



Published in final edited form as:

*Curr Dir Psychol Sci.* 2010 October ; 19(5): 319–323. doi:10.1177/0963721410383977.

## All in the Family: Comparing Siblings to Test Causal Hypotheses Regarding Environmental Influences on Behavior

Benjamin B. Lahey and  
University of Chicago

Brian M. D’Onofrio  
Indiana University

### Abstract

Psychologists in both basic and applied fields are keenly interested in the environmental influences that shape our lives. Therefore, researchers test causal hypotheses to construct models of environmental influences that can withstand attempts at refutation. Randomized experiments provide the strongest tests of causal hypotheses, but are not always feasible and their assumptions cannot always be met. In such cases, a number of quasi-experimental research designs can be used to substantially reduce confounding in tests of causal hypotheses. Sibling-comparison designs provide robust quasi-experimental tests of causal environmental hypotheses, but they are underused in psychology in spite of their power, feasibility, and convenience.

### Keywords

Sibling comparisons; environmental influences; causal models; quasi-experiments

---

Psychologists seek valid causal models of the experiences that shape our lives. Such models help us understand the experiences that make one person different from another and have profound implications for the prevention and remediation of maladaptive behavior. Randomized experiments are preferred for testing causal environmental hypotheses. In large samples, the *random* assignment of research participants to different environments virtually eliminates the possibility that each participant’s genetic characteristics and experiences are *systematically* confounded (correlated) with the experimental environment. If randomly assigned participants who experience different experimental environments behave differently afterward, one often can infer that the experience had a causal influence.

The assumptions of randomized experiments cannot always be met and they are not always feasible, however (West, 2009). For such cases, several quasi-experimental designs have been devised to test causal hypotheses by ruling out plausible alternative explanations (Shadish, Cook, & Campbell, 2002). Quasi-experiments support causal inferences in the same way as randomized experiments—by controlling genetic and environmental variables that are confounded with the hypothesized causal environment. They do so by adding design elements to observational studies that allow the researcher to compare the obtained results to both (a) the results expected if the environmental variable has a true causal effect, and (b) the results expected under alternative hypotheses of confounding of the environment with other causal influences (Shadish et al., 2002). Quasi-experimental designs rarely, if ever, control all potential confounds, but well-designed quasi-experiments substantially reduce the number of alternative explanations for apparent causal effects. Crucially, when different quasi-experiments with different flaws support the same conclusion, the causal inference is strengthened considerably.

Each quasi-experimental design has its own requirements, some of which are quite restrictive. Interrupted time-series analyses are highly useful, but require multiple repeated measures of the response variable on each participant over time and are appropriate only for relatively discrete casual events, such as a change in welfare rules implemented at a specific point in time (McDowall, McCleary, Meidinger, & Hay, 1980). Other quasi-experimental designs are challenging to implement because they are based on relatively rare events, such as the birth of identical twins who have different experiences (Rutter, 2007b). Ironically, one of the most feasible quasi-experimental designs is rarely mentioned when discussing alternatives to randomized experiments in psychology (West, 2009). We describe how simply studying two or more siblings from each family allows strong tests of causal environmental hypotheses. Sibling-comparison (SC) designs are often used in econometrics and public policy (Bjorklund, Ginther, & Sundstrom, 2007; Bohlmark, 2008), but much less so in psychology (Rodgers, Cleveland, van den Oord, & Rowe, 2000). Indeed, it is common for psychologists to randomly select one child per family to avoid the statistical implications of siblings being clustered within families. Although SC data were originally analyzed using complicated statistical models, more accessible models are now available (Neuhaus & McCulloch, 2006).

## CONTROLLING GENETIC CONFOUNDS

Environmental influences operate in a rich context of gene-environment interplay (Rutter, 2007a). Therefore, genes and environments are often confounded (correlated) and their effects must be distinguished. Such gene-environment correlation (rGE) comes about through several processes (Plomin, DeFries, & Loehlin, 1977):

1. *Passive rGE.* Genes and environments are often passively correlated—the behavior and characteristics of the individual do not *cause* the correlation. This occurs because parents provide both their children's genes and home environments. When an allele (version) of a gene of the parents is associated with their childrearing (e.g., harsh physical punishment), the same allele in the children is passively correlated with childrearing they experience.
2. *Active and evocative rGE.* Genes and environments also become correlated when the genetically influenced behavior and characteristics of individuals actively selects them into, or evokes changes in, their environments. For example, during adolescence, alcohol and drug use and other non-aggressive rule-breaking may evoke greater peer acceptance (Burt, 2009), creating a correlation between the genes of adolescents that influence rule breaking and their experience of peer acceptance.

As a result of these several forms of rGE, the confounding of genes and environments is pervasive (Rutter, 2007b) and constitute plausible alternative causal explanations for associations between putative environmental risks and outcomes. Fortunately, SC designs can greatly reduce the genetic confounding of hypothesized environmental influences without requiring specialized knowledge of genetics.

### Ruling Out Passive rGE

Because tests of environmental influences in SC studies involve comparing full biological siblings who have different experiences, sibling comparisons generally rule out passive rGE. This is because meiosis (i.e., cell division that creates sperm and eggs) *randomly* distributes alleles of the parents' genes across siblings. For example, if a mother were imprisoned during the infancy of one child but not another, the two siblings would be equally likely to passively receive any maternal alleles associated with her criminal behavior that might also be associated with offspring adjustment. Hypothetically, if allele G is associated with both

maternal criminal behavior and offspring behavior, each biological offspring of two heterozygous (Gg) parents would have the same chances of inheriting a GG (25%), Gg (50%), or gg (25%) genotype. Therefore, in sufficiently large samples, the randomization inherent in meiosis eliminates passive rGE. The same logic holds when parents are homozygous (GG or gg).

### Minimizing Active and Evocative rGE

SC designs do not automatically rule out active and evocative rGE like they do passive rGE. They can minimize active and evocative rGE, but only under some circumstances: First, exposure to the candidate environment must precede the behavior change in time so that the person's behavior cannot influence the environment (creating active or evocative rGE). For example, because each offspring's childhood behavior cannot influence how the mother eats during pregnancy, associations of maternal nutrition with offspring behavior could be free of confounding due to active or evocative rGE. It is necessary to carefully search for hidden rGE even in such cases, however. For example, genetic factors of the fetus could influence the hormones of the pregnant mother in ways that influence her eating, creating evocative rGE.

SC designs also can be used with candidate environments that occur after the offspring's birth, but additional controls are required to minimize selection factors, including active and evocative rGE. For example, one could use sibling comparisons to test the causal hypothesis that attending college influences voting behavior during middle age. Although voting cannot influence earlier college attendance, genetically influenced characteristics that differ among siblings (e.g., college aptitude scores) could influence both college attendance and voting. This active rGE would give the false impression that attending college influenced voting. When genetic confounding from active or evocative rGE is addressed (e.g., by controlling aptitude scores), SC designs can provide informative tests of causal environmental hypotheses. Although causal inferences have to be made cautiously, they are more justified than inferences based on designs that compare unrelated individuals.

## CONTROLLING ENVIRONMENTAL CONFOUNDS

SC designs also control many environmental variables that are confounded with the environment of interest. Because the statistical comparisons in SC designs are made among full siblings in the same families, SC designs automatically and completely rule out all environmental differences that vary *between* families. This means that ethnicity, socioeconomic status, neighborhoods, and the myriad fixed family characteristics shared by siblings cannot be confounded with the candidate environment. This feature of SC designs alone dramatically reduces the number of environmental confounds compared to standard comparisons of unrelated individuals.

There is a class of potential environmental confounds that SC designs do not automatically rule out, however. The candidate environmental variable to which the siblings were differentially exposed could be confounded with the "true" causal environmental variable that *also varies among siblings*. This could include environmental variables that cause the differential exposure of siblings to the putative risk factor. Fortunately, SC designs make the study of such confounds more tractable because the *only* environmental potential confounds not ruled out by SC designs must *simultaneously* meet three conditions: (1) vary among siblings within families, (2) be correlated with the target behavior, and (3) be correlated with the candidate environment *within* families. Because the number of environmental variables that meet all three requirements is limited, SC designs make it easier to systematically test potential environmental confounds. Such tests cannot definitively rule out alternative environments that do not meet these three requirements because that would require accepting

the null hypothesis. Nonetheless, systematically testing alternative environmental hypotheses could identify causal environmental hypotheses that are strong enough to justify additional quasi-experimental studies or randomized controlled trials that manipulate the candidate environmental variable to prevent maladaptive outcomes.

Consider a hypothetical example: It is possible that women who drink alcohol during pregnancy are more likely to expose their offspring to another deleterious environment (e.g., harsh punishment), which is actually responsible for their offspring's maladjustment. This could create the mistaken appearance of a causal association between prenatal alcohol exposure and offspring adjustment. A recent SC study suggested that prenatal exposure to alcohol is causally associated with offspring conduct problems (D'Onofrio et al., 2007). Nonetheless, if subsequent studies find that mothers who drink during the pregnancy of one child also are frequently intoxicated when caring for that infant (but are not intoxicated when caring for their other infants during whose pregnancies they did not drink), one could conduct further tests to determine if intoxicated infant childrearing was the causal variable rather than prenatal exposure to alcohol.

### Gene-environment Interaction

Note that tests of hypothesized environmental influences on behavior using SC designs provide no information on gene-environment interaction (GxE). GxE refers to situations in which some people are more vulnerable to environmental influences than others because of their genetic makeup (Rutter, 2007a). Therefore, any causal environmental influence inferred from an SC study could represent either (a) an environmental effect that is not moderated by the genotype of the individual, or (b) an environmental effect that is genetically moderated through GxE. In the latter case, the magnitude of the environmental effect would represent the average of the environmental effect across siblings with different genotypes.

### EXAMPLE OF A SIBLING-COMPARISON STUDY

Tests of the long-term effects of maternal smoking during pregnancy were conducted in a population-based sample (D'Onofrio et al., 2010). Standard correlational (regression) analyses that compared unrelated individuals showed that maternal smoking during pregnancy robustly predicted future adverse outcomes in the offspring, even when statistically controlling many demographic and environmental confounds (e.g., maternal drinking). Indeed, the age-related risk for conviction for a violent crime among 600,000 Swedish youth was three times greater among the offspring of women who smoked during pregnancy than women who did not (top panel of Figure 1). In contrast, comparisons among approximately 30,000 siblings in this sample (bottom panel) showed that siblings who differed in their exposure to maternal smoking during pregnancy did not differ in their risk for conviction.

The results of these SC studies are consistent with those of other quasi-experimental studies (Knopik, 2009) in indicating a lack of evidence that maternal smoking is a causal prenatal environmental risk factor for offspring antisocial behavior. This does not mean, of course, that maternal smoking during pregnancy is benign. Sibling comparisons and other quasi-experimental studies support the hypothesis that maternal smoking during pregnancy causes lower birth weight and other pregnancy-related problems (D'Onofrio et al., 2008; Knopik, 2009).

## LIMITATIONS AND ASSUMPTIONS OF SIBLING-COMPARISON DESIGNS

Like all research designs, the SC design has limitations and a set of assumptions that must be met to support valid inferences:

1. By definition, SC designs can be used only with candidate environments that vary among siblings.
2. All causal inference designs, including sibling comparisons, are based on the Stable Unit Treatment Value Assumption (SUTVA) that the effect of each participant's exposure to a risk factor does not influence other unexposed participants (Rubin, 2006). For example, if a child stops attending school dances because of being mugged afterward, the child's siblings also may stop going to school dances for the same reason. The mugging was a causal environmental event, but it would not be detected in a SC design because the behavior of all siblings changed as the result of the experience of one sibling.
3. To allow the generalization of findings of SC designs, it is necessary to determine if siblings differ in relevant ways from singletons (persons without siblings) in the population and determine if siblings who are differentially exposed to the putative risk environment are representative of all siblings (Shadish et al., 2002).
4. To satisfactorily control passive rGE, it is necessary to recruit enough full sibling pairs with the same two biological parents to have sufficient statistical power to test the causal hypothesis. If half siblings are included, differences in exposures to the candidate environment among half siblings could be confounded with differences in genes or experiences from the unshared biological parent. However, if sibling comparisons of full and half siblings yield the same results, that would fail to support the alternative hypothesis that influences from the unshared biological parents of half siblings are operating and are confounded with the candidate causal environment. For the same reason, SC designs can be strengthened by including other sibling pairs of varying genetic relatedness (e.g., dizygotic and monozygotic twins) to gain additional traction on causal influences (Rutter, 2007b).
5. As noted above, it is essential to determine why siblings were differentially exposed to the environmental risk factor. If confounded environmental factors or self-selection factors, such as evocative rGE, differentially influence exposure to the risk environment, causal inferences drawn from a SC study will likely be incorrect.

## STRENGTHENING SIBLING COMPARISON STUDIES

All quasi-experimental designs, including sibling comparisons, are based on theory-driven tests of causal hypotheses and the same basic principles of research that apply to all kinds of research. Therefore, stronger theory and stronger basic research methods (e.g., reliable and valid measurement of variables) will yield stronger SC studies. In addition, the validity of sibling comparisons can be strengthened by measuring behavior prior to environmental exposures to address sibling differences, using propensity-score matching on background variables, and by including informative nonequivalent control groups to address rival hypotheses (Rubin, 2006; Shadish et al., 2002).

To strengthen causal influences, researchers must consider and control differences in the ages, birth order, and sex of the siblings that may differentially influence exposures and outcomes. The ability of SC designs to draw valid causal inferences will be more limited when studying putative causal environments that are strongly associated with such potentially confounding factors. Nonetheless, researchers can analyze subsets of sibling

pairs with the same demographic characteristics (if the sample is large enough) or statistically control such characteristics. Furthermore, statistical methods are available to further strengthen causal inferences derived from quasi-experimental studies (Rubin, 2006).

## CONCLUSIONS

When randomized experiments cannot be conducted, quasi-experimental designs can provide highly informative tests of causal environmental hypotheses. They cannot rule out all confounds, but can greatly reduce them to strengthen causal inferences. Like other quasi-experimental designs, sibling comparisons provide considerable traction on causal inferences regarding environmental variables, and they often are more practical than most quasi-experimental designs because they are not dependent on rare occurrences and only require recruiting two or more biological siblings from each family.

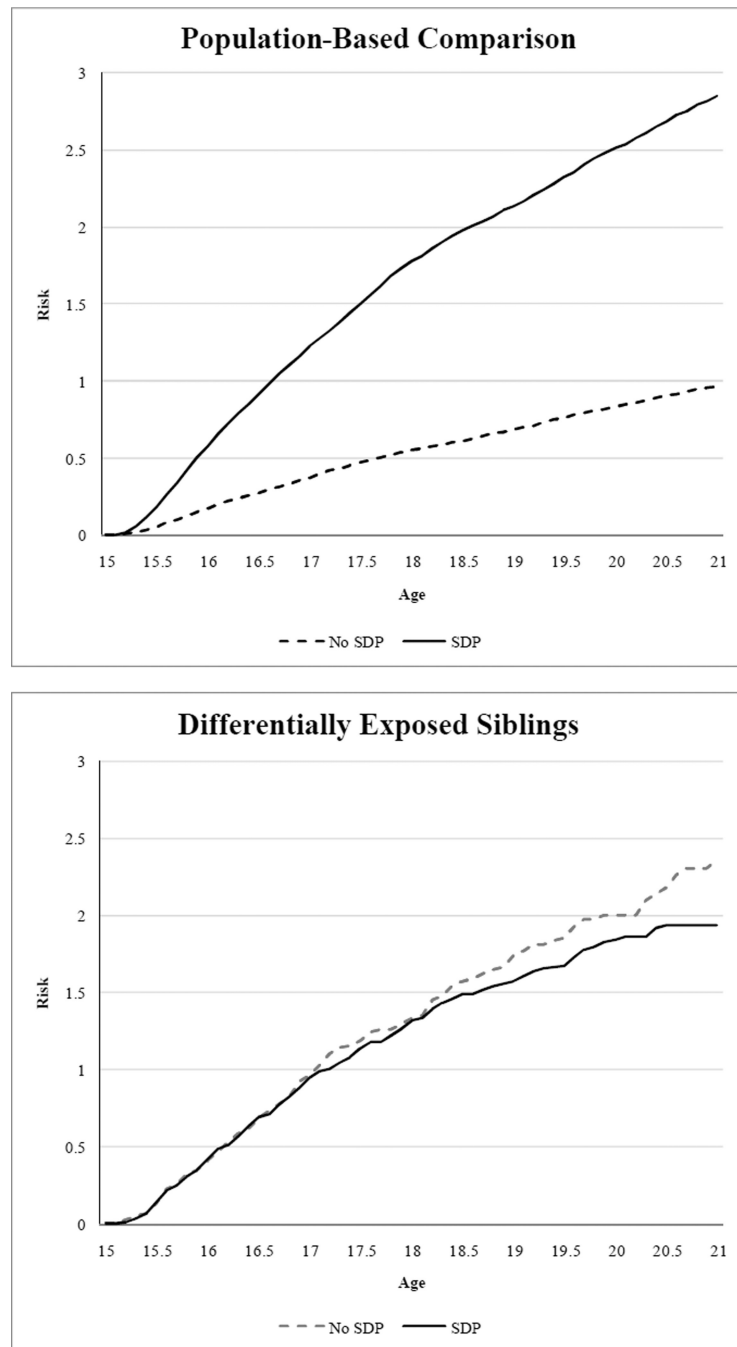
## REFERENCES

- Bjorklund A, Ginther DK, Sundstrom M. Family structure and child outcomes in the USA and Sweden. *Journal of Population Economics*. 2007; 20:183–201.
- Bohlmark A. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics*. 2008; 15:1366–1387.
- Burt A. A mechanistic explanation of popularity: Genes, rule breaking, and evocative gene-environment correlations. *Journal of Personality and Social Psychology*. 2009; 96:783–794. [PubMed: 19309202]
- D'Onofrio BM, Singh AL, Iliadou A, Lambe M, Hultman CM, Grann M, et al. Familial confounding of the association between maternal smoking during pregnancy and offspring criminality: A population-based study in Sweden. *Archives of General Psychiatry*. 2010; 67:529–538. [PubMed: 20439834]
- D'Onofrio BM, Van Hulle CA, Waldman ID, Rodgers JL, Harden KP, Rathouz PJ, et al. Smoking during pregnancy and offspring externalizing problems: An exploration of genetic and environmental confounds. *Development and Psychopathology*. 2008; 20:139–164. [PubMed: 18211732]
- D'Onofrio BM, Van Hulle CA, Waldman ID, Rodgers JL, Rathouz PJ, Lahey BB. Causal inferences regarding exposure to prenatal alcohol and childhood conduct problems. *Archives of General Psychiatry*. 2007; 64:1296–1304. [PubMed: 17984398]
- Knopik VS. Maternal smoking during pregnancy and child outcomes: Real or spurious effect? *Developmental Neuropsychology*. 2009; 34:1–36. [PubMed: 19142764]
- McDowall, D.; McCleary, R.; Meidinger, E.; Hay, RA. *Interrupted time series analysis*. Thousand Oaks, CA: Sage; 1980.
- Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B*. 2006; 68:859–872.
- Plomin R, DeFries JC, Loehlin JC. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*. 1977; 84:309–322. [PubMed: 557211]
- Rodgers JL, Cleveland H, van den Oord E, Rowe DC. Resolving the debate over birth order, family size, and intelligence. *American Psychologist*. 2000; 55:599–612. [PubMed: 10892201]
- Rubin, DB. *Matched sampling for causal effects*. New York: Cambridge University Press; 2006.
- Rutter M. Gene-environment interdependence. *Developmental Science*. 2007a; 10:12–18. [PubMed: 17181693]
- Rutter M. Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science*. 2007b; 2:377–395.
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2002.
- West SG. Alternatives to randomized experiments. *Current Directions in Psychological Science*. 2009; 18:299–304.



## RECOMMENDED READINGS

- Rodgers JL, Cleveland H, van den Oord E, Rowe DC. A powerful and highly accessible example of the use of sibling comparisons to test a causal hypothesis in psychology. 2000 (See references).
- Rutter M. This clearly written paper strongly urges psychologists to stop describing correlations and conduct tests of causal hypotheses using randomized experiments and quasi-experimental designs. 2007b (See references).
- Shadish WR. Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods*. 2010; 15:3–17. [PubMed: 20230099] This paper and the paper by [West](#) and [Thoemmes](#) (2010) below provide very readable and informative comparisons and contrasts of the two primary modern approaches to testing causal hypotheses using non-experimental designs, of which sibling-comparison designs are an example.
- West SG. A clear and thoughtful discuss of the value of quasi-experimental designs when randomized experiments cannot be conducted. 2009 (See references).
- West SG, Thoemmes F. Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*. 2010; 15:18–37. [PubMed: 20230100]



**Figure 1.**

Risk for convictions for violent offenses over increasing age in offspring who were, or were not, exposed to maternal smoking during pregnancy (SDP). The top panel presents the association between SDP and convictions in the population ( $N = 609,372$ ). The bottom panel shows the lack of association between SDP and convictions found when a subgroup of 29,482 siblings in the same families who were differentially exposed to SDP were compared. Estimates of risk are based on Kaplan-Meier estimates controlling for offspring sex and birth order (D'Onofrio et al., 2010).