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

Briefing

Joseph George Caldwell

17 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)

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.

Associational inference: how variables are *probabilistically related*.

Probability models, statistical models and 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: how variables are *causally related*.

Causal models, probability models, statistical models and statistical inference.

2. Two Main Types of Sample-Survey Inference, Categorized by Dependence on the Survey Sampling Plan: Design-Based Inference and Model-Based Inference

Design-based inference: “Direct” estimates, based on survey data and sampling plan (sampling units, selection methods and probabilities) (most sample surveys).

With or without auxiliary data (e.g., regression and ratio estimates).
Precision depends mainly on the domain sample size.
Small domains: Small or zero sample sizes; unacceptably large sampling errors.

Model-based inference: “Indirect” estimates, based on the survey data, sampling plan, *and a data generation model*.

Synthetic estimates. Use a direct estimate from a large area to make an estimate for a small area, under the assumption that the large and small areas have the same characteristics.

Standard statistical models that allow for random between-area variation.

Composite estimates (weighted average of a direct estimate and an indirect estimate).

3. Two Classes of Model-Based Inference: Model-Assisted Inference and Model-Dependent Inference

(1) *Model-assisted inference*: primary goal is still estimation of population or subpopulation characteristics (descriptive), and the estimates are based to some extent on the survey sampling plan.

Treatment of nonresponse (unit or item nonresponse)
Small-area statistics (can be handled like a missing-data problem)

Examples of model-assisted inference:

Generalized Regression (GREG) estimation (uses a regression-type model, but with error terms based on survey design)
“Fay-Herriot”-type models (“mixed” models that includes both model errors and design-induced errors from survey sampling)
Latent-variable models, such as tobit models (censoring, truncation, nonresponse)
Bayesian methods (Markov Chain Monte Carlo (MCMC) (Gibbs sampling, Metropolis-Hastings algorithm; Expectation-Maximization (EM) algorithm)

References: Rao, J. N. K., *Small Area Statistics* (Wiley, 2003) and Särndel, Carl-Erik, Bengt Swensson and Jan Wretman, *Model Assisted Survey Sampling* (Springer, 1992)

3b. Two Classes of Model-Based Inference : Model-Assisted Inference and Model-Dependent Inference (Cont'd.)

(2) *Model-dependent inference:* primary goal is estimation of parameters of a process considered to have generated the population data (i.e., of an underlying causal model). Estimates are based primarily on the survey data and a data-generation model, not on the survey sampling plan.

Types of Designs:

Experimental designs (randomization (for selection of sample units and assignment to treatment), replication, symmetry and local control)
"Broken" experimental designs and q*uasi-experimental designs* (some features of experimental design are compromised)
Observational studies (randomization not used to select sample units or assign treatment)
Analytical Survey Designs (panel surveys; pretest-posttest-with-comparison-group)

Types of Analysis:

Latent-variable models (tobit nodels)
General linear statistical model (analysis of variance and covariance, regression models)
Generalized linear model (link functions; e.g., logistic regression)
Reference: Jeffrey M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., (The MIT Press, 2010)

This briefing is concerned with model-dependent inference applied to Analytical Survey Design.

4. A Significant Problem Associated with Causal Inference: Causation Cannot Be Inferred from Data Alone

To estimate causal effects, must *specify* a causal model and base inferences on this model.

This is done in some applications (experimental design, econometric modeling, time-series forecasting) but rarely for other types of investigations (sample surveys, analysis of observational data, data analytics, data mining).

For descriptive surveys, this is not an issue (since it is not the goal to estimate causal effects).

For analytical surveys, it is a major concern, since it is the goal of these surveys is to estimate causal effects but it is generally not feasible to implement sample surveys as experimental designs.

Most statistical methods simply measure probabilistic associations among variables, not causal effects. (Except for books on experimental design, the word “causal” rarely appears in statistics texts.)

Issue: How to make causal inferences from sample surveys when an experimental design is not used?

5. Standard Statistics Texts Do Not Address the Subject of Causal Inference for Sample Surveys

Texts on Sample Survey Design and Analysis Do Not Address the Subject

Cochran, William G., *Sampling Techniques*, 3rd ed. (Wiley, 1977)

Lohr, Sharon L., *Sampling: Design and Analysis* (Duxbury, 1999)

Särndal, C-E., B. Swensson and J. Wretman, *Model Assisted Survey Sampling* (Springer, 1992)

Texts on Causal Inference Do Not Address the Subject in the Context of Sample Survey Design

Pearl, Judea, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Camb. Univ. Press, 2009)

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction* (Cambridge University Press, 2015)

Morgan, Stephen L., Christopher Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd ed. (Cambridge University Press, 2015)

Texts on Econometrics Address the Subject, but Only for the Model-Dependent Case (not for survey design, and not for the model-assisted case), e.g., Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data,* 2nd ed. (The MIT Press, 2010)

Texts on Analysis of Observational Data Do Not Address the Subject, e.g., Rosenbaum, Paul R., *Design of Observational Studies* (Springer, 2010)

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

Situation summary:

The science of statistics is focused on estimation of the strength of *probabilistic associations* between variables.

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.

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

7. 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.”

Randomized assignment of treatment enables causal inference by assuring that the probability distribution of all variables is the same for treatment and control groups (i.e., response (outcome) is independent of all variables except treatment).

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. All of them involve the concept of conditional independence of treatment assignment (selection) and treatment response, given covariates, but they differ in other respects.

We shall now discuss some of these models.

7b. Causal Inference without Experimental Designs: Major Methodologies

Neyman-Rubin Causal Model (potential outcomes, counterfactuals). Reference: Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, Vol. 81, No. 396 (Dec. 1986), pp 945-960.)

> *Rosenbaum-Rubin approach* (matching approach, balancing approach, "statistical" approach). Reference: Rosenbaum, Paul R. and Donald B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika* (1983), vol. 70, no. 1, pp. 41-55.
>
> *James Heckman approach* (regression approach, "econometric" approach). Reference: Heckman, James J. and Edward J. Vytlacil, "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation" in *Handbook of Econometrics* Volume 6B, eds. James J. Heckman and Edward E. Leamer (North-Holland / Elsevier 2007).

Judea Pearl's methodology (structural causal models; specification of causal models using Bayesian networks and Directed Acyclic Graphs (DAGs)).
Reference: Pearl, Judea, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, 2009).

7c. Causal Inference without Experimental Designs: Some Definitions

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

Key use: 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 stratify on the PS, we can obtain an estimate of the causal effect over the whole population. Key assumption: $0 < PS < 1$ for all units (every unit must have a positive probability of assignment to treatment or control).

Potential outcomes (for binary 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*.

Potential-outcome models have been subject to criticism, e.g., Dawid, A. Philip, "Causal Inference without Counterfactuals," *Journal of the American Statistical Association*, June 2000, 95, 450, pp. 407-448 / Comments by D. R. Cox et al. / Rejoinder.

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

7d. Causal Inference without Experimental Designs: Matching

Causal inference makes much use of matching, both in design and in analysis, both to increase precision and decrease selection bias.

Matching in analysis: There are a variety of different methods of matching. They are described in articles posted on Professor Gary King's website (http://gking.harvard.edu), including:

"Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," by Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, *Political Analysis*, Vol. 15 (2007), pp. 199-236, posted at http://gking.harvard.edu/files/matchp.pdf or http://gking.harvard.edu/files/abs/matchp-abs.shtml; and

"MatchIt: Nonparametric Preprocessing for Parametric Causal Inference," by Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart (July 9, 2007), posted at http://gking.harvard.edu/matchit/docs/matchit.pdf.

Matching in design: For a discussion of the drawbacks of propensity-score matching, see the article:

"Why Propensity Scores Should Not Be Used for Matching," by Gary King and Richard Nielsen, *Political Analysis* (Volume 27, Issue 4, October 2019, pp. 435 – 454) posted at https://gking.harvard.edu/files/gking/files/pan1900011_rev.pdf and supplementary material

7e. Causal Inference without Experimental Designs: Features of Major Approaches

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

- Do not specify a comprehensive causal model. (Focus is on specific estimates.)
- Check estimability / identifiability by applying tests for exogeneity (and other tests, such as rank and order tests). These exogeneity tests can be extremely difficult to apply, both from a substantive (subject-matter) and technical (statistical) perspective.
- The sample design proceeds ignorant of a comprehensive causal model.
- The exogeneity tests are often applied during the analysis phase, not in the design phase.
- Matched pairs are often formed using propensity-score matching (a terrible approach!).

Structural Causal Model Approach (Pearl):

- Specify a comprehensive 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.
- Matched pairs are formed by taking all causal variables of the causal model into account, not just matching on a single variable (i.e., on the propensity score).

7f. Causal Inference without Experimental Designs: Pros and Cons of Alternative Approaches

Rosenbaum-Rubin: Focuses on identifying groups for which the distributions of all variables except treatment are the same (conditional independence, “ignorability” of treatment); propensity score.
Good for estimating the average effect of treatment, if all variables associated with selection for treatment are observable.
No procedures for assessing estimability – just conditions (how to test?).
Recursive model (no mutual causation or “loops”).
Better for program evaluation.

Heckman: Focuses on investigation of the relationship of causal variables, not just overall effects.
Emphasizes use of design to remove influence of unobserved variables on selection for treatment.
Specifies exogeneity conditions for estimability.
No general methodology for assessing estimability.
Allows for mutually causal variables (nonrecursive, or simultaneous, causal models).
Better for policy analysis.

Pearl: Focuses on specification of a comprehensive *structural causal model* in the form of a Directed Acyclic Graph (DAG), and identifies practical tests for estimability using graphical techniques.
Greater face validity: Based on observable conditional distributions, not unobservable counterfactuals.
Recursive model.
Equally useful for program evaluation or policy analysis.

7g. Causal Inference without EDs: Assessment of Exogeneity without Graphs

1.5.4. Exogeneity

While our primary focus is on integrated series and the problems they imply for standard econometric analyses, rather than on the problems created by a failure of exogeneity (in the appropriate sense), it will be important to consider exogeneity at several points.

Econometric analysis often proceeds on the basis of a single-equation model of a process of interest. Implicitly, we assume that knowledge of the processes generating the explanatory variables would carry no information relevant to the parameters of interest. As Engle, Hendry,

and Richard (1983) indicate, concepts of exogeneity relate to the circumstances in which this assumption is valid. Rather than refer to particular variables as exogenous in general, Engle *et al.* refer to a variable as exogenous *with respect to a particular parameter* if knowledge of the process generating the exogenous variable contains no information about that parameter.

The three different concepts introduced by Engle *et al.* are called weak, strong, and super exogeneity and correspond to three different ways in which a parameter estimate may be used: inference, forecasting conditional on forecasts of the exogenous variables, and policy analysis. These different uses require that different conditions must be met for exogeneity to hold. These conditions can be examined with the following definitions.

Let $\mathbf{x}_t = (y_t, z_t)'$ be generated by the process with conditional density function $D(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\lambda})$, where $\mathbf{X}_{t-1}$ denotes the history of the variable $\mathbf{x}$: $\mathbf{X}_{t-1} = (\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{X}_0)$. Let the parameters $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ be partitioned into $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ to support the factorization

$$D(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\lambda}) = D(y_t|z_t, \mathbf{X}_{t-1}, \boldsymbol{\lambda}_1)D(z_t|\mathbf{X}_{t-1}, \boldsymbol{\lambda}_2).$$

Then $[(y_t|z_t; \boldsymbol{\lambda}_1), (z_t; \boldsymbol{\lambda}_2)]$ operates a *sequential cut* on $D(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\lambda})$ if and only if $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are *variation free*; that is, if and only if

$$(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \in \boldsymbol{\Lambda}_1 \times \boldsymbol{\Lambda}_2, \qquad \text{where } \boldsymbol{\lambda}_1 \in \boldsymbol{\Lambda}_1 \text{ and } \boldsymbol{\lambda}_2 \in \boldsymbol{\Lambda}_2,$$

so that the parameter space $\boldsymbol{\Lambda}$ is the direct product of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$. In other words, for any values of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, admissible values of the parameters $\boldsymbol{\lambda}$ of the joint distribution can be recovered. The essential element of weak exogeneity is that the marginal distribution contains no information relevant to $\boldsymbol{\lambda}_1$ (for an exposition, see Ericsson 1992).

Weak exogeneity: z_t is weakly exogenous for a set of parameters of interest $\boldsymbol{\psi}$ if and only if there exists a partition $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ of $\boldsymbol{\lambda}$ such that (i) $\boldsymbol{\psi}$ is a function of $\boldsymbol{\lambda}_1$ alone, and (ii) $[(y_t|z_t; \boldsymbol{\lambda}_1), (z_t; \boldsymbol{\lambda}_2)]$ operates a sequential cut.

Strong exogeneity: z_t is strongly exogenous for $\boldsymbol{\psi}$ if and only if z_t is weakly exogenous for $\boldsymbol{\psi}$ and

$$D(z_t|\mathbf{X}_{t-1}, \boldsymbol{\lambda}_2) = D(z_t|\mathbf{Z}_{t-1}, \mathbf{Y}_0, \boldsymbol{\lambda}_2),$$

so that y does not Granger-cause z.

Super exogeneity: z_t is super exogenous for $\boldsymbol{\psi}$ if and only if z_t is weakly exogenous for $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}_1$ is invariant to interventions affecting $\boldsymbol{\lambda}_2$.

Weak exogeneity ensures that there is no loss of information about parameters of interest from analysing only the conditional distribution; a variable z_t is weakly exogenous for a set of parameters $\boldsymbol{\psi}$ if inference concerning $\boldsymbol{\psi}$ can be made conditional on z_t with no loss of information relative to that which could be obtained using the joint density of y_t and

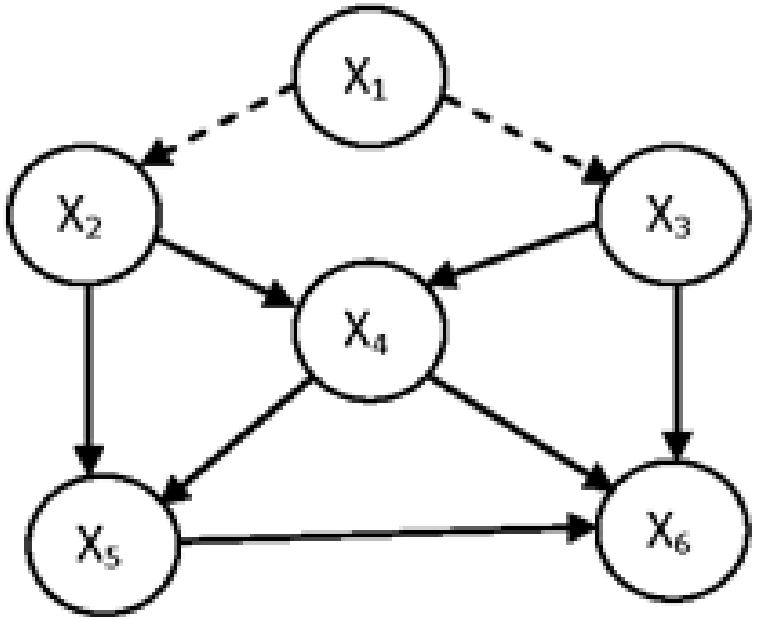
z_t. Strong exogeneity is necessary for multi-step forecasting which proceeds by forecasting future zs and then forecasting ys conditional on those zs. Super exogeneity sustains policy analysis on $\boldsymbol{\lambda}_1$ when the marginal distribution of z_t is altered.

Engle *et al.* contrast these three types of exogeneity with the traditional concepts of *strict exogeneity* and *pre-determinedness*. If u_t is the error term in a model, then z_t is said to be strictly exogenous if $E[z_t u_{t+i}] = 0 \; \forall \, i$, whereas z_t is said to be predetermined if $E[z_t u_{t+i}] = 0$ $\forall \, i \geq 0$. Engle *et al.* show that the latter concepts are neither necessary nor sufficient for valid inference since neither relates to parameters of interest.

Source: Banerjee, Aninda, Juan Dolado, John W. Galbraith, and David F. Hendry, *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, 1993.

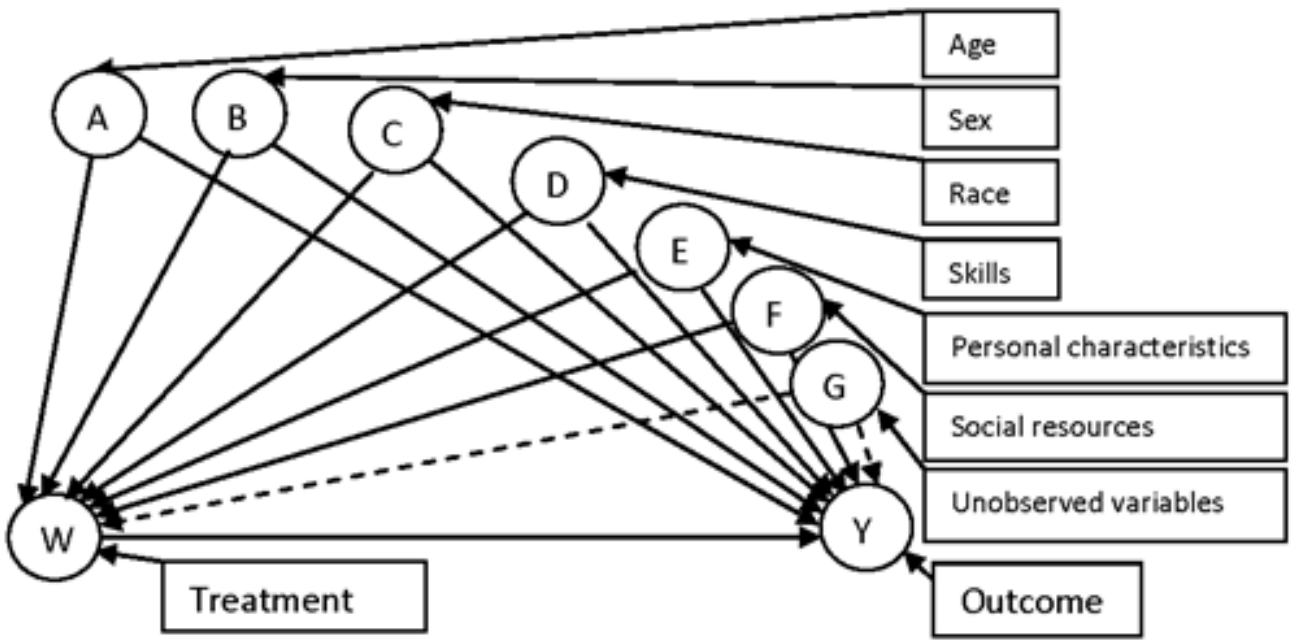
8. Example of a Causal Model Represented as a Directed Acyclic Graph

Figure 9a. Example of a "generic" causal diagram



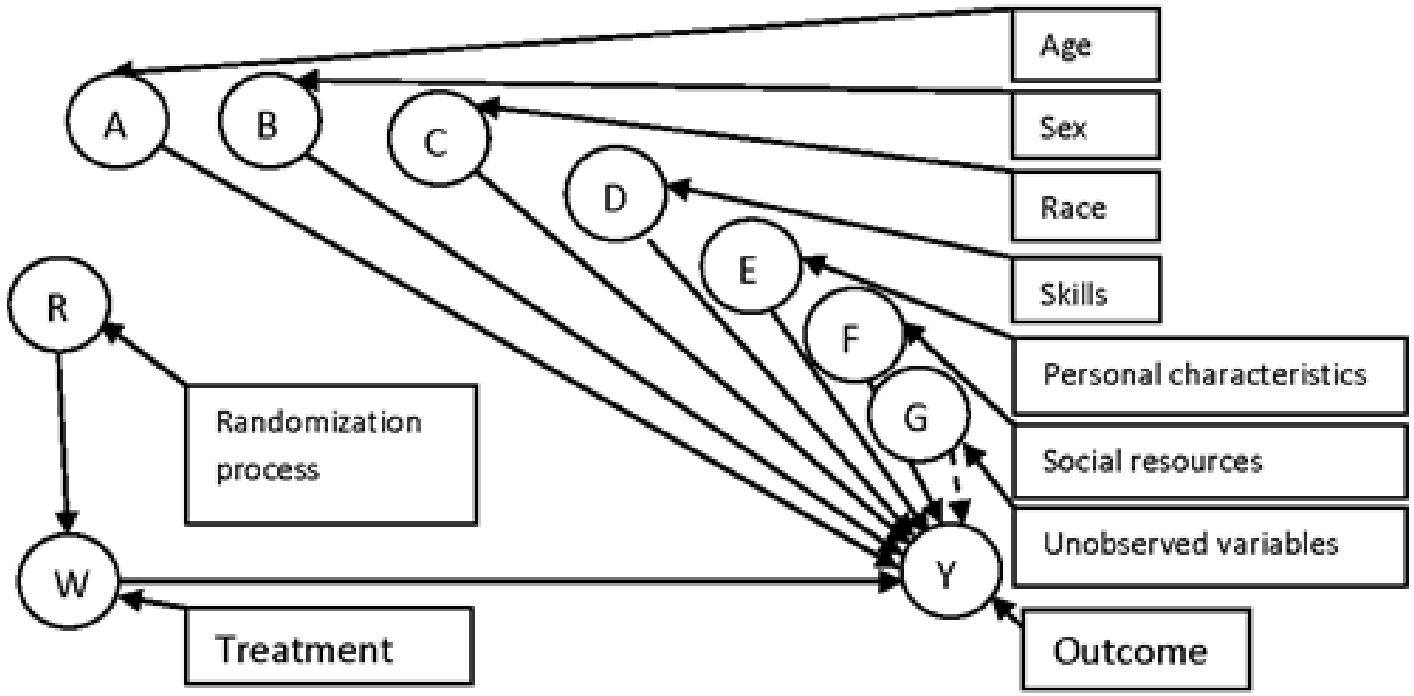
8b. Example of a Causal Model Represented as a Directed Acyclic Graph

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.)



8c. Example of a Causal Model Represented as a Directed Acyclic Graph

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.)



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

Objectives are similar to those of experimental design (randomization, replication, symmetry (orthogonality, balance), and local control.

Where randomization cannot be used, use a causal model to identify and estimate causal effects, and base matching on the causal model variables (for local control).

Use propensity scores in analysis, but not as a basis for forming matched pairs in design.

Use design features to remove effect of unobservable variables affecting both selection for treatment and outcome.

To achieve orthogonality (low correlation) and balance (spread, variation) in sampling from finite populations, use marginal stratification with variable probabilities of selection.

10. 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, of experimental design, and of sample survey design.

11. Summary of Procedures for Designing Analytical Sample Surveys

1. Construct a comprehensive 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.

11b. 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.

11c. 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.

11d. 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.

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.

12. Tips on the Analysis

1. Make certain that the analysis is consistent with the design.

   Analytical surveys may be designed to enable production of descriptive estimates as well as causal estimates. Since the person doing the analysis may not be the person who constructed the design, it is very important to verify that the analysis procedures are correct for the design. A common error in analysis of data sets involving matched pairs is to analyze the sample as if the design involved matched comparison groups, not matched pairs.

2. Check the estimability of each effect of interest relative to the causal model.

   Assessment of estimability is readily accomplished from a DAG (back-door criterion and the front-door criterion).

12b. Tips on the Analysis (Cont'd.)

Figure 11a. A causal diagram for which impact (effect of W on Y) can be estimated by conditioning on (averaging over) either all variables (S) affecting selection for treatment (W) or all variables (X) affecting outcome (Y). (All variables in S or all variables in X are observed. S and X may contain common variables. Dashed line represents unobserved variables (U) that affect X and S.)

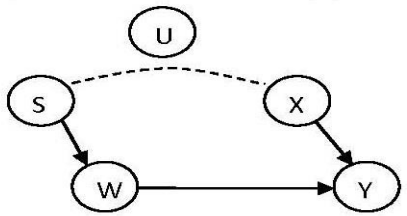


Figure 11b. A causal diagram for which impact *cannot* be estimated, without special assumptions about U. (Dashed line represents unobserved variables (U) that affect both selection for treatment (W) and outcome (Y). S is all variables affecting selection for treatment, other than U. X is all variables affecting outcome, other than U. It is assumed that all variables in S or all variables in X are observed.)

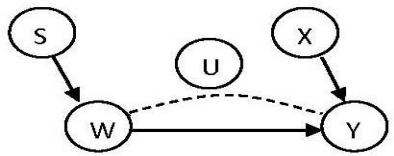
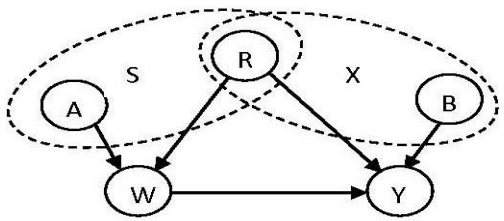


Figure 11c. impact can be estimated by conditioning over all variables (R) affecting both selection for treatment (W) and outcome (Y); all variables (S) affecting selection for treatment (W); or all variables (X) affecting outcome (Y). Conditioning on R is necessary and sufficient; conditioning on S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



12c. Tips on the Analysis (Cont'd.)

3. Determine what variables must be conditioned on from the causal model, to obtain estimates of causal effects.

> This is easily done from the DAG. For example, for estimating a treatment effect, condition on variables in such a way as to "block" all paths connecting outcome and treatment.

4. Use propensity scores to adjust for selection bias, using stratification, inverse weighting, or regression.

> Selection bias introduced by unobserved variables is handled in the design, so that the unobserved variables drop out of difference estimates.

5. When using fixed-effects estimators (such as single-difference or double-difference estimators), conduct a Hausman specification test.

> A problem with the fixed-effects (difference) estimator is that it is less efficient than the random-effects estimator. If unobserved variables are present that may be correlated with model explanatory variables, the random-effects estimator may be inconsistent, whereas the fixed-effects estimator is consistent. If the Hausman test passes (no significant difference between the estimates), the more efficient estimator, which is the random-effects estimator, may be used.

12d. Tips on the Analysis (Cont'd.)

6. Use resampling (bootstrapping) methods to estimate variances and significance levels.

    The parameter estimates are complex, and closed-form expressions will not be available for variances.

7. Consider multiple estimators ("statistical," or "balancing" estimates, such as stratification, inverse weighting, or regression using propensity scores; "econometric" estimates, such as Heckman-type models).

    Do not, however, present multiple estimates for the same parameter / effect in the final report. The Rosenbaum-Rubin "statistical" (matching, balancing) approach is used if all that is desired is to estimate an average causal effect; the Heckman "econometric" approach is used if it is desired to quantify the relationship of causal effects to policy variables.

8. Conduct a detailed *ex-post* statistical power analysis for all causal effects of interest.

9. Assess crucial assumptions, such as the "overlap" condition (that the probabilities of selection of treatment and comparison units are not equal to zero or one), and the stable-unit-treatment-value assumption (SUTVA) that responses conditional on explanatory variables are uncorrelated.

13. 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.)

13b. 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.

14. Example of Output from Software for Constructing Analytical Sample Designs

Here follows an example of output from the SurvDes program posted at http://www.foundationwebsite.org/index12-design-of-analytical-sample-surveys.htm. The SurvDes program is a Microsoft Access program that must be modified for each application.

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

[The example is included in the *Briefing Notes*, but not in the *Briefing*.]

All slides that follow are Supplemental Material, not part of the briefing.

15. Features of Descriptive Sample Surveys

Survey Purpose / Goal

To estimate overall features (means, totals) of the population being surveyed.

Not to make inferences about the process that generated the surveyed population.

Not to assess causal impact of certain variables on other variables.

To measure probabilistic associations, not causal relationships.

15b. Features of Descriptive Sample Surveys (Cont'd.)

Design Features

Design goal: To achieve a desired level of precision for estimates of interest.

Design features: clustering, multistage sampling, stratification; sampling with and without replacement.

Sample-size determination based on precision analysis, not power analysis.

Finite population correction (FPC) applies to variance estimates for non-replacement sampling.

Optimal survey design guided by Neyman allocation (allocate more sample units to strata where variability is higher and cost of sampling is lower).

Selection probabilities are generally constant within strata. Stratification is cross-stratification, not marginal stratification. There are few variables of stratification (e.g., controlled selection for two variables).

15c. Features of Descriptive Sample Surveys (Cont'd.)

Analysis Features

Estimates are design-based, design-unbiased, design-consistent.

Data analysis is straightforward, often using closed-form expressions for estimation of means and variances.

Estimates for a particular population subgroup are "direct" estimates, based only on the selection probabilities and sample data (including auxiliary data) for that subgroup.

There may be some use of model-assisted analysis, as in treatment of nonresponse or in production of small-area statistics.

Tests of hypotheses are irrelevant for comparing finite populations (any two finite-population groups will almost always have different means and totals).

16. Features of Analytical Sample Surveys

Survey Purpose / Goal

Design goal: To obtain information for program evaluation and policy analysis.

More specifically, to estimate the causal effects of certain variables (input variables, control variables) on other variables (output variables).

Estimation focus is on estimation of features of the process that generated the population data.

Focus is not on estimation of overall features of the population.

Examples:

Effects of a farmer-training program.

Effects of a change in a farm subsidy program; effects of changes in recruitment policies.

16b. Features of Analytical Sample Surveys (Cont'd.)

Relationship of Analytical Survey Design to Experimental Design:

The standard approach to estimation of impact is experimental design

Key features of experimental design (ED) are randomization, replication, local control (blocking, control groups, matched pairs) and symmetry (balance, orthogonality).

Incorporation of the features of ED into analytical sample surveys is difficult to do:

Randomization may not be possible for a variety of reasons, such as program eligibility, legal constraints, ethical constraints, or self-selection

Balance and symmetry also difficult (physical constraints (few or no population elements for certain combinations of variables)).

Lacking the methodology of ED, it is necessary to adopt a different approach to causal inference in sample-survey applications. This is the purview of causal inference (modeling and analysis).

16c. Features of Analytical Sample Surveys (Cont'd.)

Design Features:

Overall similarity to experimental design (e.g., randomization (to the extent possible), replication, symmetry, local control (control groups, comparison groups, matched pairs)

Incorporation of all aspects of descriptive survey design (clustering, multistage sampling, stratification)

Comparison groups are formed by matching, if randomization is not feasible. Matched pairs may be formed after treatment has already been assigned. The matching is based on a causal model.

Design features are incorporated that cause unobservable variables affecting selection and assignment to treatment to drop out (e.g., inclusion of the same respondents in successive survey rounds).

16d. Features of Analytical Sample Surveys (Cont'd.)

Analysis Features

Assess estimability / identifiability of estimates of causal effects relative to a causal model.

Reduce selection bias by incorporating the estimated propensity score into the analysis (by stratification, inverse weighting and regression).

Eliminate selection bias associated with unobserved variables (e.g., respondent characteristics) by using a design and difference estimators that cause these variables to drop out.

Consider both fixed-effects and random-effects estimators, use Hausman test to determine which.

17. Differences between Descriptive and Analytical Surveys

Differences in Purpose

Descriptive surveys are concerned with measurement of observed means, totals and correlations.

Analytical surveys are concerned with measurement of causal effects.

An observed treatment effect (OTE) may be quite different from an average treatment effect (ATE).

17b. Differences between Descriptive and Analytical Surveys (Cont'd.)

Differences in Design

For descriptive surveys, generally avoid correlation among sample units (e.g., low intracluster correlation, replacement of panel respondents in longitudinal surveys).

For analytical surveys, generally *introduce* correlations, such as in using the same respondents in successive survey panels, to increase the precision of differences (and regression coefficients).

For descriptive surveys, the FPC enables the use of smaller samples to achieve a specified level of precision. For analytical surveys, the FPC is not relevant.

For descriptive surveys, keep unit selection probabilities relatively uniform (for high precision of means and totals). For analytical surveys, design requirements will generally impose substantial variation in the selection probabilities.

Construct the design to remove effects of unobserved variables that might be correlated with model explanatory variables (such as selection for treatment) (e.g., by including the same respondents in both survey waves, and using a difference estimator).

17c. Differences between Descriptive and Analytical Surveys (Cont'd.)

Differences in Analysis

For descriptive surveys, standard statistical-analysis software (e.g., SAS, Stata, SPSS) can be used to quickly produce design-based estimates (means and totals) for survey data.

For analytical surveys, much custom-tailored work is required to construct model-based estimates (econometric modeling).

- Multiple estimation procedures (e.g., maximum likelihood, ordinary least squares, indirect least-squares, two-stage least squares, instrumental variables)
- Most model variables must be considered random effects, not fixed effects.
- Estimability / identifiability must be determined not just relative to a statistical (associational) model, but relative to a causal model.
- Model specification tests, such as the Hausman test
- Closed-form expressions are generally not available for variance estimates. Resampling methods ("bootstrapping") are used to estimate variances and significance levels.
- Propensity-score models are used to reduce selection bias (stratification, inverse weighting and regression using propensity scores).