Multilevel Interventions: Study Design and Analysis Issues

Paul D. Cleary, Cary P. Gross, Alan M. Zaslavsky, Stephen H. Taplin

Correspondence to: Paul D. Cleary, PhD, Yale School of Public Health, 60 College St., LEPH 210, PO Box 208034, New Haven, CT 06520-8034 (e-mail: paul.cleary@yale.edu).

Multilevel interventions, implemented at the individual, physician, clinic, health-care organization, and/or community level, increasingly are proposed and used in the belief that they will lead to more substantial and sustained changes in behaviors related to cancer prevention, detection, and treatment than would single-level interventions. It is important to understand how intervention components are related to patient outcomes and identify barriers to implementation. Designs that permit such assessments are uncommon, however. Thus, an important way of expanding our knowledge about multilevel interventions would be to assess the impact of interventions at different levels on patients as well as the independent and synergistic effects of influences from different levels. It also would be useful to assess the impact of interventions on outcomes at different levels. Multilevel interventions are much more expensive and complicated to implement and evaluate than are single-level interventions. Given how little evidence there is about the value of multilevel interventions, however, it is incumbent upon those arguing for this approach to do multilevel research that explicates the contributions that interventions at different levels make to the desired outcomes. Only then will we know whether multilevel interventions are better than more focused interventions and gain greater insights into the kinds of interventions that can be implemented effectively and efficiently to improve health and health care for individuals with cancer. This chapter reviews designs for assessing multilevel interventions and analytic ways of controlling for potentially confounding variables that can account for the complex structure of multilevel data.

J Natl Cancer Inst Monogr 2012;44:49–55

One can intervene to affect health behaviors at several levels, such as individuals (1–4), groups (1), families (5), schools (5), worksites (6), religious organizations (7), health-care plans or organizations, physicians or clinics (2,3,8–12), and neighborhoods (4) or communities (1,12–17). Many health researchers are interested in developing, implementing, and evaluating multilevel interventions (1,3,5,18–20) because they believe that we are most likely to achieve substantial and sustained change with interventions based on ecological theory that target multiple levels, or sources of influence (6,18,21–24). The expectation for a multilevel intervention usually is that the combined effect of the interventions used will at least be additive, which is the sum of the effects of what the interventions would have achieved separately. They also could have a synergistic effect, that is, an effect that is greater than the sum of the effects of the separate interventions. Such synergy could occur when a set of necessary but not sufficient conditions must be jointly present for change to take place or when an intervention at one level facilitates or reinforces an intervention at another level. Conversely, the synergy might be negative or subadditive. For example, if either of two interventions alone is sufficient to elicit the desired outcome, their combined effect might be less than the sum of their separate effects. Another motivation for a multilevel approach is the belief that interventions are more likely to transfer successfully to other settings if study designs help us understand contextual factors that affect implementation (25–27).

In this chapter, we review some approaches to assessing the impact of multilevel interventions. We focus on the types of study

designs and analytic strategies that can be used to control for potentially confounding factors and assess the impact of the different components of multilevel interventions.

Design Issues

Measuring Effects of Different Levels of an Intervention

A major design decision when evaluating a multilevel intervention is whether to assess only the combined impact of the multiple interventions compared with no intervention or the separate effects of interventions at different levels and possible interactive (synergistic) effects.

The simplest approach is to assess only the combined effect of a multilevel intervention on individuals (Figure 1; Design 1). For example, in the Community Intervention Trial for Smoking Cessation (COMMIT) (13,14), one community in each of 11 matched pairs of communities received a multilevel intervention involving public education through community wide events, health-care providers, worksites, and other organizations; and support for cessation. The main outcome was smoking cessation, but there was no assessment of the contribution of each of the components of the intervention. Similarly, in a 14-community multilevel intervention (community mobilization and education, ordinance enforcement), to reduce smoking in school-age children (15), the main outcome was adolescent smoking, but no attempt was made to assess the effect of the different intervention components. Coady et al. (19) designed and implemented a multilevel

Interventions	Dependent variable
(A) Level 1 (eg, patient)	X1-Health behavior (eg, colorectal screening)
(B) Level 2 (eg, physician)	X2-Physician provides advice
(C) Level 3 (eg, community)	X3-Promotion of health behavior change
(D) Level 4 (eg, state)	X4-Change in reimbursement policy

Design 1: Combined effect of multiple interventions on patient level outcome

Design 2: Separate effects of multiple interventions on patient level outcome

Design 3: Effects of intervening variables at different levels

Figure 1. Designs for assessing multilevel interventions.

intervention (individuals, neighborhood, and community) to increase influenza vaccination areas in New York City. They also did not attempt to assess the relative or incremental contribution of the different components but only the overall impact on individuals.

Studies that assess the effect of specific components of the intervention on outcomes at the patient level (Figure 1; Design 2) are uncommon. A review of 83 trials of interventions to prevent sexually transmitted diseases, (1) for example, found that with few exceptions, they focused on a single level. In addition to assessing the effects of interventions at different levels on individuals, one can assess the effect at different levels (eg, impact of an clinic intervention on clinic procedures), although studies that assess the influences within and across levels are uncommon (28).

Both substantive and practical concerns have limited studies that have investigated such complex multilevel influences. A substantive reason may be that researchers think that only a combined intervention will have a detectable and/or important impact. We think it is more likely, however, that the main reasons why there have not been more such studies are the complexity and expense involved in developing such designs and the difficulty and expense of launching interventions in enough units so that there is adequate variation to assess more than overall effects. For example,

the COMMIT study (13,14) had only 11 intervention communities. That was sufficient to assess the intervention, but to assess the separate effect of provider education strategies, the study would have had to use a more complex design that also assigned physicians to receive the educational intervention or not. Below, we describe designs suitable to address such questions; they are considerably more complex to specify and implement than the simple design of the COMMIT study because they require researchers to implement several combinations of the two interventions. Thus, although there are no conceptual or analytic barriers to designing and analyzing the results of such studies, there are formidable practical barriers.

Measuring Outcomes at Different Levels

In addition to assessing the independent and synergistic effects of influences from different levels on individual knowledge, attitudes, and/or behavior, another way of expanding our knowledge about multilevel interventions would be to conduct research that assesses the impact of interventions on processes or outcomes at different levels (29) (Figure 1; Design 3). For example, in a screening promotion intervention, in addition to assessing changes in screening rates, one might measure providers' priorities for cancer screening, or system changes to facilitate screening follow-up. Modeling outcomes at different levels can contribute to understanding how the components of a multilevel intervention are related to individual patient outcomes and identifying barriers to, or facilitators of, implementation.

For example, Beresford et al. (6) used the ecological model to develop a multilevel intervention to change worksite environments and randomized 34 worksites to the intervention or comparison group. They assessed the effect of the intervention on both the worksite environment and individuals. Luepker et al. (5) conducted an intervention in four sites where they randomized 56 elementary schools to an intervention and used 40 schools as controls. They measured changes at both the school and individual levels. Similarly, Fuller et al. (20) implemented a multilevel intervention to increase access to sterile syringes among injection drug users and assessed changes by asking individuals about their approval or disapproval of drug use and the impact of the program on street-discarded syringes, crime, drug use, and the spread of HIV. They also asked pharmacists about their support for the intervention and their perceptions of its impact.

To illustrate both the challenges of conducting multilevel research and ways of addressing those challenges, we discuss a hypothetical multilevel intervention designed to reduce deaths due to colorectal cancer. Decreasing colorectal cancer deaths substantially probably will require several interventions, implemented at multiple levels in a coordinated way. Our hypothetical research team has decided that to reduce the death rate, it will be necessary to address the cancer care continuum [see box 3, (30)], including detection, diagnosis, and treatment (31). The team has decided that it is important to understand which components are more effective and which have a synergistic effect. The main intervention components will be 1) state-wide waivers for the Medicaid and Medicare programs to increase reimbursement for screening and diagnosis; 2) a community education program to increase knowledge about colorectal screening; 3) clinic interventions to improve

the efficiency and effectiveness of screening, diagnosis, and treatment (32); and 4) a physician education program to train them to identify high-risk patients.

Analytic Approaches to Assessing Multilevel Interventions Effects

A major issue for impact assessment is controlling for potentially confounding variables that can affect the internal validity of comparisons between intervention and control/comparison groups. The main ways of controlling for potential confounders are randomized experiments and quasi-experimental approaches. We briefly discuss these strategies below and illustrate how they might be used in our hypothetical example.

Randomized Experiments

In randomized experiments, one randomly assigns a subset of study units (eg, patients) to receive an intervention (eg, message about colorectal screening). In some designs, the nonintervention (control) units receive nothing, but sometimes they receive a different type of intervention (eg, message about screening with a certain emphasis or just a general health message). The main benefit of randomization is that one can increase the likelihood that individuals in the intervention and control groups will be comparable with respect to potentially confounding variables, even unmeasured ones. If part of our hypothetical intervention does not have to be the same for all the patients of the same physicians, such as an electronic message to promote colorectal screening (33–35), then it would be possible to randomize patients of the same physician, as Sequist et al. (34,35) have done.

For the component of our hypothetical intervention that educates physicians about screening strategies, it is conceivable that one could ask them to use the protocol only with a random sample of patients, but usually that is impractical and/or it is likely that there would be cross-contamination: what the physician learned and applied with some of his/her patients would influence the care of all patients. There could be a similar issue with randomizing physicians within practices. It would be possible to randomize practices to a quality improvement initiative to improve colorectal screening, as Farmer et al. (32), but because all physicians in a practice could be influenced by any changes in their practice, randomization of physicians within practices would not have the desired effect (36). In such situations, it would be better to use a technique referred to as group randomization (37–39), cluster randomization (40), block randomization (8,41), or group-based designs (42). In our example, we probably would assign all patients of a physician to the same patient intervention and all physicians in a practice to the same physician intervention. When one randomizes treatments to groups of patients who are more similar to each other than a random sample of patients, this clustering typically increases standard errors, and so a larger sample size is typically required to estimate a coefficient with a given precision than with random assignment of individuals.

Quasi-experiments

Often when an intervention is implemented at multiple levels, randomization of treatment is not desirable or possible (43). Cook

and Campbell (44,45) coined the term *quasi-experiment* to refer to designs that do not use randomization to create the comparisons of interest but instead use statistical techniques to account for possible threats to causal inferences such as baseline differences between treatment settings or secular trends in treatment.

In a study to improve breast cancer screening, Otero-Sabogal et al. (46) assigned two clinics to one condition (system redesign) and one clinic to a condition with an additional component (an additional tailored call). In a situation like this, not enough clinics can be randomized to make the distribution of potentially confounding variables comparable across intervention and control groups, so general linear mixed models were used to adjust for potential confounders when assessing the impact of the interventions. In a national study of an initiative to improve the quality of HIV care, it was not possible to randomize practices because the funding agency was not willing to allocate intervention funds that way (47). One of the main ways of controlling for confounding factors, which was used in both of these studies, is to include a variable representing those factors as covariates in a regression model predicting the outcome.

Another strategy for controlling for such variables when one can only use a limited number of comparison groups is to select comparison groups that are similar to the intervention groups to reduce the potential effect of confounding variables (45). By matching intervention and control/comparison units with similar values on potentially confounding variable(s), one can make the study groups similar with respect to the distributions of those variables. For example, in the national program to improve the quality of HIV care mentioned above, in which randomization was not possible, the investigators matched intervention and control clinics on the type of site (community health center, community-based organization, health department, hospital, or university medical center), location (rural, urban), number of locations delivering care, region, and number of active HIV cases (47).

Propensity score analysis is an alternative method of controlling analytically for potential confounding variables (48) by balancing the distributions of these variables in two or more groups. The propensity score is the probability of being treated conditional on the values of covariates. It can be estimated by fitting a logistic or probit regression or discriminant model in which treatment received is the dependent variable and the potential confounding variables are predictors and estimating for each individual or entity the probability of being in the treatment group. One can then use these probabilities to match subjects/entities, stratify them, use the scores in regression models (49), or to weight the data. Each of these methods serves the objective of manipulating the data to simulate treatment and control groups that have equivalent distributions of observed covariates, which therefore no longer confound the comparisons. An advantage of propensity score methods over regression adjustment is that the propensity score model is initially developed without reference to the outcome, so complex models can be selected while avoiding the risk of intentionally or unconsciously dredging the data to produce a desired inference (49).

Another approach to controlling for differences between intervention and comparison groups is instrumental variable analyses. An instrument is a variable that is not directly related to the

outcome under study but is predictive of the intervention. In our example, assume that we want to assess the effect of attending a clinic that promotes colorectal screening and that the probability of attending such a clinic is strongly related to the distance of a patient from such a facility. Also assume that patients close to and far from such a facility are otherwise similar. Then distance could be used as an instrument. One can think of the instrument (distance) as the experimental assignment and use that variable to assess the effect of attending a clinic that promotes screening (50,51). An advantage of instrumental variable analysis, compared with other strategies discussed above, is that it has the potential of controlling for unmeasured differences between intervention and comparison individuals or unity.

A common barrier to randomization is the desire to prioritize services to those most in need. A regression discontinuity design can take advantage of the relationship between a measure of need and the dependent variable of interest (45). In this approach, only units (eg, patients, clinics, areas) whose need score exceeds a cutoff value are assigned to the intervention condition. The analysis tests whether the relationship of score to outcome has a jump at the somewhat arbitrary point at which units are switched from control to intervention. Shadish et al. (45) describe the use of this design by researchers interested in how the introduction of Medicaid affected physician visits (52).

For our example, given known disparities in colorectal cancer screening rates, one could focus services on individuals who would benefit most from a program, such as patients with a familial history of colorectal cancer, and assign patients whose risk scores were above a preestablished level to an intervention (eg, special physician information session). The key to making valid inferences is that the assignment is based on a predetermined rule rather than being left to clinician discretion, which might introduce unknown confounders.

Intervention Design

In multilevel interventions, the appropriate design depends on the effects one wants to estimate. In our example, if one wanted to test the additive and synergistic effects on colorectal screening rates of a multilevel intervention that consists of the four components described above, then there would be a total of 16 conditions ($2 \times 2 \times 2 \times 2$) necessary. If all of these conditions are represented and all treatments are assigned at the community level, this would constitute a full factorial design (41). Although such designs are extremely informative, they also are complicated to implement, and more important, may require an impractically large number of units to fill all the cells of the design, in this case, 16 randomized communities even without any replication within cells of the design.

When the interventions can be assigned at different levels, more efficient designs become possible. For example, in our hypothetical example, the state-wide intervention (the waiver) would necessarily be applied at the state level, the community education intervention would be randomly assigned to communities within the state, and the clinic interventions would be randomly assigned to clinics within the participating communities. Such a nested, or split-plot, design (so called from its origins in agricultural experimentation) greatly reduces the number of units required at the higher levels of a multilevel structure while still making it possible

to estimate interactions among the multilevel intervention components and to have a large number of randomized units at the lower levels. Furthermore, randomizing lower-level interventions within higher-level units can improve the precision of comparisons. If clinics were randomized to the clinic intervention within each community, estimates of the intervention effect would be more precise than if the intervention were assigned at the community level, so effects were subject to confounding by community characteristics.

In such a design, it is necessary to ensure that the interventions only affect the units to which they are assigned; if clinics that received the intervention shared the information with other clinics in the same community, this “contamination” would bias (likely attenuate) estimates of the effect of that intervention. For similar reasons, in a study of promotion of colorectal cancer screening, a patient reminder letter was randomized at the patient level on the assumption that communication among patients at the same clinic was minimal, but a physician reminder through the electronic medical record was randomized at the physician level because receiving reminders for some patients might influence the physician’s treatment of his or her other patients (34).

To simplify the design, one can then consider which effects and interactions are most important to estimate. One then could evaluate the feasibility and cost of different designs that allow testing of those effects (41,53). Some interactions might not be worth testing because the components involved are clearly dependent in one another. For example, in our hypothetical colorectal screening example, one might decide that it is impractical or too expensive to study the effects of the physician and clinic interventions in a control community without a waiver program because the needed services would be financially inaccessible and just test the separate and interactive effects of a physician education program and clinic interventions in the waiver state.

Instead of conducting a full study of the non-waiver state, it might be possible to use externally collected data, such as those from the National Health Interview Survey or the National Health and Nutrition Examination Survey, to assess health behaviors and compare them to similar states without a waiver. One would select a matched state; that is, one that is similar to the intervention state on major potentially confounding factors. In the intervention state, one might assign practices to a physician-only intervention or a physician-plus-healthy-system intervention. One might try to imbed those practices in communities with or without a community intervention.

These design principles are illustrated by the program of the HIV Prevention Trials Network (HPTN) (http://www.hptn.org/research_studies.asp) to assess the feasibility of a multilevel intervention to increase HIV screening and linkage to care in the Bronx and Washington, DC. This quasi-experiment uses comparable surveillance data from these areas. Each of the intervention communities will have programs to increase social mobilization, with targeted messaging to promote testing, and implementation of the universal offer of HIV testing in emergency departments and during hospital inpatient admissions. Within each community, sites will be randomized to test the effectiveness of a financial incentive. At the individual level, study participants will be randomized to test computer augmented care.

Temporal Comparisons

Another strategy for assessing the impact of an intervention is to make use of changes in temporal trends (54). Shadish et al. [(45), page 171] argue that the interrupted time series design is one of the most effective and powerful of all quasi-experimental designs. In this design, one collects data on the same variable over time and assesses the extent to which the value of the variable, or the slope of change over time, is related to the intervention. This approach is particularly powerful if one can compare the time series of those who are and are not exposed to an intervention. For example, Price et al. (55) used a variation of this approach, to assess the impact of direct-to-consumer advertising and the release of clinical guidelines on the use of human papillomavirus DNA tests. Goldberg et al. (56) used interrupted time series this approach to assess the impact of computerized clinical reminders. Michielutte et al. (57) evaluated a cancer screening program this way.

In our example, we could track changes in colorectal screening rates before and after a community education campaign and assess whether the changes among patients attending intervention clinics are similar to, or different from, changes among those in comparison clinics. When there are not enough observations to model time series effects, investigators can still make a number of measurements before the intervention to establish a baseline for the assessment (58). One can then intervene with a subset of participants. For example, if there were a small number of clinics, one could make multiple measurements and then intervene with one clinic at a time and continue measurements. Murray et al. (54) review several uses of multiple baseline designs in oncology. O'Connell and McCoach (59) also provide an accessible summary of the different analytic methods that can be used for such comparisons.

Measurement Considerations

Other treatments of multilevel intervention research address measurement issues (28,60), but when assessing predictors or changes at different levels, measurement becomes much more complex than if there is only a single intervention at one level, no assessment of intervening variables, and only patient-level outcomes (61). In our example, if one is interested in the impact of a clinic-level intervention on provider knowledge, and whether clinic changes and provider knowledge are related to screening use, then multiple levels of measurement are necessary. One might, for example, ask physicians about both changes in clinic policies, to be used as a clinic-level variable, and their own knowledge to use as a physician-level variable. Similarly, one might use patient reports as a patient-level variable (eg, knowledge) or aggregate responses to form a clinic-level measure (eg, use of screening messages). In such situations, it is important to assess reliability of measurements at higher levels, as well as individual-level assessments (62).

Analytic Issues

Before choosing an analytic model, it is important to consider the relationships between different levels. Relationships can be many to one; for example, every patient sees only one physician in a study; that is, patients are nested within physicians. If units at different levels are nested in the higher level in many-to-one relationships

(eg, each patient sees only one of the physicians, each physician is associated with only one of the study clinics), the data are hierarchical. However, the relationships among levels can be more complicated because patients have multiple physicians or physicians contract with multiple health plans.

Hierarchical multilevel data can be analyzed in models that simultaneously assess the effects of variables at different levels and the influence of variables across levels (63). Hierarchical linear models or mixed models (59,63) can simultaneously examine effects of individual- and group-level variables and their interactions on an individual-level outcome. They also can model temporal effects (eg, overall trends and differential trends in intervention and control groups) (7,15,59). Hierarchical linear models are typically used with continuous outcomes. More generalized hierarchical approaches include logistic and Cox proportional hazard models (64), but the coefficients in those nonlinear models can be difficult to interpret when random effects are present.

Hierarchical linear models can estimate temporal effects even when some data are missing and/or the time of the observations varies across individuals (59). Furthermore, the correlated errors and within-group homogeneity characteristic of most grouped data can be appropriately modeled to obtain correct standard error estimates and inferences. Such models can estimate the net effect, the additive effects at each level, and even interaction effects (the extent to which a predictor at one level affects the association at another level). They also allow one to estimate the relative amounts of unexplained variation due to different levels (eg, patient, clinic, community). Other extensions of these models can accommodate nonhierarchical data structures, such as when patients are cross classified (eg, hospital by primary care physician) with full or partial completion of the cells or even when each patient receives a portion of his or her care from each of several physicians (65,66).

Some simpler models for such data have important disadvantages. If one treats patients as the units of observation, assigning them the characteristics of a higher level (eg, the knowledge score of their physician), one often can correctly estimate the association between the physician characteristic and individual-level outcome, but this approach usually incorrectly estimates the precision of the coefficient estimate unless variance estimation methods are used that account for the clustering (63). Individual-level data can be aggregated and the analysis conducted at the group level (42). This approach, referred to as ecological regression, has several important drawbacks (36,42,63). One loses information about individual variability, decreases substantially the number of observations, and the associations at the group level may indicate different associations than are present at the individual level (42,67).

Future Directions

Some researchers have concluded that substantial and sustained impact can only be achieved with multilevel interventions. However, there is very little data supporting this belief. Thus, it is incumbent upon those arguing for this approach to conduct multilevel research that explicates the contributions that interventions at different levels make to the desired outcomes. Only then will we know

whether and to what extent multilevel interventions are better than more focused interventions.

Multilevel interventions are more expensive and complicated to implement than are single-level interventions. Methods for analyzing multilevel data and associations are well developed, but designing a study that will allow the investigator to estimate the effects of interventions at different levels and the interaction of effects is complicated. Furthermore, there are many statistical challenges for such designs. For example, without knowing how the primary outcome and predictor variables vary within and between different units at different level, it is difficult to determine the sample sizes needed. Although issues related to cost-effectiveness are beyond the scope of this chapter, it also will be important to understand the cost and benefit of different intervention components to make a convincing case that they should be used. That will require creative designs, appropriate funding, and sophisticated analyses that account for the complex structure of such data. If we address those challenges, however, we are likely to gain greater insights into the kinds of interventions that can be implemented effectively and efficiently to improve health and health care for individuals with cancer.

References

- Manhart LE, Holmes KK. Randomized controlled trials of individual-level population-level, and multilevel interventions for preventing sexually transmitted infections: what has worked? *J Infect Dis*. 2005;191(suppl 1): S7–S24.
- Greiner KA, Engelman KK, Hall MA, Ellerbeck EF. Barriers to colorectal cancer screening in rural primary care. *Prev Med*. 2004;38:269–275.
- Stewart R, Niessen WJM, Broer J, Snijders TAB, Haaijer-Ruskamp FM, Meyboom-De Jong B. General practitioners reduced benzodiazepine prescriptions in an intervention study: a multilevel application. *J Clin Epidemiol*. 2007;60:1076–1084.
- Datta GD, Colditz GA, Kawachi I, Subramanian SV, Palmer JR, Rosenberg L. Individual-, neighborhood-, and state-level socioeconomic predictors of cervical carcinoma screening among U.S. black women. *Cancer*. 2005;106(3):664–669.
- Luepker RV, Perry CL, McKinlay SM, et al. Outcomes of a field trial to improve children's dietary patterns and physical activity. *JAMA*. 1996; 275(10):768–776.
- Beresford SAA, Bishop SK, Brunner NL, et al. Environmental assessment at worksites after a multilevel intervention to promote activity and changes in eating: the PACE Project. *J Occup Environ Med*. 2010;52(1 suppl):S22–S28.
- Bowen DJ, Beresford SAA, Christensen CL, et al. Effects of multilevel dietary intervention in religious organizations. *Am J Health Promot*. 2009;24(1): 15–22.
- Henke RM, Zaslavsky AM, McGuire TG, Ayanian JZ, Rubenstein LV. Clinical inertia in depression treatment. *Med Care*. 2009;47(9):959–967.
- Parkerton PH, Smith DG, Straley HL. Primary care practice coordination versus physician continuity. *Fam Med*. 2004;36(1):15–21.
- Pham HH, Schrag D, Hargraves JL, Bach PB. Delivery of preventative services to older adults by primary care physicians. *JAMA*. 2005;294(4): 473–481.
- Wee CC, Phillips RS, Burstin HR, et al. Influence of financial productivity incentives on the use of preventive care. *Am J Med*. 2001;110:181–187.
- Task Force on Community Preventive Services. Recommendations for client- and provider-directed interventions to increase breast, cervical, and colorectal cancer screening. *Am J Prev Med*. 2008;35(1 suppl):S21–S25.
- The COMMIT Research Group. Community Intervention Trial for Smoking Cessation (COMMIT): I. Cohort results from a four-year community intervention. *Am J Public Health*. 1995;85(2):183–192.
- The COMMIT Research Group. Community Intervention Trial for Smoking Cessation (COMMIT): II. Changes in adult cigarette smoking prevalence. *Am J Public Health*. 1995;85(2):193–200.
- Forster JL, Murray DM, Wolfson M, Blaine TM, Wagenaar AC, Hennrikus DJ. The effects of community policies to reduce youth access to tobacco. *Am J Pub Health*. 1998;88(8):1193–1198.
- van Empelen P, Kok G, van Kesteren NMC, van den Borne B, Bos AER, Schaalma HP. Effective methods to change sex-risk among drug users: a review of psychosocial interventions. *Soc Sci Med*. 2003;57:1593–1608.
- Atienza AA, King AC. Community-based health intervention trials: an overview of methodological issues. *Epid Rev*. 2002;24(1):72–79.
- Schensul JJ. Introductions to multi-level community based culturally situated interventions. *Am J Community Psychol*. 2009;43:232–240.
- Coady MH, Galea S, Blaney S, Ompad DC, Sisco S, Vlahov D. Project VIVA: a multilevel community-based intervention to increase influenza vaccination rates among hard-to-reach populations in New York City. *Am J Pub Health*. 2008;98(7):1314–1321.
- Fuller CM, Galea S, Caceres W, Blaney S, Sisco S, Vlahov D. Multilevel community-based intervention to increase access to sterile syringes among injection drug users through pharmacy sales in New York City. *Am J Pub Health*. 2007;97(1):117–124.
- Trickett EJ, Schensul JJ. Summary comments: multi-level community based culturally situated interventions. *Am J Community Psychol*. 2009;43:377–381.
- Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Ann Rev Pub Health*. 2006;27:297–322.
- DiClemente RJ, Salazar LF, Crosby RA. A review of STD/HIV preventive interventions for adolescents: sustaining effects using an ecological approach. *J Pediatr Psychol*. 2007;32(8):888–906.
- Molloy JC, Ployhart RE, Wright PW. The myth of “the” micro-macro divide: bridging system-level and disciplinary divides. *J Manag*. 2011; 37(2):581–609.
- Atkins MS, Frazier S, Cappella E. Hybrid research models: natural opportunities for examining mental health in context. *Clin Psychol: Sci Prac*. 2006;13(N1):105–108.
- Zatzick DF, Simon GE, Wagner AW. Developing and implementing randomized effectiveness trials in general medical settings. *Clin Psychol: Sci Prac*. 2006;13:53–68.
- Hemmelgarn AL, Glisson C, James LR. Organization culture and climate: implications for services and interventions research. *Clin Psychol: Sci Prac*. 2006;13:73–89.
- Klein K, Kozlowski SW. From micro to meso: critical steps in conceptualizing and conducting multilevel research. *Organ Res Methods*. 2000;3(3): 211–236.
- Nastasi BK, Hitchcock J. Challenges of evaluating multilevel interventions. *Am J Community Psychol*. 2009;43:360–376.
- Taplin SH, Anhang Price R, Edwards HM, et al. Introduction: understanding and influencing multilevel factors across the cancer care continuum. *J Natl Cancer Inst Monogr*. 2012;44:2–10.
- Zapka J, Taplin SH, Price RA, Cranos C, Yabroff R. Factors in quality care—the case of follow-up to abnormal cancer screening tests—problems in the steps and interfaces of care. *J Natl Cancer Inst Monogr*. 2010;40: 58–71.
- Farmer MM, Bastani R, Kwan L, Belman M, Ganz PA. Predictors of colorectal cancer screening from patients enrolled in a managed care health plan. *Cancer*. 2008;112(6):1230–1238.
- Sequist TD, Franz C, Ayanian JZ. Cost-effectiveness of patient mailings to promote colorectal cancer screening. *Med Care*. 2010;48(6): 553–557.
- Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. *Arch Intern Med*. 2011;171(7):636–641.
- Sequist TD, Zaslavsky AM, Marshall R, Fletcher RH, Ayanian JZ. Patient and physician reminders to promote colorectal cancer screening: a randomized controlled trial. *Arch Intern Med*. 2009;169(4):364–371.
- Moerbeek M, van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multi-center intervention studies. *J Clin Epidemiol*. 2003;56:341–350.
- Murray DM. *Design and Analysis of Group-Randomization Trials*. New York, NY: Oxford University Press; 1998.

38. Murray DM, Pals SL, Bitstein JL, Alfano C, Lehman J. Design and analysis of group-randomized trials in cancer: a review of current practices. *J Natl Cancer Inst.* 2008;100(7):483–491.
39. Katz DA, Muehlenbruch DR, Brown RL, Fiore MC, Baker TB. Effectiveness of implementing the Agency for Healthcare Research and Quality smoking cessation clinical practice guideline: a randomized, controlled trial. *J Natl Cancer Inst.* 2004;96(8):594–603.
40. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research.* London, UK: Arnold; 2000.
41. Kirk RE. *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, CA: Brooks/Cole Publishing Co; 1968.
42. Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. *Eval Rev.* 1999;23(4):418–444.
43. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomized controlled trial. *Am J Public Health.* 2008;98(8):1359–1366.
44. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Boston, MA: Houghton Mifflin Co; 1979.
45. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton Mifflin Co; 2002.
46. Otero-Sabogal R, Owens D, Canchola J, Tabnak F. Improving rescreening in community clinics: does a system approach work? *J Community Health.* 2006;31(6):497–519.
47. Landon BE, Wilson IB, McInnes K, et al. Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV study. *Ann Intern Med.* 2004;140:887–896.
48. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
49. D’Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* 1998;17(19):2265–2281.
50. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Ann Rev Public Health.* 1998;19:17–34.
51. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA.* 1994;272:859–866.
52. Wilder CS. *Physician Visits, Volume, and the Interval Since Last Visit, U.S. 1969.* Rockville, MD: National Center for Health Statistics; 1972.
53. Nair V. Screening experiments and the use of fractional factorial designs in behavioral intervention research. *Am J Public Health.* 2008;98(8):1354–1359.
54. Murray DM, Pennell M, Rhonda D, Hade EM, Paskett ED. Designing studies that would address the multilayered nature of health care. *J Natl Cancer Inst Monogr.* 2010;40:90–96.
55. Price RA, Frank RG, Cleary PD, Goldie SJ. The effects of direct-to-consumer advertising and clinical guidelines on appropriate use of a cervical cancer screening test. *Med Care.* 2011;49(2):132–138.
56. Goldberg HI, Neighbor WE, Cheadle AD, Ramsey SD, Diehr P, Gore E. A controlled time-series trial of clinical reminders: using computerized firm systems to make quality improvement research a routine part of mainstream practice. *Health Serv Res.* 2000;34(7):1519–1534.
57. Michielutte R, Shelton B, Paskett ED, Tatum CM, Velez R. Use of an interrupted time-series design to evaluate a cancer screening program. *Health Educ Res.* 2000;15(5):615–623.
58. Hawkins NG, Sanson-Fisher RW, Shakeshaft A, D’Este C, Green LW. The multiple baseline design for evaluating population-based research. *Am J Prev Med.* 2007;33(2):162–168.
59. O’Connell AA, McCoach DB. Applications of hierarchical linear models for evaluations of health interventions; demystifying the methods and interpretations of multilevel models. *Eval Health Prof.* 2004;27(2):119–151.
60. Charns MP, Foster MK, Alligood EC, et al. Multilevel interventions: measurement and measures. *J Natl Cancer Inst Monogr.* 2012;44:67–77.
61. Vernon SW, Briss PA, Tiro JA, Warnecke RB. Some methodologic lessons learned from cancer screening research. *Cancer.* 2004;101(5 suppl):1131–1145.
62. Marsden PV, Landon BE, et al. The reliability of survey assessments of characteristics of medical clinics. *Health Serv Res.* 2006;41(1):265–283.
63. Bryk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage Publications; 1992.
64. Shih JH, Lu S. Analysis of failure time data with multilevel clustering, with application to the Child Vitamin A Intervention Trial in Nepal. *Biometrics.* 2007;63:673–680.
65. Hill PW, Goldstein H. Multilevel modeling of educational data with cross-classification and missing identification of units. *J Educ Behav Stat.* 1998;23(2):117–128.
66. Rasbash J, Goldstein H. Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *J Educ Behav Stat.* 1994;19(4):337–350.
67. Robinson WS. Ecological correlations and the behavior of individuals. *Am Soc Rev.* 1950;15(3):351–357.

Funding

The first author received a small stipend from the National Cancer Institute to prepare the manuscript for publication and travel expenses to participate in the meeting where this paper was discussed. He distributed the stipend among the first three authors.

Note

All opinions are those of the authors and cannot be construed to represent those of the National Cancer Institute or the federal government.

Affiliations of authors: Yale School of Public Health (PDC) and Department of Internal Medicine (CPG), Yale School of Medicine, New Haven, CT; Department of Health Care Policy, Harvard Medical School, Boston, MA (AMZ); Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD (SHT).