Using Nonequivalent Dependent Variables to Reduce Internal Validity Threats in Quasi-Experiments: Rationale, History, and Examples From Practice

Chris L.S. Coryn, Kristin A. Hobson

Abstract

Threats to the validity of inferences and conclusions regarding the effects of applied interventions have been a major dilemma for social scientists and evaluators for several decades. One mechanism for reducing threats to internal validity and improving warrants for cause-and-effect conclusions in nonrandomized investigations and evaluations is the inclusion of nonequivalent dependent variables as an element of structural design. In this chapter, the rationale for, history of, and examples from practice for using nonequivalent dependent variables to reduce internal validity threats, as well as some warrants supporting their increased use, are described. © Wiley Periodicals, Inc., and the American Evaluation Association.

Legitimate knowledge claims about causation have been a central concern among evaluators and applied researchers for several decades (Cook, Scriven, Coryn, & Evergreen, 2010). Since publication of Campbell and Stanley’s (1966) Experimental and Quasi-Experimental Designs for Research, which was followed more than a decade later by Cook and Campbell’s (1979) Quasi-Experimentation: Design and Analysis for Field Settings,
and thereafter by Shadish, Cook, and Campbell’s (2002) Experimental and Quasi-Experimental Designs for Generalized Causal Inference, alternative explanations for effects observed from studies of applied interventions have been the bane of practicing social scientists and evaluators. Collectively, these alternative explanations are generally known as threats to validity.

Threats to validity are (plausible) “reasons why an inference might be incorrect” (Shadish et al., 2002, p. 512). Such threats may take one or more forms, including those associated with internal, external, construct, and statistical conclusion validities. Although neither exhaustive nor mutually exclusive, most threats to validity can be situated into one or more of these four general categories. Contingent on the target of generalization, these validity threats also may manifest themselves in very different, sometimes conflicting ways, where efforts to thwart threats to one form of validity may simultaneously diminish the validity of another. In this taxonomy, internal validity is the approximate truthfulness or correctness of an inference or conclusion regarding whether a relationship between two variables is, in fact, causal. These types of validity are interdependent and can never be fully known (and only estimated based on logic and principled reasoning). Validity, therefore, is not an all-or-none proposition. Rather, it is a matter of degree and always associated with a particular purpose or use (Messick, 1989). Even so, in cause-probing investigations, internal validity typically is the primary priority, even if such knowledge claims are only localized in their generality (Campbell, 1986).

Irrespective of ideological, philosophical, and ultimately methodological disagreements and controversies (e.g., Cook et al., 2010; Coryn, 2009; Gargani, 2010; Mackie, 1980; Scriven, 1968, 1975, 1976, 2009), and, as Reichardt (2011) rightly notes, for the majority of evaluations, such inferences are about the effects of a given cause rather than questions about the cause of a given effect. From this perspective, a cause is that which precedes or produces an effect, and an effect is the difference between what occurred in the presence of a (presumed) cause and what would have occurred in its absence (i.e., counterfactual reasoning; Rubin, 1974, 2005). This view of causation is premised on manipulable causes that can be deliberately varied and that can generate reasonable approximations of the physically impossible counterfactual. Based on this logic, three conditions are necessary for causal inference (Shadish et al., 2002): (a) temporal precedence—that cause precedes effect, (b) covariation—that cause and effect vary together, and (c) absence of alternative causes—that no other plausible explanations can account for an observed treatment–outcome covariation.

Types and Sources of Potential Internal Validity Threats

Not enumerated in detail here, but essential for providing context, the most widely accepted internal validity threats are ambiguous temporal precedence, selection, history, maturation, regression, attrition, testing, instrumentation,
and additive and interactive threats (see Campbell & Stanley, 1966; Cook & Campbell, 1979; and, in particular, Shadish et al., 2002). And the major focus of research design is anticipating and reducing the number and plausibility of these threats to (internal) validity. By definition, experimental designs rule out selection threats (which is the predominant bias associated with nearly all other types of designs), but not the others. Although experimental designs have received far more attention, findings from recent investigations into evaluation practice suggest that quasi-experimental designs, among others (e.g., mixed-method designs, nonexperimental designs), are far more common for evaluating social programs and other types of applied interventions than experimental designs (Christie & Fleischer, 2010; Coryn et al., 2011). For quasi-experiments, Shadish et al. (2002) enumerate three closely related principles for identifying and reducing potential alternative explanations of treatment–outcome covariation: (a) identification and study of plausible threats to internal validity, (b) primacy of control by design, and (c) coherent pattern matching.

Nonequivalent Dependent Variables

Although numerous methods and techniques (both statistical and nonstatistical) are available for reducing the number and plausibility of internal validity threats, one lesser-known and underutilized design element is nonequivalent dependent variables. A nonequivalent dependent variable is a “...dependent variable that is predicted not to change because of the treatment but is expected to respond to some or all of the contextually important internal validity threats in the same way as the target outcome” (Shadish et al., 2002, p. 509). From a measurement perspective, both dependent and nonequivalent dependent variables should consist of similar or related manifest variables or latent constructs. Otherwise, the nonequivalent dependent variable is merely arbitrary.

Rationale for Using Nonequivalent Dependent Variables to Reduce Internal Validity Threats

With the exception of Shadish et al. (2002), very few research methods textbooks refer to nonequivalent dependent variables as a means for reducing the numerous internal validity threats associated with quasi-experiments. The rationale for using nonequivalent dependent variables to reduce validity threats to quasi-experimental designs is that they provide an elegant and robust, yet simple and straightforward, means for addressing the Shadish et al. (2002) coherent pattern-matching principle. In addressing this principle, nonequivalent dependent variables can be used to represent a complex prediction of causal hypotheses that few alternative explanations could explain, and “the more complex the pattern that is successfully predicted,
the less likely it is that alternative explanations could generate the same pattern . . . ” (Shadish et al., 2002, p. 105).

Borrowing from, and modifying, a recent example from Reichardt (2011), consider an educational television program intended to teach children to pronounce the letters of the alphabet. Over a 12-week period the show presents the first half of the alphabet. Prior to and after viewing the show for 12 weeks children are tested (i.e., a one-group pretest–posttest design) on their ability to pronounce letters that they were exposed to (the dependent variable) as well as letters that they were not (the nonequivalent dependent variable). In terms of validity threats, several alternative explanations for any observed effect on the dependent variable would be plausible (e.g., after-school programs, parents teaching their children at home, other television programs). Even so, the same environmental and contextual validity threats would operate on both the dependent and nonequivalent dependent variables. If such validity threats were truly warranted, then they should have an effect on both variables, and, therefore, any effects would occur concomitantly. If not, then effects would be observed on the dependent variable only and not the nonequivalent dependent variable.

**History and Examples of Using Nonequivalent Dependent Variables to Reduce Internal Validity Threats**

Historically, and although studies using nonequivalent dependent variables appeared in the literature earlier (e.g., Robertson & Rossiter, 1976), one of the most well-known examples is McSweeny’s (1978) investigation of the effect of a directory assistance charge on the number of local directory assistance calls in Cincinnati. Because the directory assistance charge was for local directory assistance only, McSweeny (1978) used the number of long-distance directory assistance calls placed as a nonequivalent dependent variable to study the effect of charging for local directory assistance on the number of calls placed for local directory assistance. Based on the a priori causal hypothesis, the charge for local directory assistance would only decrease the number of calls placed for local directory assistance, and any environmental events would affect all types of directory assistance calls and would change the numbers of local directory assistance calls as well as long-distance directory assistance calls concurrently. Plotted results, shown in Figure 5.1 (an interrupted time-series design with a nonequivalent dependent variable), suggest that only the number of local directory assistance calls changed and long-distance calls remained unchanged.

In 1980, the Philadelphia Police Department reenacted its general administrative deadly force policy that was abolished in 1974. White (2000) analyzed shooting data for on- and off-duty police officers over two periods before and after the policy revisions, 1970–1978 and 1987–1992, treating the on-duty police officers as the focal dependent variable, off-duty police...
using nonequivalent dependent variables

officers as the nonequivalent dependent variable, and the general administrative deadly force policy as the independent variable. If changes in levels of deadly force were the result of the policy, only changes in fatal force among on-duty police officers would have occurred, whereas if the effect was due to another event, the fatal force among on- and off-duty police officers would have moved in parallel. White’s (2000) analysis showed levels of deadly force between both on- and off-duty police officers moved in the same direction, but at different magnitudes, as shown in Table 5.1.

McKillip and Baldwin (1990) utilized two nonequivalent dependent variables, alcohol use and exercise, to study the effects of a sexually transmitted disease (STD) education campaign on condom usage. A college campus implemented a multimedia education campaign aimed at decreasing the prevalence of STDs through increasing condom usage. If the effect was attributable to the campaign, only college students’ attitudes toward condom usage, beliefs about the consequences of using condoms inconsistently, intentions to use condoms, and discussions with friends about condom usage would have changed. If the effect was the result of another event affecting college students’ attitudes, beliefs, intentions, and discussions, then alcohol

Figure 5.1. Effects of Response Costs in the Charging for Directory Assistance in Cincinnati

consumption and exercise habits also would have increased at the same time as the four variables associated with increased condom usage. Moreover, if the ad campaign had no effect on the observed outcome variables, condom usage attitudes, beliefs, and intentions and discussions with friends, the outcome variables in addition to the nonequivalent dependent variables, alcohol consumption and exercise habits, would have changed in the same direction with the same magnitude. McKillip and Baldwin (1990) plotted the four condom usage variables against alcohol consumption and exercise habits and found that only beliefs supportive of condom usage and discussions with friends about condom usage changed, whereas attitudes toward and intentions to use condoms did not change (see Figures 5.2 and 5.3).

Table 5.1. The Nonequivalent Dependent Variables Design With Use of Deadly Force in Philadelphia Over Time by Type of Shooting (Nonelective or Elective) and Duty Status

<table>
<thead>
<tr>
<th>Dependent variable 1</th>
<th>Postmeasurement</th>
<th>Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-duty shootings (n = 521)</td>
<td>On-duty shootings (n = 122)</td>
<td></td>
</tr>
<tr>
<td>Nonelective 64% (335)</td>
<td>Nonelective 86% (105)</td>
<td>+35%</td>
</tr>
<tr>
<td>Elective 36% (186)</td>
<td>Elective 14% (17)</td>
<td>−55%</td>
</tr>
<tr>
<td>Annual rate</td>
<td>Annual rate</td>
<td></td>
</tr>
<tr>
<td>Nonelective 5.0</td>
<td>Nonelective 2.8</td>
<td>−44%</td>
</tr>
<tr>
<td>Elective 2.8</td>
<td>Elective 0.5</td>
<td>−82%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent variable 2</th>
<th>Postmeasurement</th>
<th>Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-duty shootings (n = 122)</td>
<td>Off-duty shootings (n = 42)</td>
<td></td>
</tr>
<tr>
<td>Nonelective 61% (74)</td>
<td>Nonelective 79% (43)</td>
<td>+30%</td>
</tr>
<tr>
<td>Elective 39% (48)</td>
<td>Elective 21% (9)</td>
<td>−46%</td>
</tr>
<tr>
<td>Annual rate</td>
<td>Annual rate</td>
<td></td>
</tr>
<tr>
<td>Nonelective 1.1</td>
<td>Nonelective 0.9</td>
<td>−18%</td>
</tr>
<tr>
<td>Elective 0.7</td>
<td>Elective 0.2</td>
<td>−71%</td>
</tr>
</tbody>
</table>

Note: Annual rates are calculated per 1,000 officers using 7,388 for Time 1 (9 years) and 6,280 for Time 2 (6 years).

Conclusion

Combined with other structural design elements (e.g., control or comparison groups, multiple pretest or posttest observations, removed treatments), nonequivalent dependent variables offer a powerful means for improving warrants for certain cause-and-effect conclusions in nonrandomized investigations. What is more, nonequivalent dependent variables are amenable to nearly all types of research designs including, but not limited to, within-subjects designs, between-subjects designs, single-subject designs, removed treatment designs, and interrupted time-series designs. Additionally, Nimon, Zigarmi, and Allen (2011) and Pratt, McGuigan, and Katzev (2000) have
Figure 5.2. Mean Strength of Health Beliefs as a Function of the Health Topic and the Week of Observation

Note: The STD media education campaign occurred during week 5.

Figure 5.3. Proportion of Respondents Reporting Discussion of Health Topics With a Friend During the Current Week as a Function of the Week of Observation

provided some empirical evidence supporting the use of retrospective pretest, which also could usefully be combined with (retrospective) non-equivalent dependent variables in many evaluation contexts given that a large majority of evaluations are retrospective rather than prospective.

References

**CHRIS L. S. CORYN** is the director of the Interdisciplinary PhD in Evaluation program and assistant professor of Evaluation, Measurement, and Research at Western Michigan University.

**KRISTIN A. HOBSON** is a doctoral student in the Interdisciplinary PhD in Evaluation program at Western Michigan University.