Components of a PQE Answer / Research Proposal

1. Research Questions and Context (Don’t spend too much time on this section)
   a. Problem Statement – (Usually the PQE question or an abbreviated version.)
   b. Policy Context – Some sense of the policy context for the proposed work, but no detailed knowledge is expected.
   c. Conceptual framework / causal model – Articulates IV, DV, and covariates. Includes short statement of the theory of the treatment. What’s the theory behind the intervention? Why and how is the intervention expected to produce its intended effects? Are there other possible treatment impacts that should also be measured/considered in evaluating the intervention?
   d. Research questions (neither too nitty-gritty nor too broad)
   e. Hypotheses – Clear statement of question(s) addressed
   f. Research context (least important in PQE, since you are not expected to be familiar with prior research on chosen topic)

2. Research Methods (most important section)
   a. Basic design (Quantitative or qualitative. If quantitative: experimental, quasi-experimental, or correlational.)
   b. Source of variation (over time, across individuals, across aggregate units) – need to have variation to test anything
   c. Unit of observation and analysis (important)
   d. Sample strategy / method of assignment – Rationale for choice of that sample (those subjects) as exemplars; realistic sense of where subjects might come from. (What population does the sample represent, in space, in time, in terms of types of people, types of conditions or situations – needs to be sufficiently large to test anything). Method of assignment to treatment and comparison status discussed.
   e. Data sources / collection (surveys, administrative records, direct observation) – Realistic sense of where data might come from and what questions data can and can’t answer. (For example, you can’t ask people to recall what they did 20 years ago.) No data ex machina. Attention to spatial and temporal issues.
   f. Measurement – Brief description of how constructs are operationalized through specific measures. (Outcome (dependent) variables, treatment (independent) variables, covariates (independent, potential confounders).) Recognition that operationalization can be problematic
   g. Data analysis plan – brief and nontechnical. No advanced knowledge of statistics expected.
   h. Possible biases and how you are controlling for them (from confounding variables, etc.)
      i. Internal validity – history, maturation, testing, instrumentation, statistical regressions, selection, mortality
      ii. External validity (aka replicability, generalizability) – interactions of treatment with selection, setting, and history

3. Strengths and Weaknesses
   a. Highlight strengths of design, methods, and data – also address why you didn’t choose other designs, why yours is the best
   b. Acknowledge weaknesses in your proposed study – key threats to internal and external validity
   c. Consider/explain the feasibility of the proposed design and its ability to answer the research questions/hypothesis outline above
   d. Mention tradeoffs: internal vs. external validity, reliability vs. validity, sample size vs. better measures, cost vs. precision, etc.
## 1. RESEARCH QUESTION AND CONTEXT (don’t spend too much time)

<table>
<thead>
<tr>
<th>1.1 Context for research</th>
<th>Sense of the policy context</th>
<th><strong>Study rationale and study purpose</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1. What are the goals of the study?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. What policy problem is being addressed?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Why is this a problem? For whom is this a problem?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. State previous studies that have addressed the problem and specify by whom the studies have been done.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Why is the study relevant? Why is the study needed?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Who would the study benefit?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1.2 Conceptual framework/ casual model</th>
<th>Articulate IV, DV and covariates. Short statement of the theory of the treatment</th>
<th><strong>Research objectives and research questions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1. <strong>Objectives</strong>: a clear statement of the specific purposes of the study, which identifies the key study variables and their possible interrelationship and the nature of the population of interest.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. IV, DV and covariates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Short statement of the theory of the treatment - why and how the intervention is expected to produce its intended effects?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. <strong>Questions</strong>: specific purpose stated in form of questions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a. Connections between variables, what do you expect certain relationships to be?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Why and how do you expect the treatment to produce its intended effects?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1.3 Statement of hypothesis</th>
<th>Clear set of questions addressed</th>
<th><strong>Hypotheses</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1. A tentative prediction or explanation of the relationship between the 2 or more variables.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. A prediction to the answer to the research question (tie research questions to the hypotheses).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Hypotheses should include all variables that you are going to study.</td>
</tr>
</tbody>
</table>
2. RESEARCH METHODS (most important section)

<table>
<thead>
<tr>
<th>2.1 Design</th>
<th>Clear statement of the type of design to be used, design appropriate to the topic, definition of the unit of analysis</th>
</tr>
</thead>
</table>

**Basic design**

1. Label the design
   a. Quantitative/qualitative
   b. Intervention/descriptive
   c. Cross-sectional/longitudinal
   d. Prospective/retrospective
   e. True experiment/quasi-experiment

2. Specify the major elements of the design (diagram)
   a. Link the design to the casual model (perhaps do a diagram of the casual model)

3. State why the design suitable for the study given the constraints
   a. Is it the best design to do the study? Why?

4. Identify the source of variation
   a. Over time, across individuals, across aggregate units

5. Identify the unit of analysis

6. Time frame

7. What are the strengths and weaknesses of the design?

8. Discuss the feasibility of the design and its ability to answer the research question

<table>
<thead>
<tr>
<th>2.2 Sampling Strategy and method of assignment</th>
<th>Rationale for choice of the given sample, realistic sense of where the subjects might come from. Discussion of recruitment issues. Method of assignment to treatment and control group.</th>
</tr>
</thead>
</table>

**Subjects**

1. Who will be studied?
   a. Specify eligible subjects
   b. What population does the study sample represent?

2. How will they be selected? (sampling)
   a. Probability/non-probability
   b. What population does the sample represents?
   c. Rationale for the choice of sampling
   d. What characteristics do you want to represent in your sample?
   e. Size of the sample – How are you going to ensure a size large enough to represent each characteristic. Statistical power issues

3. How will they be recruited?
   a. Describe the methods that will be used to recruit the subjects.
   b. Discuss its feasibility and potential threats.
   c. Realistic sense of where subjects might come from

4. How will they be allocated to study groups?
   a. Randomization: why and how?
   b. Stratification, disproportionate randomization, etc. (discuss why?)
   c. Non-equivalent groups (discuss threats)
   d. Discuss costs
   e. Ethical issues

<table>
<thead>
<tr>
<th>2.3 Measures</th>
<th>Description of how constructs are operationalized through specific measures. Recognition that operationalization can be problematic.</th>
</tr>
</thead>
</table>

**How concepts operationalize**

1. Operationalization through specific measures
   a. Describe types of instruments for measurement
   b. Outcome variables: how will they be measured
   c. Treatment variables: how will they be measured
   d. Confounders: how will they be measured

2. Problems with operationalization
   a. What is lost from the concept
   b. Acknowledge threats to construct validity
   c. Discuss validity and reliability issues

3. If existing measures will be used discuss validity and reliability issues
### 2.4 Data source and data collection

Realistic sense of where data might come from, attention to spatial and temporal issues.

**Variables and data collection**

1. Specify data collection from each of the variables: where data may come from?
   - a. Dependent or outcome variables
   - b. Independent or predictor variables
   - c. Confounding variables
2. Spatial and temporal issues in data collection
3. Type of data collection
   - a. Surveys
   - b. Administrative data
   - c. Direct observation
   - d. Self reported
   - e. Other
4. Discuss if such data is already available, or how it will be collected
5. Discuss strengths and weaknesses of data (i.e. accessability).

### 2.5 Analysis plan

Brief and non-technical plan.

**Data analysis**

1. Briefly discuss how data would be recorded, stored, etc.
2. Descriptive statistics to describe shape, tendency and variability
   - a. Summarize important features of numerical data
   - b. Pick up entry errors
   - c. Characterize subjects
   - d. Determine distribution of variables
3. Inferential statistics
   - a. Estimate patterns and strength of associations among variables – what type of statistical analysis could be appropriate
   - b. Test hypotheses
4. Discuss issues about sample size (power and alpha level)
   - a. What magnitude do you expect in your results?

**Results**

1. What results do you expect, and if the results are not what you expect, how will you interpret them?
2. Given the findings, what might be the implications for policy makers and future research?

### 3. LIMITATIONS & STRENGTHS

#### 3.1. Strengths/limitations

<table>
<thead>
<tr>
<th>Measures – construct validity issues</th>
<th>Design - threats to internal validity</th>
<th>External validity issues</th>
<th>How could such validity threats and issues could be addressed in the future</th>
</tr>
</thead>
</table>

**Limitations and strengths**

1. Measures
   - a. Construct validity issues
2. Design
   - a. Threats to internal validity
   - b. External validity threats
3. Ethical considerations
   - a. Ethical principles
   - b. Beneficence (do good)
   - c. Justice (exclusion)
4. How might this be mitigated in future research
CLASSICAL and Quasi-Experimental Designs

CLASSIC EXPERIMENTS

1. Classic Experiment = Control group + Random assignment
2. What is random assignment?
   a. To randomly allocate members of the target pop to the two groups such that every unit in the target pop has the same chance as any other to be assigned to either group
      - CONTRAST THIS with self-selection!
         1. Those more motivated, informed "assign" themselves to treatment group
         2. Individuals with certain characteristics have a higher probability than others of being selected for either group
         3. Noncomparable groups
   b. Therefore, at large samples the two groups are probabilistically equivalent → High INTERNAL VALIDITY
      - They are comparable within the limits of sampling error.

Random Selection and Random Assignment

Random selection is how you draw the sample of people for your study from a population.
Random assignment is how you assign the sample that you draw to different groups or treatments in your study.

1. Random selection most related external validity because related to sampling
2. Random assignment is most related to design. In fact, when we randomly assign participants to treatments we have, by definition, an experimental design. Therefore, random assignment is most related to internal validity. After all, we randomly assign in order to help assure that our treatment groups are similar to each other (i.e., equivalent) prior to the treatment.
Variations on the classic experimental design

1. Natural experiments: Not common in evaluation
   a. The impact of floods on housing and economic development

2. Post-test only with strict random assignment
   a. Why no pre-test?
      i. Time, money, timing, sensitization
      ii. No possible pre-test
         Program for impoverished HS students go to college.
         Only way to judge is whether experimental go to college more than controls – post-test only

3. Randomized block design
   a. Group units into relatively homogenous blocks and then within each block randomly assign participants to treatment or control conditions
      i. To reduce the noise or variance in our data, because the variability within each block will be less
      ii. EXAMPLE Sample is College Students – block by student year, and then within each year, randomly assign

4. Factorial design
   a. Units are randomly assigned to multiple treatment conditions, defined by more than one manipulated variable
      i. Commonly, ONE MANIPULATED VARIABLE = treatment
      ii. MULTIPLE MANIPULATED VARIABLES
         1. Reading program variables = Hours of Instruction AND LOCATION
         2. Randomly assign participants to FOUR conditions
            H1, L1
            H2, L1
            H1, L2
            H2, L2
   b. Great flexibility explore or enhance program
   c. Only effective way to examine interaction effects.
5. **Solomon four group design**
   a. Defines four assignment groups. Two treatment; one pre-post and the other post-only. Two control; one pre-post and the other post-test only. Allows test for
      i. Pre-test learning (main effect of testing)
      ii. Potential interaction of pre-test and treatment
         Pre-test enhances effect of treatment

6. **Cross-over design**
   a. Assignment groups defined by **multiple “treatments” and multiple time periods**
   b. Same individuals receive the treatment and control treatments, just at different times.

7. **Longitudinal designs**
   a. Multiple longitudinal measurements taken during and after intervention
   b. Increase measurement reliability
   c. Permit examination of how intervention works over time.
COMPARISON GROUP DESIGNS
HOW TO...?

Identify and construct comparison groups

The major distinguishing factor in Q-E designs is way groups developed to minimize selection bias [REMEMBER BIGGEST THREAT!]

Through NONRANDOM means:

1. Matching – OR –
2. Statistical Adjustment

When?

Ex-ante or Ex-Post

Can you designate groups before intervention, or has program group already been identified and therefore have an ad hoc comparison group?

Ex-Ante: Opportunity to strengthen impact assessments.
Ex-Post: Long history in impact assessment

Do your best and then use statistical controls.

What Characteristics Matter?

In quasi-experimental designs even more important than exp groups to demonstrate comparability

Problem is know and have access to what is important in terms of comparability.

GOAL: Construct groups so comparable in terms factors relevant to OUTCOME of interest.

Failing can CONTAMINATE comparability

EX: If choose "strong" participants for treatment;
OVERESTIMATE impact. The OUTCOME will be large from
the combination of initial differences b/w T & C, and the program impacts.

**Relevant factors** based on PRIOR KNOWLEDGE and THEORETICAL UNDERSTANDING social processes in question

EX: Eval program to ↑ math competence 2nd grade kids.

Use prior knowledge characteristics individuals and settings that affect learning:
School organization
Students' sex, age, intelligence, family background

1. Literature review
   Factors known related to Outcome
2. Watch for variables potentially related to self-selection processes
3. Pre-test measure of OUTCOME
   Ideal and strongest by far,
4. Demographics important mostly for descriptive reasons

How to CONSTRUCT?
1. Matching
2. Statistical procedures (WITH OR WITHOUT MATCHING)
3: Normative or generic controls

Matching
- Must have large pool for both T & C
- Must know reasonable matching criteria and have access to that info on a CASE BY CASE basis

**Individual Matching**

To find a “partner” for each target from the unexposed pool of targets or potential targets
- Usually preferred when multiple characteristics
- Relatively more expensive, time consuming, difficult
- Can result in dramatic loss of cases

**Aggregate Matching**

Assemble groups such that the overall distributions in the treatment and comparison groups on each matching variable correspond.

EX. The same population of children by sex and age would be found in T & C schools

**BUT**

this may have been obtained by including a 12 yo girl and an 8 yo boy in comparison group to balance the aggregate distribution of the experiment group (which included a 9 yo girl and an 11 yo boy).

**Statistical Procedures**

Collect information on relevant variables for both groups and use as statistical controls.

- **PRIMARY** approach deal w/ selection bias and other unwanted differences
- Particularly common in recent decades
- Relevant to both ex-ante and ex-post

Methods:

1. Sub-group analyses
2. Multivariate Statistics

LOGIC: Create a statistical representation of the overall relationships among the control variables and the outcome variable in a statistical model that allows inferences about the relationship of the intervention to the outcome that is LEFT OVER after all the relationships of the control variables to the outcome variables have been accounted for.
YOU CAN'T CONTROL EVERYTHING

Types of relevant variables:
1. "Predictive" Variables: Characteristics of group members that are related to the outcome of interest.
2. "Selection Bias" Variables: Characteristics that are related to the selection of individuals into T vs. C groups, if also related to the outcome.

EX: Job training program evaluation
   Comparison of T & C outcomes with different degrees of statistical control: 0 controls; w/ educational attainment; w/ educational attainment and baseline employment

REGRESSION DISCONTINUITY DESIGN:
When evaluation cannot RA but can collaborate w/ program personnel to divide targets systematically on the basis of need, merit, or qualifying condition and ASSIGN those neediest/meritorious, etc. to T and those less so to C.
   - Most rigorous Q-E impact assessment that can be undertaken
   - Rarely used – perhaps evaluators don't understand!

Normative or Generic Controls
Measures that can serve to represent “control group” outcomes.
Must have data on same measure of a population it makes SENSE to compare it to

Examples:
- demographic measures
- Death and birth rates
- Sex ratios
- Labor force breakdown
- Pub. Standards various psych tests

Tempting, but dangerous compare samples from dis-advantaged pops to "normals"

Where to get your controls?

1. Different geographic location
   - Without program or policy not there
   - Often only way to go for larger-scale policies that blanket an area

2. Waiting list
   - Very strong, if no compensation
   - If possible, better to do RA though if enough flow (like BBBS study which made it a true experiment)
   - Problem in longer-term evaluations
     Waiting list eventually get program
   - History remains a threat

3. Similar units within an organization
   - Classes in schools, shifts in factory, floors in hospitals

4. Ineligibles or eligibles who didn’t participate
   - Problem? Still permits difference bw T & C groups due to selection bias.

5. Samples from earlier/later in time: Cohort Designs
   - EX – Adolescent Tobacco Prevention Program
INTERRUPTED TIME-SERIES DESIGN

A variation on quasi-experimental designs with comparison groups.
Conceptually, easy to understand, although statistically sophisticated
Need to model the structured error (autocorrelation)
Quite common in the types of research many of you will pursue
Economic analysis
Policy analysis
Some types of program evaluation
Often with large available data sets);
REFLEXIVE DESIGNS

Most **SEVERE RESTRICTION** on choice strategy whether intervention in delivered to all (or virtually all) target population.

For total coverage programs, usually impossible if anyone not receiving intervention AND who in essential ways is comparable.

↓

REFLEXIVE CONTROLS

*(some form of before-and-after comparison)*

(or variations in activities, intensity, duration of program)

Aside-

Often done in partial coverage programs

Easier and cheaper

Particularly because of volunteer bias

In a pre-post with comparison, the reflexive design exists as well, always that option

One group designs

**Simple Pre-Post**

Comparison of the same targets at two points in time, separated by period of participation in program.

Differences between T1 and T2 = Impact Estimate

One of least valid designs

**WHY?**

Can't rule out history & maturation by design

Cannot disentangle effects of extraneous factors from effects of the intervention

**Bolstering Steps/Rule out Alternative Factors**

1. Keep time frame bt pre & post short
   - Less 'history' has unfolded
- Where relevant, counters unreliability of recall

2. Employ logic, theory, previous literature to identify threats and assess relevance to your research and likely impact on findings.

3. Expert comment on historical events, maturation processes

4. Multiple data collection points/time series

Cohort Designs

Panel study or Several repeated measures
Repeated measures of targets exposed to the intervention

- More confidence that history/maturation not causing effects
- Ability specify the processes by which intervention has impacts on targets
- Only possible post-program of course

Dose-response strategy
Comparison made possible by variations within the intervention

- Differences in eligibility requirements or benefit levels, for example
- Use of “internal” comparison group [Weiss].

Estimate the effect of these variations:
- Measure how much intervention received by targets (program dosage)
- Contrast measures of outcomes for targets receiving different levels of intervention.

  EX. Compare outcomes for those who had 3 hours/week of training to those who had 8 hours/week of training.

Be careful that those who receive more service are not different on other grounds, too (more conscientious in attendance, motivation, etc.)
Shadow controls

When evaluating a program after onset, without possibility of assessing baseline characteristics, need to make assumptions about what things looked like before

How?

Examine agency records for relevant pre-characteristics

Program administrators; participants

Ask them what would the DVs have been like without the program or before program
Also ask about history – what else going on.

Experienced judgement (evaluator; experts)

biased to favor intervention
very hard to evaluate, judge

normative controls: what was the trend nationwide or with some other similar population? Helps control for broad history/external effects and maturation

Time series

Simple interrupted time series (No comparison group; reflexive design)

Repeated measures taken on target units (usually aggregate unit) with many data points preceeding and following the point in time at which a new full-coverage intervention was introduced or and old program was substantially modified.

Enable evaluator to interpret the pre-to-post changes in light of additional evidence – are the measures immediately before and after the program a continuation of earlier patterns or do they mark a decisive change?

Ex. Deterrent effect of stricter drunk driving laws on accidents related to drinking reflexive design
Not always easy to interpret, if the change at X is not striking

How? Analyze the trend before intervention
   Develop projection what WOULD have happened
   Compare projection with actual

Presence a lot of noise makes difficult assess trends
   Autocorrelation

- Biggest weakness is the **confounding of history**
  Ex. change in drunk driving laws coincided with intense PSAs;
  Try to rule out by keeping careful track of all other plausible causes
  & see if there were any noteworthy changes

- **Change in instrumentation can be a threat**
  Particularly record keeping;
  When using secondary data analysis this is a very critical issue in
  terms of reliability
  Ddefinition of "related to alcohol" narrowed/broadened)

- **Selection bias can still be a threat**
  If group changed around time of introduction of X (either because of
  X or because of totally unrelated reasons; drinking drivers moved
  out of state)

**Two group version** – Add a comparison area/group

- Comparison area helps control maturation
- Helps control threats from global history, but not local history
- Careful - Differences in instrumentation between treatment and
  comparison can be problematic
  - How different localities collect their data

**Advantages & disadvantages**

**Advantages**

- Can detect **differences between groups related to slope (rate of change)**, as opposed to just intercept (single posttreatment measure);
- Can detect **decay** in effects
- And also **delayed effects**
- Can detect or compensate for **seasonality/cyclical effects** with or w/o comparison
- In single group designs, it helps detect maturational trends prior to X

**Disadvantage**

- **Need a lot of data points**
  - Real problem in primary data collection;
  - In secondary data, do not always have exactly what you need,
    - Systematic bias
    - Hard to gain access
    - Not in the appropriate time intervals

**Not a panacea**

Time series helps with a number of the common weaknesses of the simpler quasi designs

Some people seem to think that the sheer amount of data, or the sophistication of the statistical techniques makes time series inherently superior

Do not be fooled; the problems of using archival data alone are daunting
Approaches to Outcome Evaluation

Reflexive Designs: target group program used as its own comparison group. Reasonably assumed that in the circumstances of the experiment, it is reasonable to believe that targets remain identical in relevant ways before and after the program.

1. One-Group Pretest-Posttest design
   - Most commonly used design, while least powerful because changes in those receiving the program may be caused by other events and not the program, but evaluator has no way of knowing this.

2. Simple Time-Series Design
   - Useful when program measures available on number occasions before implementation. And when there appears to be an underlying trend in the data. (Crime and accident statistics good examples) Analysis compares actual postprogram measures with projections drawn from preprogram period
   - Evaluator still no way knowing changes caused by program or some other events. Therefore appropriate when no other rival explanations effect can be entertained.

Quasi-experimental designs: Change in performance target group compared to performance comparison group; while trying to create a comparison group as close to experimental group in all relevant respects. Especially useful for retrospective analysis, when evaluation resources follow the initiation of the program.

1. Pretest-Posttest Comparison Group Design

   ![Diagram](attachment:image.png)

   - Substantial improvement over reflexive, attempt create comparison group as similar as possible to program recipients
   - Comparison group alleviate many threats to internal validity, which depends on how closely comparison group resembles program recipient group.
     - Matched-pair design assigns subjects to experimental and "control" basis some common characteristics evaluator wishes to measure, and only these factors are controlled for in the analysis.

2. Regression discontinuity design
   - Criteria for use:
     a. Distinction between treated and untreated groups must be clear and be based on a quantifiable eligibility criterion
     b. Both eligibility criterion and the performance or impact measure must be interval level variables
     - Most often met programs use family income or combination income/family size to determine program eligibility. Income maintenance, rent subsidy, nutritional and maternal health programs
     - Performance of eligible program participants is compared to that of untreated ineligibles. Assumes that the subjects just above and just below eligibility cut point are equivalent

3. Interrupted time-series comparison group design
   - Nature of design eliminates bias results when one makes only one observation of a phenomenon
- Hypothesized impact is driven by theory - knowledge of the form the impact will take
- Presence of a lot of noise makes difficult to assess trends. Some of this noise may not be nonrandom, but be correlated error terms resulting in autocorrelation (violating a fundamental assumption of OLS. Understates variance estimates of the regression coefficients - making easier to find statistically significant result, but also reduces confidence.

4. Posttest-only comparison group design
- Used primarily when evaluation proposed after the fact, one of most commonly used designs
- Weakened by data collection at single point in time - limits ability determine whether observation is really "true" or if result random fluctuations
- Sensitive to evaluator determination of outcome form, ie expecting immediate or long-term changes. Source of both Type I and Type II errors. Classic example is results early evaluations Head Start - showed little short-term impact
- Internal validity, obviously, of particular concern; therefore, important to measure characteristics treated and comparison groups at same time

Experimental designs provide best measure of the counterfactual

1. Pretest-posttest control group design
- Strength rests upon random assignment of participant to program and control groups. Given sufficiently large groups, random assignment assures characteristics subjects both groups virtually same prior to initiation program. Reasonably ensures any difference between two groups postprogram measure will be result of the program.
- Where feasible:
  - When likely high degree ambiguity whether outcome caused by program or something else.
  - When confidentiality and privacy participants can be maintained
  - When some citizens can be given different services from others without violating moral or ethical standards,
  - Findings generalizable to substantial portion population
    - For expensive programs with substantial doubts to effectiveness
  
- Prohibitive conditions
  - Ethical constraints
  - No comparable control groups for volunteers
  - Generally more costly than other designs

2. Factorial design
- To evaluate program has more than one treatment or component, allows identification of any interaction effect between variables
- Interaction describes manner in which simple effects of a variable may differ from level to level of other variables.
- Interaction effect refers to a programmatic effect of two distinct program components combined that is greater than the sum of the effects of the components separately

3. Posttest Control Group Design
- With random assignment, a pretest may be unnecessary
- Advantages
  - Time and cost savings without pretest
  - Prevent pretest sensitization problem
- Compared to pretest-posttest control group design
  - Use Pretest allows the evaluator to reduce size sample (and thus reduce costs)
  - Pretest used statistically as controlling variable in either repeated measures analysis variance or covariance
  - When subjects in limited supply, pretest recommended
Research designs compared
Implementation
Time
- Evaluation researchers hope to construct "best" design, while erring on side of expediency when decision deadline is "near."
Cost
- Remember, evaluations often first activities cut during budget slashing
- Experimental designs relatively expensive
Strength of findings
- One-group Pretest-Posttest should be used sparingly
- If experimental design not feasible, interrupted time-series or one or more quasi-experimental designs in combination (increased reliability)
Analysis [SEE NOTES .2875]
Outcome analysis
Impact analysis
1. Ordinary least squares
   • Continuous treatment and dependent variables
     • "Compared to the Treatment = 0 group, the Treatment group experienced $\beta_1$ more/less units of Y, ceteris paribus"
2. Linear probability model
   • Dichotomous outcome variable
   • Model permits calculation of 'excess' events - risk difference
   • "Compared to the T=0 group, the treatment group is ($\beta_1$ *100) percentage points more/less likely to have the outcome Y=1, ceteris paribus.
2. Logistic Regression Model
   • Dichotomous outcome, dichotomous treatment
   • "Compared with the no treatment group, members of the treatment group were e$^{\beta_1}$ times as likely to develop Y, ceteris paribus."
Sub-group analyses
- Separate analyses of impact by different groups
  - Interaction: An effect that is larger or smaller than would be predicted on the main effects of the factors.
  - Conduct subgroup analysis according to some logical division in your group (for example, their may be some reason to believe that your treatment would have a different effect on men and women). Run two separate regressions - one for men, and one for women. Compare the impact estimates.
  - If they look sufficiently different run regression on entire population after creating a new interaction term (sex*treatment). If the interaction term is significant, it should be interpreted as the difference between the impact seen on men and the impact seen on women.
Interpretability of results
Validity:
  - Internal validity: The extent to which the independent variables did, in fact, cause the dependent variables. Or the degree to which a research design allows an investigator to rule out alternative explanations concerning potential impact program on target group. Validity depends on the extent to which the two things being compared are equivalent
Threats to internal validity: Interaction on the basis of selecting control and/or experimental groups with the other factors is that the interaction may be mistaken for the effect of the treatment
- History: Events that occur during the time program that provide rival explanations for changes in target or experimental groups (coincidental history).
  - Systematic history - people's earnings go up after a spell of unemployment
- Maturation: Maturation or changes produced in the subject simply as a function of the passage of time. While history may cause changes in the measurement of outcomes attributed to the passage of time, maturation may cause changes in the subjects as a result of the passage of time
  - Especially problematic for programs directed toward any age-determined population, as maturational processes may mask program effects
- Testing: The effect of taking a test (pretest) on the score of a second testing (posttest). Difference between preprogram and postprogram scores might thus be attributed to improved test-taking or extent to which test sensitizes subject to some substantive issue, "Hawthorne effect," and placebo effect.
- Instrumentation: Internal validity threatened by change in the calibration of a measuring instrument or observers or scorers.
- Statistical regression: A regression artifact is suspected when cases chosen for inclusion in a treatment based on their extreme scores on a variable.
- Selection (Bias): Uncontrolled selection means that some individuals are more likely than others to participate in the program under evaluation. Most problematic in case of target self-selection
- Experimental mortality - concerned with why subjects self-select our program (rather than why participate in case of selection bias)
  **Without proper control, postprogram measurement may show an inflated result because it measures progress of only the principal beneficiaries
  - (Selection-maturation interaction)

- External validity: The generalizability of research findings to other sites and situations
  - Representativeness

Controlling threats to validity through randomization
- History - both experimental and control groups exposed to same events occurring during program
- Maturation - neutralized because both groups undergo the same changes
- Expect equal rates of mortality (but not necessarily true)
- Testing addressed because effects present in both groups
- Same for regression to mean and selection effect is negated.

- Trade offs between internal and external validity

Errors:
- Type II: Concluding an impact does not exists, when that impact truly exists.
  - In the case of evaluation, a program may be terminated when in fact it was beneficial
  - Relatively more common as a result of working with the "best" design and the "best" measures of variables available
- Type I: Incorrectly concluding an impact exists
Research issues
Measurement: The rules for assigning numbers to values on a variable
- Unit of analysis
- Measuring whether program had desired effect depends upon ability to measure this effect. But how does one measure the success of job training, for example.
- Need for mutually exclusive and collectively exhaustive categories of measurement
- Ecological fallacy: Problem of imputing individual level behavior from aggregate-level measures and statistics. For example, one cannot infer the voting behavior of an individual from variables measured at the city level.
- Levels of measurement
  - Nominal level: Names or numerals are assigned to classes of categories in purely an arbitrary manner - no rank order, no greater than or less than implied. Gender
  - Ordinal level: Values are arranged in a sequence from lesser to greater. "Better" or "worse" but no measure of by how much.
  - Interval level: Variables have a constant unit length between, allows one to perform mathematical functions
  - Dummy variables
    - (N-1) dummy variables used for N nominal categories to prevent overdetermination of model
- Validity: A measure is valid to the extent it captures or measures the concept or thing it is intended to measure
  - Face validity
  - Content validity (especially, achievement tests).
  - Criterion validity
  - Construct validity
- Reliability: A measure is reliable to the extent that essentially the same results are produced in the same situation, and that these results can be reproduced repeatedly as long as the situation does not change. Reliable measures are dependable, stable, consistent, and predictable. Refers to the likelihood a given measurement procedure will yield some description of a given phenomenon if that measurement is repeated.

Causal models
Variables
- Dependent variables
  - We want to choose variables we can effect
- Independent variables
- Intervening variables
- Interaction variable

Causality
Defined
- Discussed in probabilistic terms; such that there are exceptions. Certain factors make another variable or outcome more likely/less likely
- Necessary cause represents a condition that must be present for the effect to follow
- Sufficient cause represents a condition, that, if it is present, will pretty much guarantee the effect in question

Establishing
1. Covariation across units and across time
2. Exogeneity: The cause of something is not part of the effect
   - There is a directionality and a temporal ordering of variables - ideal when constructing research designs.
   - Endogeneity problem: when your outcome indirectly determines independent variable
3. Absence Alternative explanation: when you have a variable that messes up a nice relationship; for example, education may effect income, but
parental income effects both of these. Threats to validity of our designs are alternative explanations.

- **Spurious Relationship** (parental income to X and Y) When relationship between two variables appear causal when actually is not. When control for a third variable, the relationship between original two falls apart. Educational attainment would appear to effect occupational choice, but when you control for socioeconomic background, educational attainment no longer has an independent effect on occupational choice.
  - Need separate table for each combination of categories of the control variables (X and Y with Z, X and Y without Z)

- **Intervening Variables** ("how" variables)- If Z stands between X and Y in a causal sequence, it should be more highly correlated with both X and Y than X and Y are with one another. Consider X (educational level), Z (employers' selection based on ed. Levels), Y (occupational choice). Educational level is understood more accurately as an indirect cause of occupational choice, mediated by employers' selections.

- **Interaction variables** effect relationships between X and Y. These are conditional variables, a "for whom" variable. If the strength of the relationship between X and Y varies according to differing values of z.
  - Interaction effects are nonadditive effects of unusual combinations of variables

Two simultaneous causal effects: Remember example of introduction of air bags - accidents became safer having a negative impact on traffic deaths; however, at same time, air bags also associated with more accidents (possibly people more reckless) having a positive impact on traffic deaths.
- Correlated independent variables can be addressed through experimental design

Threats to valid causal inference

- **Internal:**
  1. Bias
  2. Confounding (or alternative explanations)
    - Spurious effects
  3. Chance variation
- **External**
  1. Selection criteria control nature target population, limiting external validity
  2. Participation rates effect the representativeness of the eligible population

Research designs

Classification by groups of people compared

1. Cohort and Intervention studies - groups of individuals defined in terms exposure to putative causative factor. (Cohort is a group of people with some characteristic)
   a. Observational studies - researchers observe natural events, do not influence participants
   b. Intervention studies - investigators control the assignment individuals to intervention
- Cohort studies problematic for conditions extremely rare, as will have to follow such a large number of individuals
2. Case-control studies - individuals defined in terms whether have or have not already experienced outcome under consideration, and then the exposure is measured.

Classification by time relationships
1. Prospective Study - Subjects are entered and data are collected at a point in time, and then the subjects are followed and further events recorded as they happen.
2. Retrospective cohort studies
3. Cross-sectional studies

Results of studies of causation
1. Relative risk and relative odds [SEE NOTES FROM 2875]
   - 'Relative risk' or 'risk ratio' is the ratio of the rate of disease among those exposed to the rate in those not exposed, unitless
   - 'Odds ratio' and relative risk are similar where the frequency of outcome is low; they diverge as the outcome becomes more frequent
   - Risk difference is the difference between the rates of two groups, gives the frequency of the outcome which is associated with the exposure. If there is no association present, the risk difference is zero.
      - Useful considering practical implications studies - how many lives will be saved by an intervention?
      - Particularly useful in well researched situations where implications soundly reported results are being considered. Less value in preliminary stages of assessment of a possible causal relationship.

2. Odds Ratio is primary measure of association in case-control studies
   - Cannot calculate rates of disease in the exposed and unexposed groups, nor can relative risk be directly measured
   - Very good estimates relative risk in most situations, except where outcome is very frequent
Formal Steps to Writing the Research Proposal
A Study Guide for the PQE

General Strategy
- Choose question and read it carefully, outlining question on paper: 15 minutes
- Outline proposal from start to finish on paper, using the formal steps I lay out here: 45 minutes
- Using framework, write entire proposal from start to finish: 3 hours
- Go back and edit: 30 minutes
- Read over carefully, fixing any errors: 30 minutes

Step 1: Background and Motivation
- **Time allotment:** 15 minutes
- Policy context – restate the information in the question as background
- Make a case for why it is important to study
- Why is it a problem? For whom is it a problem? Who are the major stakeholders?
- What proposals have been suggested and by whom?
- What you might learn from conducting a study
- Formally state your research question
- *Remember, this is not a mystery novel. Do not leave the reader in suspense!*

Step 2: Conceptual Framework
- **Time allotment:** 30 minutes
- Restate purpose of study.
- State theory of treatment
- You must explicitly define concept to get everyone on the same page
- Discuss the process of moving from reality of “A” to construct of “A” to measurement of “A”
- What is it you are trying to measure?
- How will you measure the construct? Articulate IV, DV, and covariates (all variables MUST appear in this section; there should be no new variables introduced in the design or methods section)
- What are the problems with your construct?
- Is there temporal precedence of IV to DV?
- What might confound the relationship (third variable) and what do you propose to do about it?
- Be clear about variation within each variable.
- What is the connection between the variables?
- What does the relationship look like? Draw it out.
- Why and how do you expect the treatment to produce its intended effects?

**Key Research Method Tools and Concepts:**
- Measurement and operationalization
- Construct validity – mainly face validity; content validity
Step 3: Research Hypotheses

- **Time allotment: 15 minutes**
- Objectives of the study
- Clearly state your research questions
- Clearly state your hypotheses; a tentative prediction or explanation of the relationship between the 2 or more variables.
- Link questions directly to hypotheses to be on the safe side – prediction to the answer to the research question
- **Make sure all variables in the study are included in your hypotheses**

Step 4: Research Design

**Design**

- **Time allotment: 45 minutes**
  - Clear statement of design type you plan to use (control group, pre post, etc)
    - Quantitative/Qualitative
    - Intervention/descriptive
    - Cross-sectional/longitudinal
    - Prospective/retrospective
    - True vs. quasi vs. nonexperimental design?
  - Link design to causal model
  - Identify source of variation (over time, individuals, aggregate units, etc.)
  - Random assignment or not? If random, make sure it is feasible in the “real world”!!!
  - Why did you choose this design? Why is it a better design than others in this context?
  - Address some internal/external validity issues here, but go into detail in limitations section; talk more about the strengths
  - What other ways might you have chosen and why did you not choose them?
  - What are some of the problems with the design (nonequivalent groups are not as equal as possible…); if no pretest is possible, make sure you explain why
  - Identify the unit of analysis
  - Feasibility of the study – are there any political/legal or ethical issues?
  - Time frame, costs, resources...

**Sampling Strategy**

- Who will be studied?
  - Specify eligible subjects
  - What population does sample frame represent?
- How will they be selected? (sampling)
  - Probability/non-probability
  - Matching
  - Rationale for choice of sampling
  - What characteristics are important to the sample?
  - How will you ensure large enough sample size
  - Sample size – relate to power issues
- How will they be recruited?
  - Describe methods used to recruit subjects
  - Feasibility/potential threats
- Realistic sense of where they might come from
- How will they be allocated to study groups?
  - Randomization
  - Stratification, etc. (why?)
  - Non-equivalent groups – discuss threats
  - Discuss costs
  - Discuss ethical issues

**Measures**
- How constructs are operationalized through specific measures
- What instruments will you use to measure constructs?
- Key outcome variable(s) and measure
- Key treatment variable(s) and measure
- Covariates and/or confounders and measurements
- What are the problems with your operationalization? What is lost in the translation?
- What are the validity and reliability issues?
- Test-retest reliability; intraobserver/rater or inter-observer/rater reliability, etc.
- How might you triangulate the measures to ensure validity (think convergent validity/discriminant validity)

**Key Research Method Tools and Concepts:**
- True, Quasi, and Non experimental designs
- Strengths and weaknesses of randomization versus nonequivalent groups
- Reasons for matching
- Construct validity – mainly convergent, discriminant
- Measurement and operationalization
- Statistical power
- Sampling techniques
- Reliability tests/sensitivity analyses
Step 5: Data and Methods

Data

- Time allotment: 30 minutes
- Remember – do not make any assumptions about data. If you think something might be included in an administrative database, say it. Be sure to include what you will do if it is not included (survey, etc.). If you propose to use a certain database, you better darned well understand the intricacies of this database.
- Realistic sense of where your data will come from (no data ex machina), type of data collection:
  - Surveys
  - Administrative data
  - Direct observation
  - Self-reported
  - other
- Will data be primary or secondary? Does it already exist?
- Any temporal/spatial issues?
- How will you collect the data? Any confidentiality/ethical issues?
- Can you supplement secondary data with a survey?
- Any qualitative data collection to supplement?
- Strengths and weaknesses of the data

Methods

- Briefly discuss how data would be recorded, stored, etc.
- Descriptive statistics to describe averages, frequencies, where the variation is
  - Summarize important features of data (what makes it unique?)
  - Cleaning – pick up entry errors, etc.
  - Imputation
  - Characterize subjects
  - Determine distribution of variables
- Inferential statistics
  - Estimate patterns and strengths of associations among variables, what type of statistical analysis would be appropriate?
  - Test hypotheses
- Keep it brief and non-technical
- Do not write down a model unless you are 100% sure it is right!!!
- Discuss statistical power
  - What magnitude do you expect in your results?

Key Research Method Tools and Concepts:

- Inter- or intra-rater reliability
- Statistical conclusion validity, power analysis

Step 6: Results and Policy Implications

- Time allotment: 15 minutes
- What might you learn from this study?
- How might you interpret the results given the theory of treatment?
- What results do you expect?
- What magnitude do you expect?
- What might be the policy implications of your results?

**Step 7: Limitations**

- **Time allotment: 15 minutes**
- Measures – construct validity, reliability?
- Design – threats to internal and external validity
- Ethical considerations – any issues?
- How might these problems be mitigated in future research? Relate back to any alternative designs you discussed.
- *This discussion should be integrated throughout the paper.*
PQE Study Guide

I. Basic concepts (be sure to be explicit):

- Quantitative or qualitative?
- Source of variation (over time, across individuals, across aggregate units); unit of variation??? Make sure IV and DV have variation so you can have at least one comparison. Anything else?
- Potential internal and external biases (from confounding variables) and how you are controlling for them. Identify major threats to validity (including social if necessary, e.g. give comparison group treatment after), and how you plan to prevent them.
- Mention tradeoffs: internal vs. external validity, reliability vs. validity, sample size vs. better measures, cost vs. precisions, etc.
- Reliability: who is collecting data? Inter-rater reliability, standardization, examiner bias (test randomly), don’t tell administrators contents of the test
- Can say: “Want to look at effects of each variable separately,” not “want to separate out the effects” (its not possible in reality)
- Mention early in paper what your design is, what kind of data you will collect (retroactive, existing, survey, etc.), what kind of analysis, etc. Just so you don’t leave reader guessing. Be brief, details will be fleshed out. (Does this sound good or not??) What are the key things to mention up-front?
- Be sure to carry through the research problem you raise to your research question to your hypothesis, to your operationalizations, to your analysis. Don’t drop anything along the way.

E.g. from my proposal last semester:

Research Question:

General: What factors are associated with garden status?
Factors ————> garden status

Because I am using a cross-sectional design with retroactive data, garden status is already set at the present moment. Therefore, the language cannot be about "predicting" the status of gardens, but certain factors are "associated with" certain garden statuses.

I say: Does the internal strength of a garden predict a garden’s status?
Specifically, are gardens with non-profit tax status, greater fundraising ability, and/or formalized governing structures more likely to become permanent? Does having an experienced leadership, or a strong sense of community among gardeners enhance the chances of becoming permanent?

I should say: Are preserved gardens more likely to have higher mean scores on internal strength measures? And sold gardens more likely to have lower mean scores on internal strength measures? Etc.
I say: Does the fact that a garden is affiliated with garden coalitions, non-profit organizations (with community garden agendas), and other community gardening groups protect them from being sold?

I should say: Are preserved gardens more likely to have higher scores on the external affiliation scales then temporary gardens? Etc.

Hypothesis:

I say: A garden with a higher public profile is more likely to become permanent. Gardens that have won dress-up-your-neighborhood awards, are well-maintained and accessible, have received positive media coverage, and/or are closely involved with the surrounding community have a better chance of being permanent.

Should say: “Preserved gardens will have higher mean scores on public access and media coverage scales than either temporary or destroyed gardens.”

I say: Gardens with non-profit status, strong gov. etc., are more likely to become permanent.

I should say: Gardens with permanent status are more likely to have strong internal organization, than gardens of other types.

Analysis Plan:

I say: “The study will be divided into two analyses. The first will look at the mean differences in the independent variables across the three statuses to test all six hypothesis listed above.” The way I’m analyzing is OK, but elaborate.

Should say: “For example, the mean scores on the public access and media coverage scales will be compared across permanent, temporary, and sold gardens.” Give other examples.

II. Outline of paper and unabridged list of issues

Context for research

- Show sense of policy context
- What policy problem is being addressed? Why is this a problem? For whom is this a problem?
- What, if any, specific proposals to address the problem are being considered currently? By whom?
- Why is a study needed?
Conceptual framework/causal model

- Articulate IV, DV and covariates; make sure variation within each variable is categorized well conceptually. (my research proposal articulated garden status as “saved” or “sold” when something more fluid would have been appropriate: “sold-uncontested,” “sold—contested.”
- Include short statement of the theory of the treatment
- Hypothesis are described and justified
- Set up some discussion of connections between variables; what do you expect certain relationship to be?
- Why and how do you expect the treatment to produce its intended effects?

Statement of the hypothesis

- Clear statement of questions to be addressed
- Clear statement of hypotheses; tie research questions to hypothesis
- Objectives of study
- Hypothesis should include all variables that you are going to study; in general the variables should be consistent throughout.
- Hypothesis should be addressed in analysis; hypothesis and analysis should have same scope

Design

- Clear statement of type of design to be used
- Is design appropriate to topic? Why is this the best design given the constraints? Is it the best—and perhaps only way—to do the study? Why?
- Identify unit of analysis, can be a tricking question be careful.
- What are strengths of design? Weaknesses?
- Consider/explain the feasibility of the proposed design and its ability to answer the research questions/hypothesis. Talk about follow-up needs/potential and strengths.
- Random assignment or not? Randomization ensures that affects are from the treatment and not some confounding variable (controls for validity problems). Is non-obtrusive. Can be expensive. Does not work when program has already started.
- Non-equivalent group design: biggest problem is ensuring that groups are as equal as possible so affects can be attributed to treatment vs. some inherent difference between groups. After participants are selected they can be assigned to groups based on matching characteristics.
- Time frame? Cost? Resources?
- Ethical issues?
- Trying to show causal relationship between treatment and outcome. Internal and external validity issues constrain the ability to draw conclusions accurately so they need to be dealt with.
• External validity: can same results be replicated? Are results generalizable to population? Issues: interaction between testing situation and experimental stimulus (test has an impact on experimental group and it cannot be ascertained whether the effect would be the same if duplicated); interaction of selection and treatment (what encourages involvement in experiment, how generalizable are results?)
• Time series design: be realistic about how many data points you can represent (how expensive/available is data?)
• If pretest is not possible, explain why.

Sampling strategy/method of assignment

• Rationale for choice of that sample (those subject) as exemplars
• Realistic sense of where subjects might come from A plus: discussion of recruitment or identification issues (selection bias, incentives to encourage participation, costs of incentives—monetary and introducing bias, etc.)
• Method of assignment to treatment and comparison status discussed
• What population does the study sample represent (in space, in time, in terms of types of people and types of situations)? (homogeneity = greater internal validity, heterogeneity = helps with external validity)
• What characteristics do you want represented in your sample? How are you going to ensure a size large enough to represent each characteristic?
• Sample size? Relates to power issues: “Sample sizes will be optimized in a way that ensures sufficient confidence at the 95% level.”

• Randomized selection (not the same as random assignment: you can randomly select non-equivalent groups)
• Matching (when you want to include certain characteristics in sample, and want to have the same representation in the comparison and treatment groups).
• Randomization is preferable to matching because you might not know a priori which characteristics are important to match.

Other sampling techniques (element = unit of analysis)
• Quota sampling: begin with a matrix describing the characteristics of the target population. Then collect data from people who have all the characteristics found in a given cell. All people in a given cell are then assigned an appropriate weight based on their proportion to total population. If properly weighted, then data should provide a reasonable representation of total population. Be wary of this method if goal is statistical description. Can be useful in a field research project to develop a representative sample from with a formal group. Its difficult to make quota matrix representative due to updating accurate group characteristics, bias may exist in who gets selected to represent certain cells.
• Simple random sampling: each element assigned a number, not skipping any number in process. Table of random numbers then used to select elements for the sample.
- Systematic sampling: not efficient, every kth element in the total list is chosen for inclusion in the sample, you select first element at random to avoid bias.
- Stratified sampling: modified systematic sampling to obtain a greater degree of representativeness. Appropriate numbers of elements are drawn from homogeneous subsets of population.
- Sometimes systematic sampling can be more accurate than simple random sampling. It's ok to leave lists in an order if it doesn't present aforementioned problems.

Measures

- Brief description of how constructs are operationalized through specific measures
- Recognition that operationalization can be problematic: if you operationalize a variable in a certain way, what is lost from the concept? Operationalization gets away from the concept to some extent because there is rarely a perfect match for your idea in the world. Acknowledge this threat to construct validity. (I have it in my notes that measures cannot be valid or invalid. Is this correct? Talk about problems of operationalization without talking about validity, per se.)
- What will the key outcome variables be, and how will they be measured?
- What will the key treatment variables be, and how will they be measured?
- What covariates (confounders) will you look at, and how will they be measured?
- If you use existing measures, be sure to discuss validity and reliability issues

Data source/data collection

- Realistic sense of where data might come from (no data ex machina)
- Attention to spatial and temporal issues
- Will data be from surveys, administrative data, direct observation, or other?
- Can you use existing data? Will you need to create data (avoid this if possible)?
- How will data be collected? (time, confidentiality, access)
- If you use existing administrative data, could you supplement with a survey?
- If you use survey, be clear about scales/indices and sample questions
- What are strengths of data? Weaknesses? (Administrative data is messy, but usually can be obtained. Collecting your own data can be time-consuming and expensive.)
- Be clear about potential problems with data: how good is it, how accessible?

Analysis plan

- Plan should be brief and non-technical
- (do not use statistical language if you don’t understand it)
- Talk about power
Results

- How might you interpret different results given your theory of treatment? (What results do you expect, and if the results are not what you expect, how will you interpret them? (This is important for quasi-experimental designs because effects are potentially due to be something other than the treatment.) What magnitude do you expect in your results?
- Given the findings that you expect, what might the implications be for policy-makers, etc., and for future research?

Limitations & strengths

- Measures: validity issues, particularly
- Design: threats to internal validity
- External validity issues
- A plus: how these might be mitigated in future research

- Note: integrate this discussion throughout paper; do not leave all issues to the end.

III. Strategy
Each Saturday the two weeks prior to exam:
  1. Do a practice run from 9a.m. to 2p.m. in Manhattan

Day of test
  1. Choose question and read it carefully outlining the question on paper—15 minutes
  2. Outline proposal from start to finish on paper—45 minutes
  3. Type out entire proposal from start to finish without going back too much—2 hours
  4. Go back and fill in strengths and weaknesses of certain choices—1 hour
  5. Read over carefully, fixing any errors—1 hour
- Are the essential components of your study clear in the beginning, so you don’t leave researcher wondering what’s coming until the last page?