

# Planning and evaluating population interventions to reduce noncommunicable disease risk – reconciling complexity and scientific rigour?

Adrian Bauman<sup>a,c</sup> and Don Nutbeam<sup>b</sup>

<sup>a</sup> School of Public Health, University of Sydney, NSW, Australia

<sup>b</sup> Vice Chancellor's Office, Southampton University, UK

<sup>c</sup> Corresponding author: [adrian.bauman@sydney.edu.au](mailto:adrian.bauman@sydney.edu.au)

## Article history

Publication date: November 2014

Citation: Bauman A, Nutbeam D. Planning and evaluating population interventions to reduce noncommunicable disease risk – reconciling complexity and scientific rigour? *Public Health Res Pract.* 2014;25(1):e2511402. doi: <http://dx.doi.org/10.17061/phrp2511402>

## Key points

- Chronic disease prevention efforts are often evaluated as 'simple' interventions
- Complex program evaluation methods provide evidence of how these programs are implemented, how they work and their effects
- Optimal research designs, including randomised trials, may not be usable in assessing comprehensive, community-wide prevention programs, but may be used to assess the efficacy of program subcomponents
- Chronic disease prevention requires new, multimode evaluation methods, and policy maker awareness that this level of evidence is useful in assessing prevention programs

## Abstract

Noncommunicable diseases (NCDs) are the leading global causes of morbidity and mortality. It is important to develop and deliver effective NCD prevention programs, but these have been difficult to evaluate. Technical approaches differ, with academic researchers, practitioners and policy makers each bringing different perspectives and priorities to the task of NCD program evaluation. Epidemiologically defined hierarchies of research evidence give preference to evaluation methods that are often unsuitable for assessing complex NCD prevention interventions. This may lead to interventions that provide the 'right answer to the wrong question', or to evaluation data that are insufficient to inform NCD prevention efforts.

This paper recommends a set of standardised stages in the planning, development and evaluation of NCD prevention programs, including the use of logic models, the expanded use of process evaluation to better understand and record the context for implementation, and the use of appropriate research designs for assessing the impact of both subcomponents and the whole program.

NCD prevention agencies and academic stakeholders need to recognise the limitations of established evaluation designs and support greater flexibility in the application of evaluation methods that are fit for purpose in describing the stages in NCD programs. This involves assessing policy development and implementation, measuring intermediate indicators, using mixed methods of evaluation, and employing population surveillance systems to assess long-term outcomes.

## Noncommunicable disease prevention – the challenge for evaluation

The prevention of chronic disease in populations is a complex challenge. It requires efforts to reduce the biobehavioural risk factors for noncommunicable diseases (NCDs) such as physical inactivity, poor nutrition and smoking, and to consider the social and economic contexts in which health-compromising behaviours occur. NCD prevention programs are 'complex public health programs' because of their multiple intervention components, delivery in different settings and prolonged time frames. Evaluation of NCD programs is correspondingly complex. Because of the complexity of NCD risk, and its myriad antecedents and determinants, people working in population-level prevention need to comprehend the complexity of understanding effective programs.

This paper describes an organising framework for evaluating comprehensive and complex intervention programs that incorporate a complicated mix of educational, environmental and policy interventions targeting risk factors in a population. The framework includes the measurement and monitoring of each of the intervention components (education, environment and policy), assessment of short-term program impact (such as change in health literacy, implementation of public policy) and assessment of the longer term outcomes (reduced behavioural, social and environmental risk).

This paper (a) outlines the history and methodological tensions in the past three decades of evaluating health promotion and disease prevention programs, (b) proposes a practitioner-relevant framework for NCD program organisation and evaluation, and (c) identifies challenges and barriers to using this framework in assessing NCD prevention efforts.

## Evaluating prevention programs – reconciling values, scientific rigour and complexity

A 2007 *Nature* paper<sup>1</sup> listed the top 20 policy and research priorities for conditions such as diabetes, stroke and heart disease. These priorities were grouped under six subheadings that included: enhancing economic, legal and environmental policies; reorienting health systems; mitigating the health impacts of poverty and urbanisation; and engaging with business and the community. These strategies appeared alongside established health system interventions that focused on modifying risk factors and raising public awareness.

In the late 1980s, such strategies would have been unimaginable. However, in the past 25 years, our understanding of NCDs has advanced, and there has

been a paradigm shift in the way public health problems are conceptualised and addressed. For many, the Ottawa Charter for Health Promotion<sup>2</sup> was the catalyst for this paradigm shift. The principles and strategies described in the charter led to the application of socioecological models to NCD development, extending beyond the individual only being responsible for unhealthy behaviour, and encompassing causal mechanisms that include the social, environmental and economic determinants of health.<sup>3,4</sup>

Prevention science has mixed origins, in part emanating from the health promotion values espoused in the Ottawa Charter.<sup>2,5</sup> The charter is more a statement of values and beliefs about how public health interventions should be organised than an empirical review of evidence and effectiveness. It places high value on participation and empowerment that tend to favour evaluation methods focused on description and case studies, and on assessing process indicators of implementation, rather than impact alone; these values and beliefs sit uncomfortably with prevailing criteria for evaluating evidence from public health interventions.

By contrast, traditional approaches to preventing NCDs have evolved from a biomedical research paradigm, with an emphasis on controlled experimental research to generate evidence of effectiveness, and the subsequent synthesis and summarising of this evidence into prevention guidelines to inform practice.<sup>6</sup> Many early prevention interventions were evaluated using this approach to changing individual health-related behaviours, especially among population subgroups at high risk for NCDs.<sup>7</sup> Experience over time has highlighted problems with this approach, including identifying effects of interventions using nongeneralisable samples, and achieving short-term effects in the research funding cycle that do not result in a sustained impact on NCD risk in populations.

There has been tension between traditional, individual, high-risk program evaluation using optimally controlled research designs, and more generalisable population approaches to NCD prevention.<sup>7</sup> Different but complementary approaches are required to assess multicomponent 'upstream' NCD prevention strategies, including assessments of policy and regulations, evaluation of financial incentives or disincentives, and assessment of the influence of environmental changes.

Continuing to contest and debate a false dichotomy between traditional evaluation methods and more complex approaches will waste much effort among prevention scientists, and does not help us to understand the complex systems that contribute to NCD risk.<sup>8</sup> It is possible to take the best from both paradigms to develop comprehensive, yet still science-based program assessment using research methods with origins in both epidemiology and health promotion. To reconcile these different approaches, we propose an integrated model that adapts both methods to comprehensively evaluate NCD prevention programs.

## A health promotion and epidemiological framework for NCD program evaluation

A framework for evaluating complex NCD prevention efforts, including the program tasks and the sequence of evaluation tasks required, is shown in Figure 1. Work starts with an assessment of the current problem and its determinants and distribution (the pre-program planning phase in Figure 1). This includes the formative evaluation tasks of understanding the magnitude of the problem and its determinants, consulting with the stakeholders and the community, and developing the conditions where policy development can be contemplated. Engaging the community from the beginning ensures that proposed large-scale solutions can be translated and implemented at the local level.

The initial steps should lead to strategic planning and associated documentation, mobilisation of resources and personnel, and development of partnerships required for successful implementation (the planning phase in Figure 1). These factors are seldom systematically measured or recorded, but can demonstrate common purpose among diverse stakeholders and provide an understanding of successful policy implementation.

Evaluation metrics could include assessments of partnership formation and maintenance among stakeholders, planned expenditures and audits of comprehensive plans against best practice.<sup>9</sup> Qualitative and quantitative measures of partnership formation can be used in this stage.<sup>9,10</sup>

Also part of the planning phase is formal cross-sectoral planning of the multiple interventions proposed in an overall NCD prevention strategy, each component of which will need its own evaluation design and methods. Planning for evaluation requires the development of a clear 'logic model' – a planning document or diagram that links each element of the intervention to the specific results that it is intended to produce.<sup>11</sup> The delineation of a logic model makes the potential links in the program explicit, and identifies metrics of success from program delivery all the way through to health outcomes. A logic model can also strengthen the population focus, with clear goals to achieve high population reach and/or community participation.

Fundamental to all NCD programs is the need for thorough process evaluation to assess the implementation and reach of each intervention component. Examples of this are described in the program implementation stage of Figure 1. In addition, measuring differences in implementation in diverse settings is important – this

**Figure 1.** Stages in the evaluation of complex NCD prevention programs: integrating health promotion and epidemiological methods

	Pre-program planning	Planning phase	Program implementation	Program impact and outcomes
<b>Stages of program evaluation challenges</b>	Identify policies, and community and epidemiological need in the population Policy/resource supports Community engagement	Develop cross-sectoral implementation plan Describe logic model Plan specific strategies (i.e. components of the overall program)	Develop timeline and accountabilities for implementation of program components Develop program components that are based on policy (regulatory), environmental change, and those that provide services or facilities to target individuals	Assess the short-term and medium-term impact of the program on meeting its specific measurable objectives
<b>Evaluation challenges</b>	Identify commitment and resources for evaluation Develop evaluation team Review evidence in literature for effective programs	Baseline measures of partnerships and coalitions Draw logic model; develop measures and research designs needed for assessing the logic model Assess resource needs for evaluation Assess need for specific qualitative and quantitative components of the evaluation	Develop process evaluation plan and indicators for each program component Develop measures of program fidelity and of local context to assess differential implementation and delivery of program elements Develop (qualitative) case studies of subcomponent implementation, and describe factors associated with high community reach or participation Audit environmental changes and policy implementation, as required	Identify regular population surveys and surveillance systems that can be used to monitor NCD risk factors Develop outcomes evaluation – decide on long-term population health monitoring of NCD incidence, mortality, morbidity and economic appraisal of program costs

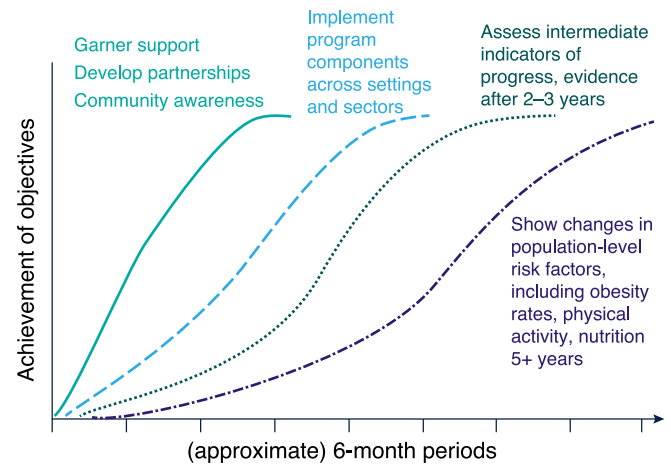
is the role of assessing 'contextual differences'. For example, a policy-endorsed nutrition program might be effective in schools in socially advantaged regions, but might not be so relevant to schools in disadvantaged regions or in culturally diverse schools without undergoing substantial modification. Measures of adaptation in different contexts are important for understanding the circumstances and conditions in which an intervention is more likely to be effective. Evaluation of context can take many forms, from assessing qualitative perceptions of stakeholders or those directly targeted, through to studies of the economic, social and physical environments in which the intervention takes place.

Finally, each intervention subcomponent needs to be considered and evaluated in terms of impact and outcomes (right-hand side of Figure 1) – that is, the endpoints for measuring success need to be defined, and the level of investment required to collect and generate these data needs to be determined. Health promotion methods that focus on case studies and program description are one way to evaluate intervention programs. Other evaluation elements may assess intermediate indicators<sup>12</sup>, including changes to the social or physical environment, or changes in service delivery and reach (shown as part of program implementation in Figure 1), that will in turn result in endpoint risk conditions and risk factor changes, that will lead to reduced NCDs.

Figure 2 illustrates how a multicomponent intervention might achieve these different outcomes over time. In this model, time units are hypothetically six months. After one year, the program impact can be measured most easily in terms of increased interagency partnerships and organisational readiness, community awareness and engagement. By the third year, good progress in achieving short-term program impact should be measurable, alongside increases in community awareness. After five years or more, major progress in intermediate indicators and in risk factors should be observable and, subsequently, some early impact on chronic incidence may be observable at a population level.<sup>13</sup> This model can be useful in understanding the long-term policy context and advocacy required to evaluate a sustained complex program to reduce NCDs.

Using the integrated model in Figure 1 allows for clear progress to be assessed in terms of policy development, partnerships and interagency responsibilities; this is accompanied by a logic model that identifies the relationships between actions and their effect on intermediate indicators and NCD risk. The use of multiple evaluation methods is an important and necessary innovation in evaluating NCD prevention efforts, with controlled designs possible to assess specific subcomponent interventions, and a mix of qualitative and quantitative methods to assess the overarching effects of policy, program implementation and intermediate indicators on improving conditions conducive to reducing NCD risk.

**Figure 2.** Hypothetical changes — the time course for assessing outcomes following complex population interventions



The principles of good research practice apply to both health promotion and biomedical methods. Further, attention to the use of reliable and valid measurements extends to all stages, from assessing partnerships through to assessing environments and risk factors. Careful focus on selection issues and generalisability is sometimes less well considered, and increased attention should be paid to process evaluation measures of population reach, engagement and participation at all stages, such that the proportion of professionals, proportion of settings implementing a policy, or proportion of people in a community who participate in the program can be documented and monitored.

Finally, the years of effort need to be monitored through routine and comprehensive use of system-level population surveillance data – 'program impact and outcomes' in Figure 1. If the overall strategy was to reduce obesity, then a surveillance system should include serial measurement of obesity rates in representative samples of the population, as well as relevant antecedent environmental indicators (such as kilojoule labelling of menus, or fat content of foods offered in community restaurants) and relevant social indicators (measures of social norms towards supersized portions or perceptions of taxation of unhealthy products). Creating a healthy environment includes changing social norms that influence community wishes, political imperatives and, eventually, program resourcing.

## Application of these evaluation methods to recommended NCD prevention actions

Examples of the kinds of intervention strategies recommended in NCD prevention reviews are shown in Table 1; these are adapted from existing reviews of NCD prevention<sup>14</sup> and physical activity.<sup>15</sup> In the middle column of Table 1, evidence-based distillations of recommended

best practices in population NCD prevention are summarised. Evaluation methods to be applied to these recommended strategies are not mentioned in these reviews.

The major NCD approaches to reducing risk factors can be evaluated as separate interventions and, eventually – through an umbrella approach – summate to an overall NCD prevention program of work. The evaluation challenges differ across NCD issues, with

surveillance methods required for salt intake, monitoring of policy implementation for diet and alcohol, assessing the reach and context of interventions for all areas, and conducting evaluation work outside the health sector, especially for addressing unhealthy food and physical inactivity.

As examples only, some of the evaluation challenges are shown in the right-hand column (of Table 1), illustrating aspects of the complexity of these

**Table 1. Examples of recommended strategies for population NCD prevention and the resulting evaluation challenges**

NCD issue	Examples of major recommended intervention approaches	Evaluation issues and challenges in assessing intervention effects	Potential issues in 'complexity' of the evaluation
Salt reduction	Develop clear reduction targets (e.g. 15% decrease), mass education, regulatory changes (e.g. reduce salt in processed food), fiscal regulation, advocacy	Evaluate the implementation of food regulations, voluntary for industry or mandatory Develop ways to monitor salt intake in population surveillance systems	Difficulty in identifying indicators to monitor salt intake at the population level, and to monitor implementation across the food industry Challenges with imported foods Need to change community understanding of salt in processed food (not just salt added to food)
Dietary intervention	Increase price of high-fat foods (taxation), reformulate carbonated drinks, policies on food advertising to children	Monitor the implementation of policy efforts, monitor industry response, monitor community food outlets	Implementation of taxation policy, other current food taxes and price elasticity induced by industry; all of these may impede assessment of 'fat tax' implementation and effects
Smoking prevention	Media campaigns, advocacy work, smoke-free policies, advertising and sales bans, taxation, packet warnings	Evaluate effects of campaign on smoking attitudes, creating antismoking perceptions in the community and among decision makers Evaluate the reach and implementation of bans, labelling and update of smoke-free policies	Need ongoing serial mass media campaigns (and funding may fluctuate); policy introduction may be resisted in some settings or organisations (such as the hotel association, which may fear losing business if smoking banned in bars, pubs)
Increase physical activity	Promote physical activity through primary care and schools, promote sport, increase active transport (e.g. cycling), build walkable environments	Challenges in assessing partnerships and implementation across sectors – does more sports participation or increased public transport contribute to increased total health-related physical activity?	Cross-sectoral policy difficult to influence from the health sector, and may occur in other sectors, but asynchronously with this NCD prevention program
Reduce harmful alcohol use	Decrease drink-driving, strengthen regulations regarding sales of alcohol or marketing to children	Assess effects of interventions to reduce drink-driving (outside health sector), assess regulation uptake in different alcohol sales contexts	Multiple separate interventions with different target groups needed to make progress Counter-marketing (advertising) by alcohol industry
Multiple risk factor programs	Integrated NCD prevention programs that consider approaches to reducing all NCD risk factors in a community	Time scale (years) to make population-level improvements in multiple risk factors is usually longer than funding cycles	Priority driven, so that some risk factors are favoured by governments at different times over other risk factors Requires even more planning and intersectoral cooperation, often a 'whole-of-government' approach

Source: Adapted from Bonita et al.<sup>14</sup> and Trost et al.<sup>15</sup>

interventions, the need for comprehensive evaluation approaches and, sometimes, the countervailing forces acting against NCD preventive actions. Here, advocacy strategies as well as intervention delivery need to be carried out and appraised.

Table 1 demonstrates a primary prevention focus to improving population health. Similar issues confront secondary and tertiary prevention strategies to reduce NCD risk among those at high risk or with existing chronic disease. The challenges of implementing effective strategies that allow access to primary care, or to effective tertiary prevention service delivery, pose similar evaluation challenges of assessing effects in different contexts and environments. Further, the research evidence generated in primary care is usually in selected samples of doctors and patients, so methods to scale up these interventions to achieve population reach is an ongoing challenge for NCD prevention.<sup>16</sup>

## Challenges and barriers to NCD program evaluation – forging a way through complexity

The greatest challenge to evaluating NCD prevention efforts is that, unlike clinical therapies, population-level prevention seldom results from a single intervention or action. NCD prevention occurs in the context of the social, economic and cultural systems that lead to unhealthy behaviours and lifestyles; these require ‘systems’ thinking to understand the problem.<sup>17</sup> Recognition of this requires program evaluations to have a broader base, with a need to understand the effects across the multiple elements and across stages of a comprehensive intervention program.

Although this complexity has become better understood and more explicitly recognised by funding agencies<sup>18</sup>, this ‘progress’ has its limitations, suffering from the problem identified by Smith and Petticrew<sup>19</sup> of using microanalysis (individuals and health services) at the expense of broader macroanalysis (societal and system) in determining interventions and their evaluation. Many effective ‘macro’ social policy interventions simply cannot be evaluated by established trial methods.<sup>20</sup> The UK Medical Research Council guidelines on developing and evaluating complex interventions<sup>21</sup> conclude that while some aspects of good practice are clear (in cases of simple interventions in manageable systems), methods for developing, evaluating and implementing complex interventions, especially those involving social policy and/or environmental change, are still being developed. On many important issues “there is no consensus yet on what is best practice”.<sup>21</sup> This is made more challenging by many NCD strategies that are implemented by sectors and agencies outside health, where the evaluation paradigm, indicator measurement and concepts of ‘evidence’ may be different.

A comprehensive program evaluation strategy, as suggested here, underpins any population prevention intervention. This includes understanding the stages in policy development, forming partnerships, and having a logic model that can help to both reconcile and include different perspectives.<sup>13</sup> These stages are schematically shown in the steps in Figure 1, but enough time is needed to complete these evaluation elements before the rollout of the multiple population intervention components. Implementing and evaluating population-wide actions are preferable to the common practice of strategic planning documents that are seldom resourced or implemented. More focused process and impact evaluation designs might then be used to assess specific subcomponent interventions within the overall program.

## W(h)ither randomised trials: evaluation designs for complex interventions

Although using a planning or logic model will provide a structured description of an intervention and its expected outcomes, it will not in and of itself resolve differences about the best research design to evaluate NCD prevention efforts. There is still tension between population-wide efforts and the values of those who currently fund research or manage scientific journals, with limited attention yet paid to the ‘quality’ of evidence from evaluations of population-level interventions.<sup>22-24</sup> Healthy debate continues about the strengths, limitations and alternatives to randomised controlled trials (RCTs) in the evaluation of scaled-up prevention interventions.<sup>18,24-26</sup> What becomes clear in the published literature (and from practical experience) is that there are circumstances in which it is not feasible to use an RCT<sup>24</sup>, or when RCTs answer narrow-cast questions – which may provide the best answer to the wrong question.<sup>25</sup>

It may be possible to embed the whole program of work in a higher-order, controlled study, such as a cluster RCT, allocating some communities to intervention and others as controls, but the degree of measurement and evaluation required and the cost may make this an infrequent choice. Some subcomponent interventions directed at single issues with well-defined but limited objectives, and interventions undertaken in more formalised systems such as schools, health clinics and workplaces are still amenable to controlled designs. New approaches have been developed for using controlled trials; these include the use of ‘realist RCTs’, which explore the intervention processes more closely, examine mediators and use factorial designs for assessing different approaches.<sup>26</sup> Further, adaptive designs may become more commonplace – for example, those that permit variations in intervention delivery or methods, and still attempt analyses using experimental methods; or step-wedge designs, which allow rollout of the

intervention differentially over time in different sites, but maintain a randomised trial approach.<sup>27</sup> In addition, the methods for evaluating natural experiments, such as the effects of regulatory policy on the food or physical activity environments, may be added to the armamentarium of NCD evaluations.<sup>24</sup>

Nonetheless, some academics still report short-term trials for complex community interventions, citing the need to report to funding agencies in short time frames. This approach runs the risk of missing effects through premature measurement of outcomes, and omitting useful process evaluation and descriptive research, which we contend is essential in understanding NCD prevention efforts. Selected samples, with short-term evidence from controlled trials, do not provide a single research design to evaluate NCD prevention programs, but may be most useful in specific subcomponent intervention appraisal. More valuable overarching program information can be obtained by assessing the implementation of each program, what activities occur under which conditions, by whom, and with what level of effort.<sup>28–30</sup> This assists with future efforts, having better articulated the conditions that need to be created to achieve successful outcomes.<sup>28</sup> Nonetheless, many prevention interventions still contain little information about describing the intervention or the factors that influence its implementation at the population level or in different contexts.<sup>8</sup> Allowing differential implementation (slightly different ‘forms’ of the intervention) in different environments is supported by a health promotion approach, but will compromise scientific rigour that uncompromisingly emphasises ‘standardisation’ of interventions.

## Conclusions

We need to focus on evaluations that provide meaningful evidence on the population-wide impact of NCD prevention efforts. Evaluations have to be customised to match the intervention types, and be capable of local adaptation. The first stage of comprehensive planning should identify elements to be pursued and the evaluation metrics required for each component. In the overall program, there can be more rigorous evaluation of subcomponent interventions, but greater attention should be paid to the external validity of the subcomponents, and relevance for scaling up to the population. Population surveillance systems should form the endpoint methods for assessing impact – that is, assessing the net sum of the prevention efforts in a population health program rather than attempting to assign causal influences to any specific component. Some progress has been made in developing methods for complex interventions in complex systems, but significant challenges remain. What is clear is that the best evaluation designs combine different research methodologies, both quantitative with qualitative, to produce a diverse range of data that will optimally

support both the integrity of the study and its external validity.

Finally, complex program evaluations are expensive, complicated, require mixed methods and involvement from different disciplines, and may score less well in attracting national-level scientific research funding. As a consequence, the research questions of greatest importance in NCD prevention are not those that researchers are often funded to answer. Some countries have made more overt efforts to expand funding opportunities and to be more responsive to opportunity as it arises, but progress is painfully slow (see, for example, the UK NIHR program: [www.nets.nihr.ac.uk/programmes/phr](http://www.nets.nihr.ac.uk/programmes/phr) or the NHMRC-funded Australian Prevention Partnership Centre: [www.phrp.com.au/issues/vol2512014/australian-prevention-partnership-centre-systems-thinking-prevent-lifestyle-related-chronic-illness](http://www.phrp.com.au/issues/vol2512014/australian-prevention-partnership-centre-systems-thinking-prevent-lifestyle-related-chronic-illness)). These initiatives will help to further improve our knowledge and understanding of effective interventions, and extend our technical capability to customise evaluation designs to fit the activity and circumstances of individual programs.

## Competing interests:

None declared

## References

1. Daar AS, Singer PA, Persad DL, et al. 2007 Grand challenges in chronic non-communicable diseases. *Nature*. 2007;450(7169):494–6.
2. World Health Organization. Ottawa Charter for Health Promotion. Geneva: World Health Organization; 1986.
3. Wilkinson RG, Marmot MG, editors. Social determinants of health: the solid facts. Geneva: World Health Organization; 2003.
4. Beaglehole R, Bonita R, Horton R, Adams C, Alleyne G, Asaria P, et al. Priority actions for the non-communicable disease crisis. *Lancet*. 2011;377(9775):1438–47.
5. Nutbeam D. What would the Ottawa Charter look like if it were written today? *Crit Public Health*. 2008;18(4):435–41.
6. Agency for Healthcare Research and Quality. The guide to clinical preventive services 2012. Recommendations of the US Preventive Services Task Force. Rockville (MD): Agency for Health Research and Quality; 2012 [cited 2014 Sep 26]. Available from: [www.ncbi.nlm.nih.gov/books/NBK115115/](http://www.ncbi.nlm.nih.gov/books/NBK115115/)
7. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427–32.
8. Green LW, Ottoson JM, Garcia C, Hiatt RA. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu Rev Public Health*. 2009;30:151–74.

9. Centers for Disease Control and Prevention. Learning and growing through evaluation: evaluating partnerships. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Environmental Health, Division of Environmental Hazards and Health Effects, Air Pollution and Respiratory Health Branch; 2012.
10. Granner ML, Sharpe PA. Evaluating community coalition characteristics and functioning: a summary of measurement tools. *Health Educ Res.* 2004;19(5):514–32.
11. Freedman AM, Simmons S, Lloyd LM, Redd TR, Salek SS, Swier L, et al. Public health training center evaluation: a framework for using logic models to improve practice and educate the public health workforce. *Health Promot Pract.* 2014;15(1 suppl):80S–88S.
12. Datta J, Petticrew M. Challenges to evaluating complex interventions: a content analysis of published papers. *BMC public health.* 2013;13(1):568.
13. Bauman A, Nutbeam D. Evaluation in a nutshell. A practical guide to the evaluation of health promotion programs. 2nd rev. ed. Sydney: McGraw-Hill; 2013.
14. Bonita R, Magnusson R, Bovet P, Zhao D, Malta DC, Geneau R. Country actions to meet UN commitments on non-communicable diseases: a stepwise approach. *Lancet.* 2013;381(9866):575–84.
15. Trost SG, Blair SN, Khan KM. Physical inactivity remains the greatest public health problem of the 21st century: evidence, improved methods and solutions using the '7 investments that work' as a framework. *Br J Sports Med.* 2014;48(3):169–70.
16. Gaziano TA, Pagidipati N. Scaling up chronic disease prevention interventions in lower-and middle-income countries. *Ann Rev Public Health.* 2013;34:317–35.
17. Froot S, Johnston LM, Matteson CL, Finegood DT. Obesity, complexity, and the role of the health system. *Curr Obes Rep.* 2013 Aug 30;2:320–26.
18. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ.* 2008;337:979–83.
19. Smith RD, Petticrew M. Public health evaluation in the twenty-first century: time to see the wood as well as the trees. *J Public Health (Oxf).* 2010 Mar;32(1):2–7.
20. Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? *Public Health.* 2005;119(9):751–7.
21. Medical Research Council. Developing and evaluating complex interventions: new guidance. London: Medical Research Council; 2008.
22. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health.* 2002;56(2):119–27.
23. Tang KC, Choi BK, Beaglehole R. Grading of evidence of effectiveness of health promotion interventions. *J Epidemiol Community Health.* 62(9):832–4.
24. Hawe P, Shiell A, Riley T. Complex interventions: how 'out of control' can a randomised controlled trial be? *BMJ.* 2004;328(7455):1561–3.
25. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med.* 2007;33(2):155–61.
26. Bonell C, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, et al. Alternatives to randomization in the evaluation of public health interventions: design challenges and solutions. *J Epidemiol Community Health.* 2011 Jun;65(7):582–7.
27. Macintyre S. Good intentions and received wisdom are not good enough: the need for controlled trials in public health. *J Epidemiol Community Health.* 2011;65(7):564–7.
28. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on programme outcomes and the factors affecting implementation. *Am J Community Psychol.* 2008;41:327–50.
29. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. *BMJ.* 2006;332(7538):413–6.
30. Armstrong R, Waters E, Moore L, Riggs E, Cuervo LG, Lumbiganon P, et al. Improving the reporting of public health intervention research: advancing TREND and CONSORT. *J Public Health.* 2008;30(1):103–9.

Copyright: 

© 2014 Bauman and Nutbeam. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which allows others to redistribute, adapt and share this work non-commercially provided they attribute the work and any adapted version of it is distributed under the same Creative Commons licence terms. See: <http://creativecommons.org/licenses/by-nc-sa/4.0/>