Commentary

Some Principles in Study Design for Preventing HIV Transmission: Rigor or Reality

Mervyn Susser, MB, BCh, FRCP(E), DrPH

Introduction

In stemming any epidemic, the main objective is to block transmission of the organism. In the human immunodeficiency virus (HIV) epidemic and many others, social norms and individual modes of behavior, as well as our technical abilities to block or destroy the virus, are at issue. Research designs for prevention must pay tribute to all of these.

Design is a means of eliciting, as best we can, the valid relationships between cause and effect. The definitive properties of causes are only three.[1] Arranged in ascending order of decisiveness, they are (1) the association between putative cause and putative effect, (2) the time order between putative cause and effect, and (3) the direction from putative cause to effect. Preventive interventions (the subject here) are of course designed to be causes in the broad sense—in this instance, causes that limit disease or block its emergence. Hence, the criteria we need to judge the effects of the interventions we design are readily at hand in the criteria we use to judge that the properties of causes are present.

Particular designs in given populations are drawn from a quiverful of different tactics to serve the same broad causal objective. Any design with human subjects cuts a swath in the demographic and social structure to isolate and display effects and their determinants under study. Thus, the emergent data describing relationships between variables for a given objective all seek to relate determinants to effects. In this respect, therefore, the array of study designs are not in principle different from one another, but they do achieve their objectives more or less well. Which is to say, they contribute more or less well to judgment about the presence of the three causal properties.

Two desirable elements of design in particular—rigor and applicability (or, as they are designated by Campbell and Stanley,[2] internal and external validity)—are problematic in the choice of interventions in the HIV epidemic. In a given study, these elements are always in conflict. Thus, choices involve necessary trading between the degree of rigor—the isolation of determinant and outcome from covariates—and the degree of applicability—the legitimate extrapolation of observed relations beyond the circumscribed study conditions.

It is essential to note, however, that a key causal criterion is the consistency of observed relations. Any single study of human subjects rarely establishes a causal relationship. Consistency is tested by two elements: survivability, defined by the number and severity of tests of association, and replicability, defined by the number and diversity of tests of association.[1] Consistency attests to a generalizable relationship as new tests probe successive situations. In other words, cause unfolds.

Rigor in Design

In quantitative human sciences, particular designs for attaining a specified objective tend to be ranked intuitively by their inherent degree of rigor and only secondarily by applicability. Rigor derives from the confidence one can place in the presence of the true and definitive properties of the putative causes elicited. This confidence resides in large part in two other elements of design:

Mervyn Susser is the Editor of the Journal.

Requests for reprints should be sent to the Journal office.

- The degree of change in the determinant—here, the intervention—mobilized by the design
- The degree of isolation of determinant and outcome from other factors (covariates, etc.)

A hierarchy of designs classified by rigor (note, *not* by aptness), each characterized by its irreducible features, is as follows:

1. Controlled experiment
   a. Planned intervention vs control untouched by the intervention
   b. Preselection of assignment to intervention and control by various means (e.g., randomization [the gold standard], matching)
2. Quasi-experiment
   a. Planned intervention vs comparable control untouched by intervention
   b. Post hoc assignment of comparison (e.g., historical controls)
3. Observation of sequential events, which appears in two main forms
   a. Longitudinal
      i. Determinants measured in preselected cohort
      ii. Outcome at follow-up compared in exposed vs comparable unexposed
   b. Case–control
      i. Determinants measured retrospectively in historical data
      ii. Exposure compared in cases (outcomes) and unaffected comparable controls, both preselected (cross-sectional observation collapses the passage of time)
4. Cross-sectional survey (nearly always the poor relation)
   a. Neither case nor comparison groups preselected but constructed post hoc
   b. At outset, size of sample and statistical power unknowable, so formal a priori hypothesis test ruled out

Applicability, Individuals, and Context

In epidemiology and, even more, in the clinical and behavioral sciences, we need to emphasize that all these designs are generally thought of as they relate to the individual level of organization—or, put simply, to persons. Designs in the population sciences, including epidemiology, may include exceptionally large numbers taken en masse. But they are nearly always populations of individuals disjoined and seen neither in connection to each other nor in an environmental context. We tend, therefore, to speak of designs as if they concerned individuals in fixed physical and social environments that are static and uninfluential. No account is taken of the dynamics of the social entities in which individuals live and work.

To step beyond this individual level of organization, we need an ecological epidemiology (eco-epidemiology).[3,4] That is, we need to understand how context affects the health of persons and groups through selection, distribution, interaction, and adaptation. More than individual-level analysis is needed to grasp the effects of any grouping, and pairings, families, peer groups, schools, communities, cultures, and legal systems are all contexts.[5,6] Without taking account of context, we can never fully explain patterns of morbidity, nor epidemic spread, nor transmission and acquisition of infection or values or behavior.

Analytically, the ecological approach reduces to the technical matter (feasible if not simple) of designating groups as units of study. The essence lies in recognizing and dealing with different levels of organization. Levels are arranged in hierarchies—for example, in ascending order, from gene or molecule to cell, to tissue, to person, to group. Complexity accrues with every ascent to a higher level. One kind of complexity resides simply in the new entities and, hence, in larger numbers of interacting variables that exist at successive levels. A more subtle kind of complexity resides in the effects of the group as a whole on its individual membership. Variables *special to groups* are present wherever groups are constituted and at all levels of the organizational hierarchy.

All will be familiar with the independent, dependent, and associated variables of regular individual-level analysis. To use these variables to characterize disjoined individuals in a group, we need only take the mean or median or proportional distribution of the given variable. But to consider such a grouped variable in terms of its effect on individuals is to introduce another idea.

To conceptualize distinctive effects of groups on the individuals within them, we need to denote at least three variables peculiar to groups:

1. Integral variables (sometimes called structural variables). These variables affect virtually all members of a group. They are conditions that vary between but not within groups. Examples that tend to affect exposed group members uniformly include discrete variables such as disasters or new laws, scaled variables such as level of hospital care, and continuous variables such as latitude and altitude.

2. Contextual variables (sometimes called derived variables). These variables have potential effects peculiar to the group level but are derived from a measured attribute of individuals within each group. For example, the cognitive performance of schoolchildren of a given social class or ethnic group is modified by the proportions of the group who have the same or different group membership.[7,8] Contextual variables are central to analyzing the dynamics both of infection and behavior. For instance, herd immunity thresholds are crucially determined by the proportion of those susceptible in a population at risk.[9]

3. Dependent happenings[10–12] or contagion. In HIV prevention, these correlated dependent variables are especially important to recognize. Contagion is a variant of the contextual variable that derives from the individual dependent variable. It applies whenever the outcome of interest is something communicable—for example, infection, violence, beliefs, or behavior in general. Thus, contagion occurs in a group whenever the prevalence of an outcome—infection, or behavior, or social norms—affects the risk of that outcome spreading to individual members of the group (and also, in consequence, the whole dynamic of group prevalence as well). Thus, contagion holds true for malaria, dengue, HIV, cognitive and criminal behavior, and much else.

Here we are clearly dealing with dynamic systems rather than static situations. The moral is, further, that individual-level analysis, which clearly cannot capture epidemic spread at the group level, cannot capture the entirety of individual effects, either.

Choosing Appropriate Designs

Keeping in mind the background of possible designs and the necessity of accounting for context, we can begin to consider how to match designs to the particular questions at issue.

First, we pay our respects to the randomized controlled trial for its unexcelled rigor. When properly done, a randomized controlled trial is no light

undertaking and has the following stringent requirements:

- A hypothesis that is well developed, narrow, and refined
- An intervention that is measurable and legitimate ethically as well as scientifically
- Adequate statistical power (A constant question is, what sample size? A handy answer is, bigger.)
- Well-defined, measurable outcomes
- Double blinding where possible (i.e., group assignments unknown to both participants and observers)
- Analysis that uses total baseline denominators ("intention to treat") and accounts for all exclusions and losses to follow-up

These principles apply to both group-level and individual-level studies. But note that numbers and statistical power are a particular problem for group-level studies. One prescription is to—at either the group or the individual level—work with very large numbers, keep both the intervention and the outcome well defined and simple, and dispense with the need for controlling covariates. It is a valuable prescription.

The Randomized Controlled Trial and HIV Prevention

Now, let us move on to a major concern for design choice: if and when to use randomized controlled trials at individual and group levels in HIV prevention.

1. At the individual level, the randomized controlled trial is, where feasible, the best vehicle for direct tests of narrow hypotheses regarding vaccines, treatments, and measurable individual behavioral change. A corollary is that the randomized controlled trial is *not* the best test for complex hypothetical pathways, which necessarily require detecting interactions and effect modifiers (unless one can enlist very large numbers that permit the use of conditional strata).

2. At the community level, the randomized controlled trial can be the best vehicle for intervention only under special conditions—namely, a sufficiency of communities for statistical power, a simple measurable intervention (e.g., a vaccine or training in cognitive skills), and a simple measurable outcome (e.g., infected vs uninfected or a change in a clearly demarcated situation or behavior). Not least, the variation between communities in individual membership needs to be reckoned with, especially when the number of community units is not abundant.

3. Again at the community level, the randomized controlled trial will seldom be a suitable vehicle for a complex community intervention—say, endeavors to change social norms, whether of values, expected behavior, or actual behavior. In such community-level intervention, two particular difficulties are often not susceptible to control. The first is contamination of the control groups by the intervention directed at the experimental group. Contamination arises in forms that differ in societies at different levels of development. The second difficulty resides in time lags in the mobilization of social process between intervention and effect.

It is, of course, always crucial to isolate the experimental intervention from all comparison groups—hence, the "blinding" procedures in individual-level trials. In today's information-laden and media-ridden world, contamination becomes a forbidding problem for ambitious community-level interventions that depend either on the diffusion of knowledge and behavior or on changing cultural norms and social situations.

A subset of the contamination problem is that of migration or, more generally, population movement. Such movement spreads information even when, as in some less developed societies, the available media are primitive. Population movement may also silently alter the supposed existing level of the state the intervention aims to change.

One should mention, however, that under special circumstances, contamination can be reasonably well controlled, even in complex community-level trials. In groups that are socially isolated, as with total institutions, or in partially segregated groups such as homeless persons in modern cities, successes are on record.

In regard to complex interventions in the HIV epidemic, one can find in the efforts to control tobacco addiction an analogy, a model, and a cautionary tale.[13] These efforts stretch over almost a half-century, since the time when the link with lung cancer was first established. In the matter of achieving and demonstrating intervention effects, social reality largely neutralized the rigor of even the best community-level interventions. If these trials demonstrated any effects, they were modest.

Nonetheless, large and dramatic social changes in both smoking norms and smoking behavior took place. Moreover, these changes went against the grain of prevailing social norms and the concerted efforts of the tobacco industry to sustain them.

Why then did the community trials fail to show noteworthy effects? The first trials began 30 years and more after efforts at control began. The whole society was already pervaded by the social movement against smoking. Dynamic change in the desired direction was well under way. If indeed the target communities were not saturated, the decline in smoking was enough to undermine the possible size of the hoped-for effects and also, therefore, the statistical power to detect them.

Ironically then, the trials were the victims of a much-delayed success, that of a sustained social movement against smoking. But success emerged only after 2 decades of research and action. Thus, the apparent failures of the trials in no way refute the hypothesis that community intervention can bring about a reduction in smoking addiction. Reality intruded in the form of both contamination and time lag between intervention and effect.

In essence, the control communities were thoroughly contaminated by the elements of the intended experimental intervention that had long pervaded society. The mass change that undermined the experiments, it is important to note, followed multiple, multilevel interventions which, in turn, were founded on causal inference from observational studies only.

Some Recommendations

What then would one advocate for appropriate intervention in the HIV epidemic?

1. Take another leaf out of the smoking story: set about building a social movement in a conscious and purposeful way. Such movements involve alliances between grass-roots and official forces. Public health agencies need first to define the nature of the threat and then to publicize appropriate measures. With all else equal, understanding will first reach those most conscious of the threat and will then become diffused. As momentum builds, all levels of the movement must engage with public policy, press hard on official agencies, and advocate legislation and regulation as needed.

With HIV in the United States, although coordination has been lacking, much is already in place. In one important respect, the US HIV epidemic differs

from the situation that existed here when smoking was first seen to present a clear and present danger. In the smoking epidemic, those with the most knowledge and the most resources were at the outset the most affected; it was they who instigated the preventive movement that diffused through society at large. In the HIV epidemic, those first affected were in well-demarcated groups, some at the margins of society. Diffusion of a preventive movement across the whole society has thus been more difficult.

2. Build a sensitive surveillance system for monitoring the diffusion not only of outcomes in terms of disease transmission but also of new norms of behavior. Again, in the United States much is in place. What we need is a purposeful, concerted address to specific goals and objectives.

3. Use designs appropriate to the purpose and pay particular attention to the appropriate level of organization. Ecological-level observation studies are appropriate to studying epidemic spread and control. Randomized controlled trials should be reserved for objectives to which they are suited—namely, well-defined a priori hypotheses, well-controlled and defined interventions, and measurable outcomes.

4. Finally, when all these things are in place and going strong, have patience. As noted, it took 10 to 20 years, depending on how one counts, even to begin to see the definitive signs of the cultural revolution against smoking. Once the problem seems to be under control, with such transmissible phenomena as infection and addiction, be prepared to pay the price one must pay for freedom, which is eternal vigilance.[14] □

References

1. Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol.* 1991;133: 635–648.
2. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research.* Chicago, Ill: Rand McNally; 1966.
3. Susser M, Susser E. Choosing a future for epidemiology: I. eras and paradigms. *Am J Public Health.* 1996;86:668–673.
4. Susser M, Susser E. Choosing a future for epidemiology: II. from black box to Chinese boxes and eco-epidemiology. *Am J Public Health.* 1996;86:674–677.
5. Susser M. The logic in ecological: I. the logic of analysis. *Am J Public Health.* 1994;84:825–829.
6. Susser M. The logic in ecological: II. the logic of design. *Am J Public Health.* 1994;84: 830–835.
7. Coleman JS, Campbell EQ, Hobson CJ, et al. *Equality of Educational Opportunity.* Washington, DC: US Dept of Health, Education, and Welfare; 1966.
8. Central Advisory Council for Education (England). *Early Learning.* London, England: Her Majesty's Stationary Office; 1954.
9. Fox JP, Elvebach L, Scott W, Galewood L, Ackerman E. Herd immunity: basic concept and relevance to public health immunization practices. *Am J Epidemiol.* 1971;94: 179–189.
10. Ross R. *Prevention of Malaria.* 2nd ed. London, England: John Murray; 1911.
11. Halloran EM, Struchiner CJ. Study designs for dependent happenings. *Epidemiology.* 1991;2:331–338.
12. Kopman JS, Longini JM, Jacquez JA, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol.* 1991;133:1199–1209.
13. Susser M. The tribulations of trials—intervention in communities. *Am J Public Health.* 1995;85:156–158. Editorial.
14. Curran JP. Speech on the right of election of lord mayor of Dublin, 10 July 1790. *Oxford Dictionary of Quotations.* 3rd ed. Oxford, England: Oxford University Press; 1979:170.