:

- ▶ key independent variable
- ▶ confounder
- ▶ effect modifier
- ▶ mediator
- ▶ background

Types of variables continued

▶ The key independent variable is the generally the predictor variable of most interest (x).

▶ Variables that simultaneously affect x and y are referred to as confounders. All potential confounders should be included in the analysis.

▶ Effect modifiers are variables that modify the association between x and y.

▶ Mediators are variables in the pathway between x and y.

▶ Background variables are basic social and demographic variables that are generally always included in analyses.

Example

Take for example a project focused on estimating the association between bednet use and child mortality

- ▶ Bednet use would be the key independent variable
- ▶ Incidence of death the outcome variable.
- ▶ An example of a confounding variable here might be mother's education, if children with more educated mothers are both more likely to use a bednet and less likely to die. This confounding could serve to artificially strengthen the association between bednet use and reduced risk of death since more educated mothers not only make sure their children sleep under bednets but also that they get their vaccinations and range of other health interventions. -Background variables in this analysis could include age, sex of child and race/ethnicity.

Types of variables continued

Variables may be broadly classified as either numeric or character.

- ▶ An example of a character variable is a unique identifier for patients in a hospital database.
- ▶ Numeric variables can be either continuous (e.g. birthweight, income) or categorical (income decile).

Categorical variables

Categorical variables can be further distinguished as ordinal (data with order) or nominal (data not ordered).

- An example of an ordinal variable is income category (low, medium, high).
- An example of a nominal variable is gender or ethnicity. They fall into categories but they are not ordered.

Dummy variables

Dummy variables are a special case of a categorical variable and are very useful for conducting analyses. These variables take on just two values: 0 and 1. For example, the dummy variable "female" would be coded 0 if the respondent is a male and 1 if the respondent is a female.

Operationalizing a variable

- ▶ Often a variable can be entered into the analysis as is. Examples include waiting time, age and gender.
- ▶ In other cases, indicators are needed to operationalize variables, because the variable cannot be measured.
- ▶ For example intelligence can be operationalized using IQ.
- ▶ Knowledge of road safety can be operationalized using a score on a test.

Identifying variables for your project

With the problem diagram as the starting point, identify for your study:

- ▶ What are the variables for each objective?
- ▶ Which are independent, dependent, and confounding variables?
- ▶ Which can be measured as they are?
- ▶ Which need to be operationalized?
- ▶ Choose indicators to measure them
- ▶ What definitions are needed for the variables and indicators chosen?
- ▶ Which need further research to be defined adequately?

Qualitative studies

- We often don't have enough information to define indicators in qualitative studies prior to undertaking research.
- In fact, in many qualitative studies identifying variables and finding indicators that make them measureable is the research goal (e.g. stigma, stress).

Annoucements

Additional resources for learning R are posted on the course website here: http://sites.bu.edu/gh811/logistics/

Reminders

- Problem diagram, objectives and variables worksheet and annotated bibliography are due on Blackboard Sunday 5 PM
- Problem Set 1 due Tuesday, February 12 at 2 PM