George M. Eliopoulos, Section Editor

The Use and Interpretation of Quasi-Experimental Studies in Infectious Diseases

Anthony D. Harris,[1,2] Douglas D. Bradham,[1,2] Mona Baumgarten,[1] Ilene H. Zuckerman,[3] Jeffrey C. Fink,[4] and Eli N. Perencevich[1,2]

[1]Department of Epidemiology and Preventive Medicine, University of Maryland, [2]Veterans Affairs Maryland Health Care System, [3]Department of Pharmaceutical Health Services Research, and [4]University of Maryland Medical System, Baltimore, Maryland

Quasi-experimental study designs, sometimes called nonrandomized, pre-post–intervention study designs, are ubiquitous in the infectious diseases literature, particularly in the area of interventions aimed at decreasing the spread of antibiotic-resistant bacteria. Little has been written about the benefits and limitations of the quasi-experimental approach. This article outlines a hierarchy of quasi-experimental study design that is applicable to infectious diseases studies and that, if applied, may lead to sounder research and more-convincing causal links between infectious diseases interventions and outcomes.

In the study of infectious diseases and, in particular, in the study of infection control and antibiotic resistance, the quasi-experimental study design, sometimes called the pre-post–intervention design, is often used to evaluate the benefits of specific interventions. We reviewed studies published in 2 journals *(Clinical Infectious Diseases* and *Infection Control and Hospital Epidemiology)* during a 1.5-year period between 1 January 2002 and 1 June 2003 and found 36 quasi-experimental studies.

Quasi-experimental studies encompass a broad range of nonrandomized intervention studies. These designs are frequently used when it is not logistically feasible or not ethical to conduct a randomized, controlled trial—the "gold standard" of causal research design. Examples of quasi-experimental studies follow. For example, if a hospital is introducing use of an alcohol-based hand disinfectant, the hospital may want to study the impact of this intervention on the outcome of acquisition of antibiotic-resistant bacteria, on the basis of surveillance culture. The intervention is implemented, acquisition rates are measured before the intervention and after the intervention, and the results are analyzed. As another example, if a hospital has an increasing rate of ventilator-associated pneumonia (VAP), the hospital personnel may design an educational intervention aimed at decreasing the rate of VAP and compare rates before and after the intervention. A third example would be a study of the effect of an antimicrobial stewardship/educational program on preintervention and postintervention antibiotic prescribing practices.

As the capacity to collect routine clinical data has increased, so has the use of quasi-experimental study designs in the study of infectious diseases and in other medical disciplines. However, little is written about these study designs in the medical literature or in traditional epidemiology textbooks [1–3]. In contrast, the social sciences literature is replete with examples of ways to implement and improve quasi-experimental studies [4–6].

In this article, we aim to review the different quasi-experimental study designs and the hierarchy of these designs with respect to their ability to establish causal associations between an intervention and an outcome. The example of an alcohol-based hand disinfectant intervention aimed at decreasing antibiotic-resistant bacteria acquisition rates will be used throughout the article to illustrate the different quasi-experimental study designs. We discuss problems that arise in quasi-experimental study designs and offer methods to improve them.

METHODS

We reviewed articles and book chapters that discuss the design of quasi-experimental studies [4–10]. Most of the articles referenced 2 textbooks, which were then reviewed in depth [4, 6].

Key advantages and disadvantages of quasi-experimental

Received 17 December 2003; accepted 3 February 2004; electronically published 12 May 2004.

Financial support: National Insititutes of Health (grant K23 AI01752-01A1) (to A.D.H.); Veterans Affairs Health Services Research and Development Service Research Career Development Award (RCD-02026-1) (to E.N.P.).

Conflicts of interest: A.D.H. is a consultant for Merck and Ortho-McNeil and a recipient of The Pfizer Scholars Grant for Faculty Development in Clinical Epidemiology. E.N.P. is a consultant for Pfizer. J.C.F. is a consultant for Ortho-Biotech and Novartis.

Reprints or correspondence: Dr. Anthony Harris, Division of Healthcare Outcomes Research, Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, 100 Greene St. Lower Level, Baltimore, MD 21201 (aharris@epi.umaryland.edu).

Clinical Infectious Diseases 2004;38:1586–91

1058-4838/2004/3811-0014$15.00

studies, as they pertain to the study of infectious diseases, were identified. Potential methodological flaws of quasi experiments in the study of infectious diseases were identified. In addition, a summary figure outlining a hierarchy of quasi-experimental study designs is provided (figure 1): designs with higher numbers have more internal validity vis-à-vis potential causation between the intervention and the outcome [4].

RESULTS AND DISCUSSION

What is a quasi experiment?

Quasi experiments are studies that aim to evaluate interventions but that do not use randomization. Like randomized trials, quasi experiments aim to demonstrate causality between an intervention and an outcome.

On the basis of this definition, it is evident that many published studies in the infectious diseases literature and, in particular, in the study of antibiotic resistance use the quasi-experimental study design. The randomized, controlled trial is generally considered to have the highest level of credibility with regard to assessing causality; however, in a hospital or public health setting, the intervention often cannot be randomized, for one or more reasons: (1) ethical considerations, (2) an inability to randomize patients, (3) an inability to randomize locations, and (4) a need to intervene quickly. Each of these reasons is discussed below.

Ethical considerations typically will not allow the withholding of an intervention that has known efficacy. If the efficacy of an intervention is not established, then a randomized, controlled trial is the design of choice to determine efficacy. But if the intervention under study incorporates an accepted, well-established therapeutic intervention, or if the intervention has questionable efficacy on the basis of previously conducted quasi-experimental or observational studies, then ethical issues concerning the randomization of patients are raised.

Interventions often cannot be randomized to individual patients. For example, in studying the effect of use of an alcohol-based hand disinfectant on vancomycin-resistant enterococcus (VRE) acquisition rates, as determined by surveillance culture, it is difficult to randomize the use of disinfectant to individual rooms or individual patients, because, once disinfected, a staff member is unlikely to agree to be recontaminated before he or she sees the next patient—nor is an IRB likely to agree to this. Similarly, an education-based intervention to decrease VAP cannot be randomized to individual patients.

Interventions often cannot be randomized to individual lo-

CATEGORY 1.

QUASI-EXPERIMENTAL DESIGNS THAT DO NOT USE CONTROL GROUPS

1. The 1-group pretest-posttest design:
   O1 X O2
2. The 1-group pretest-posttest design that uses a double pretest :
   O1 O2 X O3
3. The 1-group pretest-posttest design that uses a nonequivalent dependent variable:
   (O1a, O1b) X (O2a, O2b)
4. The removed-treatment design:
   O1 X O2 O3 removeX O4
5. The repeated-treatment design:
   O1 X O2 removeX O3 X O4

CATEGORY 3.

QUASI-EXPERIMENTAL DESIGNS THAT USE CONTROL GROUPS AND PRETESTS

1. Untreated control group design that uses dependent pretest and posttest samples:
   O1a X O2a
   O1b O2b
2. Untreated control group design that uses dependent pretest and postest samples and a double pretest:
   O1a O2a X O3a
   O1b O2b O3b
3. Untreated control group design that uses dependent pretest and posttest samples and switching replications:
   O1a X O2a
   O1b O2b X O3b

* O = Observational Measurement, X=Intervention under study, and time moves from left to right

In general, studies in category 3 are of higher study design quality than studies in category 1. Also, as one moves down within each category, the studies become of higher quality: e.g., study 5 in category 1 is of higher study design quality than study 4, etc.

Figure 1. Hierarchy of the 8 quasi-experimental study designs most relevant to infectious diseases research. Designs with higher numbers have more internal validity vis-à-vis potential causation between the intervention and the outcome.

cations. For example, it is difficult to randomize use of the alcohol-based hand disinfectant to only some health care professionals. When this design of randomized locations is employed successfully, the locations are usually geographically separated; this involves additional issues of whether other factors about the environment are different, which further complicates the design and the analysis. A compromise that has been employed is to randomize various units in the same hospital. However, it is difficult, politically, to implement use of an alcohol-based disinfectant only in certain parts of a hospital or only on certain sides of a ward. Another underused alternative is the cluster randomization trial, in which intact groups or "clusters," rather than individuals, are randomized [11].

There is often a need, when seeking to control an infectious disease, to intervene quickly, which makes it difficult to properly conduct a randomized trial. In outbreaks of infection caused by antibiotic-resistant bacteria, for example, there is often pressure to end the outbreak by intervening in all possible areas, and, thus, it is not possible to withhold care, which would occur in a randomized controlled trial in which one of the groups received no treatment. The clinical and ethical necessity of intervening quickly makes it difficult or impossible to undertake the lengthy process of implementing a randomized study. In addition, there is substantial debate in the literature about the agreement rate between randomized trials and observational studies [12, 13]. Consequently, numerous studies are carried out retrospectively, after an intervention was implemented to end such an outbreak.

What are the threats to establishing causality when quasi-experimental designs are used in the study of infectious diseases?

The lack of random assignment is the major weakness of the quasi-experimental study design. Associations identified in quasi experiments meet some requirements of causality, because the intervention precedes the measurement of the outcome. Also, the outcome can be demonstrated to vary statistically with the intervention. Unfortunately, statistical association does not imply causal association, especially if the study is poorly designed. Thus, in many quasi experiments, one is most often left with the question: Are there alternative explanations for the apparent causal association? If these alternative explanations are credible, the evidence is less than convincing. These rival hypotheses or alternative explanations arise from principles of epidemiologic study design.

The methodological principles that most often result in alternative explanations in quasi-experimental studies of infectious diseases include the following: (1) difficulty in controlling for important confounding variables, (2) results that are explained by the statistical principle of regression to the mean, and (3) maturation effects.

The difficulty in controlling for important confounding variables arises from the lack of randomization. For example, in a study aiming to demonstrate that the introduction of an alcohol-based hand disinfectant led to lower rates of acquisition of antibiotic-resistant bacteria, there are a number of important potential confounding variables that may have differed between the 2 periods (i.e., the preintervention and postintervention periods); variables include severity of illness, quality of medical and nursing care, and antibiotic prescribing practices. In a multivariable regression, the first variable could be addressed through severity-of-illness measures, but the second and third confounding variables would be difficult, if not nearly impossible, to measure and control.

Regression to the mean is a widespread statistical phenomenon [14–16]. It can result in wrongly concluding that an effect is due to treatment when it is, in fact, due to chance. The phenomenon was first described in 1886 by Francis Galton. He measured the adult height of children and their parents, noting that, when the average height of the parents was greater than the mean height in the population, the children tended to be shorter than their parents. Likewise, when the average height of the parents was shorter than the mean height in the population, the children tended to be taller than their parents.

In the treatment of many infectious diseases, what triggers the implementation of an intervention is a rise in the rate above the mean or norm. For example, statistical control charts are often used in infection control to alert infection control personnel that rates of VAP or of acquisition of antibiotic-resistant bacteria are higher than usual. The statistical principle of regression to the mean predicts that these elevated rates will tend to decline, even without intervention. However, hospital personnel cannot wait passively for this decline to occur. Therefore, hospital personnel often implement one or more interventions and, if a decline in the rate occurs, they may mistakenly conclude that the decline is causally related to the intervention. In fact, an alternative explanation could be regression to the mean.

Maturation effects are a threat to the validity of concluding that an intervention caused an outcome. These effects are related to natural changes that patients experience with the passage of time. These maturational changes can threaten the internal validity of the study. In addition, there are cyclical seasonal trends that may be a threat to the validity of attributing an observed outcome to an intervention. For example, viral infections have seasonal patterns leading to higher rates of VAP in the winter. In our example study, if the preintervention VAP rate is measured in the winter, and the intervention occurs in the spring, then the drop in the VAP rate may be due to the seasonal trend and not the intervention.

What are the different quasi-experimental study designs?

In the social sciences literature, quasi-experimental studies are divided into 3 study design categories [4, 6]:

1. Quasi-experimental study designs that do not use control groups
2. Quasi-experimental study designs that use control groups but no pretest
3. Quasi-experimental study designs that use control groups and pretests

There is a hierarchy within these categories of study designs, with category 3 studies being sounder than those in categories 2 or 1, in terms of establishing causality. Thus, if possible, investigators should aim to design studies that fall into category 3.

Shadish et al. [4] discuss 7 designs in category 1; 3 designs in category 2; and 6 designs in category 3. We determined that category 2 studies are rarely applicable in infectious diseases research, because pretest measurements are almost always available. Furthermore, we determined that most quasi experiments in the study of infectious diseases could be characterized by 5 study designs in category 1 and by 3 designs in category 3, because the other study designs were not used in the study of infectious diseases, according to the literature. Thus, for simplicity, we have summarized the 8 study designs most relevant to infectious diseases research in the following sections and in figure 1. In each symbolic notation, time moves from left to right.

Category 1: Quasi-experimental study designs that do not use control groups.

1. The 1-group pretest-posttest design.

O1 X O2

This is a commonly used study design. A single pretest observational measurement (O1) is made, an intervention (X) is implemented, and a posttest measurement (O2) is made. For example, O1 could be the acquisition rate of VRE as determined by the results of perirectal surveillance cultures, X could be the introduction of use of an alcohol-based hand disinfectant, and O2 could be the acquisition rate of acquisition of VRE following the intervention. The inclusion of a pretest provides some information about what the acquisition rates might have been had the intervention not occurred.

2. The 1-group pretest-posttest design that uses a double pretest.

O1 O2 X O3

The advantage of this study design over design 1 is that the addition of a second pretest measurement prior to the intervention reduces the likelihood that regression to the mean, maturation, and/or seasonality could explain the observed association between the intervention and the posttest outcome. For example, in a study in which use of an alcohol-based hand disinfectant led to lower VRE acquisition rates (O3 < O2 and O1), if 1 study had 2 preintervention measurements of VRE acquisition rates (O1 and O2), and they were both elevated, this would suggest that there was a decreased likelihood that O3 was lower due to confounding variables, maturation effects, seasonal effects, or regression to the mean.

3. The 1-group pretest-posttest design that uses a nonequivalent dependent variable.

(O1a, O1b) X (O2a, O2b)

This design involves the inclusion of a nonequivalent dependent variable (b), in addition to the primary dependent variable (a). Variables a and b should assess similar constructs; that is, the 2 measurements should have similar potential causal variables and confounding variables, except for the effect of the intervention. Variable a is expected to change because of the intervention X, whereas variable b is not. Taking our VAP example, variable a could be the incidence of VAP, and variable b could be the incidence of catheter-associated urinary tract infection (UTI). If an educational intervention is aimed at encouraging hospital staff to raise the heads of the patients' beds and to follow a mechanical ventilation weaning protocol, one would expect to observe a decrease in the incidence of VAP but not in the incidence of UTI. However, a number of important confounding variables, such as the severity of illness and the antibiotic prescribing practices, might affect both outcome measurements. Thus, if the VAP and UTI rates were both measured, and if the VAP rates decreased following the intervention but UTI rates did not, then the data would be more convincing than if only VAP rates were measured.

4. The removed-treatment design.

O1 X O2 O3 removeX O4

This design adds a third posttest measurement (O3) to the 1-group pretest-posttest design and then removes the intervention before a final measure (O4) is made. The advantage of this design is that it allows one to test hypotheses about the outcome both in the presence and in the absence of the intervention. Thus, if one predicts a decrease in the outcome between O1 and O2 (i.e., after implementation of the intervention), then one would predict an increase in the outcome between O3 and O4 (i.e., after removal of the intervention). A caveat is that, if the intervention is thought to have persistent effects, then O4 needs to be measured after these effects are likely to have disappeared. For example, a study would be more convincing if it demonstrated that rates of VRE acquisition decreased following an intervention with alcohol-based hand dis-

infectant (O2 and O3 less than O1) and that when use of the disinfectant was discontinued, the rates increased (O4 greater than O2 and O3 and closer to O1).

5. The repeated-treatment design.

O1 X O2 removeX O3 X O4

The advantage of this design is that it demonstrates reproducibility of the association between the intervention and the outcome. For example, the association is more likely to be causal if one demonstrates that use of an alcohol-based hand disinfectant results in decreased antibiotic resistance rates both when it is first introduced and again when it is reintroduced following an interruption of the intervention. As in study design 3, the assumption must be made that the effect of the intervention is transient. This design is not often used in the study of infectious diseases because of the ethical issues involved in removing a treatment that seems to be efficacious. However, epidemiologically, it is a better design than those previously outlined.

Category 3: Quasi-experimental designs that use control groups and pretests. The reader should note that, in all of these study designs, the intervention is not randomized. The control groups chosen are comparison groups. Obtaining pretest measurements for both the intervention and control groups allows one to assess the initial comparability of the groups. The assumption is that the smaller the difference between pretest measurements, the less likelihood there is of there being important confounding variables between the 2 groups. In each symbolic notation, the design for the intervention group is above the horizontal line and that for the comparison group is below.

1. Untreated control group design that uses dependent pretest and posttest samples.

$$\frac{\text{O1a X O2a}}{\text{O1b O2b}}$$

The use of both a pretest group and a comparison group makes it easier to avoid certain threats to validity. However, because the 2 groups are nonequivalent (that is, patients are not assigned to groups by randomization), selection bias may exist. For example, suppose that an alcohol-based hand disinfectant intervention was instituted in the medical intensive care unit (ICU) and not in the surgical ICU. If rates O1a in the medical ICU and O1b in the surgical ICU are similar, this suggests that there is little difference in the important confounding variables between the 2 units. If O2a is less than O1a, but O2b is similar to O1b, this suggests that the observed outcome may be causally related to the intervention.

2. Untreated control-group design that uses dependent pretest and posttest samples and a double pretest.

$$\frac{\text{O1a O2a X O3a}}{\text{O1b O2b O3b}}$$

In this design, the pretest is administered at 2 different times. The main advantage of this design is that it controls for potentially different time-varying confounding effects in the intervention group and in the comparison group. In our example, measurements O1 and O2 would allow one to make an assessment as to whether there were time-dependent changes in preintervention VRE acquisition rates in both ICUs and whether these changes were similar or different.

3. Untreated control-group design that uses dependent pretest and posttest samples and switching replications.

$$\frac{\text{O1a X O2a}}{\text{O1b O2b X O3b}}$$

With this study design, the researcher administers an intervention at a later time to a group that initially served as a nonintervention control group. The advantage of this design over design 2 is that it demonstrates reproducibility in 2 different groups of subjects. This study design is not limited to 2 groups; in fact, the study results have greater validity if the intervention is replicated in different groups at multiple times. In the example of alcohol-based hand disinfectant, one could intervene in the medical ICU and then, at a later time, intervene in the surgical ICU.

SUMMARY

Although quasi-experimental study designs are ubiquitous in the infectious diseases literature, particularly in the area of interventions aimed at decreasing the spread of antibiotic-resistant bacteria, little has been written about the benefits and limitations of the quasi-experimental approach. As we have outlined in this paper, a hierarchy of quasi-experimental study designs exists, with some designs being more likely than others to permit causal interpretations of observed associations. Strengths and limitations of a particular study design should be discussed when presenting data collected in a quasi-experimental study. Investigators should choose the strongest design that is feasible given the particular circumstances.

References

1. Rothman KJ, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven, **1998**.
2. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, **1987**.

3. Szklo M, Nieto FJ. Epidemiology: beyond the basics. Gaithersburg, MD: Aspen, **2000**.
4. Shadish WR, Cook, TD, Campbell, DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin, **2002**.
5. Trochim WMK. The research methods knowledge base. Cincinnati: Atomic Dog Publishing, **2001**.
6. Cook TD, Campbell, DT. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally, **1979**.
7. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. Health Technol Assess **2000**; 4:1–154.
8. Shadish WR, Heinsman DT. Experiments versus quasi-experiments: do they yield the same answer? NIDA Res Monogr **1997**; 170:147–64.
9. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract **2000**; 17(Suppl 1):S11–6.
10. Zwerling C, Daltroy LH, Fine LJ, Johnston JJ, Melius J, Silverstein BA. Design and conduct of occupational injury intervention studies: a review of evaluation strategies. Am J Ind Med **1997**; 32:164–79.
11. Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomization trials. Stat Med **2001**; 20:3729–40.
12. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA **2001**; 286:821–30.
13. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med **2000**; 342:1887–92.
14. Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. BMJ **2003**; 326:1083–1084.
15. Bland JM, Altman DG. Regression towards the mean. BMJ **1994**; 308: 1499.
16. Bland JM, Altman DG. Some examples of regression towards the mean. BMJ **1994**; 309:780.