
Available from: http://dx.doi.org/10.1017/S1368980012004958

Accessed from: http://hdl.handle.net/1959.13/1296114
ABSTRACT

Objective: The aim of this study was to review the methodological literature regarding evaluation methods for complex public health interventions broadly, and based on such methods, critically reflect on the evaluation of contemporary community based obesity prevention programs.

Design: A systematic review of the methods and community based literature was performed by 1 reviewer.

Results: The review identified that there is considerable scope to improve the rigor of community based obesity prevention programs through: prospective trial registration; the use of more rigorous research designs, particularly where routine databases including an objective measure of adiposity are available; implementing strategies to quantify and reduce the risk of selective non-participation bias; the development and use of validated instruments to assess intervention impacts; reporting of intervention process and context information; and more comprehensive analyses of trial outcomes.

Conclusion: To maximise the quality and utility of community based obesity prevention evaluations, program implementers and evaluators need to carefully examine the strengths and pitfalls of evaluation decisions and seek to maximise evaluation rigour in the context of political, resource and practical constraints.
Background

Excessive weight substantially increases the risk of a variety of chronic diseases including cancer, cardiovascular disease and diabetes\(^1\) accounting for over 2 million deaths each year\(^2\) and more than 30 million disability-adjusted life years\(^3\). While evidence regarding the effectiveness of obesity prevention initiatives is equivocal\(^4\) there have been calls from international organisations\(^5\), professional associations\(^6\) academic experts\(^7,8\) and the community\(^9\) for governments to take decisive action to mitigate the adverse impacts of obesity. Indeed, across the globe, governments have developed policy, set obesity prevalence targets\(^10\) and are implementing obesity prevention initiatives\(^11\).

Similar to efforts to reduce other chronic disease risk factors in the 1980’s and 1990’s\(^12\)\(^-\)\(^\)\(^13\), community based prevention programs which attempt to address multiple determinants of obesity through multi-component population wide strategies are frequently recommended\(^14\). Comprehensive evaluation of such initiatives is recommended to ensure that conclusions regarding program effectiveness are valid, to provide an understanding of why and how a program may have succeeded or failed, and to, facilitate program improvement, policy development and service investment\(^15\)\(^,\)\(^16\). As such, program evaluation is recognised as a key component of best practice community based obesity prevention practice\(^17\).

In order to improve the rigour of future community based obesity prevention programs to better inform policy makers and practitioners, the aim of this paper was to review the methodological literature regarding evaluation methods recommended for complex public health interventions broadly, and community based interventions specifically, and to critically reflect on the evaluations of contemporary community based obesity prevention programs in the context of such recommendations. In doing so, we sought to outline a number of issues we believe represent significant impediments to the quality of community based obesity prevention program evaluation, and offer some possible ways in which such impediments may be addressed.
METHOD

A review of the literature was performed in January 2011 to identify key elements of rigorous evaluations of community based interventions from the research methods literature and from methodological reviews of previous community based chronic disease prevention trials; and to identify the evaluation methods used in community based obesity prevention programs.

Guidance regarding rigorous evaluation of complex interventions

To identify recommended evaluation methods for community based interventions we reviewed: the United Kingdom’s Medical Research Council guidelines for developing and evaluating complex interventions\(^\text{18}\) recommendations from an expert group on research designs for complex, multi-level health interventions sponsored by the National Institute of Health and the Centre for Disease Control\(^\text{19}\) and selected research design and evaluation texts\(^\text{20,21,22}\). We also examined reviews of methodological issues of past community based health intervention trials\(^\text{12-13,23-25}\).

Community based obesity prevention programs

We employed systematic methods to identify previous community based interventions. While there have been various definitions of what constitutes a community based intervention\(^\text{13,26}\), those which characterised a ‘community’ along geographical boundaries such as cities, villages or regions are most common\(^\text{23}\) and are the subject of discussion in this manuscript. As such to be eligible for inclusion in the review, an obesity prevention intervention must have been implemented on a population basis across a defined geographic region (such as a town or city), include a measure of weight status, be published in English, and in a peer reviewed journal. Published reports of evaluation designs or study protocols of incomplete trials were also included. Community based interventions primarily focussing on reducing chronic
diseases such as cardiovascular disease or diabetes where obesity prevention was one of a number of risk factors targeted were excluded.

To identify such programs, Medline, EMBASE, Cochrane Central Register of Controlled Trials, and Google Scholar electronic data-bases were searched by one reviewer (LW) for articles published between 1990-2010. The search strategy for Medline is described in Appendix A, and was modified for other databases (a full search strategy for each database is available on request from the corresponding author). Reference lists of systematic reviews including the Cochrane review of interventions for the prevention of obesity\textsuperscript{27} and, narrative reviews or editorials on the issue of community based obesity prevention interventions\textsuperscript{28,30} were also searched.

\textit{Research synthesis}

Information from community based obesity prevention trials and texts providing recommendations regarding their evaluation was synthesised narratively in terms of issues considered to be important in the conduct of rigorous evaluations of community based interventions including\textsuperscript{13,18,22-23,25,31-32}:

1. Ethical and scientific conduct and reporting,
2. Research design,
3. Data collection,
4. Measures and,
5. Analysis.

\textit{Search results}

The electronic data-base search yielded 1197 citations. Following screening of titles and abstracts full texts of 33 manuscripts were reviewed. Of these, 12 were not population wide community based interventions, 5 targeted a broad range of chronic disease risks, 2 did not
include a measure of weight status, 1 targeted weight loss rather then prevention, and the intervention and evaluation methods of 2 trials were still under development or yet to be reported. These trials were deemed ineligible. The remaining 11 manuscripts\textsuperscript{33-43} described 10 eligible community based obesity prevention initiatives (Table 1)

1. Ethical and scientific conduct and reporting

\textit{Research ethics}

While impropriety has not been raised as an issue in the evaluation and reporting of previous community based obesity interventions, all program evaluations are recommended to begin with consideration of standards of ethical and scientific conduct and reporting. Ethics approval and monitoring of study procedures by independent ethics committees (or Intuitional Review Boards), and adherence to the World Medical Association Declaration of Helsinki Ethical Principles for Medical Research (http://www.wma.net/en/30publications/10policies/b3/index.html) are an essential part of program evaluations and a pre-requisite for publication in most peer reviewed journals. The American Evaluation Association’s Guiding Principles for Evaluators\textsuperscript{31} also provides guidance on proper professional conduct for evaluators and covers issues of competence, integrity, honesty, respect for people and responsibilities for general public welfare.

\textit{Trial registration}

The UK Medical Research Council recommends that trials are reported regardless of the ‘success’ of the intervention\textsuperscript{18}. Research has found that stakeholder interests can influence evaluation design decisions of program evaluators\textsuperscript{44} and can suppress\textsuperscript{45} or pressure evaluators to misrepresent study findings\textsuperscript{46}. The interests or inherent biases of obesity researchers have also been suggested to lead to misleading interpretations of research studies\textsuperscript{47}. Prospective trial registration is a scientific convention that can help protect against such influence by requiring evaluators to document specific details of the evaluation design, intended sample, measures and planned analysis on a publically assessable database prior to the research
evaluation taking place (Clinical Trial Registration). The requirement for registration of intervention trials was adopted by the International Committee of Medical Journal Editors in 2005 following the establishment of the WHO International Clinical Trials Registry Platform. While a number of current and planned community based obesity prevention programs have been registered with an internationally recognised trial registry, the extent to which unregistered evaluations adhered to planned evaluation protocols, or selectively reported trial outcomes is unable to be assessed. For trials in which have utilised a community-based participatory approach such as the Pacific OPIC project, prospectively describing trial methods on a register may represent a considerable challenge given evolving research methods owing to the shared decision making and collective engagement of community members, organizational representatives in each research phase (Viswanathan et al, 2004). Nonetheless, documenting changes in research methods in trial registers over the course of the study as they occur, and justifications for amendments in trial procedures may reduce the risk of selective reporting or other reporting biases.

**Scientific reporting**

The reporting of trial findings consistent with agreed standards such as CONSORT for randomised, or TREND for non-randomised evaluation designs is a requirement of many of the leading public health and general medical journals. Reporting of trial outcomes consistent with such standards has been a feature of some past published community based obesity prevention evaluations. Greater adherence to such standards improves transparent and consistent reporting of research trials and may facilitate subsequent synthesis of trial outcomes by systematic reviews.

**Protocol publication**

A reported limitation of previous community based health promotion programs has been that a lack of sufficient documentation of intervention and evaluation strategies and procedures.
has limited the ability to interpret study outcomes. Given word limit restrictions placed by non-electronic journals on submitted papers, the capacity to report the detail required to sufficiently describe evaluations of complex interventions such as community trials has been constrained. Two recent developments represent opportunities to redress this limitation. First, publication of evaluation protocols represents a mechanism whereby such detailed information is able to be provided. Publication of such protocols also serves as a further mechanism to address to scientific integrity of subsequent reporting of results through a public statement of evaluation intent. Second, publication of protocols and/or results in electronic open access journals (which do not have space restrictions) or housing additional intervention or evaluation information in other publically accessible websites has been recommended and employed successfully by some existing evaluations of community based obesity prevention programs.

### 2. Selection of research design

Previous evaluations of community based obesity prevention interventions have employed uncontrolled pre-post, post-test only comparison or quasi experimental designs with and without comparison conditions. Such designs are vulnerable to a variety of biases, particularly the influence of secular trends or confounding. The use of more internally valid research designs such as those employing methods of random allocation (experimental designs), or efforts to strengthen the internal validity of more pragmatic quasi-experimental evaluation designs will provide greater confidence in the extent to which any observed effects can be attributed to a program.

**Randomised experimental designs**

**Randomised and cluster randomised controlled trials**

Randomised (cluster) controlled trials are considered to represent the most internally valid evaluation design, and have been previously used to assess the impact of cardiovascular and cancer risk factor community based interventions. A randomised controlled trial may be particularly valuable in innovation testing phased evaluations. Ethical considerations, costs
and logistical difficulties in randomising and intervening in large geographically separated communities, however, are well documented as impediments to the use of such experimental trial designs in community-based prevention programs. Nonetheless, both the UK Medical Research Council and the recommendations from an expert group convened by the National Institutes of Health and the Centre for Disease Control recommend the use of randomised controlled trials (RCTs) for program evaluation whenever feasible. As a general rule of thumb, a sample of at least 10 intervention and 10 control communities is suggested as a minimum number of sites for the conduct of experimental trials of community based interventions.

**Stepped wedge cluster randomized trial**

A stepped wedge cluster randomized trial may overcome a number of constraints to the conduct of traditional randomised and cluster randomised community based trials. A stepped wedge design involves measurement of trial outcomes (i.e. weight status) being repeatedly undertaken simultaneously across a number of clusters (i.e. communities) over a period of time. In the context of such measurement, the delivery of the intervention occurs sequentially across communities, the order being randomly determined. Comparisons are made between groups at the intervention section of the wedge (see Figure 1.) as well as within groups before and after the intervention. Advantages of the stepped wedge design for community based prevention interventions include the receipt of intervention by all communities (overcoming ethical concerns regarding withholding intervention), the congruence of the design with intervention capacity constraints and typical schedules of program rollout, the requirement of fewer clusters than a conventional randomised cluster trial; and the capacity of the design to detect and control of underlying trends and time effects. While the design requires repeated assessment of outcomes (e.g. weight status) which may be expensive and pragmatically challenging, the use of routinely collect weight status data, such as that utilised to assess the impact of Romp and Chomp and the Healthy Living Cambridge Kids obesity prevention programs represents one approach to achieving these benefits at relatively low cost.
**Quasi-experimental designs**

**Regression discontinuity**

Where program funding permits the use of large numbers of communities but ethical or political considerations deem randomisation not acceptable, regression discontinuity designs have been recommended as a potentially attractive alternative to experimental randomised designs\(^{19-20}\). Together with randomised trials, regression discontinuity designs represent the only other trial design which can provide an unbiased estimate of intervention effect\(^{19-20}\). Rather than random assignment, in regression discontinuity designs, researchers assign communities (or participants) to intervention or comparison conditions based on their exceeding a designated cut point on a pre-intervention assignment variable (e.g. BMI). A regression line through such an outcome variable, which discontinues at the point of assignment for control group communities is taken as evidence of an intervention effect. The design may be particularly useful for community based obesity programs targeting socioeconomically disadvantaged groups, where communities are assigned to receive an intervention based on their score on a measure of disadvantage. Owing to co-linearity between the assignment and outcome variables, however, regression discontinuity designs require substantially more research participants (or communities) compared with a similarly powered randomised trial\(^{19}\).

**Interrupted time series**

In a time series evaluation of a community based obesity intervention, repeated measures of weight status are required for a period prior to, during, and following intervention delivery in one or more community. Comparisons are made between the change in level or slope of the outcome prior to and following the intervention to assess intervention effect\(^{19}\). The addition of comparison communities which do not receive the intervention improves the internal validity of the design. Like the stepped wedge cluster randomised trial design, the repeated measures allows for control of trends and time effects, and is typically a more rigorous evaluation.
design than before and after trial designs\textsuperscript{19}. While the use of routinely collected weight status data from institutions such as health services would permit a relatively low cost evaluation using this design, the lack of such obesity surveillance data in most countries and communities has been previously identified as a significant impediment to its use\textsuperscript{29}.

**Controlled and uncontrolled pre-post trials**

Pre-post designs, where outcome assessments are taken at a single point in time prior to and following the delivery of an intervention represent the most commonly utilised evaluation design in community based interventions. Such designs, however, face a number of threats to internal validity\textsuperscript{19}. Where the inclusion of controlled designs is feasible, program evaluators should seek to maximise the number of intervention and comparison communities to help account for secular trends\textsuperscript{19-20}. Stratifying, or matching communities in such designs based on factors known to be closely associated with population weight status is also recommended to reduce the risk of selection bias and to maximise statistical power by increasing precision\textsuperscript{20,23}. Identifying communities with comparable characteristics, however, represents a considerable challenge\textsuperscript{13}. Even with large numbers of communities, however, non-randomised pre post designs should not be relied upon alone to infer causal inference\textsuperscript{19}.

With the exception of the Health Promoting Communities: Being Active Eating Well initiative\textsuperscript{41} community based obesity prevention interventions reviewed in this manuscript that have employed a controlled design to date have not employed formal stratification or matching techniques. Further, a number of such obesity prevention initiatives\textsuperscript{37-39,42} selected intervention communities on the basis they had existing infrastructure to facilitate intervention delivery, unlike comparison communities. Such known non-equivalence between groups on factors such as community capacity or infrastructure represents a significant threat to the internal validity of trial findings in principal, as such factors are difficult to measure reliably and adjust for in analyses of program outcomes\textsuperscript{55}.
Strengthening quasi-experimental designs

As a possible addition to all quasi experimental designs, evaluators can maximise the internal validity of program evaluations through the use of non-equivalent dependent variables. For example, if a community based intervention is delivered to all students attending government schools only, assessments of weight status of government and non-government school students in the intervention communities pre and post intervention delivery could be conducted. As students of both schools are likely to be subject to non-intervention factors in the community which may impact on weight status (such as food marketing or macroeconomic policies or events), but only those attending government schools are exposed to the intervention, a differential impact in the change in weight status over time would provide greater evidence of an intervention effect. A second strategy to enhance the internal validity of quasi-experimental designs is via assessments of a dose response relationships between exposure to intervention activities and trial outcome (e.g. weight status). Dose response analyses have been found to be a valuable feature of past community base cardiovascular disease risk factor intervention. Neither strategy, however, appears to have been employed as an adjunct to evaluations by community based obesity prevention program evaluations employing quasi experimental designs included in this review.

3. Data collection

Survey methods

Obesity prevention interventions have typically utilised survey methods to assess intervention outcomes. Instances where survey participation rates are low, attrition is high or either rate is different between groups limit the validity of inferences regarding intervention effects.

Research participation

Despite high anthropometric survey participation rates among New Zealand and French obesity prevention programs, participation rates for programs in the U.S and Australia have
typically ranged between 29-58%\textsuperscript{38-39}. Research suggests that obese persons may be less likely to participate in research requiring anthropometric assessments\textsuperscript{56}. Examination of the potential for selective non-participation bias by comparing the characteristics of participants with non-participants\textsuperscript{48} has not, however, been reported in any obesity prevention program included in this review. Identifying non-response bias may be particularly important in repeat cross sectional evaluations of intervention effect (where independent cross sectional samples are drawn from communities pre and post intervention), as small changes in the likelihood of participation in overweight participants at follow-up could result in artefactual post-intervention reductions in estimated obesity prevalence. As a crude illustration; assuming a pre-intervention prevalence of child overweight of 25% (250/1000), a relative 20% reduction in the likelihood of participation of overweight persons at post-intervention assessments would reduce post-intervention prevalence to 21% (200/950). Hypothetically, such changes in the propensity of overweight child participation may be an unintended consequence of the intervention itself, if intervention exposure increases perceived obesity stigma and reduces children’s interest in participation. Such reductions could be misinterpreted as a public health intervention success.

Evaluation texts and methodological reviews recommend that evaluators seek to reduce the risk of non-response bias by employing intensive recruitment strategies to maximise study participation such as the pre-notification and promotion of study participation, endorsement of the research by credible individuals or organisation, participation incentives, and multiple follow-up reminders\textsuperscript{57-58}. For trials that include a comparison condition, reporting participation rates among both intervention communities and comparison communities would allow an assessment of the potential risk of differences in non-participation between groups. More rigorous strategies, such as the recording and analysis of information regarding the weight status of non-participants would provide more compelling evidence regarding the possible influence of non-participation, if permitted by ethics committees or Institutional Review Boards. For example, Booth and colleagues\textsuperscript{59}, describe a feasible, inexpensive and
valid means of collecting weight status information of children not participating in school based assessments of height and weight by having classroom teachers match the morphology of such children to those of participants.

Research attrition

Research attrition can threaten the internal and external validity of a community based intervention evaluation if characteristics of participants not providing follow-up data differ from those that do, and if the characteristics of participants are dissimilar between intervention and comparison groups. Such differences can lead to over or under estimates of intervention effect. With the exception of the Be Active Eat Well program, participant attrition (loss to follow-up) among child obesity prevention programs with a cohort design in this review ranged from 25%-45% and differed substantially between groups in the case of the Shape up Sommerville program (intervention 39%; comparison 25%). Even when loss to follow-up is small and not differing significantly between groups, as was the case for the Be Active Eat well initiative (approximately 16% loss to follow-up per group), the potential for bias to influence study findings remains present. For example, in this trial 34% of study withdrawals in the intervention group were due to parental concern regarding their child’s self-esteem following weight assessment; a reason for withdrawal which was not cited at all for children in comparison communities. Two of the reviewed cohort studies formally examined the potential for bias due to differences between participants and those lost to follow-up as recommended. The Healthy Cambridge Kids initiative reported that participants who did not provide follow-up data were more likely to be older, Asian, and less likely to pass fitness tests. Project APPLE however found no meaningful differences between participants and those lost to follow-up.

Similar to those strategies to maximise study participation, evaluators should seek to employ intensive strategies to minimise study attrition such as through offering incentives, alternate locations for data collection, and repeat reminders via mail and telephone report attrition.
rates by group\textsuperscript{48,60}, and attempt to formally assess and control for bias due to study attrition in analyses.

**Routinely collected institutional data-bases**

Given the cost and feasibility challenges of collecting survey data, the use of routinely collected data such as health services records has been suggested as an inexpensive, alternative for program evaluation that does not require the development of new data collection systems and methodologies, often allows for comparisons with other regions or jurisdictions, can be used retrospectively, and may be less likely to be subject to bias due to non-consent\textsuperscript{63-64}. Despite these potential benefits, a number of limitations exist regarding the use of routinely collected data for evaluating public health interventions, including reliability between recording personnel, changes in data recording classifications, recording practices and systems, and limited population representativeness\textsuperscript{63}. While evaluation of obesity programs using such data needs to be mindful of these limitations, routinely collected information has a capacity to provide a valuable source of data for community based obesity program evaluations, particularly where they are accessed by a large proportion of the target population. For example, routinely collected BMI data were available from maternal and child health services in Victoria, Australia for approximately 60\% of all preschool aged children and were used to evaluate the Romp and Chomp program\textsuperscript{37}. Similarly height and weight data routinely collected annually as part of the physical activity curriculum was used to evaluate the effectiveness of the Healthy Living for Cambridge Kids initiative\textsuperscript{40}.

4. Outcome, impact and process measures

While a number of program evaluation frameworks exist, the inclusion of measures which assess the outcome of the intervention (as specified by the program aim); program impacts on intermediary factors (typically described as part of program objectives), intervention processes, and context are most frequently recommended by evaluation texts and reviews of past cardiovascular community based interventions\textsuperscript{18,22,24,25,65}. Importantly, the selection of
evaluation measures should also be guided by program logic models of how an intervention is intended to produce the intended intervention effect.

**Outcome**

*Weight status*

Given the limitations of self-reported measures of weight status\(^6\) and the potential for such reports to be reactive to assessment\(^2\) objective measures of would appear important in program evaluations looking to examine the effect of a community based intervention on population adiposity. Body Mass Index is the most widely utilised measure of weight status in previous studies\(^3,34,37-41,43\) and represents a relatively inexpensive and feasible measure of adiposity\(^4\). To provide information regarding changes in both body and fat composition, additionally, assessments of waist circumference and skin-fold thickness has been recommended, particularly as interventions including physical activity promotion can actually increase lean body mass and reduce adiposity without any change in Body Mass index\(^4\). Such assessments should be conducted and reported in accordance with standard measurement protocols to reduce measurement error and facilitate comparison across trials and data pooling in meta-analyses.

*Adverse events*

In-order to evaluate the merit of community based intervention to prevent obesity, both the benefits and adverse effects of an intervention need to be considered. As such, potential harms which may arise from community based interventions should be hypothesised and assessed as part of program evaluations\(^5,18,67\). Despite a few exceptions, assessment of harms have largely been overlooked in previous or planned community based obesity prevention intervention evaluations included in this review. The Be Active Eat Well child obesity program was the only program which explicitly included measures of harm such as the prevalence of underweight, weight loss attempts, weight based teasing and unhappiness\(^39\). Similarly, assessments of the impact of the OPIC community interventions will include assessments of
perceived body image (at least in a sub sample of participants).42 While individual measures of intervention harm are important, the potential adverse impacts at the setting or community level should also be considered by evaluators. For community safety, the use of committees (which include representatives from the community) to monitor adverse intervention effects and enforce ‘stopping rules’ in instances where adverse events exceed an acceptable level has been suggested as important65.

Impact

Healthy Eating, Physical Activity and Sedentary Behaviour

As weight status changes are mediated by dietary improvements to reduce excessive energy intake, and/or increased physical activity to increase energy expenditure, assessments of these behaviours can enhance the internal validity of program evaluations and provide evidence of the mechanism behind any intervention effect. For example, greater confidence of an intervention effect would be characterised by an inverse association between weight status, and healthy eating and physical activity.

In their assessment of healthy eating and physical activity, most obesity prevention program evaluations rely on brief self-reported measures which are vulnerable to social desirable responding4. Only a few trials included in this review, indicated that such measures had been previously demonstrated to be valid or reliable33-34,39. While short item questionnaires may represent the most feasible method to assess changes in broad dietary or activity patterns, such items represent crude measures and often are not sufficiently sensitive to detect small but meaningful changes at a population level. Indeed, in community based programs, these self-reported measures have been found to contradict objective measures. In the APPLE project, for example, self-reported physical activity among children in intervention communities significantly decreased relative to control community children at the 1 year follow-up, in contrast to accelerometry assessments which significantly increased counts among intervention community children over the same period33.
The use of valid tools is key to robust program evaluations. A number of papers have been published providing guidance to program evaluators regarding the selection of measures to assess physical activity and nutrition in the context of purpose, capacity, skill and resource constraints. In principle, obesity intervention programs should seek to employ the most rigorous behavioural assessments which such constraints allow. The use of objective measures of physical activity such as accelerometry or fitness tests employed by previous interventions or more rigorous dietary assessments such as 24 hr food recall methods may be feasible for interventions with a small sample. For larger trials, the inclusion of such assessments on a sub sample of the population could be considered in addition to the broader use of validated self-reported physical activity or food frequency questionnaires as a means of measurement triangulation.

Community environment

Community based interventions often seek to reduce population adiposity through modifying community environments, such as the availability of accessibility of healthy foods or physical activity opportunities in key community settings (i.e schools). For these interventions, community environmental characteristics which are the target of intervention represent important mediating variables and program impacts in their own right. While a number of trials in this review have assessed community environments as part of program evaluations, most have not reported the validity of the methods used to do so. The psychometric properties of a number of tools to assess obesogenic characteristics of the home (e.g. family eating routines), childcare (staff prompting child activity), and neighbourhood environments (availability of physical activity facilities) have recently been published and their use would greatly strengthen the quality of future evaluations. To our knowledge, however, there remains a lack of validated instruments for assessing other key community settings such as schools. The development of valid tools suitable for population wide assessments of
community settings and organisations would be particularly beneficial for program evaluators and should be considered a research priority for the field.

Given the cost and complexity of evaluating individual health behaviour risks or chronic disease conditions noted in previous community based health promotion programs, the use of community level environmental indicators has been proposed as a more feasible and cost effective alternative when significant resource constraints exist. The use of community environment measures in this way would require the availability and selection of valid measures of environmental characteristics known to be associated with population overweight and obesity, healthy eating or physical activity. Some simple and objective environmental proxy measures already exist. For example, grocery store shelf space has previously been found to reflect changes in some community dietary indicators with similar power to individual surveys at one tenth of the cost.

**Process**

Measures of the extent to which the intervention has been delivered, it’s reach and variability, often termed process evaluation, are recommended by the UK MRC as an important means assessing intervention fidelity and exposure in complex interventions. Process evaluation can also provide insight into why and how an intervention worked, or failed, and how it may be improved. Process evaluation should be conducted to the same methodological standard and reported as thoroughly as assessments of intervention outcomes. Process evaluations reported as part of planned or previous community based obesity prevention evaluations have varied in the extent to which in they have assessed intervention delivery, reach or dose. Among the more comprehensive process evaluations have been the Romp and Chomp and Be Active Eat Well initiatives that were guided by explicit program logic models and assessed exposure of the community to the intervention through changes in community capacity, and organisational policies, practices or environments. The lack of quality process information, particularly that related to
intervention reach and dose has previously been criticised as an impediment to the interpretation of community based obesity prevention interventions.

**Context**

The effectiveness of community based interventions are undoubtedly influenced by the social, political and organisational contextual factors occurring at the time of implementation. A common criticism of process evaluation is that it presupposes and mechanistically assesses specified intervention components. While this has merit, it ignores other contextual factors which could operate as effect modifiers. As such, the UK MRC recommends that, in addition to measurement of intervention process, the evaluation of complex interventions assess, monitor and document important changes in community context over the life of the project. The Californian Endowment’s Healthy Eating. Active Communities Program explicitly states an intention to collect and synthesise context data, via interviews, surveys and focus groups with community members, program stakeholders and policy makers. While not stated as a context evaluation, other initiatives included in this review collect and reported context information such as other interventions occurring in study communities prior to, or during intervention implementation, or changes in media activity and resource contributed by local agencies. As context evaluation is a developing science, little explicit guidance is available regarding what and how such information should be collected and utilised.

**5. Analysis.**

Selection of the analytical procedure to assess the impact of a community based approach to obesity prevention should, of course, be guided by the evaluation design. As communities are typically the unit of analysis when assessing the effectiveness of community based interventions, the use of generalized linear mixed models (also referred to as multi-level models, random-effects models, hierarchical linear models or covariance component models) are often most appropriate. Such analytical techniques account for intra-class correlation and allow for individual and community level influences to be examined simultaneously.
Encouragingly, trials included in this review employed such techniques and avoided unit of analysis errors previously documented in the obesity literature.

Methodological reviews of previous community based chronic disease prevention interventions have attributed modest intervention impacts, in part, to high rates of population migration in and out intervention and comparison communities, or inconsistent intervention exposure, diluting the intervention effect. Such factors also represent a risk to evaluations of community based obesity prevention programs. The Shape-up Sommerville child obesity prevention program, for example, reported that 27% of children had moved out of intervention or comparison communities within the first 12 months of the initiative. Variable intervention implementation has also been reported. Two supplementary strategies may be useful in managing issues of intervention exposure when describing intervention effects in a community. First, a plausibility analyses, where by comparisons are made between those receiving the intervention and those who did not after adjusting for confounders. Second, a dose response analyses where the effect of the intervention is examined according to the level of exposure to the intervention. Despite the usefulness of such analyses, they have not been utilised in any of the trials included in this review. Considering issues of intervention exposure, however, children residing within intervention communities but attending school outside the region will be excluded in the analysis of intervention effects of the Fiji OPIC initiative.

**DISCUSSION**

The findings of this review illustrate that there is opportunity for greater cross disciplinary learning from the evaluation experiences of past community based chronic disease risk factor interventions, and considerable scope to improve the rigour of community based obesity prevention interventions to better inform future policy and practice. Important methodological limitations apparent in a number of included trials in this review, particularly in trial design and measurement represent considerable impediments to inferences of causal attribution in current community based obesity prevention interventions.
Internationally, government and non-government organisations invest considerable sums in community based initiatives to prevent excessive weight gain, often without or with insufficient funding for rigorous program evaluation. Quality evidence regarding the effects of community based interventions is required to assess the community benefit of such expenditure, to help maximise future investment in obesity prevention and to facilitate more timely improvements in the health of populations. Under-resourced evaluations of community based interventions offer little quality evidence order to inform public health policy and practice. Selective investment in critical-mass funding of large, rigorously evaluated community based interventions may represent a more efficient strategy to yield robust evidence of intervention effects. Funding comprehensive evaluations of targeted community based obesity prevention interventions should therefore represent a priority for governments and other funding agencies.

While the review provides useful guidance for the conduct of future community-based obesity prevention intervention evaluations, there are a number of opportunities for further methods development to advance the field, particularly for initiatives utilising community-based participatory approaches. Community based participatory research emphasizes reciprocal transfer of expertise between community and researchers, sharing of decision making power and mutual ownership of the research process and products, and is being increasingly used to address a variety of public health issues in the community (Viswanathan et al, 2004; Cargo et al, 2008). While such approaches can facilitate conceptualization of the problem, culturally appropriate intervention development and delivery, and improve data collection and analysis and interpretation, (Cargo et al, 2008) participatory approaches introduce a number of unique challenges for evaluators. For example, evolving study procedures and processes may preclude the development of study protocols a-priori, and increasing the specificity of research to a community can reduce the generalizability of the research outcomes (Cargo et
While broad guidelines exist to assist with the conduct of quality participatory research (Green et al, 1995) further development of concepts, measures to assess participatory processes and constructs and more sophisticated analytical techniques to deal with such complexity are required (Sandoval et al, 2011).

Greater examination of factors which may mediate or moderate an intervention effect may also present an opportunity to further an understanding of the casual pathways in which interventions operate, identify particular groups in the community for which an intervention may be beneficial and can facilitate the design and effectiveness of future community based interventions (MacKinnon et al, 2008; Cerin et al, 2009). Despite the benefits of such research, few trials have examined such relationships in obesity prevention research generally (Cerin et al, 2009B). Encouragingly, within community based obesity prevention research Johnson and colleagues recently published a multi-level analysis of the Be Active Eat Well initiative, in which a moderating effect of the intervention was found for the relationship between the frequency of TV during meals and BMI (Johnson et al, 2012). A greater understanding of such relationships will aid future efforts to design and evaluate community based obesity prevention initiatives.

For complex community based interventions, the conduct of rigorous evaluations undoubtedly represents a considerable challenge. Evaluators need to carefully examine and the strengths and pitfalls of decisions regarding evaluation design, data collection, measurement, and analysis and seek to maximise evaluation rigour in the context of political, resource and practical constraints. This manuscript attempts to provide guidance for evaluators to do so.

REFERENCES


7 McCormick B, Stone I. Economic costs of obesity and the case for government intervention. Obesity Reviews 2007; 8(Suppl. 1); 161-164

9 Lavery NS. Stop all further research and act. *British Medical Journal* 2008; 336:7


18 Medical Research Council. *Developing and evaluating complex interventions: new guidance*. Available at URL: www.mrc.ac.uk/complexinterventionsguidance [last accessed July 7th 2011]


25 Pirie, PL., Stone, EJ., Assaf, AR., Flora, JA. and Maschewsky-Schneider, U. Program evaluation strategies for community-based health promotion


28 Friedrich MJ. Researchers address childhood obesity through community-based programs. *JAMA* 2007; 298(23):2728-2730

29 Swinburn BA, de Silva-Sanigorski AM. Where to from here for preventing childhood obesity: An international perspective. *Obesity* 2010; 18(1):S4-S7

30 Huang TT, Story M. A journey just started: renewing efforts to address childhood obesity. *Obesity* 2010; 18 (Supl1):S1-S3


42 Swinburn B, Pryor J, McCabe M, Carter R, de Courten M, Schaaf D, Scragg R.

The Pacific OPIC Project (Obesity Prevention In Communities) – Objectives and designs. *Health Promotion in the Pacific* 2007; 14(2):139-153


44 Azzam, T. Evaluator responsiveness to stakeholders. *American Journal of Evaluation* 2010; 31: 45-65


47 Cope MB, Allison DB. White hat bias: examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. *Int J Obes* 2010; 34:84-88


50 Wolfenden L, Wiggers J, Tursan d'Espaignet E, Bell AC. How useful are systematic reviews of child obesity interventions. *Obesity Reviews* 2010; 11:159-165


56 Mellor JM, Rapoport RB, Maliniak D. The impact of child obesity on active parental consent in school-based survey research on healthy eating and physical activity. *Eval Rev* 2008; 32(3): 298-312


60 Barry AE. How attrition impacts the internal and external validity of longitudinal research. *J Sch Health* 2005; 75(7): 267-270


65 Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention. *J Epidemiol Community Health* 2004; 58(9): 788-793


### Table 1: Evaluation characteristics of population wide community based obesity prevention programs

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design</th>
<th>Outcome measures (assessments of community health status)</th>
<th>Impact measures (assessments of community environments or settings)</th>
<th>Process measures (assessments of intervention implementation)</th>
<th>Context assessment</th>
<th>Research participation and attrition</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPLE Project&lt;sup&gt;33-34&lt;/sup&gt;</td>
<td>Quasi-experimental: Pre-post with comparison. Intervention implemented in one community in New Zealand. A second geographically separate community served as a comparison (broadly comparable on socio-demographic variables)</td>
<td>Body Mass Index (height and weight) and waist circumference, pulse rate and blood pressure of children recruited through schools</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>85-92% and 81-87% response rate respectively for intervention and comparison community</td>
</tr>
<tr>
<td>The California’s Endowment’s Healthy Eating, Active Communities Program&lt;sup&gt;35&lt;/sup&gt;</td>
<td>Quasi-experimental: Pre-post with comparison Intervention delivered in 6 low income Californian communities selected as they met criteria including high rates of obesity and capacity to implement a prevention program. Selection method and number of comparison communities not stated.</td>
<td>Body Mass Index (height and weight) and physical fitness of children using existing quantitative data-sets. Diet and physical activity behaviour of children recruited through schools via survey. Psychometric properties not reported.</td>
<td>Foods and beverages sold in schools and health care institutions; physical activity programming and equipment in schools, after school programmes and parks; neighbourhood food retail and food advertising. Psychometric properties of some assessment tools not reported.</td>
<td>Not reported</td>
<td>Surveys, and focus groups assessing the extent to which the community stakeholders and policy makers were aware of, engaged in and supported the program.</td>
<td>Not reported</td>
</tr>
<tr>
<td>The Kaiser Permanente</td>
<td>Quasi-experimental: Pre-post, repeat cross sectional</td>
<td>Overweight (measure not specified) among children</td>
<td>Not reported</td>
<td>A database documenting the protocol / methods</td>
<td>Not reported</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Intervention details</th>
<th>Implementation status and the number of people reached (exposed and affected) to the intervention (no other detail provided)</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Community Health Initiative:</strong> <strong>Overview and Evaluation Design</strong>&lt;sup&gt;36&lt;/sup&gt;</td>
<td>Intervention implemented in 3 ethnic minority communities in California U.S.A. Fitness among children (measure not specified) Physical activity and nutrition behaviour via automated telephone survey (for adults) or school based survey. Specific psychometric properties not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Romp and Chomp</strong>&lt;sup&gt;37&lt;/sup&gt;</td>
<td>Quasi experimental: Pre-post repeated cross sectional with comparison Intervention implemented in one Australian community selected based on existing collaborative links with the University and health services. A sample of local government areas (LGA’s) from across the rest of the State served as a comparison Body Mass Index (height and weight) collected by nursing staff during routine 2-3.5 year old Maternal and Child Health Key Age and Stage Health Checks Physical activity and nutrition behaviour using validated survey instrument completed by parents attending child health checks.</td>
<td>Environmental audits of the physical, policy, socio-cultural and economic environments of Children’s Services; active play survey to assess staff and childcare service activities promoting active play; and community capacity building assessment. Psychometric properties of survey not reported.</td>
<td></td>
</tr>
<tr>
<td><strong>Shape Up Sommerville</strong>&lt;sup&gt;38&lt;/sup&gt;</td>
<td>Quasi experimental: Pre-post with comparison Intervention delivered in one community in the U.S.A. selected due to existing relationships between key community organisations and the research entity. Two communities with similar demographic characteristics served as comparisons. Child Body Mass Index (height and weight) based on survey of children recruited through schools. Child physical activity and nutrition assessed via parent completed questionnaire. Psychometric properties not reported.</td>
<td>Project records used to document the extent of implementation and exposure to the intervention.</td>
<td>28% participation rate in intervention and comparison communities combined. 39% attrition in the intervention and 25% in comparison communities.</td>
</tr>
<tr>
<td>Study Title</td>
<td>Study Design</td>
<td>Intervention Details</td>
<td>Outcome Measures</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Be Active Eat Well[39]</td>
<td>Quasi experimental: Pre-post with comparison</td>
<td>Intervention delivered in one community in Australia, selected based on existing community infrastructure and supportive networks. Schools randomly selected from communities in the broader region served as a comparison.</td>
<td>Body Mass Index (height and weight) and waist circumference based on survey of children recruited through schools. Nutrition and physical activity assessed using a survey based on a validated questionnaire completed by children and a parent completed telephone survey. A number of potential harms assessed including prevalence in underweight, unhappiness, attempts to lose weight, and teasing. Specific items or survey instruments or psychometric properties not reported.</td>
</tr>
<tr>
<td>Healthy Living Cambridge Kids[40]</td>
<td>Quasi-experimental: Pre-post. Intervention implemented in one community in the U.S.A</td>
<td>Child Body Mass Index (height and weight) routinely collected by the Cambridge Public School system Child fitness test score routinely collected by the Cambridge Public school system</td>
<td>Not reported</td>
</tr>
<tr>
<td>Health Promoting Communities: Be Active Eating Well[41]</td>
<td>Quasi-experimental: Pre-post repeat cross sectional with comparison</td>
<td>Intervention implemented in 5 disadvantaged communities in Australia. Comparison sample from matched schools and</td>
<td>Body Mass Index (height and weight) and waist circumference. Nutrition and physical activity behaviours assessed via survey methods. Psychometric properties not reported.</td>
</tr>
</tbody>
</table>
workplaces from non-intervention communities across the state of Victoria. Quality of life using the Assessment of Quality of Life markr2 (AQoL) assessment tool adapted from a validated tool developed for use in adults.  

**The Pacific OPIC Project**

Quasi experimental: Pre-post with comparison  
Interventions implemented in communities in Australia, Tonga, Fiji and New Zealand.  
Intervention communities selected based on criteria including amenable community capacity characteristics (such as existing project champions) geography, and ease of access to research staff and other organisations.  
Comparison communities selected to be as comparable as possible (ethnicity, socioeconomic status trajectory of weight gain) and geographically separate to minimise contamination.  

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Intervention</th>
<th>Outcome Measures</th>
<th>Quality of life assessed using established instruments.</th>
<th>Knowledge assessed (no other information provided).</th>
<th>In a sample subset, a validated assessment of perceived body image and body change questionnaire administered.</th>
<th>Audit of school environment. Validity of assessment tool or collection method not described</th>
<th>Community readiness questionnaire (validity not reported) and stakeholder interviews used to assess community capacity.</th>
<th>Co-ordinator reports on implementation activities, action plans, meeting minutes, and project records will be used to assess implementation (reach uptake barriers etc) and implementation costs.</th>
<th>Not reported</th>
<th>Not applicable (protocol/methods overview)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fleurbaix-Laventie Ville Sante study (EPODE)</strong></td>
<td>Quasi experimental: Pre-post repeated cross sectional. Comparison communities assessed at post intervention only.</td>
<td>Child Body Mass Index (height and weight) assessed during a survey of children recruited through schools.</td>
<td>Not stated</td>
<td>Not stated</td>
<td>Not stated</td>
<td>95-98% participation rate across all school children from both intervention and comparison towns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Intervention implemented in two towns in France. Two (no intervention) towns served as a comparison.
Randomly allocated groups (communities)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time Period

- **Intervention period**
- **Control period**

*Figure 1: Illustration of a stepped wedge cluster randomized trial*
APPENDIX A

*Search strategy: Medline*

1. obesity.mp.
2. prevention.mp.
3. 1 and 2
4. nation$.mp.
5. state.mp.
6. count$.mp.
7. district.mp.
8. regio$.mp.
9. communit$.mp.
10. area.mp.
11. town.mp.
12. village.mp.
13. borough.mp
14. municip$.mp.
15. province.mp.
16. shire.mp
17. 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18. randomized controlled trial.pt.
19. controlled clinical trial.pt.
20. randomized.ab.
21. randomised.ab.
22. clinical trials as topic.sh.
23. randomly.ab.
24. trial.ti.
25. double blind.ab.
26. single blind.ab.
27. (pretest or pre test).mp.
28. (posttest or post test).mp.
29. (pre post or prepost).mp.
30. Before after.mp.
31. (Quasi-randomised or quasi-randomized or quasi-randomized or quazi-randomised).mp.
32. stepped wedge.mp.
33. Preference trial.mp.
34. Comprehensive cohort.mp.
35. Natural experiment.mp.
36. (Quasi experiment or quazi experiments).mp.
37. (Randomised encouragement trial or randomized encouragement trial).mp.
38. (Staggered enrolment trial or staggered enrollment trial).mp.
39. (Nonrandomised or non randomised or nonrandomized or non randomized).mp.
40. Interrupted time series.mp.
41. (Time series and trial).mp.
42. Multiple baseline.mp.
43. Regression discontinuity.mp.
44. 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43
45. 3 and 17 and 44

46. 45

47. limit 45 to yr='1991-2010'