POLS 205
Political Science as a Social Science

Qualitative & Quantitative Research Design

Christopher Adolph

University of Washington, Seattle

April 19, 2010

Outline

Distinguishing Qualitative & Quantitative Methods

Methods of Observation

Inference in Qualitative Research

Case Selection Strategies

Qualitative vs. Quantitative: How to choose

What is *quantitative analysis*?

Use of statistical methods to understand relationships among numerically* coded variables

*Numerically coded $\neq$ Intrinsically numeric.

Examples of numerically coded variables:

- What is your personal income in dollars
- What is the consensus of an expert panel on whether a country was a majoritarian democracy ?
- What is the degree of ethnic division in a society, as defined by ethno-linguistic fractionalization?

  (Ethnolinguistic fractionalization is the probability that two randomly drawn people either self-identify as different ethnicities or speak different native languages.)

Suprising facts about numerically coded variables

1. **Can often summarize complex assessments in a single number:**

   Ethnic division sounds hard to quantify without absurd oversimplification

   But ethnolinguistic fractionalization turns “easily” collected data into comparable, quantitative measure

Suprising facts about numerically coded variables

1. **Can often summarize complex assessments in a single number:**

   Ethnic division sounds hard to quantify without absurd oversimplification

   But ethnolinguistic fractionalization turns "easily" collected data into comparable, quantitative measure

2. **Can reliably code many qualitative concepts into numbers:**

   If we can turn a qualitative variable into a systematic, binary distinction *or* a series of binary distinctions, we can make it numerical

   We do this by coding the variable as 0 or 1 (adds no assumptions)

Suprising facts about numerically coded variables

1. **Can often summarize complex assessments in a single number:**

   Ethnic division sounds hard to quantify without absurd oversimplification

   But ethnolinguistic fractionalization turns "easily" collected data into comparable, quantitative measure
2. **Can reliably code many qualitative concepts into numbers:**

   If we can turn a qualitative variable into a systematic, binary distinction *or* a series of binary distinctions, we can make it numerical

   We do this by coding the variable as 0 or 1 (adds no assumptions)
3. **Can even code subjective variables as numbers, as long as we do so consistently across cases**

Quantitative methods can study quantitative *or* qualitative data!

What are qualitative methods, then?

Use of systematic but non-statistical methods to understand relationships among qualitative or quantitative variables

Note several things:

- Qualitative methods are varied. “Non-statistical” covers a lot of ground.
- Variables need not be numerically coded. May be precisely qualitatively coded, but might not. Could be “thick”, or detailed.
- Description may be as important or more important than inference of relationships. May be laying foundation for later measurement & inference.

How much data?

Political scientists often distinguish studies by amount of data, N:

$N = 1$ These are always qualitative: *Case studies*.
What is a "case"? Chance to measure the dependent variable.
A study of a single country may not be a "case study" if multiple chances to observe the DV in that country

How much data?

Political scientists often distinguish studies by amount of data, N:

$N = 1$ These are always qualitative: *Case studies*.
What is a "case"? Chance to measure the dependent variable.
A study of a single country may not be a "case study" if multiple chances to observe the DV in that country

$1 < N < k$ for k small, e.g. 2 to 10. Usually qualitative
Comparative case study Trace out several cases in parallel
Small-N design Code & qualitatively analyze several cases in detail (Gourevitch)

How much data?

Political scientists often distinguish studies by amount of data, N:

$N = 1$ These are always qualitative: *Case studies*.
What is a "case"? Chance to measure the dependent variable.
A study of a single country may not be a "case study" if multiple chances to observe the DV in that country

$1 < N < k$ for k small, e.g. 2 to 10. Usually qualitative
Comparative case study Trace out several cases in parallel
Small-N design Code & qualitatively analyze several cases in detail (Gourevitch)

$N > k$ This is a *"large"-N* design. Usually quantitative

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

Example: Watch a peace protest, or take part as a protester to interact & observe

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

Example: Watch a peace protest, or take part as a protester to interact & observe

Overt *or* Covert Are subjects aware of your presence? (May affect behavior!)

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

Example: Watch a peace protest, or take part as a protester to interact & observe

Overt *or* Covert Are subjects aware of your presence? (May affect behavior!)

Example: Hidden camera of visitors to a Congressional office, versus press releases of visitors

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

Example: Watch a peace protest, or take part as a protester to interact & observe

Overt *or* Covert Are subjects aware of your presence? (May affect behavior!)

Example: Hidden camera of visitors to a Congressional office, versus press releases of visitors

Structured *or* Unstructured Are variables and their measures pre-determined and collected uniformly

How are observations made?

Direct *or* Indirect Do you observe the variables, or consequences of them

Example: Ask members of Congress how liberal they are, or measure their votes

Participant *or* Non-participant Are you a neutral observer, or an actor among other subjects

Example: Watch a peace protest, or take part as a protester to interact & observe

Overt *or* Covert Are subjects aware of your presence? (May affect behavior!)

Example: Hidden camera of visitors to a Congressional office, versus press releases of visitors

Structured *or* Unstructured Are variables and their measures pre-determined and collected uniformly

Example: Survey of day-laborers at a street-corner, versus hours spent in free-form conversation

Where do you get your data?

Archives Collect historical records, government data, journalistic accounts, diaries, etc. Code either qualitatively or quantitatively.

Example: Gourevitch's historical research on Keynesianism

Where do you get your data?

Archives Collect historical records, government data, journalistic accounts, diaries, etc. Code either qualitatively or quantitatively.

Example: Gourevitch's historical research on Keynesianism

Surveys Randomly or systematically select a group of people; ask a questionnaire.

Example: Gelman's research on voting in 50 states

Where do you get your data?

Archives Collect historical records, government data, journalistic accounts, diaries, etc. Code either qualitatively or quantitatively.

Example: Gourevitch's historical research on Keynesianism

Surveys Randomly or systematically select a group of people; ask a questionnaire.

Example: Gelman's research on voting in 50 states

Ethnography/Fieldwork Invest many hours in following actors in the field; collect detailed, qualitative observations; soaking-and-poking

Example: Richard Fenno's *Homestyle*; followed 17 members of Congress over 1970–76, personally observing styles of interaction with constituents over thousands of hours

Fenno, *Homestyle*

Congressman J: Person-to-person, constituent service

Congressman J comes from a hetereogenous district. . . . "It's in effect four different districts. . . "

[Congressman J's] preferred presentational style was (and is) person to person. "It doesn't fit my temperament to handshake without stopping to talk with each person. Then I feel like I can understand them and they can understand me."

. . . the segmented district he swallowed in small, sequential bites now protects him. No one can swallow it whole. . . "Each area is only interested in its own problems. That's what makes it so easy for an incumbent. He can provide services in each of the towns. . . In a totally suburban district someone can move in and take over. You can't do that with mine. It's too big and diverse.

Fenno, *Homestyle*

Congressman E: Public events, constituent service

"This district", [Congressman E] says, "is a microcosm of the nation. We are geographically southern and politically northern. We have agriculture. . . We have big business like Union Carbide and General Electric. We have unions. We have a city and we have small towns. . .

Congressman E is not a native of the area. So he does not view the district with Congressman B's ingrained local-boy sensitivity. . .

When asked to select the most important event of a busy campaign weekend, he chose walking in a parade. "Whenever you get 10,000 people, that's important.. . . But the most productive use of my time was my news conference. That got me more coverage—all the TV stations and two newspapers."

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives				

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*			

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant		

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys				

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*			

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant		

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	Structured*
Ethnography				

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	Structured*
Ethnography	Direct			

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	Structured*
Ethnography	Direct	Participant*		

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	Structured*
Ethnography	Direct	Participant*	Overt	

Classifying methods of observation

Approach	Direct?	Participant?	Overt?	Structured?
Archives	Indirect*	Non-participant	Overt*	Structured*
Surveys	Direct*	Non-participant	Overt*	Structured*
Ethnography	Direct	Participant*	Overt	Unstructured

* Usually, but not always

A large debate on inference in qualitative methods

1. King, Keohane, & Verba: Qualitative Inference is subject to all the same rules as quantitative inference

   Same pitfalls, same dangers

   So should adapt strategies from quantitative to combat them
2. Brady & Collier & others: Not so simple

   Qualitative methods may tackle some problems quantitative can't

   Examples: Better at tracing causal processes, thickly describing cases

Your instructor is a.) quantative and b.) closer to the KKV camp, but you should keep an open mind

If qualitative researchers don't feed their data into a statistics package, how do they infer causation?

Mill's Methods of Inference

A key starting point for qualitative inference is 19th c. political philosopher J. S. Mill

J. S. Mill (1843, *A System of Logic*) proposed five methods of inference:

1. Method of Agreement
2. Method of Difference
3. Joint Method of Agreement and Difference
4. Method of Concommitant Variation
5. Method of Residues

1, 2, and 4 predate Mills: Abu Ali Sina Balkhi (Avicenna), *The Canon of Medicine*, 1025 CE.

Mill's Method of Agreement

Potential causes							Potential effects						
A	**B**	**C**	**D**	**E**	**F**	**G**	**t**	**u**	**v**	**w**	**x**	**y**	**z**
•	•	•	•				•	•	•	•			
•				•	•	•				•	•	•	•

Mill's Method of Agreement says to look at the common factor in all the positive cases

Mill's Method of Agreement

Potential causes							Potential effects						
A	**B**	**C**	**D**	**E**	**F**	**G**	**t**	**u**	**v**	**w**	**x**	**y**	**z**
•	•	•	•				•	•	•	•			
•				•	•	•				•	•	•	•

Mill's Method of Agreement says to look at the common factor in all the positive cases

We infer that A causes/contributes to/jointly causes w, or *vice versa*

Mill's Method of Agreement

Potential causes							Potential effects						
A	**B**	**C**	**D**	**E**	**F**	**G**	**t**	**u**	**v**	**w**	**x**	**y**	**z**
•	•	•	•				•	•	•	•			
•				•	•	•				•	•	•	•

Mill's Method of Agreement says to look at the common factor in all the positive cases

We infer that A causes/contributes to/jointly causes w, or *vice versa*

In symbols, we infer:

$$A \Leftarrow w \qquad or \qquad A \Rightarrow w \qquad or \qquad A \Leftrightarrow w$$

Mill's Method of Agreement: Example

	Potential causes				Potential effects
Subjects	Spinach	Curry	Shrimp	Tuna	Sick?
Me	•	•	•	•	•
E.	•			•	•
S.	•	•			•
A.	•		•		•

If we apply Mill's Method of Agreement, what do we conclude?
Which potential causal variable(s) is common to all cases of sickness?
Spinach is a likely cause/contributor/effect of being Sick

Are you persuaded?

Suppose I tell you that the Method of Agreement supports my hypothesis. Are you persuaded?

Are you persuaded?

Suppose I tell you that the Method of Agreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "Americans cause murder"
My data: The nationality of the last ten people to commit murder in Seattle
My evidence: All are American!

Are you persuaded?

Suppose I tell you that the Method of Agreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "Americans cause murder"
My data: The nationality of the last ten people to commit murder in Seattle
My evidence: All are American!

I try again: "All murderers have arms."
My evidence: In a large, random sample of murderers, all had two arms.

Are you persuaded?

Suppose I tell you that the Method of Agreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "Americans cause murder"
My data: The nationality of the last ten people to commit murder in Seattle
My evidence: All are American!

I try again: "All murderers have arms."
My evidence: In a large, random sample of murderers, all had two arms.

Third try: "Murderers said they felt economically insecure."
But wait! Wouldn't a sample of non-murderers say the same during a recession?

Are you persuaded?

Suppose I tell you that the Method of Agreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "Americans cause murder"
My data: The nationality of the last ten people to commit murder in Seattle
My evidence: All are American!

I try again: "All murderers have arms."
My evidence: In a large, random sample of murderers, all had two arms.

Third try: "Murderers said they felt economically insecure."
But wait! Wouldn't a sample of non-murderers say the same during a recession?

Last attempt: "Murder is always the result of drug abuse." . . . At least in my 5 cases. . .

Problems with Method of Agreement

- Need a random or representative sample, preferably large (tough in qualitative!)

Problems with Method of Agreement

- Need a random or representative sample, preferably large (tough in qualitative!)
- May highlight trivial necessary conditions

Problems with Method of Agreement

- Need a random or representative sample, preferably large (tough in qualitative!)
- May highlight trivial necessary conditions
- How do I know that positive cases differ from negative?

Problems with Method of Agreement

- Need a random or representative sample, preferably large (tough in qualitative!)
- May highlight trivial necessary conditions
- How do I know that positive cases differ from negative?
- (The big one) What if the effect is probabilistic?

Mill's Method of Difference

Potential causes							Potential effects						
A	B	C	D	E	F	G	t	u	v	w	x	y	z
•	•	•	•	•	•	•	•	•	•	•	•	•	•
	•	•	•	•	•	•		•	•	•	•	•	•

Mill's Method of Diagreement says to look at the distinguishing factor when comparing positive and negative cases

Mill's Method of Difference

Potential causes							Potential effects						
A	B	C	D	E	F	G	t	u	v	w	x	y	z
•	•	•	•	•	•	•	•	•	•	•	•	•	•
	•	•	•	•	•	•		•	•	•	•	•	•

Mill's Method of Diagreement says to look at the distinguishing factor when comparing positive and negative cases

We infer that A causes/contributes to/jointly causes w, or *vice versa*

Mill's Method of Difference

Potential causes							Potential effects						
A	B	C	D	E	F	G	t	u	v	w	x	y	z
•	•	•	•	•	•	•	•	•	•	•	•	•	•
	•	•	•	•	•	•		•	•	•	•	•	•

Mill's Method of Diagreement says to look at the distinguishing factor when comparing positive and negative cases

We infer that A causes/contributes to/jointly causes w, or *vice versa*

In symbols, we infer:

$$A \Leftarrow w \qquad or \qquad A \Rightarrow w \qquad or \qquad A \Leftrightarrow w$$

Mill's Method of Difference: Example

	Potential causes				Potential effects
Subjects	Spinach	Curry	Shrimp	Tuna	Sick?
Me	•	•			
E.	•	•	•	•	•
S.	•			•	
A.	•	•		•	

If we apply Mill's Method of Disagreement, what do we conclude?

Mill's Method of Difference: Example

	Potential causes				Potential effects
Subjects	Spinach	Curry	Shrimp	Tuna	Sick?
Me	•	•			
E.	•	•	•	•	•
S.	•			•	
A.	•	•		•	

If we apply Mill's Method of Disagreement, what do we conclude?

Which potential causal variable(s) are only present in cases of sickness?

Mill's Method of Difference: Example

	Potential causes				Potential effects
Subjects	Spinach	Curry	Shrimp	Tuna	Sick?
Me	•	•			
E.	•	•	•	•	•
S.	•			•	
A.	•	•		•	

If we apply Mill's Method of Disagreement, what do we conclude?

Which potential causal variable(s) are only present in cases of sickness?

Shrimp is a likely cause/contributor/effect of being Sick

Are you persuaded?

Suppose I tell you that the Method of Disagreement supports my hypothesis. Are you persuaded?

Are you persuaded?

Suppose I tell you that the Method of Disagreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "English-language heritage causes economic development"
My evidence: Comparison of GDP of US and Haiti

Are you persuaded?

Suppose I tell you that the Method of Disagreement supports my hypothesis. Are you persuaded?

Suppose my hypothesis is that "English-language heritage causes economic development"
My evidence: Comparison of GDP of US and Haiti

I try again: "Long arms cause long legs."
My evidence: Measurement from a random sample of undergrads.

Problems with Method of Differences

- Need a random or representative sample, preferably large (tough in qualitative!)

Problems with Method of Differences

- Need a random or representative sample, preferably large (tough in qualitative!)
- Hard to detect spurious relationships

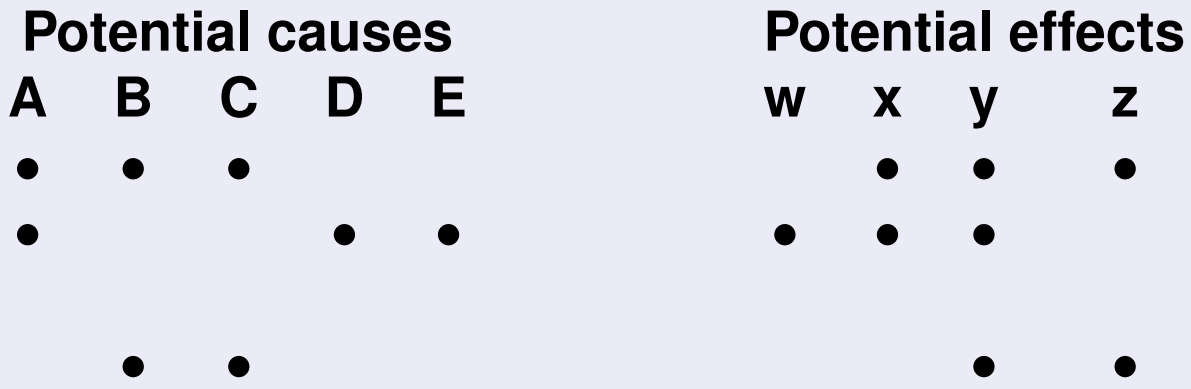
Problems with Method of Differences

- Need a random or representative sample, preferably large (tough in qualitative!)
- Hard to detect spurious relationships
- Hard to distinguish major and minor causes

Problems with Method of Differences

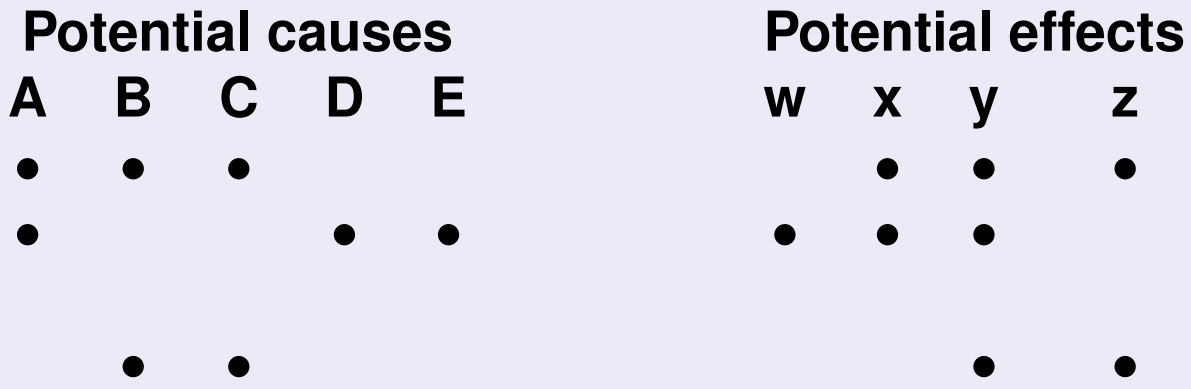
- Need a random or representative sample, preferably large (tough in qualitative!)
- Hard to detect spurious relationships
- Hard to distinguish major and minor causes
- (The big one) What if the effect is probabilistic?

Mill's Joint Method of Agreement & Difference



The joint method combines the first two: there is a potential cause present in positive cases and missing in negative ones.

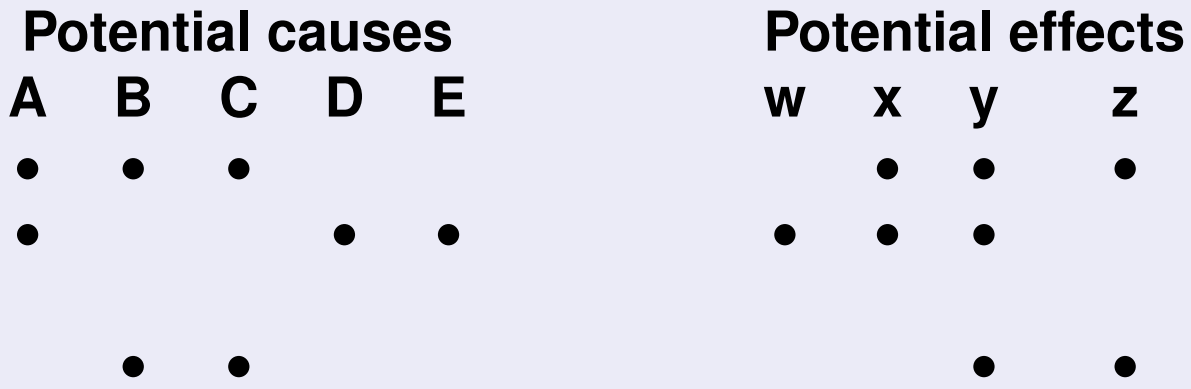
Mill's Joint Method of Agreement & Difference



The joint method combines the first two: there is a potential cause present in positive cases and missing in negative ones.

We infer that A causes/contributes to/jointly causes x, or *vice versa*

Mill's Joint Method of Agreement & Difference



The joint method combines the first two: there is a potential cause present in positive cases and missing in negative ones.

We infer that A causes/contributes to/jointly causes x, or *vice versa*

In symbols, we infer:

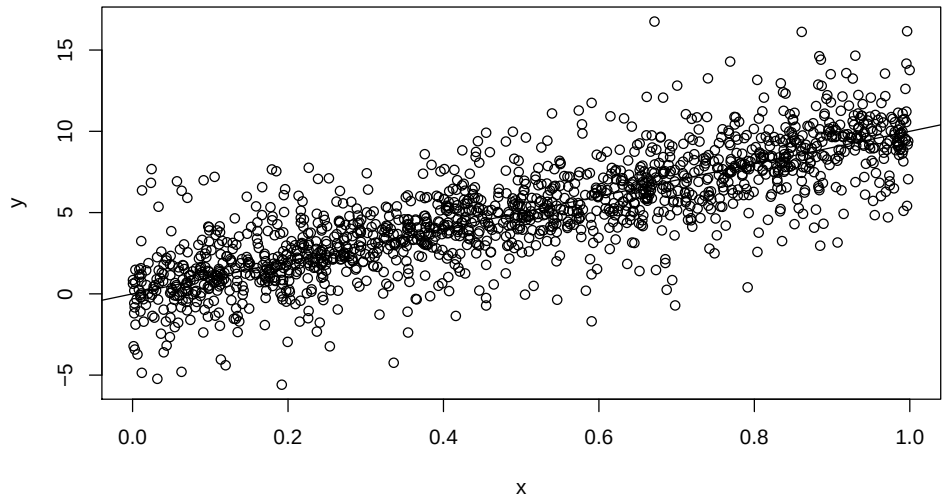
$$A \Leftarrow x \qquad or \qquad A \Rightarrow x \qquad or \qquad A \Leftrightarrow x$$

Example

Mill's Joint Method of Agreement & Difference: Example

	Potential causes				Potential effects
Subjects	Spinach	Curry	Shrimp	Tuna	Sick?
Me	•	•	•	•	•
E.	•	•		•	•
S.	•	•	•		•
A.	•			•	

If we apply Mill's Method of Difference, what do we conclude?
XXXSpinach & Curry are common to all cases of illness
If we *then* apply Mill's Method of Agreement, what do we conclude?
Only Curry is a likely cause/contributor/effect of being Sick



Mill's Method of Concommitant Variation says that if X and Y are correlated, they may be causally related

We know this one, and know how to make it *probabilistic*

Mill's Method of Residues

If we already know that $A \Rightarrow X$ *and* $B \Rightarrow Y$*, then* $C \Rightarrow Z$

Example: A doctor sees a patient who has a rash, a headache, and a stomach ache.

The previous day, that patient had oysters (which the doctor "knows" cause rashes), wine (which the doctor "knows" cause headaches) and fish.

The doctor deduces the fish caused the stomach ache.

Silly, and deeply flawed

Mill's Methods

At their best (ie, discarding "method of residue" and "direct method of agreement") Mill's Methods have some serious flaws:

1. Work poorly if causal processes are noisy
2. Work poorly if covariates are correlated
3. Tend towards monocausal explanations

However, not a terrible place to start, especially if you are using your qualitative study to build a theory, and plan to quantitatively test later (mixed methods)

Aside on Necessary & Sufficient Conditions

Notation: & (logical "and") | (logical "or")
Consider the following *deterministic* relationships:

1. $(A\&B) \mid (A\&C) \Rightarrow X$
   In words, A is a necessary but not sufficient condition for X

Aside on Necessary & Sufficient Conditions

Notation: & (logical "and") | (logical "or")
Consider the following *deterministic* relationships:

1. $(A\&B) \mid (A\&C) \Rightarrow X$
   In words, A is a necessary but not sufficient condition for X

   E.g., suppose X is "the light is on", A is "light switch is on", B is "power is on" and C is "personal generator is on"
   Note that it is easy to add theoretically trivial necessary conditions: e.g., "circuit breaker not tripped", "light bulb working", "wiring not shorted", etc.

Aside on Necessary & Sufficient Conditions

Notation: & (logical "and") | (logical "or")
Consider the following *deterministic* relationships:

1. $(A\&B) \mid (A\&C) \Rightarrow X$
   In words, A is a necessary but not sufficient condition for X

   E.g., suppose X is "the light is on", A is "light switch is on", B is "power is on" and C is "personal generator is on"
   Note that it is easy to add theoretically trivial necessary conditions: e.g., "circuit breaker not tripped", "light bulb working", "wiring not shorted", etc.
2. $A \Rightarrow X \quad \& \quad B \Rightarrow X$
   In words, A and B are both sufficient conditions for X

Aside on Necessary & Sufficient Conditions

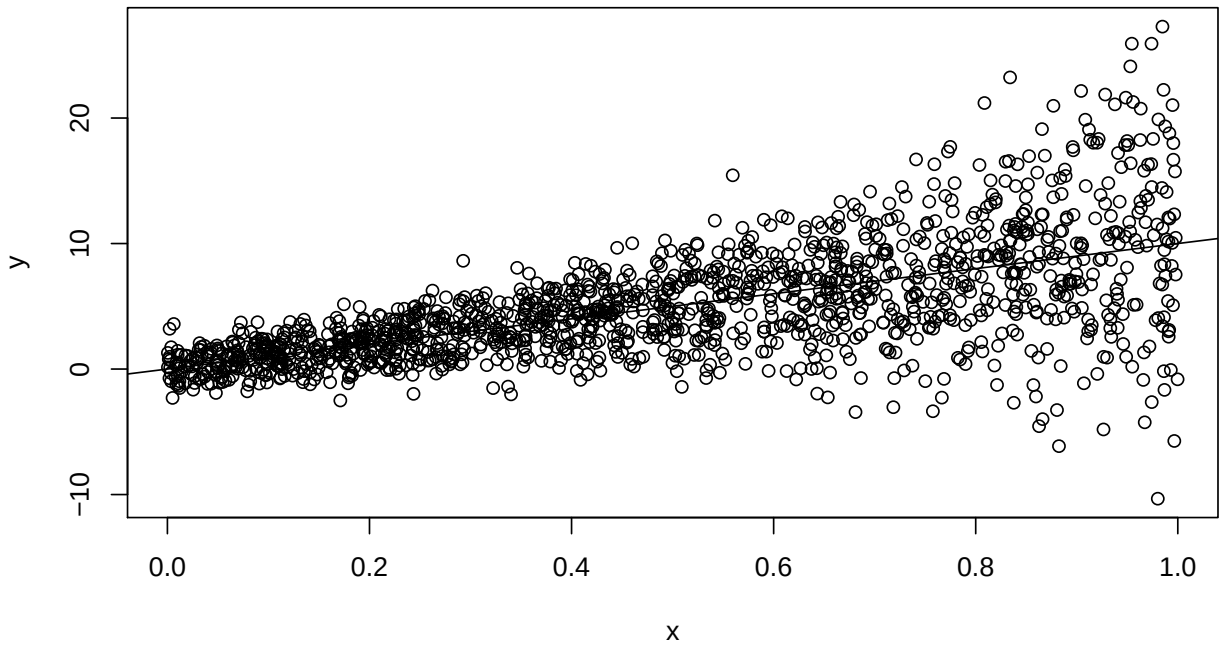
Notation: & (logical "and") | (logical "or")
Consider the following *deterministic* relationships:

1. $(A\&B) \mid (A\&C) \Rightarrow X$
   In words, A is a necessary but not sufficient condition for X

   E.g., suppose X is "the light is on", A is "light switch is on", B is "power is on" and C is "personal generator is on"
   Note that it is easy to add theoretically trivial necessary conditions: e.g., "circuit breaker not tripped", "light bulb working", "wiring not shorted", etc.
2. $A \Rightarrow X \quad \& \quad B \Rightarrow X$
   In words, A and B are both sufficient conditions for X

   E.g., X is "person voted" and A is "benefit of voting exceeded cost" and B is "social pressure to vote"

   Easy to add trivial necessary conditions here too; but only worth adding if substantively interesting



What a *probabilistic* necessary but not sufficient conditions looks like
Very common pattern in social data
Potential examples: War & power balance; Large welfare states & industrialization; Close roll call votes and controversy

Selection bias

If we select observations non-randomly from the world, our inferences may be *biased*

- Average children per marriage is 2.5. How many were in your family growing up? Are these numbers different? Who is "left out" in the second sample?

Selection bias

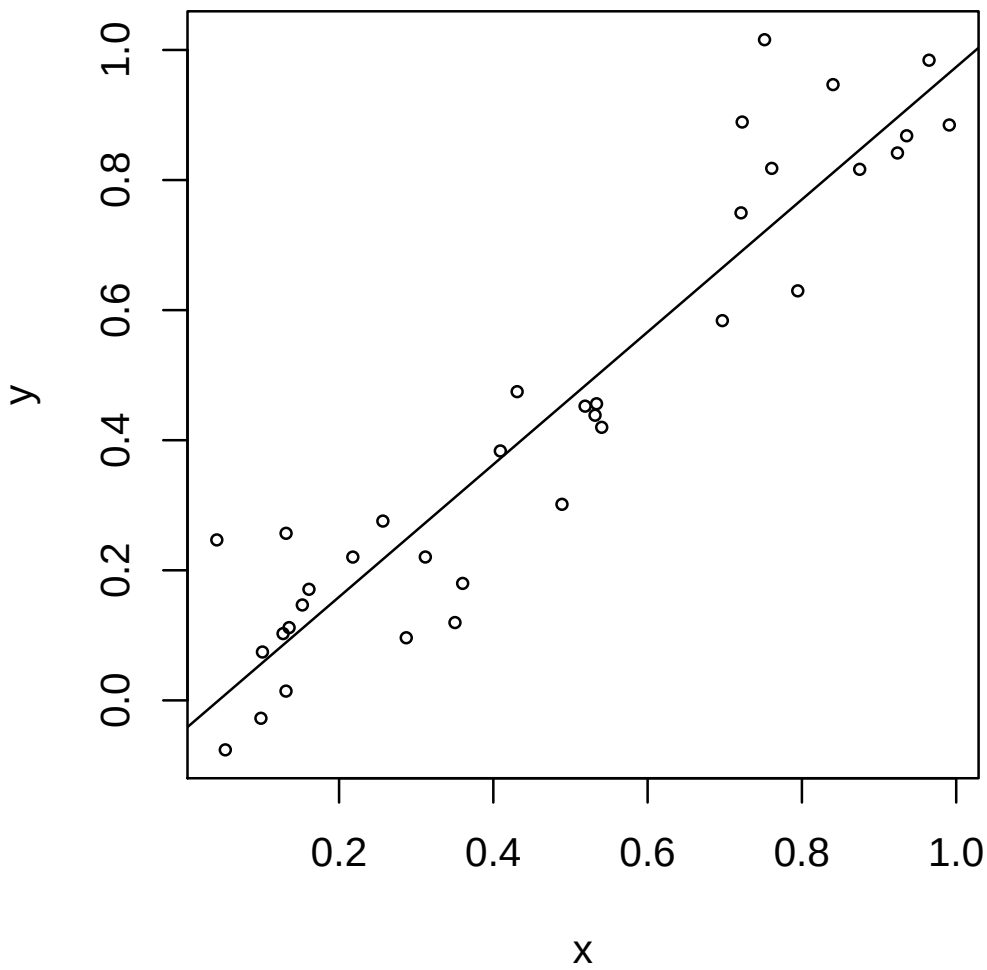
If we select observations non-randomly from the world, our inferences may be *biased*

- Average children per marriage is 2.5. How many were in your family growing up? Are these numbers different? Who is "left out" in the second sample?
- In testimony to HI state senate, motorcyclists testified that in their (multiple) crashes, helmets would not have prevented injuries. Who didn't testify?

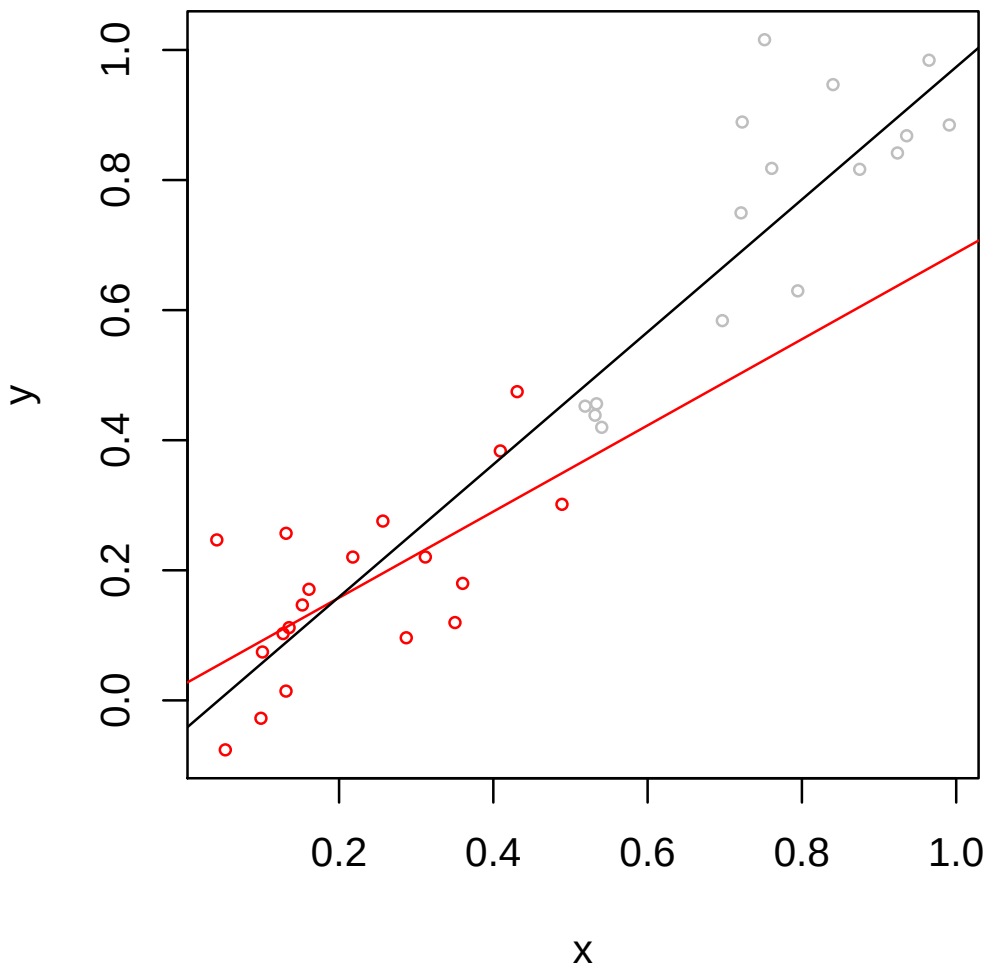
Selection bias

If we select observations non-randomly from the world, our inferences may be *biased*

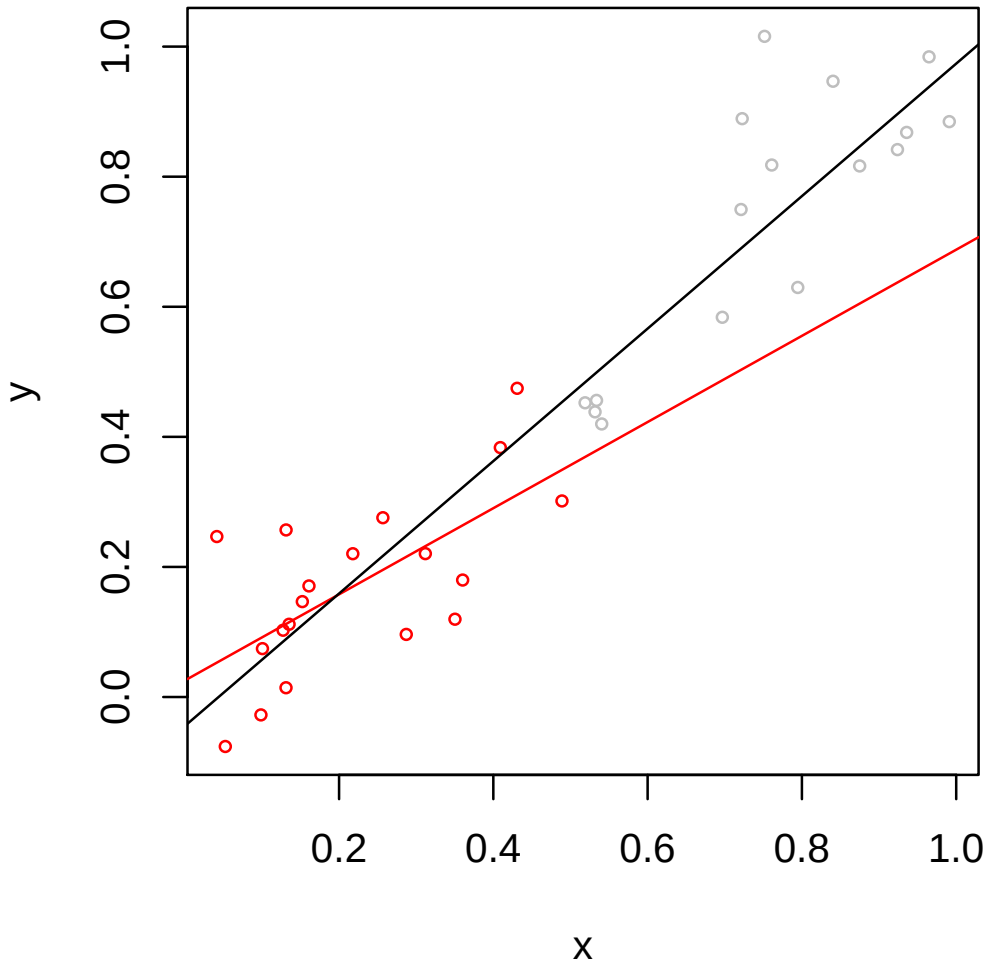
- Average children per marriage is 2.5. How many were in your family growing up? Are these numbers different? Who is "left out" in the second sample?
- In testimony to HI state senate, motorcyclists testified that in their (multiple) crashes, helmets would not have prevented injuries. Who didn't testify?
- In a study of the causes of revolution, a scholar studies only four cases: the French Revolution, the Chinese Revolution, the Mexican Revolution, & the Russian Revolution. What is missing?



Suppose we conducted a survey & asked people their income (x) and conservatism (y)
With the full range of respondents, we find a strong relationship



But suppose high income (or highly conservative) people decline to answer
Then we analyze the red dots only.
And get a result biased towards no effect of income on ideology



Try to maximize variance of covariates, and avoid selecting on response variables

Most selection is unintentional, so think hard about sources of selection bias

Critical case studies

Counter-view: Eckstein (1975) says can test with one case!

Plausibility probes Test the theory in *only* the "easiest" case

Confirmation cases Test the theory in *only* the "hardest" case

Critical case studies

Counter-view: Eckstein (1975) says can test with one case!

Plausibility probes Test the theory in *only* the "easiest" case

Confirmation cases Test the theory in *only* the "hardest" case

Problems with "critical case studies":

1. Any randomness or measurment error risks huge bias
2. Omitted variable bias potentially enormous
3. Lots of space between minimally plausibility and 100% confirmation. Evidence for scientific theories probably mostly inside the range

Making the most of your cases

Critical case studies can be illuminating, but risky

We get more out of small-*N* studies, but how to choose our cases?

Huge debate. If look to King, Keohane & Verba:

Maximize variance on covariates, and avoid selection on dependent variable

1. Identify key covariates and controls

Making the most of your cases

Critical case studies can be illuminating, but risky

We get more out of small-*N* studies, but how to choose our cases?

Huge debate. If look to King, Keohane & Verba:

Maximize variance on covariates, and avoid selection on dependent variable

1. Identify key covariates and controls
2. Select cases that cover at least the extremes of these variables

Making the most of your cases

Critical case studies can be illuminating, but risky

We get more out of small-*N* studies, but how to choose our cases?

Huge debate. If look to King, Keohane & Verba:

Maximize variance on covariates, and avoid selection on dependent variable

1. Identify key covariates and controls
2. Select cases that cover at least the extremes of these variables
3. Ideally, have at least one case for each combination of high and low values on important covariates

Making the most of your cases

Critical case studies can be illuminating, but risky

We get more out of small-*N* studies, but how to choose our cases?

Huge debate. If look to King, Keohane & Verba:

Maximize variance on covariates, and avoid selection on dependent variable

1. Identify key covariates and controls
2. Select cases that cover at least the extremes of these variables
3. Ideally, have at least one case for each combination of high and low values on important covariates
4. *Then* code dependent variable, and see which covariates predict

   Vital not to look at DV in advance on this strategy to avoid selection bias

Making the most of your cases

Suppose you are studying countries—lots of effort for each observation

Need a result for all that work:
But what if random variation or measurement error mask your findings?

KKV suggest expanding the number of observations for each country:

1. Find additional testable implications of theory to create add'l dependent variables on the same cases

Making the most of your cases

Suppose you are studying countries—lots of effort for each observation

Need a result for all that work:
But what if random variation or measurement error mask your findings?

KKV suggest expanding the number of observations for each country:

1. Find additional testable implications of theory to create add'l dependent variables on the same cases
2. Perhaps your dependent variable varies over time; could expand to time series

Making the most of your cases

Suppose you are studying countries—lots of effort for each observation

Need a result for all that work:
But what if random variation or measurement error mask your findings?

KKV suggest expanding the number of observations for each country:

1. Find additional testable implications of theory to create add'l dependent variables on the same cases
2. Perhaps your dependent variable varies over time; could expand to time series
3. Perhaps your dependent variable is reproduced in subsidiary or parallel cases

Example: Want to know the effect of term limits on legislative productivity in US states. Can afford to code 8 states. What should we do?

Making the most of your cases

Suppose you are studying countries—lots of effort for each observation

Need a result for all that work:
But what if random variation or measurement error mask your findings?

KKV suggest expanding the number of observations for each country:

1. Find additional testable implications of theory to create add'l dependent variables on the same cases
2. Perhaps your dependent variable varies over time; could expand to time series
3. Perhaps your dependent variable is reproduced in subsidiary or parallel cases

Example: Want to know the effect of term limits on legislative productivity in US states. Can afford to code 8 states. What should we do?

Could code data for several states with/without term limits, and varied on key factors like partisanship, professionalism, length of term.

Making the most of your cases

Suppose you are studying countries—lots of effort for each observation

Need a result for all that work:
But what if random variation or measurement error mask your findings?

KKV suggest expanding the number of observations for each country:

1. Find additional testable implications of theory to create add'l dependent variables on the same cases
2. Perhaps your dependent variable varies over time; could expand to time series
3. Perhaps your dependent variable is reproduced in subsidiary or parallel cases

Example: Want to know the effect of term limits on legislative productivity in US states. Can afford to code 8 states. What should we do?

Could code data for several states with/without term limits, and varied on key factors like partisanship, professionalism, length of term.

Then look at:
a.) both chambers, b.) multiple years, c.) committees and subcommittees?

Making the most of your cases

What if my dependent variable is rare, like war or revolution?

Then applying the above strategy may turn up *zero* positive cases!

Obviously we have to "select on the dependent variable" in some sense here, so what do we do? Borrow from medical studies of rare disease!
Proposed qualitative strategy (KKV & King Zeng):

Study every extent positive case, then randomly sample the same number from the universe of negative cases ("shadow cases")

Selecting the universe is hard, and controversial. Skocpol tried something similar (but non-random); looked at "near-misses" of revolution.

Hold constant the obvious necessary conditions: sufficient economic development, global availability of revolutionary paradigm.
Then askwhat tipped France into revolution but not Britain?

When should you use Qualitative Methods?

One view: strength of small-*N* qualitative studies is developing theory
Things qualitative methods may do better than quantitative:

1. "process trace" (ie, study history!)
2. sift through hundreds of possible variables
3. "thick" description of cases, variables, process (versus "thin" in quantitative)
4. investigate plausibility of conjunctural/contingent causation
   contra common misperception, quant can *test* conjunctural/contingent causes as well as simple ones

When should you use Qualitative Methods?

One view: strength of small-*N* qualitative studies is developing theory
Things qualitative methods may do better than quantitative:

1. "process trace" (ie, study history!)
2. sift through hundreds of possible variables
3. "thick" description of cases, variables, process (versus "thin" in quantitative)
4. investigate plausibility of conjunctural/contingent causation
   contra common misperception, quant can *test* conjunctural/contingent causes as well as simple ones

Weakness of qualitative:

1. Hard to do large *N*, so hard to cope with probablistic effects (& everything's probabilistic)
2. Hard to measure uncertainty of results
3. Easier to measure lots of variables, but harder to adjudicate among them

When should you use Quantitative Methods?

One view: strength of large-*N* qualitative studies is testing theory
Things quantitative methods may do better than qualitative:

1. Uncover probabilistic relationships
2. Minimize subjectivity & circularity in coding
3. Test in broad range of cases: out of sample tests
   Avoids danger of only analyzing cases you know intimately

When should you use Quantitative Methods?

One view: strength of large-*N* qualitative studies is testing theory
Things quantitative methods may do better than qualitative:

1. Uncover probabilistic relationships
2. Minimize subjectivity & circularity in coding
3. Test in broad range of cases: out of sample tests
   Avoids danger of only analyzing cases you know intimately

But why not combine qualitative & quantitative?
Mixed methods give best of both worlds!

1. Develop a theory based on detailed understanding of cases: qualitative
2. Check if that theory works outside original case, with controls: quantitative
3. Return to critical and interesting cases to recheck intuition of final quantative theory: qualitative
4. Repeat!