Design and Analysis of Analytical Sample Surveys for Program Evaluation and Policy Analysis

Briefing
(Short version)

Joseph George Caldwell

18 April 2022

(*Briefing*: Microsoft PowerPoint file design-and-analysis-of-analytical-sample-surveys-briefing.pptx, .pdf
Briefing Notes: Microsoft Word file design-and-analysis-of-analytical-sample-surveys-briefing-notes.docx, .htm, .pdf
Posted at http://www.foundationwebsite.org/index12-design-of-analytical-sample-surveys.htm)

1. Context: Two Main Types of Sample Surveys, Categorized by Purpose

Descriptive surveys:

Estimate population characteristics, such as means and totals for the population and subpopulations of interest.

Statistical inference.

Analytical surveys:

Estimate parameters of models, such as the social and economic impact of a government program, or the effects of changes in government policies.

Causal inference (causal modeling and analysis).

2. A Major Problem in Analytical Survey Design: Lack of Technical References

Situation summary:

Except for texts on experimental design, most statistics texts do not address the issue of estimating *causal effects.*

There does not exist a body of literature on the subject of Analytical Survey Design

This briefing will describe a methodology developed and used by the author for design and analysis of analytical sample surveys.

This methodology builds on existing theory of experimental design, sample survey design, and causal inference.

Before proceeding, we shall review the general theory on the subject of causal inference without experimental designs.

3. Causal Inference without Experimental Designs: Must Be Based on a Causal Model

George Box once asserted (1966), “To find out what happens to a system when you interfere with it you have to interfere with it (not just passively observe it).”

Paul Holland and Donald Rubin coined the aphorism (1986), “No causation without manipulation.”

Experimental design uses randomized intervention to assess causal effects.

In the absence of randomized intervention, causal inference about a system must be based on assumptions about the causal nature of the system, i.e., on a causal model of the system.

If the causal model is reasonable, then inferences based on the model should be reasonable.

A number of causal models have been developed.

We shall now discuss some of these models.

3b. Causal Inference without Experimental Designs: Major Methodologies

Neyman-Rubin Causal Model (Potential-Outcomes Model, Counterfactuals Model). Neyman in 1920s for experimental data; Rubin in 1980s for observational data.

> *Rosenbaum-Rubin approach* (matching approach, balancing approach, “statistical” approach): estimation of Average Treatment Effect, such as in program evaluation.
>
> *James Heckman approach* (regression approach, “econometric” approach): estimation of relationship of treatment effects to policy-relevant variables.

Judea Pearl’s methodology (Structural Causal Models); specification of causal models using Bayesian networks and Directed Acyclic Graphs (DAGs)).

3c. Causal Inference without Experimental Designs: Basic Concepts

Propensity Score (PS): the probability of selection for treatment (in the case of two treatment levels).

> *Key use of the propensity score:*
>
> In groups of sample units having the same PS, the difference between means of the treated and untreated units is an unbiased estimate of the causal effect of treatment for the group.
>
> So, if we can obtain a good estimate of the PS and stratify on it, we can estimate the causal effect over the whole population.

Potential outcomes (for binary-treatment case): For each unit of the population, there are two hypothetical outcomes, corresponding to the two treatment levels (treatment and control). After the experiment, one of them is observed. The unobserved one is called a *counterfactual outcome*.

The potential-outcomes approach involves the assessment of estimability using the concepts of “ignorability” (conditional independence of outcome (response) and selection for treatment, given covariates) and tests of exogeneity based on the structure of joint probability distributions.

The structural-causal-model approach bases assessment of estimability on properties of the DAG.

3d. Causal Inference without Experimental Designs: Features of Major Approaches

Potential-Outcomes Approach (Rubin, Rosenbaum, Heckman):

- Do not specify a structural causal model. (Focus is on specific estimates.)
- Check estimability / identifiability by assessing the reasonableness of “ignorability” and exogeneity assumptions. These assessments can be extremely difficult to make, both from a substantive (subject-matter) and technical (statistical) perspective.
- The exogeneity tests are often applied during the analysis phase, not in the design phase.
- The propensity score is used in design (matching) and in the analysis.

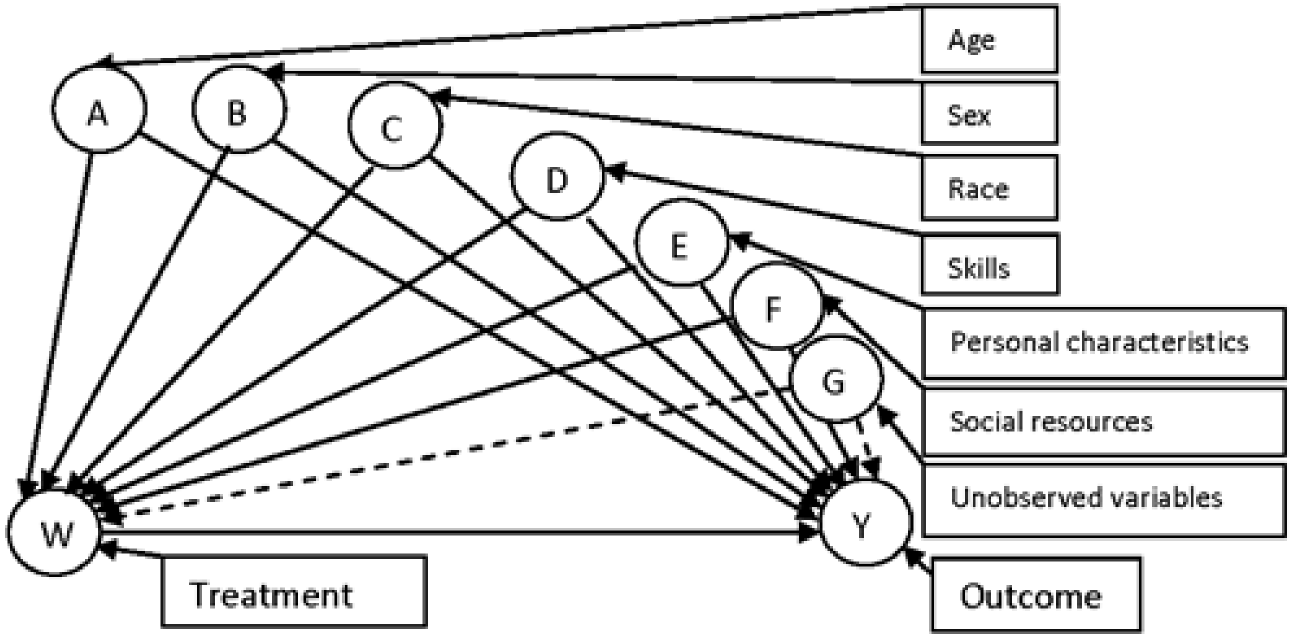
Structural-Causal-Model Approach (Pearl):

- Specify a structural causal model, represented as a Bayesian Probability Network and a Directed Acyclic Graph (DAG).
- Check estimability / identifiability from the DAG.
- The sample design is consistent with and guided by the causal model.
- The propensity score may be used in the analysis, but it is not used in design.

Which model is correct? None of them. George Box: “All models are wrong, but some are useful.”

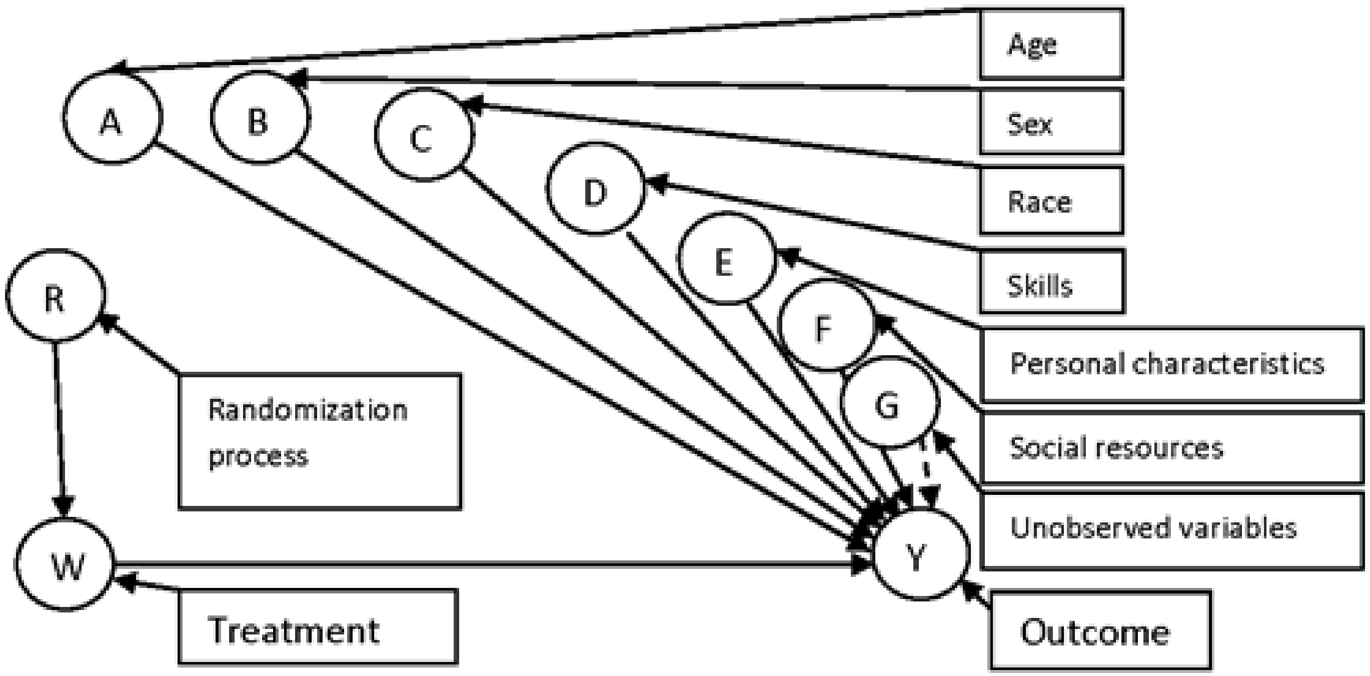
3d. Example of a Causal Model Represented as a Directed Acyclic Graph (DAG): Observational Data

Figure 9b. A causal model corresponding to observational data. (The variable names are enclosed in boxes; in this figure, a variable name may represent several variables.)



3e. Example of a Causal Model Represented as a Directed Acyclic Graph (DAG): Randomized Intervention

Figure 9c. A causal model corresponding to a designed experiment with randomized assignment to treatment. (The variable names are enclosed in boxes; in this figure, a variable name may represent several variables.)



3f. Some Points

Causal effects cannot be inferred from data alone.

To make causal inferences, must specify a causal model (data generation process).

Even for a randomized experiment, must specify the selection process (it cannot be inferred by examining a sample unit).

In this briefing, we are concerned with estimation of the *effects of causes*, not with inferring the *causes of effects*.

Effects of causes: Apply a fertilizer treatment, what is the (expected) change in yield? (Program evaluation, policy analysis.)

Causes of effects: Was a worker's lung cancer caused by smoking, genetics, or chemical fumes? (Litigation.)

4. Challenges in Applying Causal-Inference Theory to Analytical Survey Design

The objectives of Analytical Survey Design are those of experimental design (ED): high precision, low bias for estimates of interest.

Techniques of ED: randomization, replication, symmetry (orthogonality, balance), and local control.

Techniques of Analytical Survey Design:

> Where randomization cannot be used, use the causal model to assist survey design and to identify and estimate causal effects.
>
> Use matching based on causal model variables to increase the precision of causal estimates.
>
> Use design features to remove effect of unobservable variables affecting both outcome and selection for treatment. (Estimated propensity scores aren't sufficient!)
>
> To achieve orthogonality (low correlation) and balance (spread, variation) in sampling from finite populations, use marginal stratification with variable probabilities of selection.
>
> Use propensity scores in the analysis (to reduce selection bias), but not as a basis for forming matched pairs in design.

5. A Methodology for Designing Analytical Sample Surveys

Most of the published material on causal inference is concerned with *analysis*, not with *design*.

There is no standard reference text that presents a detailed or comprehensive description of procedures or general methodology for constructing analytical survey designs.

This author presents a general methodology in the paper:

> *Sample Survey Design for Evaluation (The Design of Analytical Surveys)* posted at Internet website http://www.foundationwebsite.org/SampleSurveyDesignForEvaluation.htm.
>
> Additional material is presented in lecture notes for the courses:
> *Causal Inference and Matching*, at
> http://www.foundationwebsite.org/StatCourse4and5CausalInferenceAndMatching.htm; and
>
> *Statistical Design and Analysis for Evaluation*, at
> http://www.foundationwebsite.org/StatCourse6and7StatisticalDesignAndAnalysisForEvaluation2DayCourse.htm.

That methodology will now be summarized. It includes elements of all major approaches to causal inference, experimental design, and sample survey design.

6. Summary of Procedures for Designing Analytical Sample Surveys

1. Construct a comprehensive structural causal model for the process under investigation.
   Represent it as a DAG.
   Classify variables as observable and unobservable.
   Construct a survey design such that unobservable variables will drop out of estimates of interest (e.g., interviewing the same subjects in successive survey rounds of a panel survey if selection is associated with personal characteristics).

2. Identify causal effects of interest, and minimal detectable effect sizes for each (i.e., effect sizes that are to be detectable with high probability).

3. Use statistical power analysis to determine sample sizes for the survey design.
   Allow for nonresponse.

4. A computer program for determining sample sizes for evaluation designs (e.g., pretest-posttest-comparison-group design) is posted at http://www.foundationwebsite.org/SampleSizeEstimationProgram.htm.
   Summary information about the program is posted at http://www.foundationwebsite.org/SampleSizeEstimationAnalyticalSurveysGeneric.htm.
   Lecture notes on a course in determination of sample size for evaluation surveys are posted at http://www.foundationwebsite.org/StatCourse8SampleSizeDetermination.htm.

6b. Summary of Procedures for Designing Analytical Sample Surveys (Cont'd.)

5. Identify variables that are causally related to output variables of interest, and for which data are available prior to the survey data collection (i.e., that can be used for design).

    Do this for each stage of sampling.

6. Define strata for these variables.

    The stratification for each variable is a *marginal stratification*, not a cross-stratification or nested stratification.

    Cross-stratification (such as Kish's controlled selection) and nested stratification are not feasible since, for 5-10 variables, there would be a very large number of stratum cells, leading to few or one or no population items in many cells (leading to large sample sizes and highly variable selection probabilities).

7. Select from this set of variables a subset having low correlations.

    As a measure of association, use the Cramér phi (φ_c, V) correlation coefficient, applied to the stratum cells.

    This set typically contains 5-10 variables.

8. For each variable, allocate sample units to the stratum cells in such a way as to achieve a high degree of variation.

6c. Summary of Procedures for Designing Analytical Surveys (Cont'd.)

9. Determine selection probabilities for each sample unit to achieve the desired marginal stratifications.

    Keep variation in probabilities as low as possible.

    If the survey is to produce descriptive estimates as well as analytical estimates, it may be desirable to place a "floor" on how small the unit selection probabilities may be.

10. If matching is used to construct matched pairs, then base matching on a distance measure that takes into account the relative importance of each variable of stratification on output measures of interest.

    Use strata that are sufficiently "coarse" that there are lots of reasonable match candidates.

    Note: The use of importance weights in the matching distance function increases the precision of causal estimates and does not introduce bias.

    Do not form matched pairs using propensity-score matching (see *Briefing Notes* and King/Nielsen article for discussion).

11. If a "treatment" sample has not yet been selected, use matching to define matched pairs, select the pairs with probabilities such that the marginal-stratification sample allocations are reasonable, and randomly allocate one member of each pair to treatment and one to control.

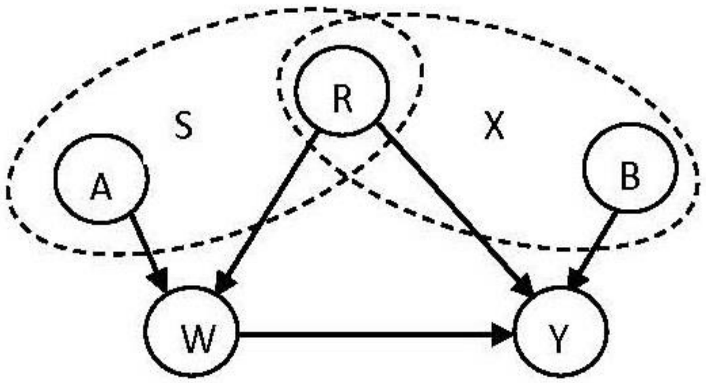
12. If a treatment sample has already been selected, use matching to define matched pairs.

6d. Summary of Procedures for Designing Analytical Surveys (Cont'd.)

13. In the analysis, to obtain consistent estimates of causal effects, we must condition on (average on) either: (1) all variables affecting output; or (2) all variables affecting selection; or (3) all variables affecting both output and selection.

Make sure that such variables, if observable, are reflected in the variables of stratification.

Figure 11c. impact can be estimated by conditioning over all variables (R) affecting both selection for treatment (W) and outcome (Y); all variables (S) affecting selection for treatment (W); or all variables (X) affecting outcome (Y). Conditioning on R is necessary and sufficient; conditioning on S or X is sufficient, but not necessary (if there exist variables that affect just one of W and X, but not both).



6e. Summary of Procedures for Designing Analytical Surveys (Cont’d.)

14. For unobserved variables (e.g., farmer characteristics that might affect selection for treatment), configure the survey design so that these variable “drop out” of difference estimates.

    All causal variables involved in estimation of a causal effect must be conditioned on or “drop out.”

15. Explicitly describe the inferential scope of the study.

    For example, if selection for treatment is random and countrywide, the scope of inference will be the causal effect of the project / program intervention relative to the entire country.

    If a treatment group has already been selected (e.g., by political means) prior to the sample design and selection, then the scope of inference will be the causal effect of that particular already-selected project.

7. Examples of Analytical Survey Designs Constructed Using the Method Described Above

Impact Evaluation of the Farmer Training and Development Activity in Honduras, Millennium Challenge Corporation. Project final report at http://www.foundationwebsite.org/MCCFTDAEvaluationFinalReportRevisedNov15-2013.htm.

Honduras Road Transportation Improvement Project, Millennium Challenge Corporation. Project final report at http://www.foundationwebsite.org/MCCTransportationProjectEvaluationFinalReportRevisedDec12-2013.htm.

Impact Evaluation of the Competitive African Cotton for Pro-Poor Growth Program ("COMPACI", "Cotton Made in Africa"), Deutsche Investitions und Entwicklungsgesellschaft GmbH (DEG), in six African countries: Benin, Burkina Faso, Côte d'Ivoire, Zambia, Ghana and Malaŵi. (Separate surveys in each country.)

Monitoring and Evaluation of the Competitive African Cashew Value Chains for Pro-Poor Growth Program", Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, in five African countries: Benin, Burkina Faso, Côte d'Ivoire, Ghana and Mozambique. (Separate surveys in each country.)

7b. Examples of Analytical Survey Designs Constructed Using the Method Described Above (Cont'd.)

Impact Evaluation of the Programme of Advancement through Health and Education (PATH), Jamaica. Government of Jamaica.

Evaluation des performances et de l'impact de l'activité de rehabilitation et d'intensification des plantations d'oliviers au niveau des zones pluviales," Agence du Partenariat pour le Progrès, Millennium Challenge Account – Maroc, Project Arboriculture Fruitière.

Impact Evaluation of Agricultural Development Projects in the Sourou Valley and Comoé Basin, Millennium Challenge Account – Burkina Faso.

Impact Evaluation of Conservancy Support and Indigenous Natural Products, Millennium Challenge Account – Namibia.

Impact Evaluation of Ghana Water Supply Activity, Millennium Development Authority – Ghana.

Impact Evaluation of Feeder Roads Activity, Millennium Development Authority – Ghana.

8. Software for Constructing Analytical Sample Designs

Software for implementing the preceding methodology for designing analytical surveys is posted at http://www.foundationwebsite.org/index12-design-of-analytical-sample-surveys.htm.

The software (SurvDes) is a Microsoft Access program that is tailored to each application.

An example of output from the program is presented in the *Briefing Notes*, but not in this *Briefing.*

This example draws from the survey design constructed for the Impact Evaluation of the COMPACI (Cotton Made in Africa) Benin Project.