The Multiple Baseline Design for Evaluating Population-Based Research

Nathan G. Hawkins, BA (Hons), Robert W. Sanson-Fisher, PhD, Anthony Shakeshaft, PhD, Catherine D'Este, PhD, Lawrence W. Green, PhD

Abstract: There is a need for pragmatic and rigorous research designs to evaluate the effectiveness of population-based health interventions. The randomized controlled trial (RCT) has limitations in its practicality, ethical appropriateness, and cost when evaluating population-based interventions. Like RCTs, the multiple baseline design can demonstrate that a change in behavior has occurred, the change is a result of the intervention, and the change is significant. Especially important practical advantages over the RCT are that this design requires fewer population groups and communities may act as their own controls. Advantages and methodologic limitations of the multiple baseline design are discussed, and where feasible, strategies to minimize the impact of its limitations are suggested. Recommendations for future research are included.

(Am J Prev Med 2007;33(2):162–168) © 2007 American Journal of Preventive Medicine

Introduction

Improving the health of populations requires interventions which target the population as a whole, in addition to those tailored to individuals.[1] Populations can be defined in a number of ways including socioeconomic class, workplace, behavior, geography, age or gender.[2] For example, cardiovascular health promotion programs such as the Minnesota Heart Health Program,[3] North Karelia Project,[4] and Stanford Five-City Project[5] used geographically defined population groups as the unit of analysis. In contrast, the Midwestern Prevention Project defined the population of interest by age, implementing community-wide interventions to reduce drug use in at-risk school children.[6]

Why Evaluate Population-Based Health Interventions?

In order to accurately determine an intervention's efficacy (the extent to which an intervention engenders a beneficial outcome under optimally controlled conditions) and effectiveness (the extent to which an intervention engenders a beneficial outcome under usual, as opposed to optimal, conditions) within a population unit, methodologically rigorous evaluation is needed.[7,8] The Cancer Action in Rural Towns project combined multiple interventions with previously demonstrated efficacy in 10 rural Australian communities in an attempt to improve cancer control.[9] However, when these efficacious interventions were applied to actual communities, no statistically significant reduction in cancer risk behaviors was observed.[10] This example highlights the need to rigorously test the effectiveness of an intervention at the population level before often-scarce resources are invested in their wide-scale implementation.

The efficiency of an intervention (the extent of an intervention effect relative to the cost of its implementation) is also an important consideration.[8,11] Economic analyses, such as total cost, cost effectiveness, cost–benefit, and cost–utility, provide indicators of efficiency.[12] An economic analysis of the Drink-less program, a brief primary health care intervention for problem drinking, found that the program had an estimated average cost per life-year saved of $645. This varied from $581 per life-year saved if Drink-less was implemented without follow-up or support, to $653 per life-year saved if practitioner follow-up and support were provided.[13] In this example, implementing the intervention without follow-up would be more cost effective, since any benefit of follow-up is outweighed by its additional cost. The most efficient intervention, however, might not be the most appropriate in all situations.

Evaluation of population interventions also potentially enhances the understanding of how and why interventions work, which may enable them to be improved.[8,14] Long-term assessments of school-based

From the Health Behavior Unit, School of Medicine and Public Health, Faculty of Health, University of Newcastle (Hawkins, Sanson-Fisher), Newcastle, and National Drug and Alcohol Research Centre, University of New South Wales (Shakeshaft), Sydney, New South Wales, Australia; Health Surveillance Unit, Centre for Military and Veterans' Health, University of Queensland (D'Este), Herston, Queensland, Australia; and Department of Epidemiology and Biostatistics, School of Medicine and Comprehensive Cancer Center, University of California (Green), San Francisco, California

Address correspondence and reprint requests to: Rob Sanson-Fisher, University of Newcastle, Room 267 David Maddison Building, Royal Newcastle Hospital, King Street, Newcastle NSW 2300, Australia. E-mail: rob.sanson-fisher@newcastle.edu.au.

© 2007 American Journal of Preventive Medicine • Published by Elsevier Inc.
0749-3797/07/$–see front matter
doi:10.1016/j.amepre.2007.03.020

smoking prevention programs have found programs based on providing young people with tools to resist social pressures to smoke[15,16] have greater long-term benefits than programs that educate young people about the health risks of smoking.[17,18] Such findings have provided a solid foundation for the further development of interventions in this area.

What Designs Are Suitable for Evaluating Population-Based Interventions?

Considering the importance of evaluating population-based interventions, appropriate evaluation methods must be sought. At present, the accepted gold standard for the evaluation of interventions in health care is the randomized controlled trial (RCT).[19] However, applying RCTs in population-based settings is challenging. Population groups, as opposed to individuals, become the units of analysis and recruitment of the typically large sample sizes required may be difficult and expensive.[20] Furthermore, the costs of implementing interventions across multiple entire populations are often prohibitive.[21]

Given these limitations, the Cochrane Effective Practice and Organization of Care Group endorses three alternative methodologies for evaluating population interventions: (1) the non-RCT, (2) the controlled before-and-after study, and (3) the interrupted time series design.[22] A number of other alternative designs may be potentially suited to population-based intervention analysis, such as case studies,[23] Baseline–Intervention–Withdrawal,[24] and Baseline–Intervention–Withdrawal–Intervention designs.[25] These are not widely endorsed, however, due to potential internal and external biases,[26] ethical and logistical challenges of removing an intervention from a population group,[27] and the potential lack of generalizability of findings.[28]

In a non-RCT, individuals or groups are allocated to experimental conditions using a quasi, or nonrandom, method.[22,29–31] While nonrandom allocation may be more convenient in some circumstances, it increases the probability that unmeasured characteristics that may influence the outcomes are not distributed randomly among comparison groups, introducing a systematic bias that could artificially exaggerate, or reduce, true intervention effects.[32,33]

In controlled before-and-after studies, experimental and control groups are not allocated randomly and single points of data collection occur before and after implementation of the intervention.[22,31,34,35] An advantage of this design is that it may be used with as few as two groups,[36] making it relatively inexpensive to implement. However, using single points of data collection does not allow assessment of the impact of other simultaneously occurring events on the outcomes.[35]

An interrupted time-series design allows the same group to be compared over time by repeatedly measuring and analyzing data.[22,31] As few as one population group may be used, and a baseline measure is used for control comparison.[37] The interrupted time-series design is limited, however, by its inability to assess the impact of concurrent events on the outcomes of interest.[21,22,37] A strategy to increase confidence that the intervention was responsible for a change in outcome is to conduct multiple time-series in multiple population units, each of which deliberately receives the intervention at a different point in time.[21,38–40] This staggered design is known as the multiple baseline design.

Multiple Baseline Design

Although identified as potentially useful as far back as 1968,[39] very little descriptive literature has focused on the multiple baseline design and few population-based researchers have implemented the design. While there are variations of the multiple baseline design,[40] the remainder of this paper will focus on assessment of interventions aimed at a single behavior across multiple population groups, such as targeting smoking cessation across separate geographic communities. Figure 1 provides a hypothetical example of such a multiple baseline design.

In this example, interventions are staggered across time and intervention populations. The repeated pattern of a reduction in the measured outcome following the implementation of the intervention in each community, along with an absence of substantial fluctuations in the data at other time points, suggests that the intervention is having an effect.[41] Conceptually a multiple baseline design may use as few as two groups to test an intervention,[42] reducing costs and alleviating some of the difficulties in obtaining a sufficient sample size required in RCTs. These benefits make the multiple baseline design a pragmatic tool for the assessment of population-based health interventions. However, the methodologic rigor of the multiple baseline design must also be considered.

To What Extent Is the Multiple Baseline Design Methodologically Rigorous?

To establish basic methodologic rigor, the multiple baseline design must be able to show that (1) a change in behavior has occurred, (2) the change is a likely result of the intervention, and (3) the change is statistically and practically significant.

A change in behavior has occurred. It is possible to test for changes in the outcome of interest by comparing means during baseline and post-intervention phases.[21,22] For example, in communities A, B, C, and D, in Figure 1, the mean level of behavior was lower during the post-intervention period (after April), compared to the

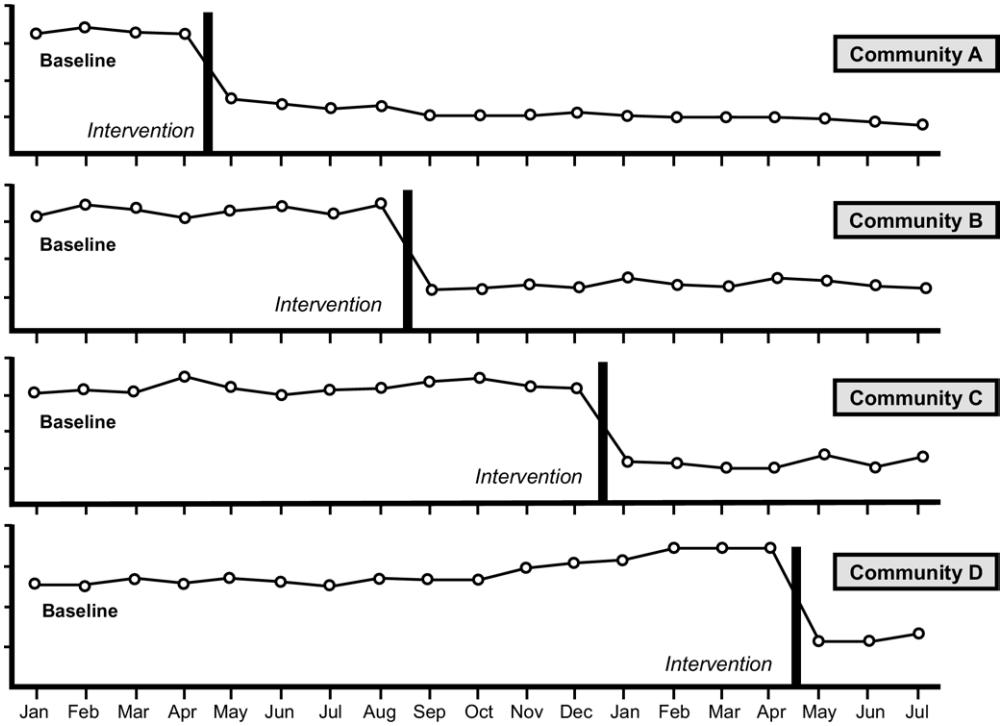

Figure 1. Hypothetical example of a multiple baseline design used to assess behavior change following an intervention in four communities.

baseline period, indicating that a change in the behavior has occurred.

The change is a likely result of the intervention. Examination of whether a change in behavior is a result of the intervention can occur in two ways. First, the repeated measurement of the defined outcome variable allows its trend to be determined.[21] A baseline trend that is neutral, or in the opposite direction to an observed behavior change, adds strength to the conclusion that the change in behavior resulted from the intervention.

Second, staggering the implementation of the intervention across time in different population units allows the researcher to monitor the possible influence of extraneous variables on the outcome of interest.[40] A change in behavior following intervention in one population group, coupled with the absence of change in other groups yet to receive an intervention, suggests that the change resulted from the intervention.[41]

The change is significant. Statistical techniques appropriate for analyzing repeatedly measured data are readily available and are suitable for multiple baseline data.[21,43] The major statistical challenge with repeatedly measured data is that they are typically autocorrelated. That is, the value of a measure at any given time may contain part of a value measured at an earlier time that can lead to inaccurate estimates of the intervention effect.[43–45] Statistical tools for removing autocorrelation, such as Auto-regressive Integrated Moving Average[43,44,46,47] or Independent Time Series Analysis of Autocorrelated Data[47,48] modeling, can be used for multiple baseline data and are designed to be applicable whether levels of autocorrelation are high or low.

In addition to statistical significance, it is important to consider the practical implications of intervention outcomes. A recent review of brief interventions for alcohol misuse implemented in primary care settings showed that approximately 2.5% of patients screened required an intervention for their alcohol misuse, and that approximately 10% of those receiving the intervention showed a reduction in alcohol use as a result of the intervention.[49] Despite the statistical significance of study results in this review, the value of committing resources to achieve a modest reduction in consumption among problem drinkers is, in practical terms, debatable.[49]

Additional Tests to Reinforce Findings from Population-Based Multiple Baseline Designs

Given that the multiple baseline design is currently a relatively novel, under-used methodology, one strategy to increase confidence in its scientific rigor is to examine how basic rules of causality complement data resulting from multiple baseline evaluations. The following eight rules of causality are derived from nine rules formulated by Sackett et al.[50]

Is there supporting evidence from other data sources? Complementary research findings have greater strength than inconsistent findings. For example, using a multiple baseline design, Biglan et al.[51] combined merchant education with other interventions to reduce the number of tobacco outlets willing to sell to young people. Their interventions and results were consistent with previous merchant education research,[52,53] adding credibility to the findings.

Is the intervention effect strong? Intervention effects in public health can be small and difficult to detect. It can be difficult to determine whether small changes are a result of the intervention or other confounding factors. Although the multiple baseline design allows comparison between groups at baseline, to monitor the potential impact of confounders, a strong intervention effect further reduces the probability that the change was due to a confounding factor. For example, following the introductions of random breath testing for alcohol consumption in Australian drivers, a 22% reduction in road fatalities was observed in the first year.[54] The size of the effect provided little doubt that random breath testing was the cause for the reduction in fatalities.

Are the findings consistent across each time series? It is desirable for an intervention to produce similar effects in the entire population group of interest. For example, in the hypothetical study represented in Figure 1, the experimental effect was consistent across all communities. Conversely, Biglan et al.[51] aimed to reduce tobacco sales to young people in four communities. Non-aggregated data showed the intervention effect was observed in only three out of the four communities. Greater confidence in the effect would be possible if the effect were consistent across all communities.

Is the intervention effect temporally consistent across each time series? The time lag of intervention effects may vary from immediate, such as a 29% reduction in local violent crime in the first year following a licensed premises violence prevention program,[55] to delayed, such as reductions in mortality from school-based anti-smoking campaigns that may not occur for decades.[56,57] Although it can be difficult to identify an appropriate time-frame within which an intervention effect should be expected, confidence in the relationship between the intervention and its effect can be increased if observed effects are temporally consistent across groups. In the hypothetical example in Figure 1, all four communities experienced an immediate reduction in behavior following the intervention, a consistent result that adds strength to the findings.

Is the intervention effect consistent with the complexity of the intervention? Public health problems are often multifaceted in nature.[58] Harms related to alcohol abuse, for example, affect individuals and the wider community. The potential causes for this harm vary from individuals' attitudes and beliefs[59] to societal norms about acceptable drinking behavior.[60] Given this complexity, a multifaceted, multicomponent approach might reasonably be expected to have more potential to reduce such harms. Projects such as the Preventing Alcohol Trauma Community Trial, have used a complex combination of behavioral change strategies, social support networks, and policy changes to reduce a range of alcohol-related behaviors.[61,62]

Although some interventions may have an effect that is disproportionately greater than their apparent complexity (e.g., taxation to increase the price of high-strength alcoholic beverages is commonly regarded as an effective policy to reduce problem drinking in developed countries),[63] the effect of most interventions is likely to be consistent with their complexity. That is, the greater the number of behavior and harm dimensions addressed, the greater the likely community-level effect.

Is the intervention effect consistent with fundamental behavioral knowledge? Interventions consistent with fundamental behavior change theories have added face validity. For example, media campaigns to encourage smoking cessation have used principles of Fishbein and Ajzen's theory of reasoned action.[64] Such theories lend weight to ideas that attitudes of individuals toward smoking may be targeted by outlining health problems related to smoking,[65] social pressures to quit smoking may be emphasized by focusing on the impact of secondhand smoke on nonsmokers (such as family members),[65,66] and smokers' beliefs about their ability to stop smoking may be bolstered by examples of their peers who have successfully quit.[66,67]

Is the intervention effect specific? If the effect is observed only in the targeted population, the likelihood that it is attributable to the intervention is increased. For example, evaluations of workplace bans on smoking in California during the early 1990s found that people working in "smoke-free" workplaces experienced less than half of the tobacco smoke exposure than did those from workplaces without "smoke-free" policies.[68] Of course, it would be unreasonable to expect that workplace bans on smoking would affect youth smoking rates, given that they were not the target population.

Is the intervention effect consistent in related fields? Some public health initiatives are developed following the successful implementation of similar strategies in other fields. For example, community prevention trials had been successfully used to reduce heart disease–related risk behaviors[5] before being used to reduce alcohol-related harms in the Preventing Alcohol Trauma Community Trial.[61,62] Engendering similar effects by implementing a simi-

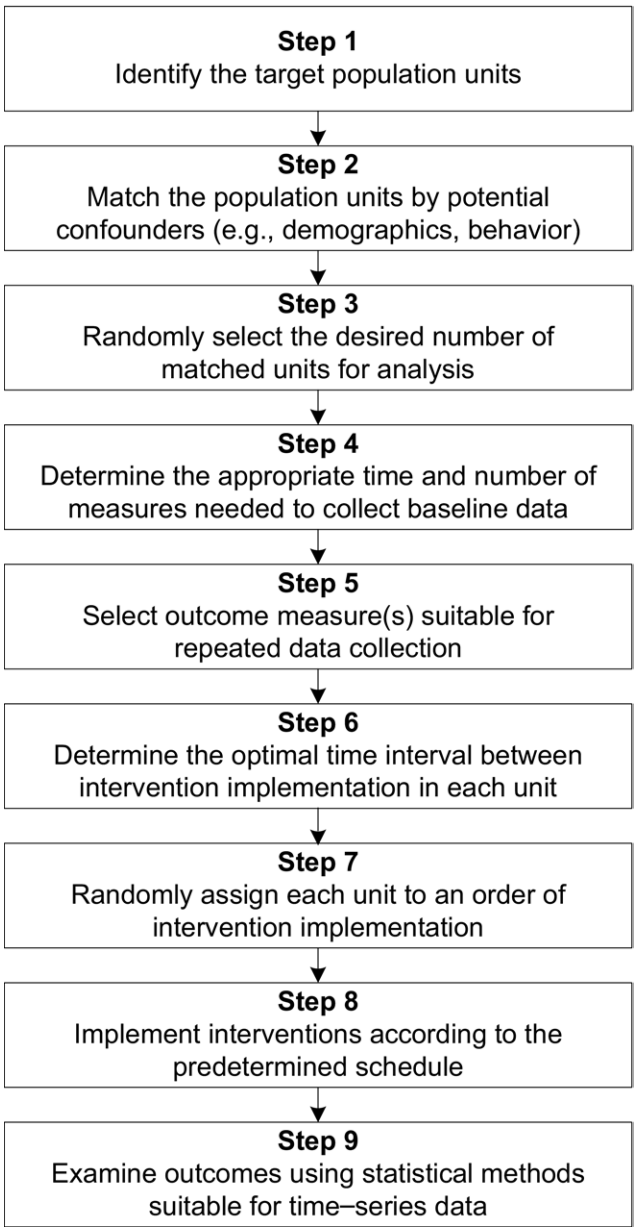

Figure 2. Algorithm of the decision-making process necessary to successfully implement a multiple baseline design in population-based settings.

lar intervention strategy for separate, but related, health problems, adds confidence that the intervention was effective.

Specific Methodologic Recommendations for the Multiple Baseline Design

A number of refinements to the basic principles of the multiple baseline design can be applied to improve both the external and internal validity of the results. These basic principles are briefly outlined in Figure 2.

Improving external validity. The ability of population-based research outcomes to be generalized is largely dependent on the extent to which experimental population groups differ from the wider populations of interest. In RCTs, matched samples and randomized selection and allocation procedures are typically used to minimize systematic differences between the units of analysis and the population of interest. Similar techniques can be applied to the multiple baseline design. Matching eligible population groups into clusters according to demographic characteristics or a behavior of interest allows similar groups to be identified and reduces possible systematic variability between them. Groups for inclusion in the study can then be randomly selected from each matched cluster.[69–71] This increases the probability that the selected populations are representative of the characteristics of the eligible population. Finally, population groups can be randomly allocated into the order in which they will receive the intervention. This minimizes the potential for bias where the intervention order reflects the extents to which participating groups are amenable to the intervention or the ease of researcher access.

Improving internal validity. The ability of the multiple baseline design to demonstrate that observed changes in the behavior of interest are a result of the intervention may be maximized by (1) ensuring the stability of data at baseline, and (2) identifying an optimal period of time between implementation of the intervention in each experimental group.

The stability of data at baseline can influence how data are interpreted; however, defining a stable baseline can be difficult. From the point of view of statistical analysis, a minimum of three data points are required for plotting any trend,[47] and a minimum of 10 data points are needed for statistical procedures such as the ITSACORR analysis.[21] From a practical point of view, data instability can be a result of either random variation or predictable, seasonal patterns. The latter needs to be taken into account in any baseline data collection. For example, surveys about recent alcohol consumption have found that persons interviewed in January are more likely to report at-risk consumption due to holiday-related celebrations in the month of December.[72,73] It may be necessary to define a time period for baseline measurement sufficient to account for such variations.

Stability in baseline data is likely to be optimized by using frequent and repeated measures.[21,43] An efficient method of obtaining such data is to utilize data routinely collected by the population of interest. This offers two advantages: a stable baseline can be established over an extensive time period, and data collected using resources external to the project may reduce data collection costs. Problematically, routinely collected data may have unknown reliability and validity, necessitating reliability and validity studies within the context of a project. At the very least, it is important to ensure measures are consistent across multiple groups of anal-

ysis and throughout the time period of interest. Unknown record-keeping shifts, differences in recording standards between groups, or even legislative changes can have unintended consequences, such as producing false trends or inflating error variance.

Improving the internal validity of the multiple baseline design also requires the identification of an optimal period of time between implementing the intervention in each experimental group. Too small a time interval between implementation in sequential population groups may obscure the full effect of the intervention, and too large a time interval may not allow sufficient time for completion of the study or may provide too much opportunity for other variables, extraneous to the intervention, to have an effect.[41] For example, educating students about the dangers of illicit drug use may take too long to disseminate through staff and students in order to have an effect, while implementing an intervention in one school every month may be too frequent to observe and attribute an effect to the intervention.

Conclusion

The multiple baseline design is a practical and methodologically rigorous design for the evaluation of population-based health interventions. It is a viable alternative to the RCT and offers researchers a design that can be conducted in a relatively small number of groups at potentially lower cost than alternative approaches, yet can be evaluated with rigorous statistical analyses. While the multiple baseline design has methodologic limitations, these can be minimized through careful planning. With the growing importance of population-based health interventions, there is a need for funding agencies and review committees to recognize methodologically adequate alternatives to the RCT, such as the multiple baseline design.

This research was supported by the National Health and Medical Research Council of Australia (project grant), National Cancer Institute, U.S. Department of Health and Human Services (Comprehensive Cancer Center grant).

Thanks to Marian Shanahan, health economist at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia, for comments on the economics section of the paper.

LWG was employed by CDC from 1999 to 2004, and has received various honoraria, reimbursements for chairing panels, consulting, speaking, since 2004. He served as a member of Board of Scientific Counselors for the National Human Genome Research Institute; and was a speaker, expert panel member, and consultant for other NIH, SAMHSA, and AHRQ units and contractors.

All of these agencies have some stake in the allocation of resources to the various types of research and evaluation discussed and criticized in the three papers on which I am a co-author and the introduction to them. Some of my university colleagues at UCSF could gain, others lose, resources for their research if the allocation of resources to specific types of research is influenced by this set of papers.

No other authors reported financial disclosures.

References

1. Lindholm L, Rosen M. What is the "golden standard" for assessing population-based interventions? Problems of dilution bias. J Epidemiol Community Health 2000;54:617–22.
2. Hingson RW, Howland J. Comprehensive community interventions to promote health: implications for college-age drinking problems. J Stud Alcohol Suppl 2002;14(suppl):226–40.
3. Luepker R, Murray D, Jacobs DJ, et al. Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program. Am J Public Health 1994;84:1383–93.
4. Vartiainen E, Puska P, Jousilahti P, Korhonen H, Tuomilehto J, Nissinen A. Twenty-year trends in coronary risk Factors in North Karelia and in other areas of Finland. Int J Epidemiol 1994;23:495–504.
5. Winkleby MA, Taylor CB, Jatulis D, Fortmann SP. The long-term effects of a cardiovascular disease prevention trial: the Stanford five-city project. Am J Public Health 1996;86:1773–9.
6. Pentz M, Dwyer J, MacKinnon D, et al. Multicommunity trial for primary prevention of adolescent drug abuse. JAMA 1989;261:3259–66.
7. Sanderson C, Haglund B, Tillgren P, et al. Effect and stage models in community intervention programmes;and the development of the model for management of intervention programme preparation. Health Promot Int 1996;11:143–56.
8. Ovretveit J. Action evaluation: what is it and why evaluate? Action evaluation of health programmes and changes: a handbook for a user-focused approach. Abingdon, Oxfordshire: Radcliffe Medical Press, 2002:1–17.
9. Hancock L, Sanson-Fisher R, Redman S, et al. Community actions for cancer prevention: overview of the Cancer Action in Rural Towns (CART) project, Australia. Health Promot Int 1996;11:277–90.
10. Hancock L, Sanson-Fisher RW, Perkins J, McClintock A, Howley P, Gibberd R. Effect of a community action program on adult quit smoking rates in rural Australian towns: the CART project. Prev Med 2001;32:118–27.
11. Russell LB. Opportunity costs in modern medicine. Health Aff 1991; 11:162–9.
12. Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, 1997.
13. Wutzke S, Shiell A, Gomela M, Conigrave K. Cost effectiveness of brief interventions for reducing alcohol consumption. Soc Sci Med 2001;52: 863–70.
14. Puska P. Do we learn our lessons from the population-based interventions? J Epidemiol Community Health 2000;54:562–63.
15. Sussman S. School-based tobacco use prevention and cessation: where are we going? Am J Health Behav 2001;25:191–9.
16. Vartiainen E, Paavola M, McAlister A, Puska P. Fifteen-year followup of smoking prevention effects in the North Karelia youth project. Am J Public Health 1998;88:81–5.
17. U.S. Department of Health and Human Services. Preventing tobacco use among young people: a report of the Surgeon General. Atlanta GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 1994.
18. Thomson EL. Smoking education programs 1960–1976. Am J Public Health 1978;68:250–5.
19. Augurzky B, Schmidt C. The evaluation of community-based interventions: a Monte Carlo study. Bonn: Institute for the Study of Labor, 2001.
20. Kirkwood BR, Cousens SN, Victoria CG, de Zoysa I. Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. Trop Med Int Health 1997;2:1022–9.
21. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. Prev Sci 2000;1:31–49.
22. Cochrane Effective Practice and Organisation of Care Group. The data collection checklist. Ottawa: Cochrane Effective Practice and Organisation of Care Group, 2003.
23. Loos C, Oldenburg B, O'Hara L. The planning of a community-based program to injury control and safety promotion in a rural community 2001;9:222–8.

24. Lavelle J, Hovell M, West M, Wahlgren D. Promoting law enforcement for child protection: a community analysis 1992;25:885–92.
25. Honnen T, Kleinke C. Prompting bar patrons with signs to take free condoms 1990;23:215–7.
26. Delgado-Rodriguez M, Llorca J. Bias 2004;58:635–41.
27. Resnik D, Zeldin D, Sharp R. Research on environmental health interventions: ethical problems and solutions 2005;21:69–101.
28. Paolo A, Bonaminio G, Gibson C, Partridge T, Kallail K. Response rate comparisons of e-mail and mail-distributed student evaluations. 2000;12: 81–4.
29. Deeks J, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:1–173.
30. MacLehose R, Reeves B, Harvey I, SHeldon T, Russel I, Black A. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. Health Technol Assess 2000;4:1–154.
31. National Health and Medical Research Council. How to use the evidence: assessment and application of scientific evidence. Canberra: National Health and Medical Research Council, 2000.
32. Beller E, Gebski V, Keech A. Randomisation in clinical trials. Med J Aust 177:565–7.
33. Schulz K, Grimes D. Generation of allocation sequences in randomised trials: chance, not choice. Lancet 2002;359:515–9.
34. Robson L, Shannon H, Goldenhar L, Hale A. Guide to evaluating the effectiveness of strategies for preventing work injuries: how to show whether a safety intervention really works. Cincinnati OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2001.
35. Dimitrov D, Rumrill P. Pretest-posttest designs and measurement of change. Can Fam Phys 2003;20:159–65.
36. Campbell DT, Stanley JC. Experimental and quasi-experimental design for research. Chicago: Rand McNally College Publishing Company, 1963.
37. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract Suppl 2000;17(suppl 11):S11–6.
38. McGuigan FJ. Experimental psychology: methods of research. Englewood Cliffs NJ: Prentice Hall, 1993.
39. Baer DM, Wolf MM, Risley TR. Some current dimensions of applied behavior analysis. J Appl Behav Anal 1968;1:91–7.
40. Barlow DH, Hersen M. Multiple baseline designs. In: Goldstein AP, Krasner L, eds. Single case experimental designs: strategies for studying behavior change. New York: Pergamon Press, 1984:210–51.
41. Albert P. Longitudinal data analysis (repeated measures) in clinical trials. Stat Med 1999;18:1707–32.
42. Schmidt C, Bultussen R, Sauerborn R. The evaluation of community-based interventions: group randomization, limits and alternatives. Bonn: Institute for the Study of Labor, 2000.
43. Kazdin AE. Statistical analysis for single-case experimental designs. In: Goldstein AP, Krasner L, eds. Single case experimental designs: strategies for studying behavior change. New York: Pergamon Press, 1984:285–324.
44. Wickizer TM, Kopjar B, Franklin G, Joesch J. Do drug-free workplace programs prevent occupational injuries? Evidence from Washington State. Health Serv Res 2004;39:91–110.
45. Skog O. Alcohol consumption and fatal accidents in Canada 1950–89. Addiction 2003;98:883–93.
46. Abeyshinghe T, Balasooriya U, Tsui A. Small-sample forecasting regression or ARIMA models? J Quant Econ 2003;1:103–13.
47. Crosbie J. Interrupted time-series analysis with brief single-subject data. J Consult Clin Psychol 1993;61:966–74.
48. Mullins R, Borland R, Hill D. Smoking knowledge, attitudes and behaviour in Victoria: results from the 1990 and 1991 Household Surveys. Melbourne, Victoria, Australia: Victorian Smoking and Health Program, 1992.
49. Beich A, Thorsen T, Rollnick S. Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ 2003;327:536–42.
50. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown and Company, 1985.
51. Biglan A, Ary D, Koehn V, Levings D. Mobilizing positive reinforcement in communities to reduce youth access to tobacco. Am J Community Psychol 1996;24:625–38.
52. Landrine H, Klonoff EA, Alcaraz R. Asking age and identification may decrease minors' access to tobacco. Prev Med 1996;25:301–6.
53. Feighery E, Altman DG, Shaffer G. The effects of combining education and enforcement to reduce tobacco sales to minors. A study of four northern California communities. JAMA 1991;266:3168–71.
54. Grabosky P, James M. Random breath testing in NSW. In: Grabosky P, James M, eds. The promise of crime prevention. Griffith, Queensland: Australian Institute of Criminology, 1995:20–1.
55. Australian Institute of Health and Welfare. A guide to Australian alcohol data. Canberra: Australian Institute of Health and Welfare, 2004.
56. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004;328:1519.
57. Peto R, Lopez A, Boreham J, Thun M, Heath CJ. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. Lancet 1992;339:1268–78.
58. Holder H. Community prevention of alcohol problems. Addict Behav 2000;25:843–59.
59. Simons JS, Gaher RM. Attitudes toward alcohol and drug-free experience among college students: relationships with alcohol consumption and problems. Am J Drug Alcohol Abuse 2004;30:461–71.
60. Mattern JL, Neighbors C. Social norms campaigns: examining the relationship between changes in perceived norms and changes in drinking levels. J Stud Alcohol 2004;65:489–93.
61. Voas RB, Holder HD, Gruenewald PJ. The effect of drinking and driving interventions on alcohol-involved traffic crashes within a comprehensive community trial. Addiction 1997;92(suppl 2):S221–36.
62. Holder HD, Gruenewald PJ, Ponicki WR, et al. Effect of community-based interventions on high-risk drinking and alcohol-related injuries. JAMA 2000;284:2341–7.
63. Hall W. British drinking: a suitable case for treatment? 2005;331:527–8.
64. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Reading MA: Addison-Wesley, 1975.
65. Lavack A. Canadian anti-tobacco campaigns: the past 10 years. Regina, Saskatchewan: Faculty of Business Administration, University of Regina, 2001.
66. Grey A, Owen L, Bolling K. A breath of fresh air, tackling smoking through the media. London: Health Education Authority, 2000.
67. McBride N, Midford R, Farringdon F, Phillips M. Early results from a school alcohol harm minimization study: the School Health and Alcohol Harm Reduction Project. Addiction 2000;95:1021–42.
68. Patten C, Pierce J, Cavin S, Berry C, Kaplan R. Progress in protecting non-smokers from environmental tobacco smoke in California workplaces 1995;4:139–44.
69. Kratochwill TR, Levin J. Single subject research. Strategies for evaluation change. New York: Academic Press, 1978.
70. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues. Boston: Houghton Mifflin, 1979.
71. Atienza AA, King AC. Community-based health intervention trials: an overview of methodological issues. Epidemiol Rev 2002;24:72–9.
72. Weir E. Seasonal drinking: let's avoid the "January effect." CMAJ 2003;169:1186.
73. Carpenter C. Seasonal variation in self-reports of recent alcohol consumption: racial and ethnic differences. J Stud Alcohol 2003;64:415–8.