

Experimental Political Science

EPS Lecture 6

Rebecca B. Morton

NYU

EPS Lectures

Validity and Experimental Manipulations

- Suppose that we have conducted some empirical research with either experimental or nonexperimental data.

Validity and Experimental Manipulations

- Suppose that we have conducted some empirical research with either experimental or nonexperimental data.
- We ideally want a research design that will provide us with *valid* results, results that are true for the population we are analyzing, and *robust* results, results that *generalize* beyond our target population.

Validity and Experimental Manipulations

- Suppose that we have conducted some empirical research with either experimental or nonexperimental data.
- We ideally want a research design that will provide us with *valid* results, results that are true for the population we are analyzing, and *robust* results, results that *generalize* beyond our target population.
- Lot of confusion/debate over these issues ...

What is Validity?

Definition (Validity)

The approximate truth of the inference or knowledge claim.

What is Validity?

Definition (Validity)

The approximate truth of the inference or knowledge claim.

Definition (Internal Validity)

The approximate truth of the inference or knowledge claim within a target population studied.

What is Validity?

Definition (Validity)

The approximate truth of the inference or knowledge claim.

Definition (Internal Validity)

The approximate truth of the inference or knowledge claim within a target population studied.

Definition (External Validity)

The approximate truth of the inference or knowledge claim for observations beyond the target population studied.

Deconstructing Internal Validity

Statistical Validity

Definition (Statistical Validity)

Whether there is a statistically significant covariance between the variables the researcher is interested in and whether the relationship is sizeable.

Deconstructing Internal Validity

Statistical Validity

Definition (Statistical Validity)

Whether there is a statistically significant covariance between the variables the researcher is interested in and whether the relationship is sizeable.

- Can be more important than traditionally thought, need to use statistical replication ...

Deconstructing Internal Validity

Statistical Validity

Definition (Statistical Replication)

When a researcher uses a different sample from the same population to evaluate the same theoretical implications as in the previous study or uses the same sample from the same population but comparing statistical techniques to evaluate the same theoretical implications as in the previous study. **Relevant to internal validity**

Definition (Scientific Replication)

When a researcher uses a different sample, a different population to evaluate the same theoretical constructs with the same theoretical implications or uses the same or a different sample from either the same or a different population to evaluate different theoretical implications from these constructs. **Relevant to external validity**

Deconstructing Internal Validity

Causal Validity and the Identification Problem

Definition (Causal Validity)

Whether the relationships the researcher finds within the target population analyzed are causal.

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- A closely related concept to causal validity is the notion of *identification* in econometrics.

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- A closely related concept to causal validity is the notion of *identification* in econometrics.
- As Manski (1995, 2003) explains, econometricians have found it useful to separate out the concerns of identifying relationships from those of the concerns in estimating these relationships.

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- A closely related concept to causal validity is the notion of *identification* in econometrics.
- As Manski (1995, 2003) explains, econometricians have found it useful to separate out the concerns of identifying relationships from those of the concerns in estimating these relationships.
- Estimation problems, have to do with statistical issues of whether, given the dataset analyzed and the assumptions made about the relationship between the dataset and the population, the parameters of interest are efficiently and consistently estimated, or statistical validity.

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- A closely related concept to causal validity is the notion of *identification* in econometrics.
- As Manski (1995, 2003) explains, econometricians have found it useful to separate out the concerns of identifying relationships from those of the concerns in estimating these relationships.
- Estimation problems, have to do with statistical issues of whether, given the dataset analyzed and the assumptions made about the relationship between the dataset and the population, the parameters of interest are efficiently and consistently estimated, or statistical validity.
- Manski remarks (2003, page 12): “Statistical inference seeks to characterize how sampling variability affects the conclusions that can be drawn from samples of limited size.”

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- An identification problem exists when it is problematic to establish causal inferences even if the researcher has an unlimited sample from the population. Identification problems exist in many contexts.

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- An identification problem exists when it is problematic to establish causal inferences even if the researcher has an unlimited sample from the population. Identification problems exist in many contexts.
- For example, distinguishing between cultural effects and preference effects?

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- An identification problem exists when it is problematic to establish causal inferences even if the researcher has an unlimited sample from the population. Identification problems exist in many contexts.
- For example, distinguishing between cultural effects and preference effects?
- Difficulty in measuring altruism versus fairness in some contexts ...

Deconstructing Internal Validity

Causal Validity and the Identification Problem

- An identification problem exists when it is problematic to establish causal inferences even if the researcher has an unlimited sample from the population. Identification problems exist in many contexts.
- For example, distinguishing between cultural effects and preference effects?
- Difficulty in measuring altruism versus fairness in some contexts ...
- Inability to observe people in multiple states of the world simultaneously.

Deconstructing Internal Validity

Construct Validity

Definition (Construct Validity)

Whether the inferences from the data are valid for the theory (or constructs) the researcher is evaluating in a theory testing experiment.

Deconstructing Internal Validity

Construct Validity and the Generalizability of Results

- Construct validity is also about generalization.

“ ... valid knowledge of constructs that are involved in a study can shed light on external validity questions, especially if a well-developed theory exists that describes how various constructs and instances are related to each other.

Deconstructing Internal Validity

Construct Validity and the Generalizability of Results

- Construct validity is also about generalization.
- The generalization is to a theoretical construct that ideally the researcher does not view as limited to the particular empirical analysis, but a more general theory.

“ ... valid knowledge of constructs that are involved in a study can shed light on external validity questions, especially if a well-developed theory exists that describes how various constructs and instances are related to each other.

Deconstructing Internal Validity

Construct Validity and the Generalizability of Results

- Construct validity is also about generalization.
- The generalization is to a theoretical construct that ideally the researcher does not view as limited to the particular empirical analysis, but a more general theory.
- Because of this, being able to establish construct validity can actually help answer external validity questions about the theory and any analysis of the theory.

“ ... valid knowledge of constructs that are involved in a study can shed light on external validity questions, especially if a well-developed theory exists that describes how various constructs and instances are related to each other.

Deconstructing Internal Validity

Construct Validity and the Generalizability of Results

- Construct validity is also about generalization.
- The generalization is to a theoretical construct that ideally the researcher does not view as limited to the particular empirical analysis, but a more general theory.
- Because of this, being able to establish construct validity can actually help answer external validity questions about the theory and any analysis of the theory.
- **As SSC argue (page 93):**

“ ... valid knowledge of constructs that are involved in a study can shed light on external validity questions, especially if a well-developed theory exists that describes how various constructs and instances are related to each other.

Deconstructing Internal Validity

Construct Validity and the Generalizability of Results

Medicine, for example, has well-developed theories for categorizing certain therapies (say, the class of drugs we call chemotherapies for cancer) and for knowing how these therapies affect patients (how they affect blood tests and survival and what their side effects are). Consequently, when a new drug meets the criteria for being called a chemotherapy, we can predict much of its likely performance before actually testing it (e.g., we can say it is likely to cause hair loss and nausea and to increase survival in patients with low tumor burdens but not advanced cases). This knowledge makes the design of new experiments easier by narrowing the scope of pertinent patients and outcomes, and it makes extrapolations about treatment effects likely to be more accurate."

Deconstructing External Validity

External, Statistical, and Ecological Validity

Definition (Ecological Validity)

Whether the methods, materials, and settings of the research is similar to a given target environment.

- External validity often confused with ecological validity

Deconstructing External Validity

External, Statistical, and Ecological Validity

Definition (Ecological Validity)

Whether the methods, materials, and settings of the research is similar to a given target environment.

- External validity often confused with ecological validity
- *Increasing ecological validity for one target population does not necessarily mean that the results generalize to another population and setting.*

Deconstructing External Validity

External, Statistical, and Ecological Validity

Definition (Ecological Validity)

Whether the methods, materials, and settings of the research is similar to a given target environment.

- External validity often confused with ecological validity
- *Increasing ecological validity for one target population does not necessarily mean that the results generalize to another population and setting.*
- External validity can only be established by generalizing beyond the target population and any target environment or setting.

Deconstructing External Validity

External, Statistical, and Ecological Validity

Definition (Ecological Validity)

Whether the methods, materials, and settings of the research is similar to a given target environment.

- External validity often confused with ecological validity
- *Increasing ecological validity for one target population does not necessarily mean that the results generalize to another population and setting.*
- External validity can only be established by generalizing beyond the target population and any target environment or setting.
- **Also, political scientists often say an experiment isn't externally valid because the sample drawn is not random from a target population the researcher thinks is relevant.**

Deconstructing External Validity

External, Statistical, and Ecological Validity

Definition (Ecological Validity)

Whether the methods, materials, and settings of the research is similar to a given target environment.

- External validity often confused with ecological validity
- *Increasing ecological validity for one target population does not necessarily mean that the results generalize to another population and setting.*
- External validity can only be established by generalizing beyond the target population and any target environment or setting.
- Also, political scientists often say an experiment isn't externally valid because the sample drawn is not random from a target population the researcher thinks is relevant.
- **If the solution is drawing a random sample, then this is a problem of statistical validity which is part of internal validity.**

Deconstructing External Validity

Establishing External Validity

- Without further empirical study a researcher can only conjecture or hypothesize that his or her result has external validity based on similar studies or assumptions about the relationship between the population initially analyzed and the new population to be considered.

Deconstructing External Validity

Establishing External Validity

- Without further empirical study a researcher can only conjecture or hypothesize that his or her result has external validity based on similar studies or assumptions about the relationship between the population initially analyzed and the new population to be considered.
- The proof of external validity is always empirical.

Deconstructing External Validity

Establishing External Validity

- Without further empirical study a researcher can only conjecture or hypothesize that his or her result has external validity based on similar studies or assumptions about the relationship between the population initially analyzed and the new population to be considered.
- The proof of external validity is always empirical.
- Debates about external validity in the absence of such empirical proof are debates about the similarity of a study to previous studies or population similarities, but there can never be a resolution through debate or discussion alone.

Deconstructing External Validity

Scientific Replication

- Scientific replication is all about establishing external validity.

Deconstructing External Validity

Scientific Replication

- Scientific replication is all about establishing external validity.
- Less easily mandated by journals than statistical replication since it involves taking the same theoretical constructs and applying them to new populations or evaluating new theoretical implications or taking causal inferences based on fact searching & determining if they can be identified and estimated in a different dataset.

Deconstructing External Validity

Scientific Replication

- Scientific replication is all about establishing external validity.
- Less easily mandated by journals than statistical replication since it involves taking the same theoretical constructs and applying them to new populations or evaluating new theoretical implications or taking causal inferences based on fact searching & determining if they can be identified and estimated in a different dataset.
- Often a researcher has taken considerable effort to find, build, or create, as in an experiment, the dataset for a study of a target population.

Deconstructing External Validity

Scientific Replication

- Scientific replication is all about establishing external validity.
- Less easily mandated by journals than statistical replication since it involves taking the same theoretical constructs and applying them to new populations or evaluating new theoretical implications or taking causal inferences based on fact searching & determining if they can be identified and estimated in a different dataset.
- Often a researcher has taken considerable effort to find, build, or create, as in an experiment, the dataset for a study of a target population.
- Usually a researcher has sought all the data that he or she could find that was relevant and leaves establishing external validity through scientific replication to other researchers.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One way to establish some external validity for one's own empirical results is through the use of *nonrandom holdout samples* as advocated by Keane and Wolpin (2007) and Wolpin (2007).

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One way to establish some external validity for one's own empirical results is through the use of *nonrandom holdout samples* as advocated by Keane and Wolpin (2007) and Wolpin (2007).
- A nonrandom holdout sample is one that differs significantly from the sample used for the estimation along a dimension over which the causal inference or theoretical prediction is expected to hold.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One way to establish some external validity for one's own empirical results is through the use of *nonrandom holdout samples* as advocated by Keane and Wolpin (2007) and Wolpin (2007).
- A nonrandom holdout sample is one that differs significantly from the sample used for the estimation along a dimension over which the causal inference or theoretical prediction is expected to hold.
- If the empirical results from the original estimation are supported with the nonrandom holdout sample which involves observations that are well outside the support of the original data, then the results will have more external validity along this dimension.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One way to establish some external validity for one's own empirical results is through the use of *nonrandom holdout samples* as advocated by Keane and Wolpin (2007) and Wolpin (2007).
- A nonrandom holdout sample is one that differs significantly from the sample used for the estimation along a dimension over which the causal inference or theoretical prediction is expected to hold.
- If the empirical results from the original estimation are supported with the nonrandom holdout sample which involves observations that are well outside the support of the original data, then the results will have more external validity along this dimension.
- As Keane and Wolpin remark, this procedure is often used in time series analyses and has been used in the psychology and marketing literature.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- Used by McFadden (1977).

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- Used by McFadden (1977).
- McFadden estimated a random utility model of travel demand in the San Francisco Bay area before the introduction of the subway system & then compared his estimates to the actual usage after the subway was introduced.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- Used by McFadden (1977).
- McFadden estimated a random utility model of travel demand in the San Francisco Bay area before the introduction of the subway system & then compared his estimates to the actual usage after the subway was introduced.
- The observations after the subway was introduced was the nonrandom holdout sample.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- Used by McFadden (1977).
- McFadden estimated a random utility model of travel demand in the San Francisco Bay area before the introduction of the subway system & then compared his estimates to the actual usage after the subway was introduced.
- The observations after the subway was introduced was the nonrandom holdout sample.
- Keane and Wolpin point out that experiments can provide an ideal opportunity for analyses with nonrandom holdout samples.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One can imagine that treatments can be used as subsets of the population just as in the cross-valuation procedure above.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One can imagine that treatments can be used as subsets of the population just as in the cross-valuation procedure above.
- Suppose a researcher conducts K treatments on different dimensions.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One can imagine that treatments can be used as subsets of the population just as in the cross-valuation procedure above.
- Suppose a researcher conducts K treatments on different dimensions.
- Then the researcher can estimate the effects of the treatments on each of the possible groups of $K - 1$ subsets as separate target populations and then assess the predictive accuracy on the subset omitted on the dimension omitted.

Deconstructing External Validity

Scientific Replication and Nonrandom Holdout Samples

- One can imagine that treatments can be used as subsets of the population just as in the cross-valuation procedure above.
- Suppose a researcher conducts K treatments on different dimensions.
- Then the researcher can estimate the effects of the treatments on each of the possible groups of $K - 1$ subsets as separate target populations and then assess the predictive accuracy on the subset omitted on the dimension omitted.
- In this fashion the researcher can gain some traction on the external validity of his or her results.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- Outside of the nonrandom holdout sample idea, statistical replication that involves drawing a new sample from the same population requires that a new experiment be conducted using subjects from the same target population with the same experimental protocols, etc.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- Outside of the nonrandom holdout sample idea, statistical replication that involves drawing a new sample from the same population requires that a new experiment be conducted using subjects from the same target population with the same experimental protocols, etc.
- Oftentimes experimentalists do this as part of their research, conducting several independent sessions of an experiment using different samples of subjects from the same pool.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- Outside of the nonrandom holdout sample idea, statistical replication that involves drawing a new sample from the same population requires that a new experiment be conducted using subjects from the same target population with the same experimental protocols, etc.
- Oftentimes experimentalists do this as part of their research, conducting several independent sessions of an experiment using different samples of subjects from the same pool.
- Scientific replication through experimentation can occur when subjects from a different target population are used with the same experimental protocols to evaluate the same theoretical implications or subjects from the same or different target population are used to evaluate different theoretical implications with sometimes a change in experimental protocols (maintaining the same theoretical constructs).

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- For example, Potters and Van Winden (2000) replicated an experiment they had conducted previously with undergraduate students, Potters and Van Winden (1996), with lobbyists.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- For example, Potters and Van Winden (2000) replicated an experiment they had conducted previously with undergraduate students, Potters and Van Winden (1996), with lobbyists.
- One advantage of laboratory experiments is that usually statistical verification with different samples from the same target population can be reasonably conducted as long as researchers make publicly available detailed experimental protocols.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- For example, Potters and Van Winden (2000) replicated an experiment they had conducted previously with undergraduate students, Potters and Van Winden (1996), with lobbyists.
- One advantage of laboratory experiments is that usually statistical verification with different samples from the same target population can be reasonably conducted as long as researchers make publicly available detailed experimental protocols.
- Such explicit publicly available protocols are also required for effective scientific replications as well, particularly if the experimenter seeks to replicate with a sample from a new target population using the same experimental design.

Deconstructing External Validity

Scientific Replication Beyond Nonrandom Holdout Samples

- For example, Potters and Van Winden (2000) replicated an experiment they had conducted previously with undergraduate students, Potters and Van Winden (1996), with lobbyists.
- One advantage of laboratory experiments is that usually statistical verification with different samples from the same target population can be reasonably conducted as long as researchers make publicly available detailed experimental protocols.
- Such explicit publicly available protocols are also required for effective scientific replications as well, particularly if the experimenter seeks to replicate with a sample from a new target population using the same experimental design.
- It should be the norm of experimentalists in political science to provide access to these protocols for such replication.

Deconstructing External Validity

Stress Tests and External Validity

- A stress test is also a way in which an experimentalist can explore issues of external validity when evaluating a formal model.

Deconstructing External Validity

Stress Tests and External Validity

- A stress test is also a way in which an experimentalist can explore issues of external validity when evaluating a formal model.
- For example, suppose a researcher has tested a theory of legislative bargaining in the laboratory. The model is one of complete information.

Deconstructing External Validity

Stress Tests and External Validity

- A stress test is also a way in which an experimentalist can explore issues of external validity when evaluating a formal model.
- For example, suppose a researcher has tested a theory of legislative bargaining in the laboratory. The model is one of complete information.
- However, the researcher relaxes some of the information available to the subjects to determine if the behavior of the subjects will be affected. The researcher has no theoretical prediction about what will happen.

Deconstructing External Validity

Stress Tests and External Validity

- A stress test is also a way in which an experimentalist can explore issues of external validity when evaluating a formal model.
- For example, suppose a researcher has tested a theory of legislative bargaining in the laboratory. The model is one of complete information.
- However, the researcher relaxes some of the information available to the subjects to determine if the behavior of the subjects will be affected. The researcher has no theoretical prediction about what will happen.
- If the theory's predictions hold despite this new wrinkle, then the researcher has learned that the results of the first experiment can generalize, under some circumstances, to a less than complete information environment.

Deconstructing External Validity

Stress Tests and External Validity

- A stress test is also a way in which an experimentalist can explore issues of external validity when evaluating a formal model.
- For example, suppose a researcher has tested a theory of legislative bargaining in the laboratory. The model is one of complete information.
- However, the researcher relaxes some of the information available to the subjects to determine if the behavior of the subjects will be affected. The researcher has no theoretical prediction about what will happen.
- If the theory's predictions hold despite this new wrinkle, then the researcher has learned that the results of the first experiment can generalize, under some circumstances, to a less than complete information environment.
- **Should the researcher resolve the theory?**

Deconstructing External Validity

Stress Tests and External Validity

- Another example would be to conduct the same complete information legislative bargaining theory with different subject pools by conducting lab in the field versions of the experiment to determine how robust the results are to changes in who participates in the experiment.

Deconstructing External Validity

Stress Tests and External Validity

- Another example would be to conduct the same complete information legislative bargaining theory with different subject pools by conducting lab in the field versions of the experiment to determine how robust the results are to changes in who participates in the experiment.
- Again, if the theory's predictions hold, we say that the results are robust to this change, and vice-versa.

Deconstructing External Validity

Stress Tests and External Validity

- Another example would be to conduct the same complete information legislative bargaining theory with different subject pools by conducting lab in the field versions of the experiment to determine how robust the results are to changes in who participates in the experiment.
- Again, if the theory's predictions hold, we say that the results are robust to this change, and vice-versa.
- Or the experimentalist may vary the frame of the experiment – perhaps the original experiment used a neutral frame where subjects were told they were players in a game without any political context.

Deconstructing External Validity

Stress Tests and External Validity

- Another example would be to conduct the same complete information legislative bargaining theory with different subject pools by conducting lab in the field versions of the experiment to determine how robust the results are to changes in who participates in the experiment.
- Again, if the theory's predictions hold, we say that the results are robust to this change, and vice-versa.
- Or the experimentalist may vary the frame of the experiment – perhaps the original experiment used a neutral frame where subjects were told they were players in a game without any political context.
- The experimentalist could introduce a political context to the experiment by telling the subjects they are legislators and they are bargaining for ministerial positions and see if this frame difference affects the subjects' choices.

Deconstructing External Validity

Stress Tests and External Validity

- The beauty of stress tests is that the experimentalist can incorporate new features of the experimental environment on a piecemeal basis and investigate each aspect of the change in an effort to test the limits of the external robustness or validity of the results.

Deconstructing External Validity

Stress Tests and External Validity

- The beauty of stress tests is that the experimentalist can incorporate new features of the experimental environment on a piecemeal basis and investigate each aspect of the change in an effort to test the limits of the external robustness or validity of the results.
- Stress tests, then, are important tools for experimentalists to test whether their results are externally valid or robust and where in particular the robustness or validity might break down.

Deconstructing External Validity

Theory, Replication, and Validity

- Suppose a researcher finds a situation where empirical research is considered to have construct validity & theory's behavioral predictions are not supported.

Deconstructing External Validity

Theory, Replication, and Validity

- Suppose a researcher finds a situation where empirical research is considered to have construct validity & theory's behavioral predictions are not supported.
- Should we always change the theory once we find a single negative result?

Deconstructing External Validity

Theory, Replication, and Validity

- Suppose a researcher finds a situation where empirical research is considered to have construct validity & theory's behavioral predictions are not supported.
- Should we always change the theory once we find a single negative result?
- It is unlikely that a single negative result would be seen as decisive in determining the merits of the theory. Why?

Deconstructing External Validity

Theory, Replication, and Validity

- Suppose a researcher finds a situation where empirical research is considered to have construct validity & theory's behavioral predictions are not supported.
- Should we always change the theory once we find a single negative result?
- It is unlikely that a single negative result would be seen as decisive in determining the merits of the theory. Why?
- All theories & models are abstractions from the DGP & therefore, all have parts that are empirically false & can be proven so when confronted with some observations.

Deconstructing External Validity

Theory, Replication, and Validity

- Suppose a researcher finds a situation where empirical research is considered to have construct validity & theory's behavioral predictions are not supported.
- Should we always change the theory once we find a single negative result?
- It is unlikely that a single negative result would be seen as decisive in determining the merits of the theory. Why?
- All theories & models are abstractions from the DGP & therefore, all have parts that are empirically false & can be proven so when confronted with some observations.
- The question is not whether a theory can be proven empirically false, but when do empirical inconsistencies with the theory matter enough for the theory to be modified or even discarded?

Deconstructing External Validity

Theory, Replication, and Validity

- Similarly, suppose a researcher, again conducting empirical research considered to have construct validity, finds that theory's behavioral predictions *are* supported.

Deconstructing External Validity

Theory, Replication, and Validity

- Similarly, suppose a researcher, again conducting empirical research considered to have construct validity, finds that theory's behavioral predictions *are* supported.
- Should we unconditionally accept the theory? No.

Deconstructing External Validity

Theory, Replication, and Validity

- Similarly, suppose a researcher, again conducting empirical research considered to have construct validity, finds that theory's behavioral predictions *are* supported.
- Should we unconditionally accept the theory? No.
- Theory evaluation in the social sciences is a cumulative process that occurs through replication & complementary studies.

Deconstructing External Validity

Theory, Replication, and Validity

- Similarly, suppose a researcher, again conducting empirical research considered to have construct validity, finds that theory's behavioral predictions *are* supported.
- Should we unconditionally accept the theory? No.
- Theory evaluation in the social sciences is a cumulative process that occurs through replication & complementary studies.
- However, since any theory can be disproved with enough data, the evaluation of theory is not purely an empirical question.

Deconstructing External Validity

Theory, Replication, and Validity

- Similarly, suppose a researcher, again conducting empirical research considered to have construct validity, finds that theory's behavioral predictions *are* supported.
- Should we unconditionally accept the theory? No.
- Theory evaluation in the social sciences is a cumulative process that occurs through replication & complementary studies.
- However, since any theory can be disproved with enough data, the evaluation of theory is not purely an empirical question.
- As with Fudenberg (200x), we believe that theory should be judged on Stigler's (1965) three criteria: accuracy of predictions, generality, & tractability. Scientific replication is useful and important, but if the goal is complete external validity in some sort of ideal sense, no theory will satisfy that goal.

Analyses of Multiple Studies

Narrative and Systematic Reviews

Definition (Narrative Review)

Reviews of existing literature focusing on a particular research question.

Analyses of Multiple Studies

Narrative and Systematic Reviews

Definition (Narrative Review)

Reviews of existing literature focusing on a particular research question.

Definition (Systematic Review)

A narrative review that includes either a qualitative or quantitative synthesis of the reviewed studies' results.

Analyses of Multiple Studies

Narrative and Systematic Reviews

Definition (Narrative Review)

Reviews of existing literature focusing on a particular research question.

Definition (Systematic Review)

A narrative review that includes either a qualitative or quantitative synthesis of the reviewed studies' results.

Definition (Meta-analysis)

A quantitative systematic review using statistical methods where the researcher uses study results as the unit of observation or to construct the unit of observation.

Analyses of Multiple Studies

Narrative and Systematic Reviews

Definition (Narrative Review)

Reviews of existing literature focusing on a particular research question.

Definition (Systematic Review)

A narrative review that includes either a qualitative or quantitative synthesis of the reviewed studies' results.

Definition (Meta-analysis)

A quantitative systematic review using statistical methods where the researcher uses study results as the unit of observation or to construct the unit of observation.

Definition (Pooled Analysis)

A quantitative study that pools data from multiple studies to examine a particular research question.

Issues in Meta-Analyses

Criteria for Including a Study

- In a meta-analysis a researcher first has to decide what will be the criteria for including a study.

Issues in Meta-Analyses

Criteria for Including a Study

- In a meta-analysis a researcher first has to decide what will be the criteria for including a study.
- **Setting the criteria raises a lot of questions for the researcher.**

Issues in Meta-Analyses

Criteria for Including a Study

- In a meta-analysis a researcher first has to decide what will be the criteria for including a study.
- Setting the criteria raises a lot of questions for the researcher.
- For example, suppose that the researcher is more suspect of the statistical or causal validity of some studies than others, should the researcher include all studies, but use statistics to control for these differences or simply exclude studies with less valid results?

Issues in Meta-Analyses

Criteria for Including a Study

- In a meta-analysis a researcher first has to decide what will be the criteria for including a study.
- Setting the criteria raises a lot of questions for the researcher.
- For example, suppose that the researcher is more suspect of the statistical or causal validity of some studies than others, should the researcher include all studies, but use statistics to control for these differences or simply exclude studies with less valid results?
- As Bangert-Downs (1986) discusses, in psychology there has been much debate over whether low quality studies should be included in meta-analysis; whether meta-analysis is simply “garbage in-garbage out” in such cases.

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.
- Suppose that the treatment in a study is similar to the treatments given in other studies, but distinctive, to what extent can dissimilar studies be combined in an analysis that makes theoretical sense?

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.
- Suppose that the treatment in a study is similar to the treatments given in other studies, but distinctive, to what extent can dissimilar studies be combined in an analysis that makes theoretical sense?
- One of the more seminal meta-analysis in psychology is Smith and Glass' (1977) study of the effects of psychotherapy.

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.
- Suppose that the treatment in a study is similar to the treatments given in other studies, but distinctive, to what extent can dissimilar studies be combined in an analysis that makes theoretical sense?
- One of the more seminal meta-analysis in psychology is Smith and Glass' (1977) study of the effects of psychotherapy.
- In this study the authors combined studies of a wide variety of psychotherapy from gestalt therapy to transactional analysis.

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.
- Suppose that the treatment in a study is similar to the treatments given in other studies, but distinctive, to what extent can dissimilar studies be combined in an analysis that makes theoretical sense?
- One of the more seminal meta-analysis in psychology is Smith and Glass' (1977) study of the effects of psychotherapy.
- In this study the authors combined studies of a wide variety of psychotherapy from gestalt therapy to transactional analysis.
- What does such research tell us when so many different types of psychotherapy are combined?

Issues in Meta-Analyses

Criteria for Including a Study

- Studies also vary in the types of treatments and manipulations considered.
- Suppose that the treatment in a study is similar to the treatments given in other studies, but distinctive, to what extent can dissimilar studies be combined in an analysis that makes theoretical sense?
- One of the more seminal meta-analysis in psychology is Smith and Glass' (1977) study of the effects of psychotherapy.
- In this study the authors combined studies of a wide variety of psychotherapy from gestalt therapy to transactional analysis.
- What does such research tell us when so many different types of psychotherapy are combined?
- The “apples-and-oranges” problem of meta-analysis.

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?
- Should the researcher compare significance and probabilities or sizes of effects?

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?
- Should the researcher compare significance and probabilities or sizes of effects?
- How does the researcher deal with publication biases?

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?
- Should the researcher compare significance and probabilities or sizes of effects?
- How does the researcher deal with publication biases?
- Suppose studies differ substantially in sample sizes which has implications for comparisons across studies, how can a researcher control for these differences?

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?
- Should the researcher compare significance and probabilities or sizes of effects?
- How does the researcher deal with publication biases?
- Suppose studies differ substantially in sample sizes which has implications for comparisons across studies, how can a researcher control for these differences?
- Are there statistical techniques to estimate how robust the results of the reported studies are to unreported negative results?

Issues in Meta-Analyses

Statistical Questions

- What measures from the different studies should the research compare?
- Should the researcher compare significance and probabilities or sizes of effects?
- How does the researcher deal with publication biases?
- Suppose studies differ substantially in sample sizes which has implications for comparisons across studies, how can a researcher control for these differences?
- Are there statistical techniques to estimate how robust the results of the reported studies are to unreported negative results?
- **What happens if there is statistical dependence across different output measures?**

Issues in Meta-Analyses

Statistical Questions

- Fortunately, the statistical methods used in meta-analysis are advanced enough in medicine and in psychology that researchers in political science who would like to conduct a meta-analysis can find a large literature on the methods that have been used to address these and many other methodological concerns.

Issues in Meta-Analyses

Statistical Questions

- Fortunately, the statistical methods used in meta-analysis are advanced enough in medicine and in psychology that researchers in political science who would like to conduct a meta-analysis can find a large literature on the methods that have been used to address these and many other methodological concerns.
- There are a number of textbooks on the subject, see for example Hunter and Smith (1990).

Issues in Meta-Analyses

Statistical Questions

- Fortunately, the statistical methods used in meta-analysis are advanced enough in medicine and in psychology that researchers in political science who would like to conduct a meta-analysis can find a large literature on the methods that have been used to address these and many other methodological concerns.
- There are a number of textbooks on the subject, see for example Hunter and Smith (1990).
- **SSC also discuss meta-analysis at length in chapter 13.**

Issues in Meta-Analyses

Statistical Questions

- However, given the research interests of the other disciplines, sometimes their answers are not appropriate for political science questions since many of the questions in medicine and psychology focus on particular isolated treatment effects of manipulations on individuals while much of political science research examines effects at both the individual and group level and the interactions between the two.

Issues in Meta-Analyses

Statistical Questions

- However, given the research interests of the other disciplines, sometimes their answers are not appropriate for political science questions since many of the questions in medicine and psychology focus on particular isolated treatment effects of manipulations on individuals while much of political science research examines effects at both the individual and group level and the interactions between the two.
- Furthermore, in the other disciplines, especially in medicine, it is likely that there are many studies which examine a common treatment and can be easily placed on a common metric for quantitative analyses, whereas doing so in political science may be more problematic.

Issues in Meta-Analyses

Statistical Questions

- However, given the research interests of the other disciplines, sometimes their answers are not appropriate for political science questions since many of the questions in medicine and psychology focus on particular isolated treatment effects of manipulations on individuals while much of political science research examines effects at both the individual and group level and the interactions between the two.
- Furthermore, in the other disciplines, especially in medicine, it is likely that there are many studies which examine a common treatment and can be easily placed on a common metric for quantitative analyses, whereas doing so in political science may be more problematic.
- See book for example of Lau's meta-analyses of the effects of negative campaign ads.

Is External Validity Possible Without Satisfying Internal Validity?

- What do you think?