Strengthening Nonrandomized Studies of Health Communication Strategies for HIV Prevention

Calum Davey, MSc,* Marc Boulay, PhD,† and James R. Hargreaves, MSc, PhD*

INTRODUCTION

Health communication (HC) is a central component of the HIV prevention agenda. Early in the epidemic, policymakers identified promoting awareness as a priority in the global response to HIV,1 and most national programs responded with information campaigns.2 HIV-related HC efforts have evolved from straightforward media campaigns to encompass a range of communication activities seeking to influence behaviors associated with disease transmission3,4 and characteristics of the broader social environment in which these behaviors are embedded, such as stigma5,6 and gender norms.7

Several factors complicate impact evaluation of HC interventions. Individuals are often able to self-select for exposure to these interventions, potentially biasing comparisons of outcomes between those exposed and those unexposed.8 Also, evidence suggests that these messages may diffuse through informal community networks to influence individuals who do not directly see or hear program materials.9,10 Furthermore, many HIV-related communication initiatives have multiple simultaneous elements using overlapping messages and channels. Cluster randomized controlled trial designs (cRCTs) respond to many of these issues. In such trials, the outcome distribution among those allocated to the control arm can be interpreted as the potential outcome distribution that would have been observed in the intervention arm if the intervention had not been allocated. A comparison of outcome distributions between places with and without the intervention can therefore be interpreted as the average causal effect of allocation to the intervention.11 Integrated process evaluation is an essential part of such studies since HC programs delivered in real-life settings do not guarantee that randomization alone will ensure useful results.12-14 Although cRCTs offer the least-biased and simplest approach to estimating and understanding the intention-to-treat (ITT) effect,15 investigator-controlled random allocation of HC interventions is often not an option.

Broadly speaking, randomized trials control for confounding by design, whereas observational studies do so by analysis. Observational studies of HC interventions can have great value, and in some cases they are the only option available. Analytic approaches are described in the literature to adjust for confounding in such studies.16 However, evaluators also worry about confounding by unmeasured and poorly understood factors. Instrumental variable approaches offer the chance to control confounding without measurement of all confounding variables, but in practice naturally occurring valid instruments are rarely identifiable.17

If evaluators do not control allocation of the intervention, then evaluation “design” principally refers to making decisions
This article/C211 rather than to individuals. We refer to these
Often, many “...S272 gous to a cRCT is residual confounding by unknown, unmea-
...tenent of such data. The most significant problem for an eval-
analysis of such data. The most significant problem for an eval-
...18–21 This article aims to improve and encourage the use of quasi-experimental
designs in evaluations of HC strategies.

METHODS

HC interventions are naturally delivered in clusters—
groups or areas—rather than to individuals. We refer to these
units as “clusters,” which may be districts, towns, schools, or
any other politically or physically determined unit. We
emphasize situations where the primary aim of an evaluation
is to estimate the causal effect of a defined program on HIV-
related end points using the ITT principle.22 Often, many
people within intervention clusters will not be exposed, but
the primary concern of an ITT-analysis is to estimate the
overall effect in the target population. The scope of this article
does not permit discussion of several other critical design
elements, for example, population sampling, data validity,
or sample size estimation. Rather, we hope to help teams
navigate design options and recognize critical decision points
where impact evaluation may be strengthened.

We start by outlining 2 extreme but recognizable
scenarios and associated evaluation designs, anticipating that
readers will recognize both and agree that often neither design
will match their needs or the real-life conditions in which evaluations are planned (Fig. 1). At one extreme is the cRCT,
which can produce internally valid ITT estimates of interven-
tion effects. At the other extreme is the observational study,
susceptible to bias, that must rely on associations between
self-reported exposure and end points to estimate effects.

We argue that there is a “middle ground” of cluster-level
quasi-experimental designs. These designs can be adapted for
use in HC intervention rollout scenarios commonly encountered
by implementers and can give rise to valid effect estimates. The designs will require evaluators and implementers to work
together and make informed compromises. Implementers should
consider evaluation as part of intervention planning. Evaluations
will be better able to produce valid estimates of impact, without
randomization, where, in advance of deploying the intervention,
the following are clearly documented:

1. The intervention components;
2. Criteria that determine which clusters are eligible to
   receive the intervention; and
3. Criteria that determine which eligible clusters will actu-
   ally receive the intervention; we refer to this as the
   presence of an allocation scheme.

Our call mirrors concerns in the causal inference
literature, where it is argued that a counterfactual approach
in public health requires that causal effects are defined in
terms of contrasts between health outcomes corresponding to
different “well-defined” intervention conditions, and
where analysis and design strategies allow, appropriate con-

Not all situations are amenable to evaluation, but when
evaluation is a major concern, then, satisfying these con-
ditions should be feasible. Where these conditions are met,
we discuss 4 quasi-experimental research designs that are
based on implementation scenarios, which are commonly
encountered in real life. These are shown in Table 1 and
described in more detail below.

Design 1: Nonrandomized
Controlled Comparison

The implementation plan may allocate some eligible
clusters to receive the HC intervention but not others (see row
1, Table 1). The evaluation design may exploit this variation
between clusters. For example, community-based HC pro-
grams, such as those that involve community drama; peer
educators, or other change agents recruited in the community;
or other interpersonal channels of communication, are typi-
ically implemented in a subset of communities within an over-
all project area. Although mass media programs that use
national broadcast channels would not fit this scenario, pro-
grams relying on community radio stations, with circums-
cribed broadcast areas, may reach just a subset of clusters.
For example, a current trial in Burkina Faso is testing the
effectiveness of a community radio-based intervention by
defining the nonoverlapping geographic catchment areas for
14 community radio stations and randomly allocating 7 areas
to receive messages on key health issues.

Random allocation is not always feasible and other
considerations, such as a desire to target areas with less favorable
health or economic indicators, may influence the selection of
areas for implementation. An evaluation will be strengthened
when the factors determining whether or not eligible clusters are
allocated to receive the interventions are determined in advance
and/or easily measured. Causal attribution is much harder in
situations where allocation is driven by unknown factors or is
chaotic and unplanned. The challenge facing evaluation teams is
to measure outcomes in places that are and are not allocated to
receive the interventions, and crucially, to be able to convinc-
ingly argue that differences in the outcome distributions between
these places, after adjusted analysis, arise because of the
intervention allocation in question, that is, that the difference
is not confounded by other factors.

In some cases, the rules determining allocation may be
complex. When the clusters allocated to receive the interven-
tion are defined in advance by such rules, the design
challenge is to identify other eligible clusters that will not
receive the intervention that can act as controls. The intention
is to select clusters that are alike, before intervention, in
respects relevant to the outcome distribution. Matching may
be used, incorporating geopolitical factors such as cultural,
health system, and political contexts. For example, the Young
People’s Development Programme in the United Kingdom
was an intervention delivered to schools that were selected
through a competitive tendering process.24 After sites were
allocated to receive the intervention, comparison sites were
drawn from among unsuccessful applications, matched to the
intervention sites by region, local deprivation, teenage pregnancy rates, urban/rural/seaside residence, and sector (voluntary or statutory). The evaluators compared outcomes in intervention and matched schools.

It is often convenient to match on a small number of strong predictors of the outcome; however, where no single factor is strongly predictive of the end point, a “propensity score” approach may be used. A propensity score is calculated for each eligible cluster, usually using a logistic regression of potential confounding factors that “predict” whether or not a cluster will actually be allocated to receive the intervention. Clusters with similar scores to those of intervention clusters—that is, judged to have a similar propensity to be allocated to the intervention—are considered, along with the intervention clusters, to have been effectively randomly allocated to receive the intervention or not. The propensity score can then be used in both design (eg, for matching or defining eligibility) or analysis (eg, as an independent variable). These approaches are described in more detail in a wide literature on the subject.25,26 Matched studies are more complex to analyze than unmatched designs and may have less statistical power.27

In other situations, deterministic rules may be applied to define whether clusters actually receive the intervention. For example, Arcand and Wouabe evaluated an HIV education training module for school teachers in Cameroon.28 For pragmatic reasons, although there were villages with between 1

---

**FIGURE 1.** Two extreme scenarios and evaluation designs.
More recently, in 2005, a delivery-fee exemption was rolled out, then the interruption may be a break in the prevailing outcome trend coinciding with the start of the Stop AIDS Love Life communication program.

An evaluation design for this scenario cannot make comparisons between clusters to estimate impact. Instead, the outcome time trend before intervention is used to estimate the outcome trend if the intervention had not been implemented. This is distinct from a simple before-after comparison, which does not account for temporal changes in the outcome distribution. Evidence of the effect of the intervention comes from an “interruption” in the prevailing outcome trend coinciding with the intervention. The interruption may be a break in the trend line or a change in the gradient of the trend. For example, in Ghana, a time-series design was used to investigate the effect of 2 policy decisions on the proportion of pregnant women having deliveries that were assisted by a skilled attendant. In 2005, a delivery-fee exemption was rolled out, then in 2008, the government exempted pregnant women from national insurance fees so that they were entitled to antenatal, childbirth, and postnatal care without charge. Data on time trends in the proportion of women giving birth in a facility were plotted over time. Although there was an upward secular trend in the outcome, it was possible to convincingly isolate the impact of the policy changes on the outcome of interest.

This approach requires multiple data points before and after the introduction of the intervention; the number needed depends on a range of factors. Because the design relies on a good characterization of the prevailing trend in the outcome, the evaluators may need to draw on routine or surveillance data, for example, antenatal clinic data on HIV infections, clinic registers, or data gathered for another study. Evaluators may look for specific places where sufficient data have been collected so that this design can be used and then work with implementers to balance the requirements of the design against their priorities for rollout.

### Design 2: Interrupted Time Series

HC intervention implementation plans may result in all eligible clusters being allocated to receive the intervention at a given time, that is, that there is variation in allocation status in time, but not between clusters (see row 2, Table 1). This scenario may occur for programs relying primarily on national mass media channels, which typically have well-defined phases separated by periods of time with little or no activity, but little or no planned variation geographically. For example, in Brazil PRO-PATER implemented 3 separate mass media campaigns promoting vasectomy in 1983, 1985, and 1989, and the number of vasectomies increased markedly during each campaign period. More recently, a time-series analysis of condom sales in Ghana demonstrated an abrupt upward shift corresponding with the start of the Stop AIDS Love Life communication program.

Letters A–D represent clusters or groups of clusters. Numbers indicate intervention allocation status, 0 = no intervention, 1 = intervention, numbers between 0 and 1 represent variation in intensity of intervention. Time is shown along the horizontal axis.

<table>
<thead>
<tr>
<th>Implementation Scenario</th>
<th>Evaluation Design Option</th>
<th>Design Schematic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Some clusters are nonrandomly allocated to the program, but others are not</td>
<td>Nonrandomized controlled comparison</td>
<td>A 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C 0 0 0 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D 0 0 0 0</td>
</tr>
<tr>
<td>2. All clusters are allocated to start the program at the same time</td>
<td>Interrupted time-series</td>
<td>A 0 0 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B 0 0 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C 0 0 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D 0 0 1 1</td>
</tr>
<tr>
<td>3. Clusters are allocated to start the program at different times</td>
<td>Nonrandomized stepped-wedge or phased implementation</td>
<td>A 1 1 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B 0 1 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C 0 0 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D 0 0 1 1</td>
</tr>
<tr>
<td>4. All clusters allocated some of the program but at different intensities</td>
<td>Dose–response or implementation–strength</td>
<td>A 0.2 0.2 0.2 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B 0.8 0.8 0.8 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C 0.3 0.3 0.3 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D 0.5 0.5 0.5 0.5</td>
</tr>
</tbody>
</table>

Letters A–D represent clusters or groups of clusters. Numbers indicate intervention allocation status, 0 = no intervention, 1 = intervention, numbers between 0 and 1 represent variation in intensity of intervention. Time is shown along the horizontal axis.
Because the dynamics of infectious diseases rarely conform to simple linear trends, mathematical models can draw on other data to help predict trends. A lag between the interruption and a change in outcomes can make analysis more complicated and also widens the period when other events that potentially explain changes in the outcomes could have taken place. Optimal interventions for a time-series approach will be implemented at a defined time point, rapidly taken up by the target populations, and could feasibly cause changes in outcomes quickly.

**Design 3: Phased Implementation**

The implementation plan may allocate clusters to initiate the intervention at different times, with eventual initiation in all clusters (see row 3, Table 1). A typical HC case is one where a mixture of community-level and mass media programs are initiated at different times in different places, possibly because of limitations in an NGO’s capacity to train community leaders and produce locally relevant health messages. As an example, the Bridge Project in Malawi, between 2001 and 2008, used mass media and community-level interventions to communicate HIV prevention messages. Initially implemented in 8 of Malawi’s 28 districts, it has since expanded to 11 more districts. When randomization determines when places are allocated to receive interventions, the evaluation design is known as a “stepped-wedge” or “phased-implementation” cRCT design.

As in nonrandomized controlled studies (design 1), a challenge for evaluations of this scenario is ensuring that during the time periods where clusters do not change allocation status (phases) there is “balance” between arms on important characteristics. As in the time-series design (design 2), phased implementation studies often include multiple measurements over time. Analysis of the data from such studies can thus be thought of in 2 ways. A “horizontal” approach estimates the secular trend in the outcome in the clusters that are not changing intervention condition, and accounts for this trend in the before and after comparison of outcome data in clusters changing intervention allocation status. The challenge is ensuring that the measured trend is a valid estimate of the expected trend in the clusters that change intervention status. The second approach compares clusters with and without the intervention within phases and combines the within-phase estimates, making no assumptions about the nature of the secular trend—a “vertical approach.” The analyses of stepped-wedge trials with randomized start times will often combine horizontal and vertical approaches.

The probability of detecting an effect may be reduced if there is a lag between the introduction of the intervention and a change in the outcomes, as with HIV prevention, or if the full intervention is not realized during the time between steps.

**Design 4: Implementation Strength**

The implementation plan may entail variation in the strength of the intervention allocated to clusters, with some clusters allocated a greater “dose” of activities than others (see row 4, Table 1). An evaluation design that relies on this variation could be appropriate in situations where a program uses a single channel, for instance, small-group activities at the community level, to communicate messages, with activities occurring more often in some communities than in others. Alternatively, because most large-scale programs now use multiple channels to communicate health messages, this design may also apply when the number of program channels differs across clusters. For instance, the COMMIT project in Tanzania used both mass media and community-based activities to communicate messages promoting behaviors to reduce the transmission of malaria. The program’s mass media messages reached all communities, but only some clusters had community-based group activities, and in even fewer communities the project recruited community members to serve as local change agents promoting malaria prevention. The allocation of these channels across communities would allow program evaluators to measure the dose of the intervention for each cluster.

As for a nonrandomized controlled study (design 1), the evaluation design and analysis will need to account for potential confounding arising from differences between clusters that receive different strengths of intervention. Ideally, the variation in implementation strength will be planned, so that the results are in keeping with the ITT principle. However, where this variation is not planned, the next best option will be to estimate variations in implementation strength as it happens. Developing an index of implementation strength involves a numerator, for example, money spent on interventions, and a denominator, for example, the size of the target population. Few research studies using this design with an ITT approach are found in the literature. An example with a measured index of intensity comes from an evaluation of the impact of Avalahan, a large, targeted HIV prevention intervention in India. Ng et al estimated the intensity of the intervention using the money spent in each district per year on targeted interventions ($/PLHIV) in each district. The cumulative HIV allocation intensity ($/PLHIV) was summed from the start of the program until year t and regressed against HIV prevalence among individuals attending antenatal care clinics in year t. Using a multilevel regression analysis approach, they estimated the association between cumulative resource allocation for interventions ($/PLHIV) in a district and the odds of a particular woman at an antenatal clinic being HIV positive. In this design, detailed plans and budgets will be useful, and the evaluation may benefit from following how allocation intensity changes with time.

An additional challenge with this design is interpreting the dose effects. We have suggested that intensity can be indexed using a continuous variable such as frequency of radio transmissions or in terms of overlapping components. Although we are more concerned here with identifying simple dose effects of increasing intensity on outcomes, the interpretation of a dose effect may include combination effects from different components acting together. Process evaluation, as well as a comprehensive theory of change, may help evaluators interpret their results.

**Combining Designs**

How well the assumptions of each design are met may inform the choice of design for a particular situation. It is unlikely that all of the assumptions of any one of the designs will be completely satisfied, and practical factors such as cost...
and the availability of data may make 1 option stand out over another. Combining methods can balance the limitations of each design. However, if different methods find different results, interpretation can be difficult. A combination of methods should not be viewed as mutually exclusive routes but rather as mutually supporting options for evaluating an intervention.

**DISCUSSION**

HC evaluation teams should more commonly deploy quasi-experimental study designs, as these studies can yield greater validity than purely observational studies. Designing such studies can be organized around common HC rollout scenarios. Maximizing the utility of these designs will require collaboration from the outset between those primarily concerned with implementation and those primarily concerned with evaluation. Such HC evaluations will be strengthened if, in advance of implementation: (1) the planned intervention components are described, (2) cluster eligibility criteria are defined, and (3) intervention allocation criteria are defined and are driven by predictable and measurable factors.

We advocate better and closer communication between evaluators and implementers, up to and including having evaluators influence rollout of the intervention. We recognize that this may be difficult when the evaluation is strictly “external” to the implementation, for example, when evaluators and implementers are based at separate institutions and when there is a mindset that the “independence” of the evaluation is based largely on the separateness of these 2 groups. We argue that the face-validity of the evaluation is increased with good design, and that procedures such as protocol registration and preanalysis plans can increase the transparency of the method. We do not wish to argue against the merits of external evaluations, but rather that this should not be pursued at the expense of the simple ways that collaboration can improve the evaluation design.

For each intervention allocation scenario, there will be many possible evaluation designs. We have focused on the problem of identifying the ITT effect. The proposed approaches emphasize evaluation questions seeking to identify whether the program had an effect, and in themselves, may not necessarily inform questions seeking to identify how the programs may have influenced changes in behavior. Comprehensive evaluations of HC programs ideally include assessments of the applicability of the theoretical hypotheses informing the messages used by a program. Assessing the theory of change associated with an HC program provides insight into the relative effectiveness of the specific messages and informs program refinements. The validity of an evaluation is further determined by such factors as monitoring of the intervention as it is delivered, data on intervention availability in comparison places, and data on intermediate factors in the theory of change.

One potential limitation of these approaches arises from the difference between intervention allocation and intervention exposure. Although HC programs allocate intervention messages at the cluster level, exposure to these messages occurs at the individual level. In situations with high levels of exposure to intervention messages, a high level of correspondence will exist between membership in an intervention cluster and exposure to a program’s messages. When exposure to a program’s messages is relatively low, it may be harder to detect ITT effects simply because of the low exposure levels. However, in communities with a cohesive social structure and a high level of interpersonal communication about health topics, the diffusion of program messages through peer networks may mitigate the problems associated with low levels of direct message exposure.

Evaluations in real-life contexts may struggle to achieve the internal validity of a cRCT, but quasi-experiments have advantages in terms of their external validity. Evaluations delivered at scale and with the budget and oversight of real-life implementation may have greater external validity than a cRCT performed in limited conditions with an unrealistic implementation budget.

Overcoming the barriers to timely communication between implementing and evaluating partners will go a long way in strengthening evaluation results. Moreover, since donors, civil society, governments, and are increasingly interested in knowing “what works,” we hope that the vision and funding will be available to ensure that implementers and evaluators work as partners.

**REFERENCES**


