PAI 705
McPeak
Lecture 4
RESEARCH DESIGN

Purposes of research.

- Exploratory.
    - To answer (or potentially ask) a question the researcher has and develop a better understanding.
        - Do gender roles matter in income generation strategies in this area?
        - Do gender role patterns change when looking at different sub-groups?
        - What are the broad outlines of roles I can ask questions about? Production? Marketing? Input provision? Labor?
    - To test the feasibility of conducting a more extensive investigation
        - Is there enough variation across households that I would learn something from asking about it further?
        - Is it a topic people will feel comfortable answering questions about or is it a sensitive issue?
    - To develop the methods that can be used in a subsequent study
        - How do I ask questions about 'income generation strategies' that give me information I

can use and people can understand what I am asking and sensibly respond?
        - Income over what time; daily? Weekly? Monthly? Seasonal? Annual? Different kinds of years?
        - Production meaning what? What is a way to ask this that has salience?
- Description
    - Observe and describe what was observed.
        - Census, demography, crime statistics, ethnography, chronology – time line.
        - What, where, when, and how
- Explanation
    - What factors lead to an observed outcome?
    - Why did this happen?

Criteria for Nomothetic Causality.

- The variables are correlated.
    - This is a statistical question. Do the values of the variables co-move in some kind of generalizable way?
- The cause takes place before the effect.
    - With temporal ordering, this can be clearly identified.
    - But there are also outcomes that may be jointly determined by independent variables.
        - At the cross country level, high income, good public education, good health care, good

governance; each is both an outcome and a possible cause of the other outcomes.

- The relationship cannot be explained by a third, omitted, variable – the relationship is not spurious.
    - As ice cream sales go up, drowning deaths go up at beaches on Long Island. Therefore we will be banning the sale of ice cream at Long Island beaches.

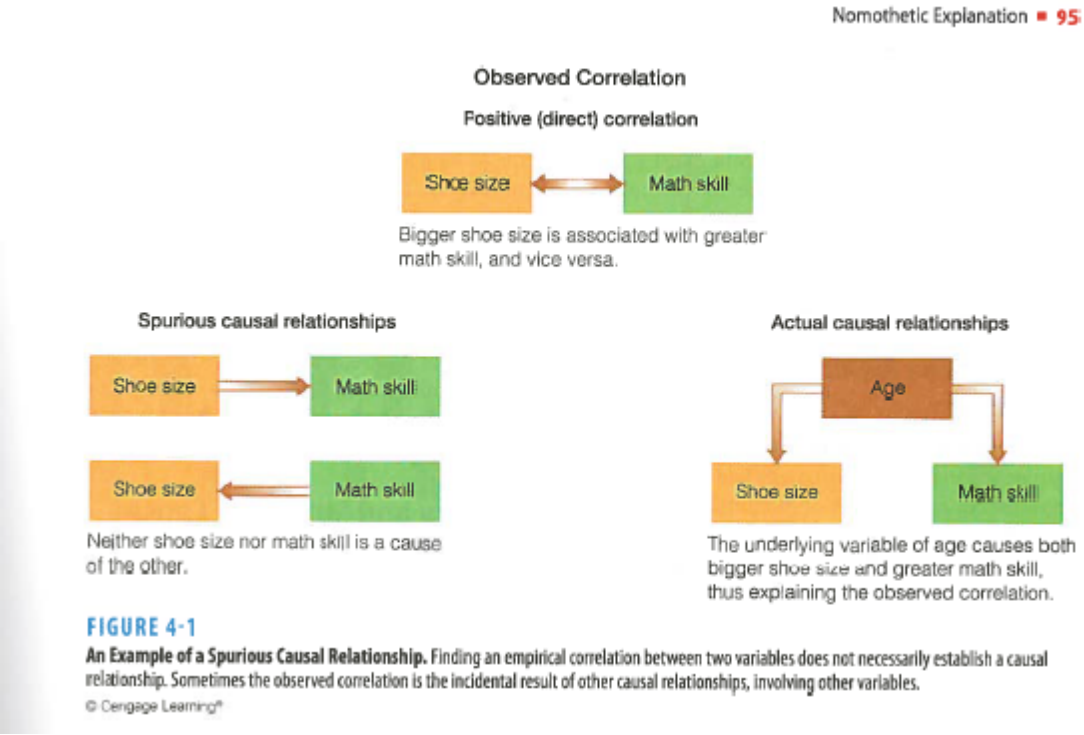


FIGURE 4-1

An Example of a Spurious Causal Relationship. Finding an empirical correlation between two variables does not necessarily establish a causal relationship. Sometimes the observed correlation is the incidental result of other causal relationships, involving other variables.

This type of reasoning leads one to hypothesis testing.

- What is the proposed relationship? Positive? Negative? Strong? Weak?
- What is my standard for a statistical test?
    - 90% sure, accepting a 10% chance I am saying it is there when it truly is not?
    - 95% sure, accepting a 5% chance?

    - 99% sure, accepting a 1% chance?
    - We can't ever get 100%, so there is always at least some chance we say it is there when it is not, and we also run a risk saying it is not there when it truly is. More on these later.

When we say we have a causal relationship between two variables we need to be clear what we are NOT claiming.

- Complete Causation.
    - The causal relationship is probabilistic and incomplete. It may be one of the causes, not the only cause (and then only probabilistically),
- X always leads to Y.
    - The causal relationship is not going to happen all the time. There are exceptions. It is a probability statement so some of the time it is not true that X leads to Y, but most times it is true. If X is a cat, then Y is hates baths. Overall, true. Odd cats exist who love baths.
- X leads to Y in the majority of cases.
    - The causal relationship may not hold for a majority of the cases. It just has to hold up as an empirical generality. Only 23% of households are female headed, but they approached traditional authorities for dispute resolution more than male headed households.

- In a regression of pet attitudes towards bathing, the significance of the cat dummy variable was negative and significant for 1= like and 0=hates coded responses. Cats were only 12% of the sample. The goldfish dummy in contrast was positive and significant.

Necessary and sufficient causes.

- Necessary – a precondition, one of the things that must hold true. If a necessary condition is not met, the outcome in question will not happen.

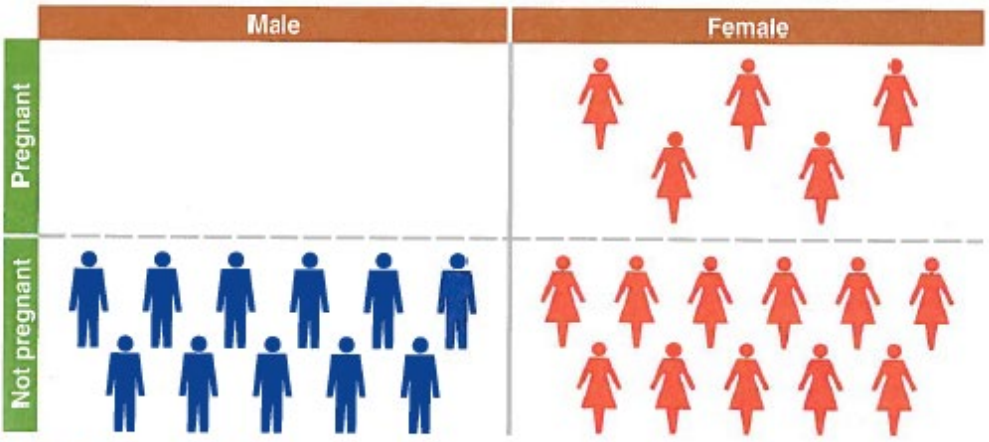


FIGURE 4-2
Necessary Cause. Being female is a necessary cause of pregnancy; that is, you can't get pregnant unless you are female.

- In 4-2, being female is a necessary condition of being pregnant. Is being pregnant a necessary condition of being female? Is it just being female that leads one to be pregnant or do other things have to happen as well?

- Sufficient – a condition that if present guarantees the proposed outcome. If you don't take the exam, you will get a failing grade for the exam.

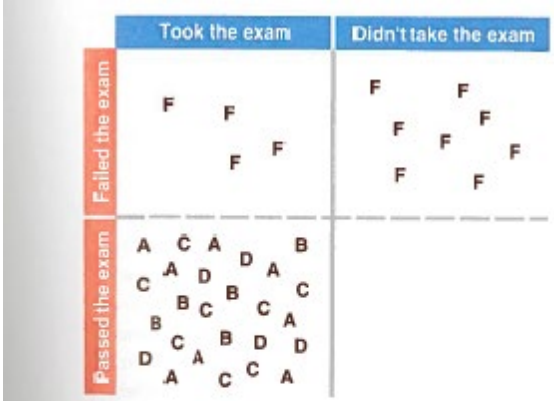


FIGURE 4-3
Sufficient Cause. Not taking the exam is a sufficient cause of failing it, even though there are other ways of failing (such as answering randomly).

  - Note that in figure 4-3 not taking the exam is sufficient but not necessary for receiving an F on the exam. Some took it and got an F so it is not the case that all F grades resulted from skipping the exam – there is another path to that outcome.

Units of analysis, units of observation.

- The what or who studied.
  - Nation, region, county, village, ethnic group, clan, program, project, intervention, case.
  - Household, individual within the household, female headed households, female heads of households.
    - Study individuals as members of groups; a particular set of responses from a person in the 20-30 Female registered Independent category.

- Study of groups. Syracuse Common Council is composed of individuals, and we could look at the group's gender composition, age composition, educational background, professional status, race, religion,…

The group is characterized by aggregates of the attributes of the individuals within the group.

The unit of analysis and the unit of observation may or may not be the same.

- Studying attitudes of IR students to the orientation program. The unit of observation is the individual student, the unit of analysis is the individual student.
- The unit of analysis is the household, the unit of observation is the household head. The head is reporting on behalf of the household since you are not sitting will all household members when you ask your questions.

- The unit of analysis is the infant, the unit of observation is both the parent reporting information and direct observation of the infant.

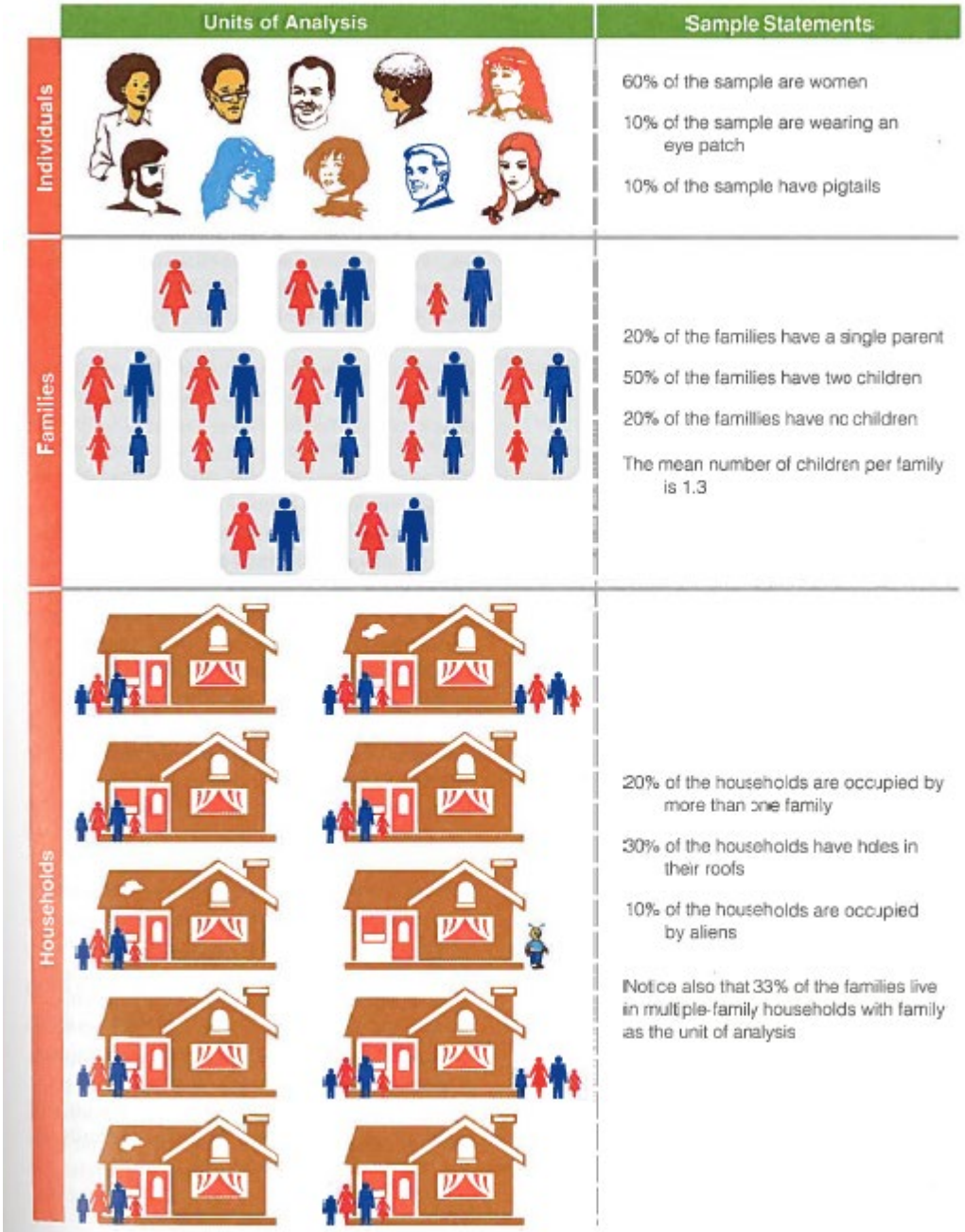
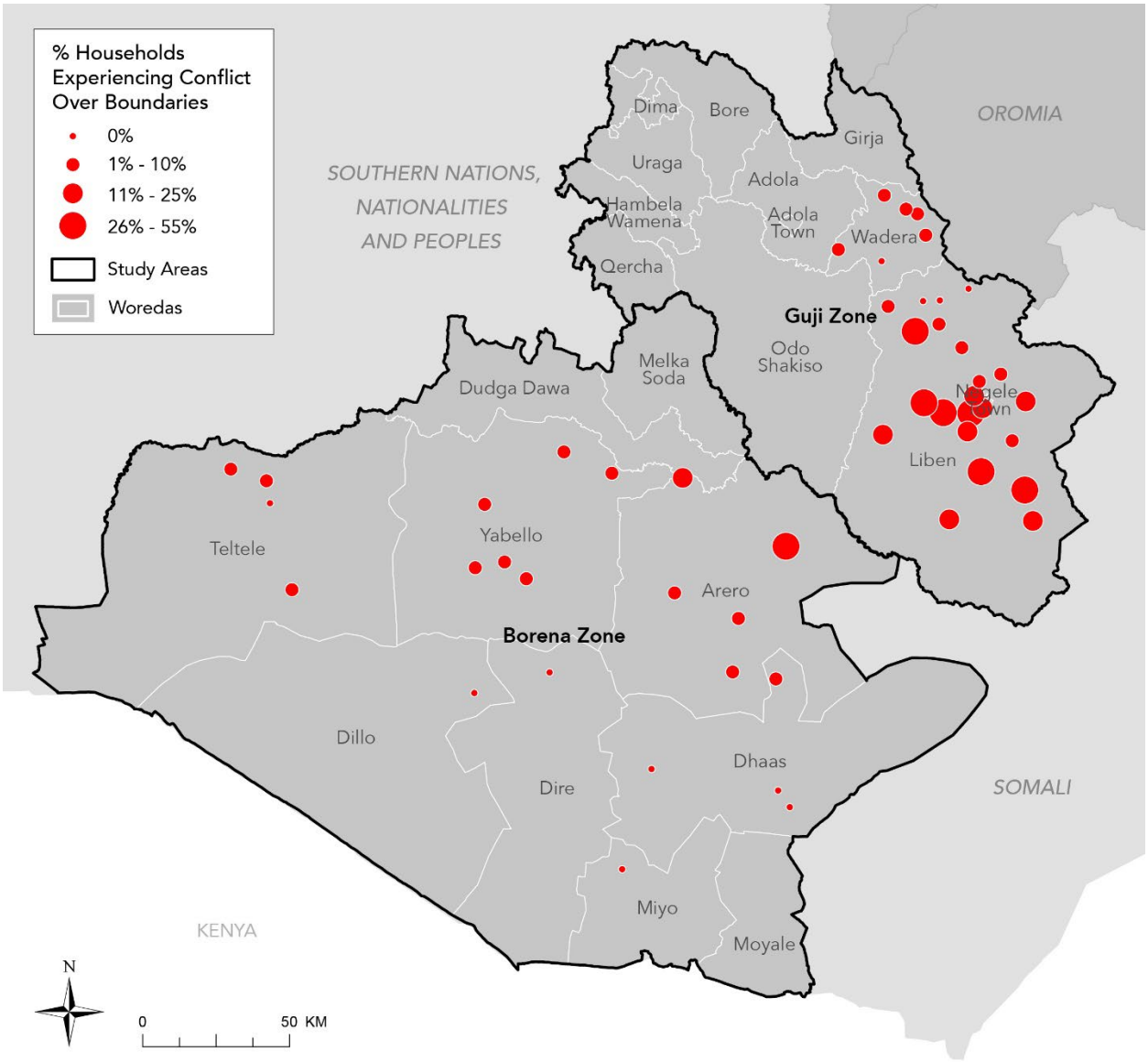


FIGURE 4-4

Illustrations of Units of Analysis. Units of analysis in social research can be individuals, groups, or even nonhuman entities.

- Start by looking at household level reports of conflict; was any in your area. Unit of analysis is household, unit of observation is household head.
    - Analysis may lead to considering it by community to identify patterns.
- Of those who say yes, how was the conflict resolved. Unit of analysis is conflict event, unit of observation is household head.



Units of observation are drawn from a particular population that you have some ability to define. The attributes required to be in the population and sample are part of your research construction.

- Residential: Households resident in 13210 zip code
- Filtered: Households in 13210 with children in Ed Smith School
- Further refined filtered: Households in 13210 with children in Kindergarten at Ed Smith School.
- Yet more: Households in 13210 headed by female divorced parents of children in Kindergarten at Ed Smith School who are left handed (the kids or the moms?).
- The text also points out that we may need to draw a conceptual distinction between 'family' and 'household'.

Statements can be made about patterns in the attributes.

Research can also be done on ‘social artifacts’.

- Gender roles in nursery rhymes.
- Barbie advertising over time.
- Elvis outfits.

The ecological fallacy.

- Applying what we learned about the whole to the individuals making up the whole.
- Drawing conclusions about individuals solely from observations of groups.
    - Made up example. PAIA cohorts as the group. The course evaluations for PAIA overall are higher when the female to male student ratio is higher. I can't conclude (though I might be tempted to) that female students give higher course evaluations than do male students.
        - My unit of observation is the student cohort, not the individual student. I am applying the cohort level observation to the individual.
        - It might be true, but other possible explanations might be reasonable
            - Classes are objectively better when the ratio is higher.
            - All the change could be coming from males changing their evaluations in response to the changing environment while females don't change – males are happier when there are more females around.

Can also operate in the opposite direction with ‘the individualistic fallacy”.

- Lifetime expected earnings increase with the highest level of education obtained.
- But Bill Gates dropped out of Harvard and is rich.
- So I am dropping out of college because Bill Gates shows you don’t have to go to college to be rich.

Reductionism

A reduction of the kinds of concepts to be considered relevant to understanding the phenomenon under study. Studying why Gabra herders decide to migrate from one point to another.

- Clearly it is a result of constrained optimization of a utility maximizing individual subject to a binding budget constraint.
- No, it is due to their interpretations of the visibility of clouds that reflects meteorological conditions and flowering plants that predict future vegetation.
- No it is due to social and clan dynamics that are rooted in the Proto Rendille Somali culture that is the foundation for Gabra culture.
- No it is due to class oppression where the elite ruling class in conjunction with neo colonial powers has alienated vast areas of dry season grazing reserve thus giving them very little option on where to move.

- It has to do with their relationship with their mother when they were little.
- The camels peed in the water here, and they did not pee in the water over there.

Reductionism is to some degree inevitable if we are to place our analysis in a disciplinary methodology.

The challenge is to keep in mind that what we are producing is an analysis and an explanation, not necessarily the explanation.

- I produced an economic analysis of migration decisions. That is not the only valid analysis of migration decisions that is possible of producing insights and generating understanding.

TIME DIMENSIONS

Cross Section Studies

- One round of use of a survey instrument within a given time period on a sample population. A census. A household survey. Multiple units asked the same question at a (more or less) given point in time. In September- October 2014, we asked 3827 heads of households in Borana and Guji areas of Ethiopia questions using a questionnaire.
- This is potentially representative of that point in time.
    - Dry season or wet season?
    - Good year or bad year?
    - Position of key slow moving variables, markets, politics, expansion of infrastructure, recent conflicts, Al Shabab, climate change….

Longitudinal Studies

- Observations of the same phenomenon over an extended period of time.
    - Trend studies. Looks at changes in a population over time. Census data. Married couples. Racial profile. Demographic pyramid.
        - https://www.census.gov/popclock/
        - Example in the book; a study of political knowledge over time finds that people in 1989 were better informed than people in 1947 about politics.

        - However, when they factored in that education levels were higher in 1989 to 1947 they found political knowledge actually decreased in particular sub groups.
    - Cohort Studies. Take a specific population and survey them repeatedly as they change of time. A sample of people born in 1960 is sampled every 5 years from 1980 on. So a sample of them when 20. A sample when 25. A sample when 30,….Not the same people from round to round unless by chance.

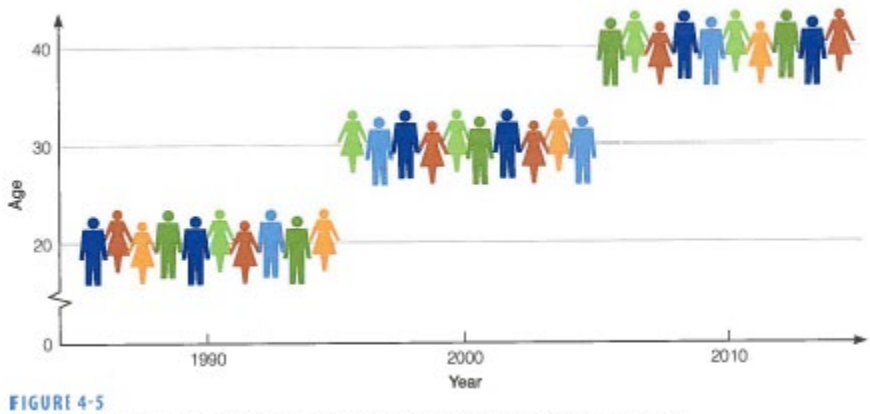


FIGURE 4-5
A Cohort Study Design. Each of the three groups shown here is a sample representing people who were born in 1970.

    - Panel Studies. Repeated survey of the same (as much as possible) sample over time.
        - Common to have a baseline and repeat rounds in impact evaluation.
            - Treatment and control.
        - Panel mortality / attrition. Do the missing units differ systematically from the ones you were able to track down.

        - Panel at times may be at the community level. They change the sample in the community, but looking at the evolution of the mean from the community over time.
        - A cross section can give panel like data if done using retrospective questions.
            - Herd structure and milk production.

Issues of using cross sectional data to make over time predictions.

- Kuznet's curve.
- Lifetime income-earning profile from a cross section survey.
- Cross country comparison to identify what will lead to income growth.

Mixed modes.

- In Kenya and Ethiopia, we conducted the panel study of households and we were able to identify empirically some poverty traps.

- An anthropologist who was part of our team went and conducted interviews with selected households in different quadrants of our poverty and wealth categories. What led to a fall into poverty? What led to a climb out of poverty? What kept a household in a poverty trap? What kept a household from falling into a poverty trap? Life

histories to give more meaning to ‘bins’ drawn from survey data.

- Survey data mixed with Key Informant interviews and focus group discussions.

How to design a research project? Idealized version in an image:

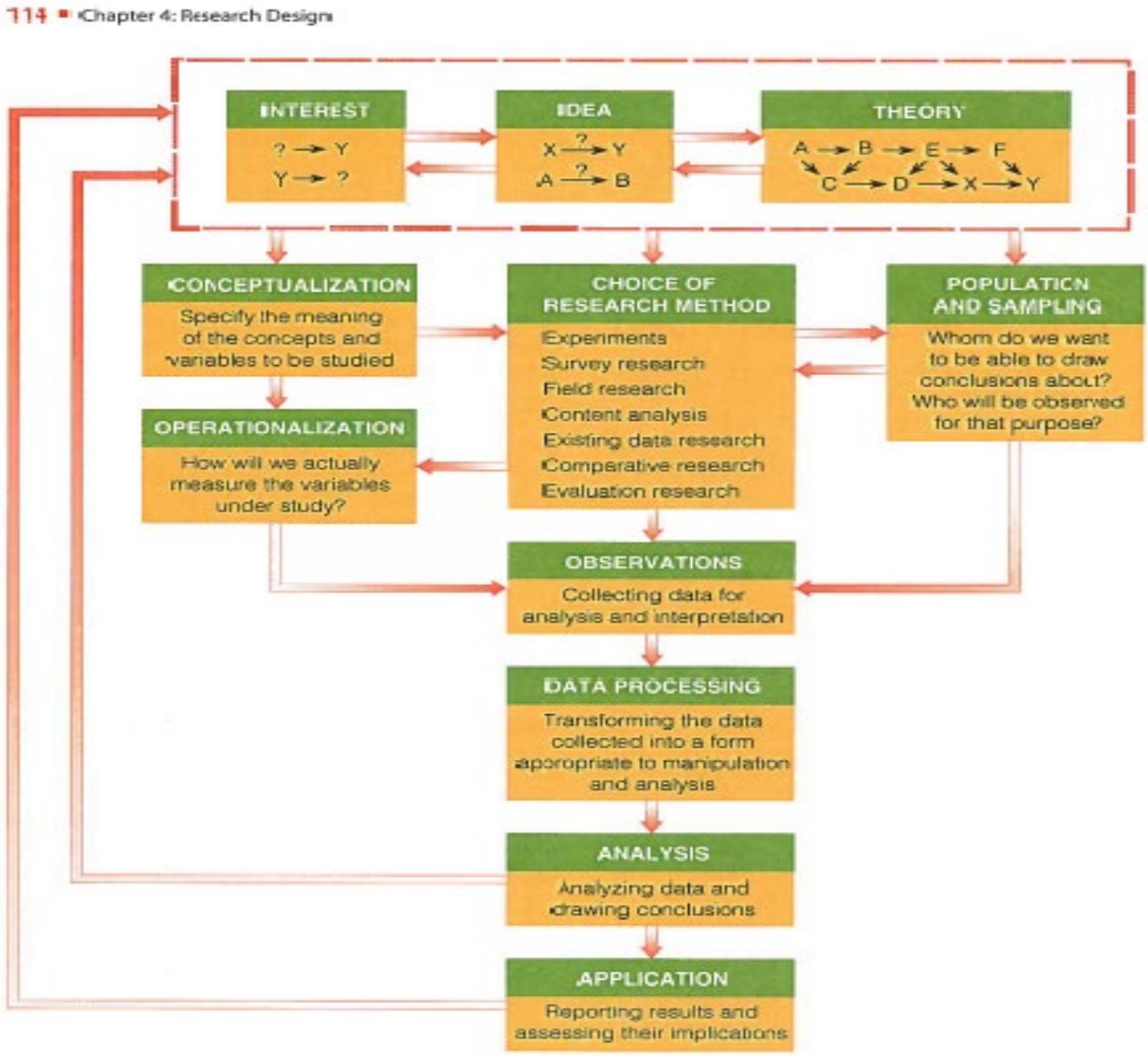


FIGURE 4-7
The Research Process. Here are some of the key elements that we'll be examining throughout this book: the pieces that make up the whole of social research.

An example of this: What explains the difference in living standards observed in different countries of the world today?

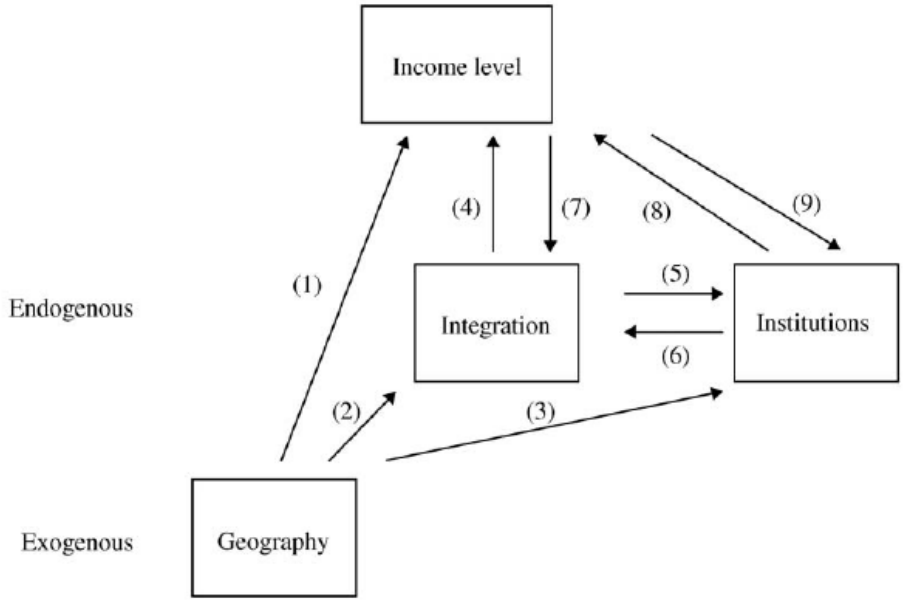


Figure 1. The ''deep'' determinants of income.

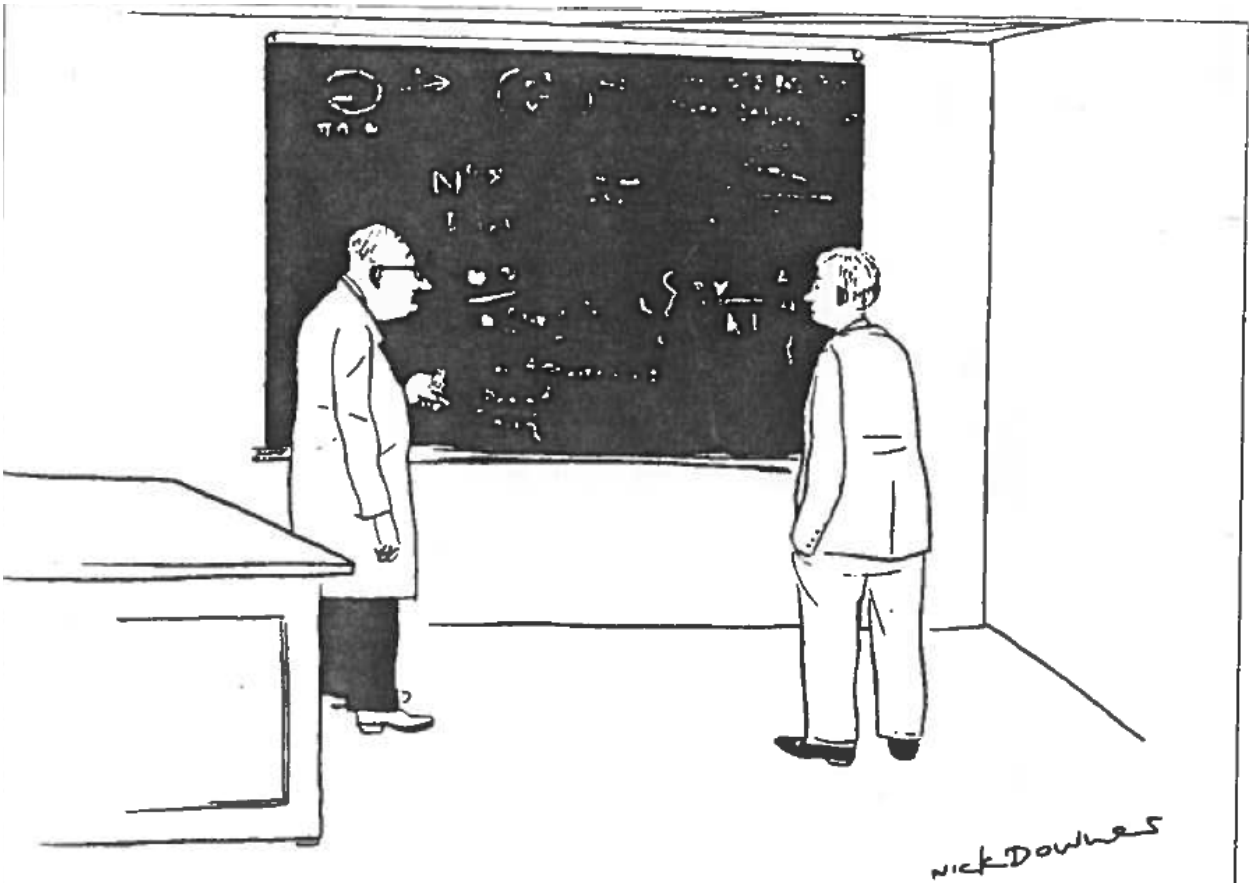
Getting Started. Literature review.

http://apps.webofknowledge.com/DRCI_GeneralSearch_input.do?product=DRCI&search_mode=GeneralSearch&SID=3DepZUGKL3LbsM5coK1&preferencesSaved=

What has been done by whom already? What empirical evidence exists? What is the history and context of the area I am studying? What theories can I draw on to frame my analysis?

What kind of a study is this that I am doing; exploratory, descriptive, explanatory? What am I trying to accomplish with this study? Who is the audience for this study? Why is the question I am asking interesting? What will I be able to say when this is finished that I am not able to say now?

For more academic research, we sometimes talk of it passing the ‘mom and dad test’. Could I explain what I am doing and why it is important in a way mom and dad could understand? If I cannot I may need to further clarify my thinking and approach to refine my understanding of what I am trying to do.

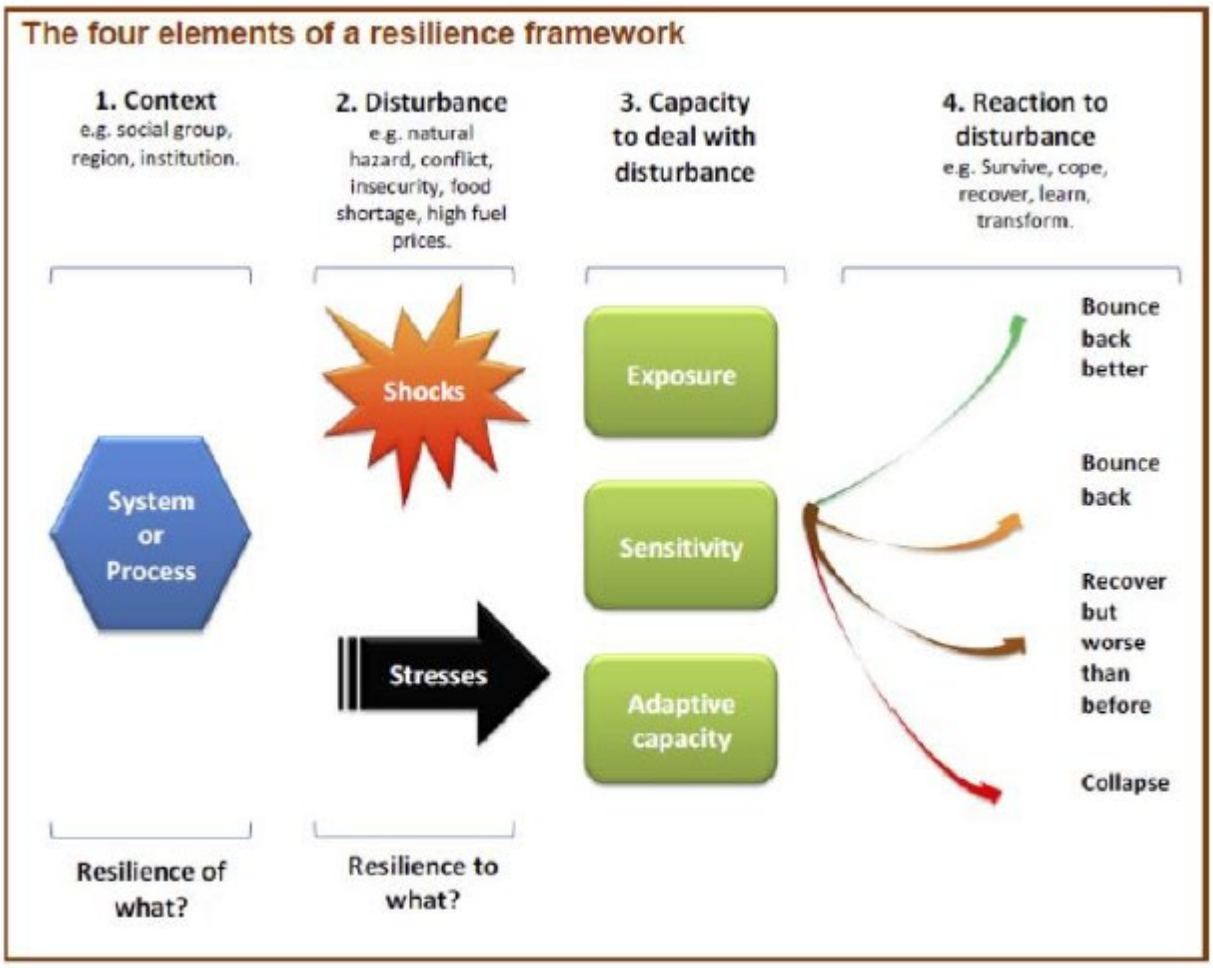


"In layman's terms? I'm afraid I don't know any layman's terms."

Conceptualization.

Example from Rodrik et al above. Difference in institutions, openness to trade, and geography lead to differences in income per capita across countries. Describe how each is logically related to the outcome in question.

If I am going to ask in the survey, what are the dimensions I will measure? Recall the resilience example from earlier in the semester.


The four elements of a resilience framework
1. Context
e.g. social group, region, institution.
2. Disturbance
e.g. natural hazard, conflict, insecurity, food shortage, high fuel prices.
3. Capacity to deal with disturbance
4. Reaction to disturbance
e.g. Survive, cope, recover, learn, transform.
Shocks
Exposure
System or Process
Sensitivity
Stresses
Adaptive capacity
Bounce back better
Bounce back
Recover but worse than before
Collapse
Resilience of what?
Resilience to what?

Box 2. Dimensions of resilience

A number of studies define 'dimensions' of resilience, which have similarities to the five dimensions or 'capitals' defined in earlier livelihood frameworks. For example, a study by Oxfam GB defines five dimensions of resilience which were applied to a study of disaster risk reduction in Ethiopia's Somali region[6]. A study commissioned by DFID and undertaken by the authors of this guidance reviewed a number of methodologies for measuring resilience, and identified nine, very broadly defined, 'dimensions' of resilience based on these methodologies[7]. These are listed below. Dimensions 1-5 were common to all the methodologies reviewed that defined dimensions of resilience. Dimensions 6-9 represent factors that were identified by a subset of the methodologies reviewed. It is not recommended that these dimensions are used in a prescriptive manner. However, they may be useful as a loose framework for guiding the process of identifying contextual factors that are important in influencing resilience.

1. **Assets,** including physical and financial assets, food and seed reserves, and other assets that can be deployed or realised during times of hardship to help people absorb losses, and recover from stresses and shocks. Debt could be considered as a negative asset.
2. **Access to services**, including water, electricity, early warning systems, public transport, and knowledge and information that helps people plan for, cope with and recover from stresses and shocks, and how vulnerable these services are themselves to shocks and stresses.
3. **Adaptive capacity**, including factors that specifically enable people to anticipate, plan for and respond to changes (for example by modifying or changing current practices and investing in new livelihood strategies). The ability to adapt to changes in any of the other dimensions listed here might also be included.
4. **Income and food access**, including the vulnerability to shocks and stresses of income sources and food supplies (including food prices/ability to purchase or otherwise access food, and the vulnerability of food supply chains to local and remote shocks and stresses).
5. **Safety nets**, including access to formal and informal support networks, emergency relief, and financial mechanisms such as insurance.
6. **Livelihood viability**, in terms of the extent to which an individual's livelihood can be sustained in the face of a shock or stress, or the magnitude of shock or stress that can be accommodated before a livelihood ceases to be viable.
7. **Institutional and governance contexts**, including extent to which governance processes, institutional mechanisms, policy environments, conflict, and insecurity constrain or enable coping and adaptation. It can include community level capacity to cope with and adapt to shocks and stresses and to support those living within it.
8. **Natural and built infrastructural contexts**, including extent to which coping and adaptation is facilitated or constrained by the quality of built infrastructure (e.g. roads), the quality/functioning of environmental systems/natural resources (e.g. health of ecosystems providing livelihoods), and geographical factors (e.g. remoteness) and the vulnerability of the infrastructure to shocks and stresses.
9. **Personal circumstances**, including any factors not covered by other dimensions that might make an individual more or less able to anticipate, plan for, cope with, recover from, or adapt to changes in stresses and shocks. These might include psychological resilience, past experience of coping, personal connections (social capital), health, socio-economic status, etc.

Coping capacity versus adaptive capacity

- Food Security
- Self-reported resilience
- Access to services
- Access to markets
- Seasonal differences in this
- Access to information to prepare and minimize impact
- Experience of shocks in past year
- Techniques to minimize vulnerability to shocks identified.

Applying the Concept of Resilience to Pastoralist Household Data. With Peter Little. *Pastoralism: Research, Policy and Practice* (2017)

Poverty

- Stages of progress exercise
  https://sites.duke.edu/krishna/files/2013/06/SoP.pdf
- Relative poverty assessments
- $1.25 USD per capita in purchasing power parity adjusted units (or more recently $1.90).
- US Census bureau
  http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html
  https://www.census.gov/data-tools/demo/saipe/saipe.html?s_appName=saipe&map_yearSelector=2016&map_geoSelector=aa_c
- UNDP multidimensional poverty index
  https://hdr.undp.org/sites/default/files/2021_mpi_report_en.pdf

Choice of Research Method.

- Does data already exist?
- What method of gathering data gives me the kind of information I need to answer the questions that interest me?

Operationalization.

- What is the measure to proxy for quality of institutions? For trade openness? For geography?
- Where does my income measure comes from?
  - Cash income?
  - Full income?
- Are you poor, yes = 1, no = 0?
- UN data set? Census bureau?
- Drafting a survey instrument that is going to work.

- Testing and revising instruments.
- Language issues in posing questions.

Population and sampling.

- What is the population of interest from whom I will draw a sample? The large N question. To whom can I extrapolate my findings?
    - Random sample of village residents.
        - How did I select my villages?
            - Market access
            - Agroecology
            - Easy to get to?
    - PAIA students
        - Currently in Syracuse?
            - Full time?
            - MPH / EMPA / PhD / MAIR / MPA / MAIR / JDMPA/ Public Diplomacy / Econ IR…..
            - First year – second year – third year
            - On e-mail
            - Into responding to survey monkey
    - How do I draw the sample?
        - Snowball
        - Random from a list
        - Transect
        - Anybody willing to answer my questions?

Observations

- Pencil hits the paper
- Tablet

- Interview
- Observation
- Getting data on variable responses to similar questions across your sample.
- Enumerator management
- Enumerator verification
- Language issues

Data processing

- How do I get the information on these pieces of paper into a form that can be analyzed?
- Training data entry people
- Software for data entry
- Data cleaning
- Missing observations
- Clearly wrong answers
    - 100.00 or 10000?

Analysis

- Descriptives as entry point
- Range of responses also worth looking at early in the process (and reporting later in the process).
- Move on to explanatory analysis and causation/ hypothesis testing.
- Analysis of subgroups and subtopics

Application and reporting out.

- How do you succinctly summarize what you have discovered?
- Why is this important?
- What does this lead us to consider as a program / policy response to what you have discovered?
- What did you discover that we still don’t know and need to know?