

EVALUATION REPORT

Evaluating Place-Based Approaches: a review of methods used

Dr Stephanie Smith, Dr Michelle Irving,
Georgina Mann, Prof Arild Bjørndal and
Jane Lewis

August 2023



Centre for
Evidence and
Implementation



YOUTH
ENDOWMENT
FUND

About the Youth Endowment Fund

The Youth Endowment Fund (YEF) is a charity with a mission that matters. We exist to prevent children and young people becoming involved in violence. We do this by finding out what works and building a movement to put this knowledge into practice.

Children and young people at risk of becoming involved in violence deserve services that give them the best chance of a positive future. To make sure that happens, we'll fund promising projects and then use the very best evaluation to find out what works. Just as we benefit from robust trials in medicine, young people deserve support grounded in the evidence. We'll build that knowledge through our various grant rounds and funding activity.

And just as important is understanding children and young people's lives. Through our Youth Advisory Board and national network of peer researchers, we'll ensure they influence our work and we understand and are addressing their needs. But none of this will make a difference if all we do is produce reports that stay on a shelf.

Together we need to look at the evidence and agree what works, then build a movement to make sure that young people get the very best support possible. Our strategy sets out how we'll do it. At its heart it says that we will fund good work, find what works and work for change. You can read it [here](#).

For more information about the YEF or this report please contact:

Youth Endowment Fund

C/O Impetus

10 Queen Street Place

London

EC4R 1AG

www.youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

Registered Charity Number: 1185413

About the Evaluators

Centre for Evidence and Implementation (CEI)

CEI (www.ceiglobal.org) is a global, not-for-profit evidence intermediary dedicated to using the best evidence in practice and policy to improve the lives of children, families, and communities facing adversity. Established in 2015, CEI is a multi-disciplinary team across four offices in London, Singapore, Melbourne and Sydney. We work with our clients, including policymakers, governments, practitioners, programme providers, organisation leaders, foundations and funders, in three key areas of work:

- Understanding the evidence base
- Developing methods and processes to put the evidence into practice
- Trial, test and evaluating policies and programmes for more effective decisions and better outcomes

Dr Stephanie Smith, Senior Advisor, stephanie.smith@ceiglobal.org

Dr Michelle Irving, Senior Advisor, michelle.irving@ceiglobal.org

Georgina Mann, Advisor, georgina.mann@ceiglobal.org

Prof Arild Bjørndal, Associate Director, arild.bjorndal@ceiglobal.org

Jane Lewis, Managing Director UK & Europe, jane.lewis@ceiglobal.org

Contents

About the Youth Endowment Fund	2
About the Evaluators	3
Glossary of terms	4
Executive Summary.....	5
Chapter 1 - Introduction	9
Chapter 2 - Addressing challenges in Evaluation of PBAs	12
Chapter 3 - Planning PBA evaluations.....	20
Chapter 4 - Designs for Evaluating PBAs.....	27
Chapter 5 - 10 key messages for PBA evaluation	47

Glossary of terms

Attribution: A statement about causation, i.e. that an intervention caused the observed effects.

Contribution: The idea that an intervention is one of several or many factors that have contributed to observed change, distinguished from attribution.

Cost-effectiveness: The extent to which an intervention is effective (see below) in relation to its cost.

Counterfactual: What would have happened or been the prevailing conditions if the intervention had not been implemented.

Effectiveness: The extent to which an intervention achieved the intended impacts. Usually used to describe quantitative measurement based on causation.

Implementation: The process of putting a decision or plan into effect. The activities and strategies involved in the adoption and integration of an innovation or intervention in service systems to achieve improvements.

Mechanism of change: A theory-driven explanation of the process or series of events through which change is expected to occur.

Place-based approaches (PBAs): Collaborative long-term approaches to addressing a social issue (e.g. youth violence). They operate in a defined geographic location, focus on local needs and the attributes of the locality, involve multiple agencies, and engage local communities.

Theory of change: A description and illustration of how and why a desired change is expected to come about, as a result of activities and inputs.

Theory of place: The rationale for working in a particular way in a particular locality.

Triangulation: The application and integration of multiple research methods or sources of data to develop a comprehensive understanding of an issue.

Executive Summary

- This review summarises approaches to evaluating place-based approaches (PBAs). It has been produced to inform the Youth Endowment Fund (YEF)'s evaluation of PBAs and wider PBA work, although it is hoped it will be of value to external audiences too. The review draws on key texts on evaluation of PBAs and complex interventions, and advice from a small group of international experts, as well as bringing in the authors' own experience and commentary.
- The evaluation methods literature drawn on does not focus specifically on PBAs in youth violence (as there is little literature with this specific focus), and instead addresses PBAs, or evaluation of complex interventions, more widely.
- PBAs are defined for the purposes of this review as collaborative long-term approaches which operate in a defined geographic location, focus on local needs and the attributes of place, involve multiple agencies, and engage local communities in design and delivery.
- There is no single acceptable best practice approach to evaluating PBAs, and there are debates about what should be prioritised in evaluation and about the role of different methods.

Challenges in evaluation of PBAs

- The inherent qualities of PBAs pose some challenges to evaluation. PBAs bring together groups of stakeholders (typically local people, service providers, organisation leaders, systems leaders) to address entrenched problems that have roots in local contexts (Bellafontaine & Wisener, 2011). The 'wicked problems' they address generally require collective efforts. However the challenges they aim to address can stem from quite fundamental differences in understanding between key stakeholders which pose challenges in agreeing defined objectives for initiatives and their evaluation. PBAs are not usually neatly defined from the start: their focus and form emerges through the work itself, and can vary from initial plans and between sites (Skivington et al., 2021). This makes developmental evaluation a particularly valuable approach, with exchange of learning between evaluators and those involved in programme development and implementation (Egan et al., 2019a; Quinn Patton, 2011). Evaluations may need to take into account the roots from which wicked problems stem and the multi-factorial nature of PBA work. Unpredictable systems interactions make attribution and causality hard to pin down (Bicket et al., 2020).
- Evaluation will usefully start with work defining, conceptualising, measuring and assessing context, hypothesising how it interacts with the PBA and its implementation, and studying this over time (Kelly, 2010; Skivington et al., 2021). The quality, quantity and impacts of community engagement may be an important part of the PBA, and thus of the

evaluation (Taylor & Buckley, 2017). The multiple levels at which PBAs operate will also often need to be taken into account.

- Constructing local area datasets that map the locality of PBAs is also challenging (Bellafontaine & Wisener, 2011): PBAs often focus on areas that have social but not necessarily administrative meaning. It may be necessary to include adjacent areas to assess displacement of the problem being addressed, as well as diffusion of positive change.
- Timeframes for PBAs are often long, and evaluations would therefore need to capture short, medium and long-term change that emerges at multiple points in time.
- Attribution and causation are particularly challenging and the 'counterfactual' - what would have happened in the absence of the PBA - is hard to construct. Some would argue that attribution is neither feasible nor meaningful, and that assessment of the *contribution* made by a PBA should instead be the focus. The counter view, and the view held by the authors, is that there are available methods to measure impact robustly, with some form of comparison, and that the measurement of outcomes is challenging but not impossible.

Asking the right questions

- PBA evaluations fundamentally ask questions about what happened (processes), what difference was made (outcomes and impacts) and value for money. The review uses the RE-AIM framework (Holtrop et al., 2021) to illustrate typical evaluation questions in each of these areas, and YEF evaluation guidance provides further details of evaluation questions for different types of evaluation. Given their long term nature, evaluations will also often ask different questions at early and later stages of the PBA implementation.
- The planning stage of PBA evaluation is particularly important. It will often be useful to begin with evaluation questions about the rationale for the PBA, the theory of change and the 'theory of place' that is, how the local area and context is conceptualised in relation to the problem being addressed (how it stems from or is entrenched by place, and how it therefore needs to be addressed in a way that reflects the local place) (Davies, 2019; Taylor & Buckley, 2017). These questions may surface deep-rooted differences and tensions in the perspectives held by those involved.
- The purpose, intended uses and scale of ambition for the evaluation will also shape the evaluation approach. A robust theory of change and logic model may be aids to the development of a detailed evaluation plan because they set out or inform hypotheses about how the intended change will occur, which provide a useful framing for evaluation (Bellafontaine & Wisener, 2011; Dart, 2018; Sridharan, 2011; Taylor & Buckley, 2017).

- There is some criticism of the value of theories of change for complex interventions as they may not capture complex contexts sufficiently. Systems mapping is also proposed as a helpful approach at intervention design stage, evaluation planning or as an early evaluation activity. It is used to build a detailed understanding of the system, its context and how the PBA interacts with it and can be undertaken with stakeholders (Barbrook-Johnson & Penn, 2021; HMT, 2020a; Skivington et al., 2021) and systems thinking is a valuable element of PBA evaluation (Egan et al., 2019a and 2019b).
- Developmental and learning evaluation approaches, where early learning is fed back to programme developers to support adaptation (Bellefontaine & Wisener, 2011; Harrison-Evans et al., 2016; Kelly, 2010; Quinn Patton, 2011), may be particularly helpful in PBA evaluation, because of the evolving nature of PBAs. Early learning about whether intended activities are being used and seen as acceptable, and whether expected short and medium term changes are being seen, will aid course correction and refinement of approaches. The evaluation thus becomes a core tool for change and may become an integral part of the PBA itself (Kelly, 2010).
- Community involvement in the evaluation is also seen as important (Taylor & Buckley, 2017). There is value in involving multiple stakeholders in the design and reviewing of evaluation approaches, and in participatory and emancipatory approaches to involve local people in research activity.

Designs for evaluating PBAs

- Implementation and process evaluation is of heightened importance in PBA work because of the complexity and evolving nature of PBA contexts, designs and activities. Key elements of PBA evaluations often address activities and strategies, barriers and enablers (determinants), and implementation outcomes ('how well' the PBA was put into practice). A wide range of forms of data will be valuable to assess different types of change, from multiple perspectives (Bicket et al., 2020; HMT, 2020a; Skivington et al., 2021).
- Impact evaluation involves measuring change in intermediate (proximal) and final (distal) outcomes, at multiple levels - typically for individuals, for targeted groups and for the community population, as well as in organisations and local systems. Assessing whether the changes observed can be attributed to the PBA is usually a fundamental aspect of evaluation. Broadly, three sets of designs are available. These are not mutually exclusive. They can be brought together in combinations to test and strengthen analyses and provide a more rounded picture of change.
- Experimental and quasi-experimental methods involve either randomisation (often challenging for a range of reasons) or a counterfactual involving comparison between areas, over time, or both. Potentially, high quality designs that, well executed, can

attribute change to a PBA with high levels of certainty include cluster randomised trials, stepped-wedge designs, interrupted time series, difference-in-difference estimations, geographical matching, propensity score and other statistical matching methods, synthetic controls and instrumental variable approaches (Bicket et al., 2020; HMT, 2020a; Skivington et al., 2021).

- Approaches such as Social Network Analysis, modelling and stimulation approaches can be used to explore changes in systems (Bicket et al., 2020; HMT, 2020a; Skivington et al., 2021).
- Theory-based approaches test whether evidence supports hypothesised mechanisms for change. They do not provide estimates of effect size but enable assessment from multiple perspectives of whether the PBA might plausibly be said to have contributed to any observed change. Available methods include realist evaluation, qualitative comparative analysis, process tracing, contribution analysis, Bayesian updating, contribution tracing, Most Significant Change and outcome harvesting (Bicket et al., HMT, 2020a; Skivington et al., 2021).
- Commonly used methods for studying value for money are cost-effectiveness analysis and cost-benefit analysis. They are particularly challenging in PBA evaluation as they require that benefits and costs across the system are accounted for, recognising that a cost occurred in one service area may produce a benefit in a different one (HMT, 2020a).
- The most rounded and full evaluations will involve bringing together multiple methods and approaches in hybrid designs.
- The report summarises considerations guiding choice of methods and sets out 10 key messages for PBA evaluation.

Chapter 1 - Introduction

Purpose and methods

This report provides an overview of approaches and methods used in evaluations of place-based approaches (PBAs). It was produced to inform the Youth Endowment Fund (YEF)'s approaches to evaluating place-based approaches (PBAs), and to aid those commissioning, undertaking and taking part in PBA evaluations. It should be read and used alongside YEF's own evaluation guidance resources¹, to help to inform evaluation strategies in PBAs.

This report is linked with a synthesis of evidence about PBA models, implementation, and impact on youth violence outcomes (Baidawi, Valdebenito et al., 2023). In both that and the current report, PBAs are defined as collaborative, long-term approaches that:

- are based in a geographically bound location
- focus on local needs, local solutions, and the unique attributes of a place
- involve more than one statutory agency (e.g. schools, police, hospitals, health services, child welfare, youth justice); and
- demonstrate meaningful engagement and involvement of local communities (including private citizens, local services and organisations, and local community groups) in the design and/or delivery of the approach.

PBAs may be undertaken in a single area but may also part of a wider initiative where a number of sites are funded or supported to implement a PBA, using the same model and/or working to the same goals (Baidawi, Valdebenito et al., 2023). As well, PBAs vary in how sharply focused they are, from those targeting a specific measurable outcome for a defined population (e.g. reducing violent crime, or gang membership, among specific age groups) to those with more loosely described goals (see Baidawi, Valdebenito et al., 2023 for examples). YEF's investment in PBAs, involves supporting multiple local areas to design, plan and implement a PBA focused on reducing youth violence.

PBAs are complex initiatives for a number of reasons, including that they are used to address deep-rooted problems that are themselves complex, and because of inherent qualities such as their involvement of multiple organisations and individuals, long-term and adaptable nature, and the fact that they may emerge through interaction rather than the application of a fixed design. There is no single accepted 'best practice' approach to evaluation of PBAs, and there are debates and opposing views about what should be prioritised and the

¹ <https://youthendowmentfund.org.uk/resources-for-evaluators/>

appropriateness of different methods. In this report we describe the approaches and methods that are options, and the issues that funders or managers of evaluations, stakeholders in PBAs, and evaluators themselves, may need to consider in planning, designing and conducting evaluations.

The report draws on the following:

- A review of texts that provide guidance on evaluation methodologies for PBAs or for systems change and complex interventions. (We looked beyond guidance texts on PBAs because the literature on PBA evaluation is limited, and the review did not focus on youth violence as this literature is also limited.) Included texts were identified in several ways. We drew on the searches undertaken for a synthesis of evidence about PBA models, implementation, and impacts on youth violence (Baidawi, Valdebenito et al., 2023). We supplemented this with further searching for guidance texts (using Google and Google Scholar). We also followed up references in the reviewed texts, and references and pointers provided by a group of experts (see below). In prioritising texts for inclusion, we aimed to represent the diverse perspectives on evaluation and methods. The texts initially selected were summarised within a systematic framework, and further texts were consulted and incorporated into the report to expand or fill gaps. The reviewed texts we drew on are set out in the References.
- Discussion with an international group of nine experts, selected for their involvement in the conduct or evaluation of PBAs and complex interventions. Their input was particularly important to ensure the report is based on up to date approaches to evaluating PBAs. Their involvement varied, and included virtual and face-to-face contact for consultation about the texts to draw on, discussion of key themes emerging from the literature and to reflect in the guidance document, and review of an earlier draft of this guidance. The names of the eight experts involved are shown in the Appendix.
- The perspectives and experience of the authors, particularly where published guidance was limited or contradictory, and following recommendations from the international experts and YEF about issues to address.

This report consolidates ideas and recommendations from all three sources. Where content was drawn from reviewed texts, this is indicated in chapters through referencing; elsewhere content comes from the perspectives of the international experts and/or the authors.

The report begins by outlining how the features of PBAs raise challenges in evaluations. We set out considerations for initial evaluation planning, including the types of questions that might be addressed in evaluations, and the value of a clear evaluation framework. We then outline potential methods for PBA evaluations, and the considerations that inform selection of approaches. We finish with a set of summary key messages for the evaluation of PBAs.

Limitations

As noted, the report does not draw specifically on literature and views about evaluating PBAs relating to youth violence. The texts drawn on were not identified through systematic and replicable approaches as the study resources did not allow for this. Our intention was to identify and draw on 8-10 key texts, although many more than this were in fact used. The texts reviewed often diverge in the approaches recommended, there was not always consensus among the international experts consulted, and (as with evaluation methods in other contexts) there are no fixed and agreed rules about the 'right' or 'best' way to evaluate PBAs. The review provides recommendations, suggestions and pointers rather than a set of requirements. As noted the review should be read alongside YEF's evaluation guidance, and more work is needed to consider how to align them. Overall, it will be important to continue to keep under review the methods being used in PBA evaluation, to stay abreast of innovations and development, and to continue to reflect and learn.

Chapter 2 - Addressing challenges in Evaluation of PBAs

PBAs by their nature respond to the differing needs and resources of each 'place', meaning they often provide a tailored intervention that is developed by, for and with the community of interest. They are designed to meet unique conditions, emerge through interactions rather than a fixed design, and are flexible and adaptable. These inherent qualities of PBAs are their unique 'offer' and seen as their strengths. However they also create a number of challenges in evaluation. We begin by describing wider features of PBAs before moving to more specific evaluation challenges raised.

The discussion draws where possible on the reviewed literature (indicated with references), but also reflects the input of the international experts and the authors' experience and analysis.

Wicked problems

Wicked problems are problems that do not have clearly defined or finite causes, formulation and solution, and which are not easily or adequately resolved through individual services, organisations or programmes. They often involve multiple stakeholders with different views and perspectives, and cannot be addressed through 'trial and error' as their solutions are not easily reversible (Rittel & Webber, 1973; Bellefontaine & Wisener, 2011). PBAs are particularly used to address these deep-rooted, pervasive, and enduring social problems. Such problems require the 'collective' to become organised and to work together. There can be fundamental differences in understanding of root causes, whether particular populations (e.g. parents, young people or a local organisation) are part of the problem or part of the solution, what success would look like (e.g. a reduction in youth violence, or improved wellbeing, or community cohesion) and the approach that should be taken (Bellefontaine & Wisener 2011). There will also be differences in what different agencies, groups or parts of the system have to gain or lose - that is, what is at stake for different stakeholders and how they experience and view the results of PBAs.

Evaluation approaches therefore will often need to include methods to explore and include multiple perspectives, to capture diverse views about the problem and about changes observed and experienced, to understand the involvement and experiences of multiple players, and to better understand local systems and contexts (Egan et al., 2019a and b).

Emergence and variability

The emergent nature of PBAs means that it may take some time before the intended goals, impacts and ways of working are agreed across stakeholders, and they may change over time as the PBA adapts to changes in the context, new opportunities, or in light of learning about progress (Skivington et al., 2021). The approach may not be in a final state when evaluation

planning begins, and the evaluation approach may need to be adaptive and agile (Bellefontaine & Wisener, 2011).

In PBAs that involve multiple sites, there is also inevitable variation between settings. Although the expectation would be that there is some initial common ground in terms of needs, goals and target groups, these and the way in which initiatives develop will inevitably differ from place to place as they are influenced by local circumstances (Bellefontaine & Wisener, 2011). There is thus unlikely to be the same attempts at standardisation across PBAs that would be found in implementation of individual interventions, or there may standardisation of the purpose or function of different PLA components rather than on their form (Hawe et al., 2009).

Variation and evolving approaches make evaluation particularly challenging but, well handled, these features of PBAs can enrich learning. They provide scope for exploring diversity in implementation and outcomes across contexts and 'what works, for whom, in what contexts, how and why'. This requires methods to address context and its interaction with the PBA (see below), adaptive evaluation plans that are kept under regular review, documenting key changes in the intervention, its delivery and the context in which it is being applied. It also underscores the value of a developmental evaluation approach with two-way communication and learning between evaluators and those involved in implementation, underpinned by strong working relationships and trust (Egan et al., 2019a).

Complexity

Wicked problems arise from, and are held in place by, complexity. In a complex system, outcomes are created by a multitude of interdependent elements and forces within a connected whole (Bicket et al., 2020). For example, a relevant local system might involve young people, families, schools, practitioners and leaders of different organisations, services, social networks and groups, policies, funding etc. The causes and effects of changes in such systems are harder to evaluate than single interventions which aim to create a particular set of effects in a group of individuals in a linear fashion. In complex systems, changes may lead to other changes that may be hard to predict, because of feedback loops and long causal chains across many highly interconnected factors (Bicket et al., 2020; Parolini et al., 2019; Skivington et al., 2021). Behaviour may also adapt in response to attempts to intervene in such systems (Skivington et al., 2021). PBAs are usefully understood as an 'event' within a larger system (Hawe et al., 2009). For example, introducing new preventive programmes as part of a PBA might mean children are identified earlier with other needs. This might lead to changes in other services or new opportunities for information and joint work, and might lead to changes in the services involved in PBA governance and delivery. Such dynamics might also raise conflicts between organisational priorities, for example creating a pressure for a less punitive approach to policing that would compromise police clean-up rates.

Egan et al. (2019a) describe three levels of complexity that evaluators might need to address:

- the complexity of the intervention: its multiple components, flexibility etc.
- the complexity of the environment: the range of relevant people, groups and aspects of the local context, interacting as parts of a system
- the complexity of the consequences: the range of impacts it may have, expected and unexpected, for individuals, organisations and parts of the system and how they interact.

For evaluation, this means that it will often be useful to draw on wide-ranging data and analysis to explore issues across, and interactions within, the systems involved (Egan et al., 2019a and b). Evaluation may need an adaptive and flexible approach to take account of changes in the activities involved in a PBA, or to respond to new questions and issues, and a focus on close working and shared learning between evaluators and implementers is often useful. These issues can also make attribution and causality particularly difficult to pin down (Bicket et al., 2020) and hybrid methods (see further below) will be particularly useful.

Interactions with context

All places differ according to geographical, social, cultural, economic, political, and other dividing lines (Skivington et al., 2021). Particularly in the case of PBAs, context involves both observable features (e.g. geographical boundaries, population characteristics, measures of poverty) and features that are relational and dynamic (e.g. the history and quality of relationships, extent of collaborative working, quality of partnerships, local cultures, and feelings about the local area) (Greenhalgh & Mazano, 2021; Rogers et al., 2020).

Even with simple interventions, context always matters. It matters all the more with PBAs because multiple aspects of context may be enablers, inhibitors, and important prognostic factors (i.e., a contextual factor that strongly influences the outcome). Furthermore, the aims of PBAs will often include influencing within and between contextual dimensions (for example, aiming to change local cultures, develop new partnerships and reduce poverty), and PBA mechanisms of change will involve activating or changing contextual factors.

Context is shaped and changed by the PBA, and the PBA is shaped and changed by the context, making relationships dynamic and multi-directional. The evaluation itself will also impact on, and become part of, the context. Thus, context may affect every part of a PBA – goals scope, operations, and outcome – in unpredictable ways (Kelly, 2010; Skivington et al., 2021).

How 'to handle' context, both theoretically and practically, is therefore an important consideration in PBA evaluations.

There are frameworks that help to define context. For example, the EPIS implementation framework distinguishes between the 'inner context' of individual 'adopter' characteristics and intra-organisational characteristics (culture, climate, leadership, role, values, knowledge, readiness for change etc.) and the 'outer context' of inter-organisational networks,

legislation, policies, funding, leadership and collaboration (Moullin et al., 2019). The Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2022) similarly distinguishes between inner and outer settings, the former including organisational dimensions such as infrastructure, relationships, culture and tension for change, and the latter including attitudes, conditions, partnerships, policies, law and external pressures.

A recent scoping review of how context is included in implementation frameworks describes context dimensions at micro level (target population preferences, attitudes etc.), meso level (organisational culture, climate, readiness to change, support, structure), macro level (policies, regulation, networks) and multiple levels (social relations and support, funding, leadership, time availability and the physical environment) (Nilsen and Bernhardsson, 2019).

These issues are important considerations in evaluations. Evaluation methods that involve mapping or exploring local systems, and monitoring how they change over time, are a valuable component of evaluations (see further below). Because context is so important and multi-faceted, including it in evaluation is usefully shaped by hypotheses - developed ahead of time, as context is explored and changes, as well as after the fact (post hoc) testing. This conceptualising is key for assessing how context might be measured and its influences explored. Measurement of context is challenging but practical methods of assessment are needed if it is to be understood (Rogers et al., 2020).

Detailed data about key aspects of context will often need to be collected, both to be studied in their own right and to be used in the interpretation of other evidence, including understanding how intervention effects are modified by context. Such investigation of these questions will often lead to refined or new hypotheses, rather than firm conclusions.

Approaches for assessing context may involve describing the characteristics and possible influences of context, exploring correlation between contextual factors and other findings, using contextual prognostic factors at baseline to identify comparison areas, and exploring how context interacts with the PBA and vice versa. Context is also time sensitive, and a temporal perspective is often needed.

Community and stakeholder engagement

One feature of the variability inherent to PBAs, and a particularly important part of context, is community and stakeholder involvement, which is likely to involve a range of formal and informal organisations, groups and individuals (Taylor & Buckley, 2017) - any of which can change membership or character over time.

The involvement of multiple stakeholders brings contrasting perspectives on goals, activities and approaches of the PBA, and contrasting perspectives on what success would look like and on evaluation priorities and approaches. Shared ownership means that objectives and activities need to be renegotiated and hence change over time, driven by stakeholder

interests, deliberation, and local opportunities. In PBAs that are part of a wider initiative, there may also be tensions in views about goals and approaches between national and local level stakeholders, as well as in how change is experienced, and power dynamics and sensitivities at play (HMTreasury, 2020a). Developing a full understanding means exploring these different experiences and perspectives.

If community and stakeholder involvement is an important component of a PBA, it will be relevant to find ways to measure, or assess, whether and how it happens, and perceptions of its contribution to the observed changes. This might involve measuring the quality and quantity of stakeholder participation in decision-making, formal meetings, co-design activity, consultation and participative exercises, and delivery. The relevant groups might include statutory agencies, formal and informal organisations, networks, community groups, different groups of local people, local influencers etc. Effective community and stakeholder involvement requires people and organisations to be ready for this work (e.g. for foundational relationships and trust to be in place, with supportive conditions and cultures) and it may therefore also be helpful for evaluations to assess the readiness of communities for this collaborative work as an aid to understanding its success or otherwise. The centrality of community involvement in PBA theory and implementation also argues for community involvement in evaluation design, conduct and sharing learning (see further below).

Operating at multiple levels

A further challenge in evaluating PBAs is the multiple levels at which PBAs operate, and at which changes may be sought. PBAs typically aim to achieve change not only for the individual people who receive or use services but also for the neighbourhood population as a whole (or for sub-populations such as young people, but including those not directly touched by PBA activities). PBAs may also aim to stimulate changes in services, organisations and systems, and to improve neighbourhood conditions. PBAs target multiple levels for change, and they are enacted or implemented through work at multiple levels.

Baidawi, Valdebenito et al., (2023) provides examples of targets for change in PBAs at multiple levels:

- Individual: impacts for individual people (e.g. young people)
- Microsystem: that is, things that have direct impacts on those in the immediate environment (e.g. parents and schools)
- Mesosystem: interactions between microsystems (e.g. between schools and families)
- Exosystem: impacts on formal and informal social structures which affect the microsystem (e.g. neighbourhood connectiveness, the local service system)
- Macro system: cultural elements (e.g. poverty or attitudes to young people).

Frameworks such as EPIS (Moullin et al., 2019) and CFIR (Damschroder et al., 2002) both noted above, identify the multiple levels that influence implementation and aid decisions about what might need to be measured and how. Whatever the framing, evaluations need to focus on multiple levels and to consider the interactions between them with both influencers and outcomes in mind.

Challenges in measuring change

There are a number of challenges in measuring or assessing change in evaluations of PBAs.

First finding ways of measuring features of contexts and of change, consistently and in a way that can be replicated across time (and if relevant sites) is a key challenge in evaluations of PBAs (Kelly, 2010). It is particularly challenging to define and measure common but abstract PBA concepts such as partnerships, capacity building, and participation, and to capture systems change, e.g. changes in policies, processes, relationships, and power structures (Bellefontaine & Wisener, 2011). Likewise, norms, values and intentions may be important enablers or hindrances in a changing system, but are hard to measure robustly. However if these are part of the intended outcomes of PBAs, there is scope for ambition about measuring them. Identifying validated standardised measures and assessing their fit to the intended outcomes of PBAs, or implementation components, will be a first step. Where they do not exist, careful design and testing of new measures may be required. Qualitative exploration will also be particularly valuable here.

Second, there will often be challenges in the availability of data that align completely with the PBA locality (Bellefontaine & Wisener, 2011). Especially when PBAs are implemented in areas that do not coincide with local administrative borders (streetcorners, 'hot spots', neighbourhoods or other hyper-local areas) it might be difficult to find or capture data that maps directly to locality. Mapping as closely as possible is necessary for the precision and trustworthiness of estimates of effects, and for comparisons between areas.

In addition, it may be valuable to look beyond the immediate locality of focus to examine displacement and spill-over effects. For example, in the context of PBAs to reduce crime, criminal activity may be displaced to neighbouring localities, and the positive impacts of change (e.g. improved collaboration, evidence sharing or civic pride) may also ripple out to neighbouring areas.

Box 1. Study example: Analysis of spillover effects

Name of intervention	City of Los Angeles Mayor's Office of Gang Reduction and Youth Development (GRYD)
Location	Communities in Los Angeles with high levels of gang violence.
Intervention	GRYD incorporates community engagement, youth prevention programming, direct service intervention programming including life skills and problem-solving strategies, and collaborative violence interruption efforts involving community

	intervention workers and police. It aims to reduce the allure of gangs, gang membership and embeddedness, and retaliation and targets at risk 10-15 year olds and their families and gang involved youth and adults aged 14-25.
Methods	Changes in crime and disorder were measured over a 13-year period in GRYD intervention areas, control areas and displacement areas. Displacement areas were immediately adjacent to intervention areas, control areas were slightly more distant but selected to be ecologically similar in terms of demographics, violence levels and socio-economic conditions. The evaluation used a difference-in-difference approach and had to take account of extension of GRYD intervention areas into control and displacement areas.
Findings	The analysis found a statistically significant reduction in violent crime in intervention areas and no evidence for displacement to adjacent areas (nor of diffusion of possible benefits).
Reference and link	Brantingham, P., Tita, G. & Herz, D. (2021). The Impact of the City of Los Angeles Mayor’s Office of Gang Reduction and Youth Development (GRYD) Comprehensive Strategy on Crime in the City of Los Angeles. <i>Justice Evaluation Journal</i> , 4:2, 217-236. To access the full report, visit: https://www.tandfonline.com/doi/full/10.1080/24751979.2021.1887709

Third, evaluations need to align with expected timelines for PBAs to produce change. Because PBAs involve coordinated planning and work across multiple levels and agencies to tackle deep-seated issues, their timeframes may stretch to 10–20 years or further into the future (Bellefontaine & Wisener, 2011). The timeframe for evaluation needs to be considered carefully, with recognition that assessing impacts on ultimate goals will require a long term commitment. In the short or medium term, the focus will be on whether programmes are stimulating the changes the initiative seeks, i.e., as laid out in a theory of change. Different forms of change emerge at different time points, as changes in partnerships, structures, services and policies within a community typically precedes changes in population outcomes. Evaluations would also ideally assess if actual, observed change is durable, and whether approaches that show some promise in the short run go on to demonstrate lasting effects. Evaluations therefore need to allow enough time for it to be feasible to capture change, and a phased approach, discussed further below, is therefore helpful.

Challenges in attribution

A final, and key, challenge is the challenge of attribution, one of the most contested aspects of PBA evaluation. Understanding whether observed changes have occurred because of the intervention, and would not have occurred without it, is a fundamental aspect of evaluation: arguably the hardest aspect of evaluation of any intervention, but particularly challenging in PBA evaluations (Bellefontaine & Wisener, 2011; Bicket et al., 2020).

Counterfactual approaches are widely viewed as the most robust (or only) way of assessing attribution (Bicket et al., 2020). They compare what happened in the PBA area with what

happened in an equivalent of the area absent the PBA. Using a counterfactual depends on being able to control or account for context so that the two units of analysis are identical except for the presence of the PBA. There are a range of possible designs, described in Chapter 4, but which essentially involve either comparison between PBA and non-PBA areas, or comparison between PBA and non-PBA periods of time.

Counterfactual designs are used to measure the impact of PBAs (see Baidawi, Valdebenito et al., 2023, which includes a synthesis of twenty impact studies, 17 using quasi-experimental designs and three using RCTs). However, a key challenge for PBAs is that context cannot easily be controlled for, partly because PBAs interact with and change the local context. In addition, in a multi-site PBA initiative, comparisons across areas are more robust if the intervention is standardised, but standardisation may be in tension with the goal of reflecting the local area and local community engagement (Bicket et al., 2020; Hawe et al., 2009).

These challenges mean that some would argue that attribution is both infeasible in the context of PBAs, and not meaningful. The alternative approach put forward is to assess 'contribution' rather than 'attribution', that is, to develop plausible hypotheses for the way in which the PBA has played a part in bringing about change, rather than using statistic estimates of causality to measure its influence. Contribution approaches (Bellefontaine & Wisener, 2011; Bicket et al., 2020; Mayne, 1999, 2008) involve developing and testing hypotheses about outcomes and how they might arise, and assessing whether the change has occurred and whether the PBA can be said to have contributed. The approach aims to establish the plausibility of cause and effect by creating evidence for the hypothesised change mechanisms, using multiple lines of evidence, and assessing whether there are alternate explanations.

The possible designs for these approaches are discussed in Chapter 4. As we note there, given the potential limitations of either contribution or attribution approaches, we would strongly recommend that evaluators seek to combine both and use triangulation to strengthen their analyses and the confidence with which claims can be made.

The need for ambition and creativity to meet challenges

In summary, it is the inherent features of PBAs, which many would see as core to their promise and value, that raise challenges in evaluation, and it is therefore recommended that evaluators are ambitious and creative about how they might be met. PBAs involve very substantial public and social investment, and robust evaluation of their ability to achieve meaningful social change is vital. If done well, evaluation can provide rich insight into 'what works, in what contexts, for whom, how and why', and can be a key contribution to the PBA effort as well as generating learning for application elsewhere. As further chapters sets out, there is a range of potential evaluation methods and approaches available to tackle these challenges.

Chapter 3 - Planning PBA evaluations

This chapter provides an overview of the considerations involved in planning PBA evaluations. We begin by reviewing the types of research questions that are typically asked in evaluations of PBAs. This is important both because the nature of PBAs means that a wide-ranging set of questions is likely to be relevant, and to provide context for the following chapter which discusses evaluation methods. We also discuss key considerations in planning evaluations, and the value of developmental evaluation and a focus on learning.

As with other chapters, the content draws where possible on the reviewed literature, but also reflects the input of the international experts and the authors' experience and analysis.

Key questions in PBA evaluations

Broadly, the aim of evaluation is to learn about successes and failures to support decisions about improving, sustaining and replicating PBA efforts, and to build the evidence base for future initiatives. This means asking questions about processes and implementation, and about effectiveness and cost-effectiveness. As well as existing YEF evaluation guidance, several of the resource reviewed provide examples of questions that might be asked in a PBA evaluation (e.g. Dart 2018; HMT 2020a; Smith 2011; Skivington et al. 2021). In the table below we have also used the RE-AIM framework (reaim.org; Holtrop et al., 2021) to provide a structure for thinking about relevant questions. The RE-AIM framework identifies five dimensions (Reach, Effectiveness, Adoption, Implementation and Maintenance) which give rise to a range of questions set out in Table 1.

Table 1. Example questions in PBA evaluations

RE-AIM dimension	Definition	Example questions in PBA evaluations
Reach ²	Number and representativeness of individuals participating	<ul style="list-style-type: none">• Did the PBA reach the target groups? Including their most marginalised members?• Who was involved and in what?• How strong was community involvement?• To what extent did participants engage with it?
Effectiveness ³	Impact on outcomes, including unintended and negative and economic outcomes	<ul style="list-style-type: none">• What were the observed changes, including proximal short term effects and distal longer term effects? How do these relate to the theory of change?• What changes were observed at different levels (e.g. for individuals, at a local population level, for organisations or the wider system)?• Were the observed changes equitably experienced?• Were there unintended or negative effects?• Were different groups impacted in different ways?• Was there variation in outcomes across sites?

² Aligned with reach and responsiveness in YEF evaluation guidance.

³ Aligned with evidence of promise and cost effectiveness in YEF evaluation guidance.

		<ul style="list-style-type: none"> • To what extent can outcomes be attributed to the PBA? How much confidence can be placed in that assessment? • At what timepoints did different outcomes occur? • How did contexts influence the results? • What features of the PBA were most relevant to its impacts? • What were the costs of implementing the PBA, compared with benefits? • Where in the local system did costs and benefits fall? • How does the cost-effectiveness of PBAs compare with alternative approaches?
Adoption ⁴	Number and representativeness of settings and individuals initiating the intervention	<ul style="list-style-type: none"> • This would be most relevant to multi-site PBA initiatives and would ask whether all intended sites initiated the PBA
Implementation ⁵	Delivery as intended, adaptation, costs of delivery	<ul style="list-style-type: none"> • What constituted the PBA? What mix of investments and activities were involved? • Was it implemented as intended? What adaptations were made to the intended model? • How was it differentiated from previous ways of working locally? • How was it implemented? What range of implementation strategies were used? • What were the barriers and how were these addressed?. What facilitating factors existed and how were they leveraged? • How well was it implemented? What worked well, where and why? Was it implemented equally well for all sub-groups, including the most marginalised or disadvantaged? • Was their readiness and capacity for the PBA, with supportive cultures, coordination and leadership? • Was there sufficient readiness for a PBA from the start? What preparatory work was needed and undertaken? • How did aspects of the context influence strategies and their outcomes? Were there important changes in the social, economic, or political context as the PBA was implemented? How did such changes influence how the PBA was implemented?
Maintenance ⁶	Sustainment and institutionalisation of implementation, sustainment of impacts	<ul style="list-style-type: none"> • Is the PBA still being implemented, and/or is it seen as sustainable? Are core agencies and people still involved? • Is there (a need for) continued funding for the PBA? • Is the PBA incorporated into strategies and priorities of key agencies? Has it become 'business as usual'?

⁴ Not specifically referenced in YEF evaluation guidance.

⁵ Aligned with e.g. fidelity/adherence, dosage, quality, differentiation, adaptation and factors affecting implementation in YEF evaluation guidance.

⁶ Not specifically referenced in YEF evaluation guidance.

Stages of implementation

For PBAs as for any other intervention, it is helpful to see implementation as a staged approach and to align evaluation activity with stage of implementation. Given their duration and complexity, PBA evaluations are often designed as phased approach, addressing different questions at different points in time. For example, the widely used EPIS model (Aarons et al., 2011) describes four implementation phases:

- Exploration: identification of the issue/s to be addressed and of a PBA as a potential solution
- Preparation: decision to adopt the PBA, development of the intended approaches, and o resources, capacity, infrastructure and 'readiness'
- Implementation: putting the PBA into practice, likely over an extended period of time
- Sustainment: maintaining, institutionalising and potentially also scaling the PBA to new geographical areas or issues.

Table 2 shows how the focus of evaluation questions changes across stages of implementation (repeating some questions shown in Table 1 to illustrate their fit with specific phases of activity).

Table 2. Example evaluation questions at different stages of implementation

Exploration phase questions	Preparation stage questions	Implementation stage questions	Sustainment stage questions
<ul style="list-style-type: none"> • What is the problem? Who and how many are affected, how are they affected, and where are those people located? • How has the problem developed over time? What are the causes of the problem and what needs to change to address it? • Why is it important to address this social problem in a place-based way? • What is the current best evidence about effective approaches? What has been tried in the local settings? 	<ul style="list-style-type: none"> • How was the PBA developed - what tensions, compromises and trade-offs were involved? • What are the key hypotheses about how change will occur and what evidence supports this thinking? • What and where are the risks and uncertainties? • What baseline evidence is available, and can it be used to measure future change? • Is the suggested PBA sufficiently well-developed and described? • What is known about 'readiness for change' on the ground? • What would success look like? What could go wrong and what could stall the initiative? 	<ul style="list-style-type: none"> • Is the PBA being implemented as intended? • What are early and later indications of change and how do these relate to the theory of change? What is the scale of change? • How do impacts vary for different groups • Are there any unintended consequences? • Which hypotheses are supported or otherwise? • What new hypotheses are surfaced? • How can design or implementation be improved? • Were the results, fully or partially, caused by the PBA? What confidence is there in this assessment? • What works for whom, how and why, and why does this vary across contexts and population groups? • What were the costs involved, and how do these compare with benefits achieved? • What have we learned about what works from this way of trying to achieve social change? • How transferable are the lessons? 	<ul style="list-style-type: none"> • Is the PBA still being implemented? Are core agencies and people still involved? • Are the resources and capacity needed still in place? Has the source of funding changed and is it sustainable? • Is the PBA incorporated into strategies and priorities of key agencies? Has it become 'business as usual'? • Have the impacts for individuals and groups been sustained at periods of time after initial outcome measurement?

Theory of place and theory of change

It is usually valuable to understand the rationale for the PBA and for working in this way in, this place, or this set of places (the 'theory of place': Davies, 2019; Taylor & Buckly, 2017) early in the evaluation process. This involves assessing the distribution and seriousness of the problem, asking questions about the nature and magnitude of the problem, the characteristics of the groups affected, the outcomes viewed as desirable by different stakeholders, and similar foundational issues.

Documenting a clear and robust theory of change and a more detailed logic model is seen as being of particular significance in PBA evaluation planning, because of the multiple stakeholders involved and because of the potentially wide-ranging focus and work. These models also help to surface the trade-offs and prioritisation that are often needed in planning PBA work and provides an important foundation to theory-based evaluation approaches. Involving multiple stakeholders in developing a theory of change helps to create shared expectations about intended outcomes (at different time points and levels), the activities that are envisaged, assumptions about how they will bring about the intended changes (mechanisms of change), the timelines required and the necessary preconditions (Bellefontaine & Wisener, 2011; Dart, 2018; Taylor & Buckly, 2017).

It is helpful in early evaluation planning to test the theory of change, asking questions about the intended outcomes (immediate, short-term, and ultimate), testing whether the theory is sound and supports expectations of achieving such outcomes, and considering whether the theory still holds when multiple perspectives have been heard and included. The theory of change is most usefully viewed as a live document, kept under review and revised during the evaluation to reflect learning and adaptation (Skivington et al., 2021).

There is some criticism of the value of theories of change for complex interventions as they may not capture complex contexts sufficiently. Systems mapping is also proposed as a helpful approach at intervention design stage, evaluation planning or as an early evaluation activity. It is used to build a detailed understanding of the system, its context and how the PBA interacts with it and can be undertaken with stakeholders (Barbrook-Johnson & Penn, 2021; HMT, 2020a; Skivington et al., 2021). Systems thinking is a valuable element of PBA evaluation (Egan et al., 2019a and 2019b).

Other planning considerations

The complexity of PBAs, and the range of questions asked at different levels and points in time mean that the planning stage for PBA evaluation - a key part of any intervention evaluation - is of heightened importance. Achieving alignment between the PBA and

evaluation requires purposeful effort, and the involvement of multiple stakeholders and partners.

It is also helpful at this stage to clarify the overall purpose of evaluation and the scale of ambition and resources. Is the purpose primarily to support design and ongoing implementation of the PBA, to test whether it has achieved its aims, to assess whether the approach should be used elsewhere, to add to the evidence base about PBAs, or for accountability to a funder? Who is the key audience, and when and how is it intended that they will use the findings and learning?

There is also a need to honestly discuss ambitions (Sridharan, 2011). Evaluation methods and resources required differ significantly between seeking early insight into whether a PBA can engage the local community and seems feasible to implement, and testing whether a completely implemented initiative reduces youth violence or achieves other population level goals. Clarity about expectations, the resources available and other constraints and the implications for evaluation designs, is key. A further key element of planning is to consider expectations about the timeline over which results might be observable (Dart, 2018; Sridharan, 2011).

There are likely to be differences among stakeholders and partners in all these areas, which need to be surfaced and, where possible, resolved. Ultimately there will be trade-offs, between different stakeholders' preferences, between different potentially prioritised questions, and when ambitions collide with realities. These, and the resulting decisions, need to be made explicit in the evaluation plan.

The evaluation plan⁷ will usefully set out:

- the main evaluation questions and sub-questions, focusing on the outcomes and assumptions that will be tested and why these were prioritised
- indicators and measures for outcomes, at multiple levels and for multiple parts of the system
- the points in time at which changes are hypothesised to occur
- the data needed, and how it will be sourced, analysed, triangulated and interpreted.

Developmental and learning evaluation approaches

Because PBA approaches are iterative and adaptive, developmental evaluation is likely to be particularly relevant. Developmental evaluation is used where a new approach is being tested

⁷ The evaluation plans would need to align with the YEF protocol template in YEF-funded evaluations.

and provides rapid real time feedback to programme developers to support adaptation (Bellafontaine and Wisener, 2011; Harrison-Evans et al., 2016; Kelly, 2010; Quinn Patton, 2011). It usually focuses on implementation and on early indicators of change rather than final outcomes, and can provide rapid feedback from multiple perspectives and support early insight into whether hypothesised changes and linkages are observed or not.

Early learning should be shared to influence the PBA design and how it will be implemented, and may inform adjustments to the theory of change and decisions about measurement and data collection as well identification of risks and uncertainties (Taylor & Buckley, 2017). Evaluation during implementation provides important learning about whether intended activities are being used and expected short and medium term changes are observed, which contributes to understanding of the implementation process, possible mechanisms of change and early indicators of impact. Later stages of evaluation typically focus on final outcomes and on the sustainability of the PBA and the changes it has brought.

The evaluation thus becomes a core tool for change, supporting continuous adaptation of the PBA and its implementation, and may become an integral component of the PBA itself.

There is also a need for a strong mutual learning orientation between evaluators and those involved in implementation. This needs to be underpinned by strong relationships based on mutual respect and trust which can challenge conventional notions of hierarchy and expertise.

Community involvement

It will often be valuable to involve multiple stakeholders in planning and reviewing evaluation approaches. Participatory and emancipatory approaches, in which local people lead or are involved in research studies including scoping objectives and methods, undertaking fieldwork and analysis, are also sometimes used in PBA evaluation, particularly to surface local sub-cultures, priorities and perspectives (Taylor & Buckley, 2017).

Chapter 4 - Designs for Evaluating PBAs

This chapter provides an overview of designs for evaluating PBAs. We begin by discussing implementation and process evaluations. We then discuss approaches to assessing the impact of PBAs, both methods that allow the *attribution* of impacts (experimental and quasi-experimental methods) and those focused on assessing the *contribution* of PBAs to observed changes (including theory-based approaches). We conclude with commentary on the value of combining both sets of approaches.

The chapter aims to provide insights to support the commissioning, planning and undertaking of PBA evaluations. As with previous chapters, the content is drawn from the reviewed literature (highlighted with references) and is also informed by the international experts' input and the authors' own experience and perspectives.

Implementation and process evaluation

Implementation and process evaluation is of heightened importance in PBA evaluations because of the complexity of PBA contexts and activities, and their evolving nature. It provides valuable insight into how well the PBA is operating, for course correction and learning for wider application and scaling. It also provides an opportunity to determine how implementation elements may influence the impacts and outcomes of the PBA. Finally, in multi-site evaluations, implementation and process evaluations also provide key information about how the PBA intervention differed between sites, which provides insight into what works in different contexts and into external validity (or transferability) of findings. The approach will usefully be shaped around hypotheses formulated and key assumptions about how change will occur, such as the engagement of communities.

Key elements of implementation and process evaluations of PBAs are likely to include:

- Implementation determinants (barriers and enablers): implementation and process evaluations typically describe the context in detail and identify how aspects of the context (at multiple levels), the PBA approach itself, and the ways in which it was implemented presented barriers, and how (well) these were addressed or leveraged. They aid understanding of the necessary conditions for success, and assessments of 'readiness' for place-based activity. Barriers and enablers will emerge from (Damschroder et al., 2022):
 - the design and nature of the PBA initiative itself
 - the people involved: young people, families, local community, professionals, organisational and PBA leaders
 - the organisations involved in the PBA: e.g. structures, cultures, networks, compatibility of the PBA with priorities and workflows

- the outer setting e.g. the wider community, network, system, policies, funding
- the implementation processes involved.
- Implementation strategies: the activities involved to put the PBA into practice across the course of delivery. This will include examining what was actually done and the approaches used, such as convening, community engagement, collaboration, devolved decision-making, developing common agendas, shared use of evidence and data, governance etc., as well as the specific services delivered as part of the PBA (Powell et al., 2012). Exploring strategies involves assessing who was involved, the roles played by different organisations or groups, and the timing and intensity of activity: how much was done, when, for how long, and in what sequence (Presseau et al., 2019).
- Implementation outcomes: or assessments of 'how well' the PBA and its various components were put into practice, from multiple perspectives (Proctor et al., 2011). If aims are not achieved, this is crucial analysis for distinguishing intervention failure from implementation failure. Key considerations here will include whether the PBA model was:
 - feasible
 - implemented as planned, and adaptations required
 - taken up by and reached the populations, groups and parts of the local system as intended
 - acceptable and engaged stakeholders as intended
 - integrated into local systems
 - sustained or viewed as sustainable
 - scalable.

One of the challenges in PBA evaluation is that it is not always easy to distinguish conceptually between implementation and intermediate goals in theorising and planning for change. For example, shared vision, shared evidence use, improved partnership working, and community engagement may be both fundamental aspects of the PBA approach and intermediate goals. The theory of change may help to determine how to treat them.

Implementation and process evaluations utilise a multitude of methods to gather information, both quantitative and qualitative, and many data sources, including existing administrative and project specific data (see Table 3).

Table 3. Potential methods for PBA implementation and process evaluation

Method	Short description	Examples of use
Surveys	Surveys in the form of questionnaires provides quantitative data about e.g. behaviours, experiences, preferences, and attitudes among informants, and provide measurement for assessment of outcomes and effects. They may be one-off or longitudinal (repeat cross-sectional or panel) and might involve any of the populations involved in or touched by PBAs	<ul style="list-style-type: none"> • Experiences of engagement and participation • Measures of local people's sense of social cohesion • Young people's victimisation or weapon carrying
Qualitative interviews with individuals or groups	Provide data in the form of text where key informants describe and explain views and experiences about processes that they are part of.	<ul style="list-style-type: none"> • Capturing preferences for forms of engagement • Understanding drivers of local cohesion
Observation or participation	Offers evaluators a direct encounter with people engaging in processes, with or without active participation.	<ul style="list-style-type: none"> • Assessing extent of shared decision-making • Observing how young people interact in a hot-spot area
Deliberative processes	Moving beyond experiences and opinions, involve information sharing and discussion to get stakeholders to make suggestions, appraise and prioritize.	<ul style="list-style-type: none"> • Shared decision-making about spending priorities • Service or policy co-design
Management and monitoring data / shared measurement systems	Data collected centrally within the PBA or by individual agencies about activity relevant to PBA. May be set up as a shared measurement system across agencies to track selected performance indicators consistently	<ul style="list-style-type: none"> • Tracking progress in PBA activity, reach, timings, costs etc
Document analysis	A variety of documents may inform the process evaluation, e.g. meeting agendas and minutes, policies, procedures, case descriptions	<ul style="list-style-type: none"> • Assessing shared decision-making and governance • Monitoring changes in policies and procedures or whether they reference PBA work
Ethnography	A set of methods (including interviews, observation, analysis of documents and artefacts e.g. photographs) that allow for detailed investigation of the structures, interactions, and beliefs in a group, eliciting information about their norms and culture.	<ul style="list-style-type: none"> • Understanding local relationships and behaviours and how they change • Understanding community cohesion or belonging • Exploring working relationships and cultures between PBA groups or organisations
Systems mapping, casual loop diagrams and social network analysis	A range of methods for generating a map of key systems components and interactions or of social networks	<ul style="list-style-type: none"> • Understanding underlying causes and potential levers for change • Assessing which systems components are involved in or influence the PBA • Assessing systems level impacts and changes • Can particularly identify unexpected effects

Box 2. Study example: implementation and process evaluation

Name of intervention	Youth Violence Prevention Training and Technical Assistance (YVP TTA) Initiative
Location	Communities with high rates of youth violence (YV) across the USA
Intervention	A 5-year training and technical assistance (TTA) initiative designed to increase the capacity of local health departments (LHDs) to coordinate a multi-sectoral, public health approach to preventing youth violence (YV) at the community level. The initiative included support to build and sustain a multi-sector coalition for YVP; provision of training and materials to support the development of local expertise and leadership in YVP; and support in the development of a comprehensive local YVP plan.
Methods	The evaluation explored (1) the level of engagement with TTA offered, (2) progress towards the development of local YVP plans, (3) improvements in local capacity and infrastructure and (4) how active a role LHDs played in leading YVP efforts locally. The evaluation used mixed methods approach, including administrative data from an online system tracking requests for TTA, semi-structured group interviews with sector leaders and surveys.
Findings	Results indicated variation in uptake of TTA across sites, however, several target outcomes were achieved including increased representation and engagement of diverse perspectives in local YVP efforts, and strengthened infrastructure and integration of YVP at LHDs.
Reference and link	Dymnicki, A., Katz, J., Young, X., Thorngren, M., Orazi, J., Marshall, K. & Lumpkin, C. (2021). Supporting Local Health Departments to Lead Multisectoral Youth Violence Prevention Efforts. <i>Health Promot Pract</i> , 22(6), 863-872. doi: 10.1177/1524839920947766. To access the full report, visit: https://pubmed.ncbi.nlm.nih.gov/32762377/

Impact evaluation

Impact evaluation involves both measuring change and efforts to determine whether the change was caused by the PBA. As outcomes also are affected by events and influences other than the PBA, such changes cannot directly be interpreted as a PBA effect. How to account for this is noted earlier, a particularly contested aspect of PBA evaluation. Important questions also concern the size, consistency, and variability of effects, whether different groups or places experienced different effects, and how effects develop over time.

Measuring change

Impact evaluation measures changes in outcomes that represent the intermediate and ultimate goals, at different levels. The distinction between intermediate and final outcomes is an important one. For the target population, intermediate goals might involve changes in awareness, knowledge, attitudes, behaviours or skills of the individuals that participate in the intervention. They may also involve antecedents (e.g. mental health problems, exposure to family violence) or proxy variables (e.g. availability of weapons, gang involvement and

development of positive behaviours and attitudes) that are closely linked to, but not, the ultimate goal.

For organisations and systems, the composite, collaborative and multi-level nature of many PBAs make it important to measure contextual and systemic changes e.g. the quality of collaboration, the comprehensiveness of service provision, the allocation of resources etc. how people and organisations come together.

Such intermediate or proximal effects are valuable, and different from distal endpoints that directly describe the social problem addressed, which also need to be measured.

Distal outcomes may be measured in relation to targeted individuals (e.g. those receiving interventions that are part of the PBA), the target population or the whole local community. Measurement needs to reflect the targeted area (which may be a neighbourhood or street corner), but will ideally also include adjacent areas to detect possible displacement or the diffusion of benefits.

Data for measuring distal impacts may be generated from surveys (among target population or the whole community), or from administrative data from one or more agencies. For example in the context of youth violence this might involve police and crime data on recorded cases, arrests, charges, convictions, reconvictions etc; youth offending services on referrals and disposals; hospital data on serious injuries etc. Administrative data will assist in the establishment of the intervention baseline and in testing equivalence of intervention and control areas (see below). Such data can also be used in propensity score matching, where controls are artificially created from large databases of individuals with the same statistical propensity as the target population, and access to historical administrative data is essential in time series analysis or difference-in-difference designs. In practice working with administrative data is a challenging endeavour requiring extensive discussion, negotiation and considerable expertise to structure and analyse the data well.

Assessment of outcomes (both intermediate outcomes and ultimate outcomes) also needs to involve consideration of unintended consequences, both positive and negative.

Assessing attribution or contribution

Assessing what difference the PBA itself has made to the changes observed is a fundamental aspect of evaluation. All social problems change in form, intensity and reach over time because of a myriad of causes and influences, and assessing whether the change caused by the PBA is larger than would be expected if the PBA had not been in place is a key consideration.

There are broadly three sets of designs for making these important judgements: experimental and quasi-experimental designs, modelling and simulation approaches, and theory-based approaches.

Experimental and quasi-experimental designs

Comparison with areas unaffected by the PBA allows for reasoning about what would have happened in the PBA areas in the absence of introducing the PBA initiatives (i.e. the counterfactual). For robust and 'fair' comparisons, this requires comparison with areas that are similar in other relevant ways (e.g. level of poverty, level of crime, social cohesion) except that PBAs were introduced in some places and not in others.

In principle, where it is feasible, randomisation provides the strongest support for inferences about causation, and attributing change to the PBA becomes more trustworthy. The two key design options are cluster randomised controlled trials (where groups of areas are assigned to the PBA as intervention or to a control group) and stepped wedged designs, where all the areas are assigned to the PBA condition but at controlled times e.g. randomly assigned in rounds.

Box 3. Study example: a randomised control trial (RCT)

Name of intervention	Communities That Care (CTC)
Location	24 communities across the USA
Intervention	Communities That Care (CTC) is a manualised framework to guide community coalitions to improve child and adolescent development outcomes, using evidence-based preventive interventions tailored to a community's specific needs. It involves: 1) recruitment of community leaders, 2) formation of a local coalition, 3) use of data to prioritise risk and protective factors, 4) selection of appropriate interventions, and 5) implementation and monitoring of selected interventions. It aims to produce community-level changes in prevention service systems characteristics (e.g. increased collaboration), leading to reductions in risk factors in the community and reduced adolescent delinquent behaviours and substance use among young people.
Methods	Areas were randomised to intervention and control. Youth survey data were collected from a panel of students in intervention and control areas (pre-intervention baseline assessment at grade 5; interim assessment at grade 6; endline assessment at grade 7 to assess (1) risk factors, (2) delinquent behaviour, and (3) substance use.
Findings	The evaluation found significantly lower mean levels of targeted risks and initiation of delinquent behaviour in CTC communities compared with controls, but no significant difference in substance use initiation.
Reference and link	Hawkins, J., Brown, E., Oesterle, S., Arthur, M., Abbott, R. & Catalano R. (2008). Early effects of Communities That Care on targeted risks and initiation of delinquent behavior and substance use. <i>J Adolesc Health</i> , 43(1), 15-22. doi: 10.1016/j.jadohealth.2008.01.022.

To access the full report, visit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3867289/

Randomisation is challenging for several reasons. It requires a large enough number of areas to allow for differences in contexts to be taken into account. Ideally places would be at similar starting points in terms of readiness for a PBA approach, or difference in these factors can be measured and controlled for in analyses. It is also difficult to constrain what happens in the control areas. If areas are at similar stages of readiness it is unlikely that, if assigned to the control group, they would do nothing to take these and other similar aspirations forward. Randomisation also requires a degree of consistency across sites in the intervention being used so that effects across PBA initiatives can be collated and the results generalised. Finally there is also a sample size problem. There are only so many comparable areas, and power to detect statistical differences in outcomes can be easily compromised if sufficient places and their comparisons cannot be obtained.

A stepped wedged approach (involving random allocation to 'rounds' of PBA activity) may be more feasible than cluster randomisation, although the timeframes for change may require significant time gaps. Randomisation might also be more feasible if the focus for evaluation is comparison between PBA approaches, for example between a codified approach such as (in the context of youth violence) Communities that Care (Hawkins et al., 2014) and a non-codified approach developed through local collaboration and co-design, with sites randomised to PBA type.

The review of impact evaluations (Baidawi, Valdebenito et al., 2023) found three instances of an experimental design to measure the effectiveness of PBAs in reducing youth violence and 17 using quasi-experimental designs (QEDs). These involve creating a counterfactual involving either comparison areas or comparison between periods of time. The comparability of the control areas or time frames determines the strength of the approach, and these are usually challenging to construct. Identifying control areas involves matching areas on the basis of known and measurable criteria that are hypothesised as relevant to the PBA activity (e.g. poverty, level of crime, degree of social cohesion, readiness for a PBA approach etc.). The approach also depends on being able to 'control' PBA-related activities and conditions which, as with RCTs, is challenging.

A number of different QED methods are available (see Table 3), and there is merit in including more than one method of matching to account for selection biases and differences, and then combining it with a broader approach (e.g. difference in difference analysis with propensity score matching at the individual or geographic level). Designs may also be strengthened by identifying multiple control areas for each PBA area, to avoid a situation where significant and unexpected changes undermine the quality of the initial match. The impact evaluations in Badaawi, Valdebenito et al., 2022, variously involved matched control groups, non-matched

control groups, propensity score matching, one group or interrupted time series, and difference in difference.

Box 4. Study example: QED using matched control groups

Name of intervention	Michigan Youth Violence Prevention Center (MI-YVPC)
Location	Michigan, USA
Intervention	The Michigan Youth Violence Prevention Center (MI-YVPC) is a community-academic partnership aimed at reducing youth violence based on public health principles. It includes six distinct evidence-based programs designed to reach both at-risk and general youth populations across individual, social relationship, and community ecological levels, to reduce violent crime and injury among 10–24 year olds.
Methods	The study identified intervention and matched neighbourhood (with similar crime rates and demographic characteristics). The evaluation measured changes in number of assaults before and after the intervention began, using two independent sets of geo-coded data to conduct the evaluation: (a) crime incidents provided by the local Police Department and (b) youth presenting with an assault injury in the only public Emergency Department and regional trauma center.
Findings	The evaluation found lower number of assaults in the intervention area relative to the comparison area, sustained over time.
Reference and link	Heinze, J., Reischl, T., Bai, M., Roche, J., Morrel-Samuels, S., Cunningham, R. & Zimmerman, M. (2016). A Comprehensive Prevention Approach to Reducing Assault Offenses and Assault Injuries among Youth. <i>Prev Sci</i> , 17(2), 167-76. doi: 10.1007/s11121-015-0616-1. To access the full report, visit: https://pubmed.ncbi.nlm.nih.gov/26572898/

Box 5. Study example: QED using propensity score matching

Name of intervention	Communities That Care (CTC)
Location	Pennsylvania, USA
Intervention	Communities That Care (CTC) (see description in Box 3)
Methods	This study evaluated the effect of widespread diffusion of CTC across Pennsylvania on adolescent substance use, delinquency, and depression by using a repeated cross-sectional design. Propensity score weighting was at the school district level based on variables selected from US Census Bureau data that were theorised to affect the likelihood of CTC being effective. The study collected survey data from 6th, 8th, 10th, and 12th grade students every other year for 10 years. Outcome measures were lifetime and past 30-day alcohol,

	tobacco, marijuana, and other drug use; lifetime and past year participation in delinquency, and current depressive symptoms.
Findings	CTC school districts had significantly lower levels of adolescent substance use, delinquency, and depression.
Reference and link	Chilenski, S., Frank, J., Summers, N. & Lew, D. (2019). Public Health Benefits 16 Years After a Statewide Policy Change: Communities That Care in Pennsylvania. <i>Prev Sci</i> , 6, 947-958. doi: 10.1007/s11121-019-01028-y. To access the full report, visit: https://pubmed.ncbi.nlm.nih.gov/31214854/

Box 6. Study example: QED using interrupted time series

Name of intervention	Phoenix TRUCE
Location	Phoenix, USA
Intervention	The Phoenix TRUCE project is a Chicago CeaseFire replication site. The Ceasefire program is a strategy for reducing the risk that the youth will engage in gun violence and changing attitudes toward violence in the community. It involves implementation of several core strategies to impact the decision-making process of those involved in shootings, including identification and detection, community mobilisation and outreach workers and violence interrupters.
Methods	The study compared local Police Departments crime data (dispatched calls for service, officer-initiated events and callbacks) for 41 months pre-implementation and 19 months post-implementation. A number of time-series models were used to assess the effects of dosage, controlling for the trends in the comparison area.
Findings	A significant decrease in overall levels of violence, assaults and shootings per month was found.
Reference and link	Fox, A., Katz, C., Choate, D. & Hedberg, E. (2015). Evaluation of the Phoenix TRUCE Project: A Replication of Chicago CeaseFire. <i>Justice Quarterly</i> , 32:1, 85-115, DOI: 10.1080/07418825.2014.902092 To access the full report, visit: https://doi.org/10.1080/07418825.2014.902092

There is some disagreement about both how robust QED designs are compared with RCTs, and about their appropriateness in assessing impacts in the complex circumstances of PBAs. However they are a rigorous alternative if randomisation is not feasible, and under favourable conditions, or systematic errors, a well-executed QED study will produce results aligned to those of an RCT. For both approaches, considering at an early stage in evaluation design the sample size required to detect the effect of specific outcomes is necessary to protect studies from either underpowered or wastefully overpowered samples which affect the statistical testing of outcomes as a result of Type I and Type II errors.

Table 4 lists several relevant designs and describes key strengths and weaknesses.

Even if not feasible as a way of measuring the impact of the PBA itself, experimental and quasi-experimental approaches may be feasible *within* a PBA evaluation, for example to test the effectiveness of different implementation strategies or of individual interventions or activities that form part of the PBA, where intervention and control groups are established within the PBA population.

Table 4. Commonly used experimental and quasi-experimental impact evaluation and assessment designs, based on the Magenta Book (HM Treasury, 2020b, p13-24)

Method	Short description	Strengths and weaknesses
Methods involving randomisation		
Cluster-randomised trials	Areas (or groups) are allocated in a random way to intervention or control.	Achieves ‘fair’ comparisons by accounting for known and unknown prognostic factors. Allows for stronger reasoning about causality. Not always feasible for practical reasons.
Stepped-wedge design	If all areas eventually will get the intervention, but not at the same time, e.g. because of resource constraints, it is possible to randomize for a place in the queue.	Strong design when staggering of intervention is inevitable. Contamination and other changes happening as the trial goes on, are two possible problems.
Quasi-experimental methods		
Interrupted time series	Time-series data are utilised to estimate trend and to describe what happens when the trend is “interrupted by” an intervention. Using data from the preintervention period to model a prediction for the postintervention period creates a counterfactual.	Considered a robust design. Often possible to implement and therefore widely used. The time series must be of some length to be able to account for any secular trend and for serial correlation (data collected over time tend to be correlated).
Difference-in-difference	Builds on interrupted time series. By estimating trends in control areas, it is possible to strengthen the inference by comparing differences before and after an intervention period.	Intuitively simple method and results are easy to explain. As above, sufficient data points are needed. Cannot control for confounders.
Regression discontinuity design	Sometimes a cut-off threshold is introduced e.g. to restrict access to a programme offered to people, groups, or areas. Those just above and just under the threshold are probably very similar in all other respects (except being offered the PBA or not). Comparing their results offer an estimate of impact.	Inferences about causal effects of interventions or exposures using real-world data. The critical area is just around the cut point, and only those data points are useful. Those further away may not be as similar. The two groups may also differ in additional “treatment”.
Use of concurrent control areas with pre- and post-measurements	Helps in contrasting findings.	Often possible to find broadly comparable areas. Areas will nevertheless be different in many ways. Thus, pre-, and post-measurements with unmatched controls

(simple pre- and post measurements)		is a very weak design for causal inferences due to biased comparisons.
Propensity score matching	A selection algorithm is used to improve the selection of control areas instead of using a “manual” procedure (exact matching). Data from the intervention sites and comparison sites are combined, the probability of being selected as an intervention site is estimated (called propensity scores) and those scores can be used in matching. PSM can also be used at an individual level. That is, sites can be selected but individuals within sites can be matched with individuals who actually receive an intervention within the PBA. However, this requires individual level data and this may be difficult to acquire.	Propensity score matching at local area level improves comparisons between areas according to many (but not all) researchers. Several caveats may necessitate sensitivity analysis.
Synthetic control methods	A pool of potential comparable observations, using historical data, is used to model how areas would have fared without the intervention. Divergence between the actual observations and the “synthetic” control gives the impact estimate.	May be used with a small number of observations. Can account for changes in confounders over time. Do not rely on parallel pre-implementation trends. Historical data are needed. The synthetic control must be built from a pool of potential controls that are similar to the treated unit.
Identifying and using Instrumental variables	Natural experiments where the selection for an intervention (by an instrumental variable) do not influence the outcome. Used mostly in econometric analysis and epidemiology.	Can be used to estimate causal relationships in observational data between the intervention and the outcome of interest. Natural experiments that meet the assumptions are few and far between and generally require substantial content expertise and familiarity with data sources

Modelling and simulation approaches

Another way of understanding what would have happened if the PBA had not been introduced is *predictive approaches* using informant opinion-based, statistical or simulation modelling. Simulation modelling combines evidence from different sources using conceptual theory of how the system responds to change. They create mathematical representations of multi-level scenarios, based on the hypothesised impact pathway. The model simulates a set of outcomes, based on modelled inputs. However, the inherent difficulty of building models of sufficient complexity, with good enough data, and simulation logic that can be understood and evaluated, limits learning. In practice simulation modelling is not often used for assessing impacts of complex interventions in complex settings, and was not used in any of the studies identified in the wider review.

Theory-based approaches

A third set of approaches for assessing impacts and effects, commonly used in PBAs, are theory-based approaches (Bicket et al., 2020; HMT, 2020a; Skivington et al. 2021). There are contested views about whether these are always the optimal approach or should be used only when experimental or quasi-experimental designs are deemed unrealistic

Theory-based impact evaluations specifies the causal chains that it is hypothesised will bring about change, and then tests whether there is sufficiently strong evidence to support this line of reasoning. Evaluators, usually working with stakeholders, interpret data within the context of the theory of change to establish whether the expected changes are occurring through the expected pathways. A well-developed theory of change can be a valuable aid to evaluation, with expected results and pathways clearly set out. The interpretation also involves considering alternative explanations for the set of relevant outcomes. Multiple evidence sources (as per Table 2) may be used, and need to be specific and reliable enough to test the theories. The use of multiple sources of data and reasoning (triangulation) strengthens inference, as do explicit discussions of alternative causes, through critical reflection and external peer-review.

Theory-based methods do not provide precise estimates of effect sizes but may be able to confirm whether there was an observed effect in the desired direction and whether the PBA could plausibly be said to have contributed in a definitive way to the outcomes. The argument for contribution will be more persuasive if a chain of hypotheses can be supported, alternative explanations have been considered, and multiple sources of data have been used. They can be a way of attempting to address the complexity of, for example, contexts, partnerships, strategies and observed changes. Rather than seeking to 'null out' context, these methods pay explicit attention to it.

Theory-based approaches are often proposed to help explain why, where, for whom and under which circumstances there was an effect as well as judging the transferability of the results. As with the approaches described above, there are different views about the robustness and appropriateness of these approaches and whether they can be viewed as assessing impact as they do not have a counterfactual.

There are a number of different methods and approaches for theory-based evaluation, and Table 6 gives a summary of some theory-based evaluation methods.

Table 5. Commonly used theory-based impact evaluation designs, based on Magenta book (HM Treasury, 2020a p.45, HM Treasury 2020b p4-11)

Method	Short description	Strengths and weaknesses
Realist evaluation	Articulates specific, hypothesised causal mechanisms, in context, and gathers evidence on them, using a Context +	Can inform impact assessment, causal mechanisms, and theory development.

	<p>Mechanism = Outcome framing. Takes 'motivation' and 'intent' into consideration, as outcomes are determined by the programme as an opportunity <i>and</i> people's responses to these resources ('mechanisms').</p> <p>Hypotheses links intervention, context and mechanism with outcome. Data are evaluated to judge whether the hypotheses hold.</p>	<p>Often used for new initiatives, in scaling, and when results have been mixed.</p> <p>Time consuming and resource intensive. Content expertise is needed. Does not provide quantitative estimates of effect.</p>
Qualitative comparative analysis	<p>Systematic post-hoc comparisons within and across cases, based on qualitative analysis of features of intervention and context, to identify characteristics (or combinations) associated with outcomes. Based on binary or more nuanced measurement of attributes. Aim to improve understanding of what leads to desired or undesired outcomes.</p>	<p>Pragmatic method that attempts to explain how a certain outcome is produced when causality is complex. Software tools exist. Makes qualitative reasoning more transparent.</p> <p>Difficult to account for confounding. Needs similar data across (typically 10–50) cases.</p>
Process tracing	<p>Structured method to develop and test theories about how outcomes arise. Examine a single case to develop hypotheses about causal mechanism and outcomes. A wealth of information may be gathered from a systematic, detailed investigation of a single case. Various formal tests are used, looking for evidence that only would be present if the hypothesis was true or false.</p>	<p>May support or overturn alternative explanatory hypothesis and inform on causal direction. Based on 'real world' cases.</p> <p>Data may be missing, comparative evidence may be lacking and causal reasoning is tentative and uncertain</p>
Contribution analysis	<p>Structured way of assessing the likelihood of if, and how much, a programme has contributed to an outcome. Develops a contribution narrative. Is quite similar to process tracing, but without formal testing. Intended to be done in an iterative manner. Steps involve setting out the cause-effect issue, developing the theory of change, assessing the contribution story, gathering evidence, reassessing and challenging the contribution story, and continuing to iterate between evidence and the story.</p>	<p>Helpful in revising a theory of change. May increase confidence in hypotheses about causal chains.</p> <p>Dependent on how 'good' the theory of change is. Susceptible to a large degree of variability (in implementation or outcomes)</p>
Bayesian updating	<p>Supports theory-based methods by utilising Bayes' theorem in improving the level of confidence in contribution claims. The probability of a contribution claim being true is estimated from probability absent the evidence, probability of</p>	<p>Strengthens hypotheses in light of new evidence.</p> <p>Relies on robust probabilities. The calculations can make the arguments less transparent and accessible.</p>

	observing the evidence if the claim is true, and if the claim is false.	
Contribution tracing	Combines process tracing and Bayesian updating in a participatory process with all stakeholders. Takes participants through a contribution 'trial' to establish what would prove or disprove the claim, identification of alternative causes, and application of Bayesian updating.	Will help assess confidence about impact. Must explore other potential causes and pathways.
Most Significant Change	Participatory method. Significant change stories are collected, selected, and prioritised by panels of stakeholders.	Can be used throughout the evaluation. Builds understanding and agreement among participants. Cannot estimate or predict in any detail or with any certainty. Resource intensive, including skilled facilitating.
Outcome harvesting	Participatory method. Works backwards from evidence of change (outcome mapping) to assessing contributions to such change.	Helps foster agreement. Resource demanding, including facilitation.

Box 7. Study example: Qualitative comparative analysis

Name of intervention	Learning Together (LT)
Location	Secondary schools in south-east England
Intervention	A three-year, whole-school intervention that aimed to reduce bullying and improve physical and mental health among mainstream secondary-school students through provision of curricula for social and emotional learning (SEL) and use of restorative practices (RP) for prevention of, and response to, incidents of bullying or conflict.
Methods	The study drew on data collected through surveys for the RCT impact evaluation, completed by staff and Year 7 students (age 11-12) at baseline and again at two- and three-years post-baseline. The surveys included measures relating to the hypothesised mechanisms of change and outcomes. Building on the theory of change and qualitative analysis from the RCT, the authors developed a set of hypotheses about intervention mechanisms (M) and how these might interact with features of context (C) to generate outcomes related to bullying reduction (O). A data table was constructed with individual conditions and outcomes assigned values between 0 and 1 based on either direct evidence or researcher knowledge of the subject area. The analysis then moved from looking at individual cases to understanding the “pathways” (complex combinations of contextual features and mechanisms) that combined to generate outcomes. Pathways were then assessed for consistency (i.e., the proportion of schools in which a given pathway was associated with the observed outcomes, for example, reduction in bullying) and coverage (i.e., how much of the outcome was explained by the pathway). Conditions were removed from a combination if neither its presence nor absence was found to affect the emergence of the outcome.
Findings	Analyses suggested that the intervention worked via three mechanisms: (1) improving student commitment to school; (2) improving student pro-social skills; and (3) de-escalating conflict and bullying. The findings also suggested that there were multiple possible pathways to the same outcome. Contextual features were

	identified, such as a pre-existing ethos of involving students in school decision-making, that may be important for activating change mechanisms.
Reference and link	Warren, E., Melendez-Torres, G. & Bonell, C. (2022). Using fuzzy-set qualitative comparative analysis to explore causal pathways to reduced bullying in a whole-school intervention in a randomized controlled trial. <i>Journal of School Violence</i> , 21:4, 381-396, DOI: 10.1080/15388220.2022.2105856 To access the full report, visit: https://www.tandfonline.com/doi/abs/10.1080/15388220.2022.2105856

Box 8. Study example: Contribution analysis

Name of intervention	Coronavirus Community Support Fund (CCSF)
Location	England
Intervention	A funding programme involving distribution of grants to small- and medium-sized community sector organisations supporting vulnerable people affected by the COVID-19 crisis.
Methods	A set of hypotheses was developed and tested using contribution analysis to assess the extent to which the CCSF contributed to its intended outcomes (increased service delivery, reduced service closure and reduced demand for other services). A theory of change was first developed, and an initial review of background documentation and consultation with relevant stakeholders carried out to identify hypothesised mechanisms of change. Qualitative and quantitative data were collected including via online surveys and interviews with grant holders and service users. Evidence from across the various strands of data collection was then synthesised and mapped to each of the evaluation hypotheses using a thematic analysis approach. An assessment was made of the degree of confidence for each of the hypotheses (using a three-point scale) based on robust direct evidence and conflicting or alternative explanations.
Findings	Grant holders were increased community support to vulnerable people affected by the COVID-19 crisis and there was promising evidence that grant holders had reduced closure of services. Evidence was less clear regarding lower demand for public services, with some cases suggesting that advice and signposting may have increased demand for public services in the short term. The findings supported the hypothesis that CCSF funding contributed towards the positive outcomes and alternative explanations were identified that might have accounted in part or in full.
Reference and link	Ipsos MORI (2021), Impact Evaluation of the Coronavirus Community Support Fund: Final Report, September 2021 To access the full report, visit: https://www.tnlcommunityfund.org.uk/media/insights/documents/CCSF-Impact-Eval-Final-Report.pdf

Evaluation methods for studying cost-effectiveness

In *evaluations of cost and benefits*, positive gains are contrasted with the resources needed to achieve them and the costs incurred, including unintended negative effects. The evaluation forms the basis for judgements about whether there is a reasonable balance between outcomes and costs. Thus, good information on both effects and costs are needed, ideally compared with other ways of achieving the same results

This is an important assessment because a PBA may be effective in achieving its intended goal, but there may be unintended consequences which move costs to somewhere else in the system, or the costs of achieving the effect may have been high. The balance of benefits and costs needs to be described and judged.

Value for money analyses of PBAs are challenging because they need to take into account benefits, and the costs of achieving them, across the system, recognising that a cost occurred in one service area may produce a benefit in another.

The two most used value-for-money evaluation methods are cost-effectiveness and cost-benefit analysis, the latter expressing the balance exclusively in monetary terms. In addition to describing effects in a precise and reliable way, such analyses require costing in a similarly reliable way, across the system.

Table 6. Commonly used evaluation strategies for assessing return-on-investment (HM Treasury 2020b)

Method	Short description	Strengths and weaknesses
Cost-effectiveness analysis	Compares costs and effect sizes (as measured in real terms of social benefit, e.g. reduction in crime rates).	Allows for comparing options in a systematic way. Comparability is hampered by lack of a common metric for benefits. Heavily dependent on trustworthy data.
Cost-benefit analysis	Both effects and costs are monetised. Allows for comparing interventions with different outcome measures.	Allows for comparing options in a systematic way with maximum comparability. Heavily dependent on available data to monetise in a sensible way. Seemingly orderly comparisons might hide underlying difficulties in including social costs.

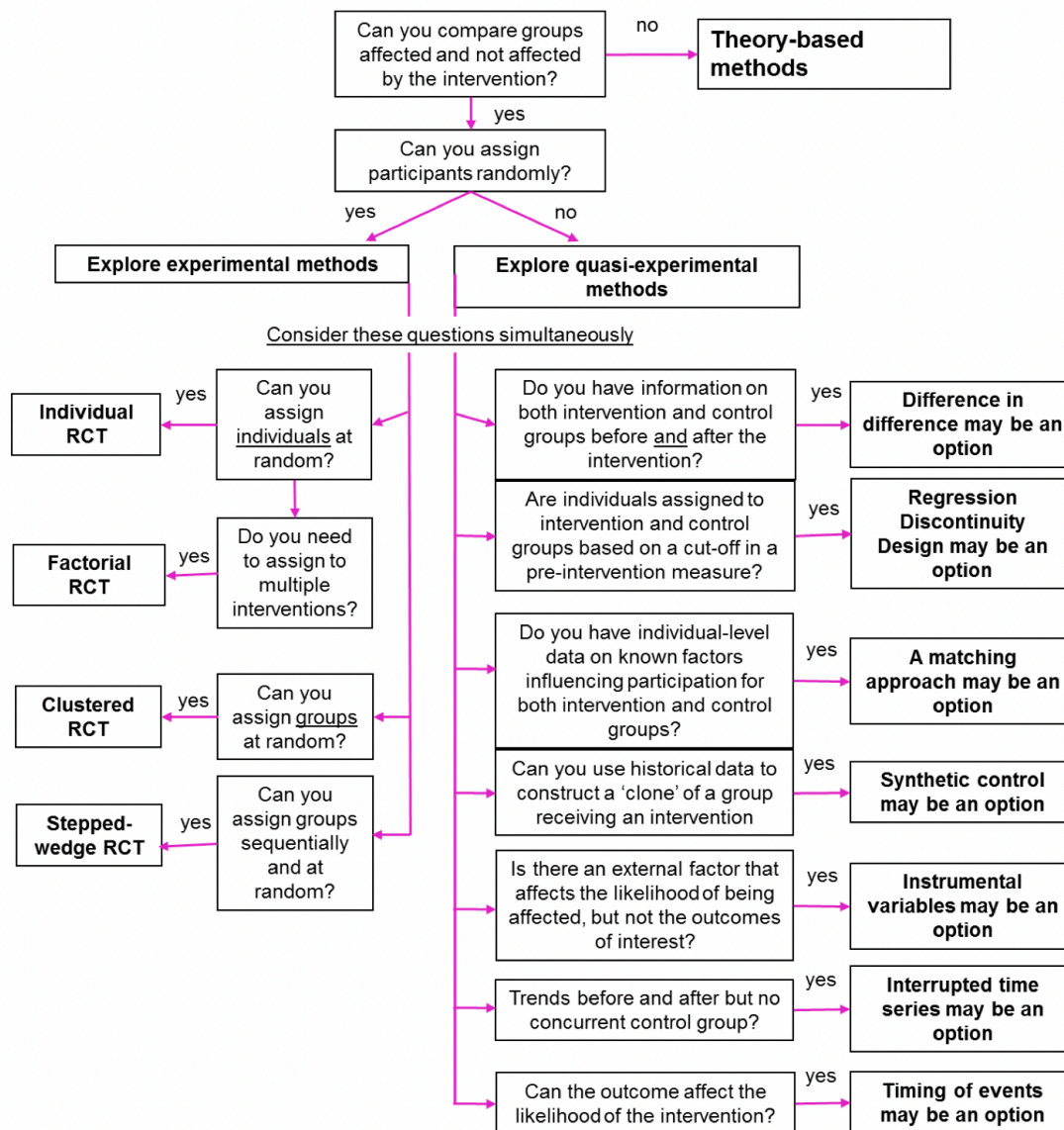
Deciding on design

Because of the complexity of the activity and changes involved, the number of moving parts and measurement challenges, evaluations of PBAs sometimes struggle to demonstrate change in outcomes that matter in a reliable way, and in explaining results. However ultimately it is important to know whether the social problem addressed is alleviated or not, and whether the PBA was instrumental in achieving such change. This requires both process and impact evaluation, with impact evaluation methods that involve quantitative

measurement and analysis of the counterfactual providing more robust assessments of impact. The unpredictable nature of systems change is seen as supporting the adoption of a more interpretive way of working with evaluation, although there is disagreement as to the validity and, hence, usefulness of theory-based attempts to understand whether a PBA contributed to observed effects.

The diagram provides pointers for making choices between different experimental and quasi-experimental methods. It is taken from a text that provides guidance on evaluation generally, without specific reference to PBAs or other complex interventions, and the reference to individual RCTs is unlikely to be relevant to PBAs. The diagram does not discuss theory-based approaches in any detail and proposes that these should be used only if experimental and quasi-experimental designs are not feasible, although we would suggest that evaluation approaches could and should combine quasi/experimental approaches and theory-based approaches.

Figure 1. Example of a flow-chart for choosing among experimental and quasi-experimental designs, from the Magenta Book (HM Treasury (2020a), p.47)



If none of these methods seem appropriate, consider Theory-Based methods

The value of hybrid approaches

Given the breadth of issues in focus for PBA evaluations and the strengths and limitations of different evaluation approaches, the strongest evaluation designs are likely to use hybrid approaches. We use this term to refer to designs that evaluate implementation outcomes and outcomes of the PBA itself (Curran et al., 2022; Landes et al., 2019), and to refer to designs that combine theory-based and quasi-/experimental approaches to assessing change (Bicket et al., 2020). Using a hybrid design can strengthen evaluations as it means that different types of data about change can be triangulated, exploring where they converge and diverge, how they help to explain observed changes and variation, and thus refining and strengthening the conclusions drawn. These designs also enrich considerations of whether evidence from one evaluation might be transferrable to other PBAs contexts.

Box 9. Study example: Hybrid design

Name of intervention	Violence Reduction Units
Location	Police force areas with highest rates of serious violence (SV) across England
Intervention	Violence Reduction Units (VRUs) are multi-agency units designed to support a public health approach to tackling serious violence (SV) and its root causes by providing leadership and facilitating strategic co-ordination of all relevant agencies across the police area. The four core elements of the whole-system approach taken by VRUs are: (1) developing local multi-agency working; (2) supporting data sharing and analysis; (3) community engagement, and (4) commissioning and delivering evidence-based interventions.
Methods	<p>The process evaluation aimed to understand how VRUs were delivering the four elements of the approach. A comprehensive document review was carried out to develop a theory of change (TOC) for each VRU. Qualitative research, including interviews with stakeholders and beneficiaries, was undertaken to explore views and experiences of implementation.</p> <p>The impact evaluation (IE) applied multiple quasi-experimental designs, with hospital admissions for knife assaults, knife-enabled SV non-domestic homicides as primary outcomes. The IE methods adopted included (1) Synthetic control methods (SCMs) and (2) interrupted time series analysis (ITS). SCMs involved a weighted pool of comparator areas, and ITS examined pre- and post-intervention trends.</p>
Findings	The evaluation found strong evidence to indicate that VRUs had made progress towards implementing, and delivering on the aims of, a whole-systems approach to SV reduction, especially evident in multi-agency working and data sharing. The evaluators concluded that in some instances VRUs had made substantive changes to prevention of SV, whilst in others they had accelerated the speed of change.
Reference and link	<p>Violence Reduction Unit (2022). Violence reduction unit year ending March 2021 evaluation report. Home Office, UK Government.</p> <p>To access the full report, visit: https://www.gov.uk/government/publications/violence-reduction-unit-year-ending-march-2021-evaluation-report/violence-reduction-unit-year-ending-march-2021-evaluation-report#references</p>

Other design considerations

Beyond these considerations, decisions about the optimal design will also turn on issues such as:

- The primary purposes of evaluation
- The scale of ambition and the prioritised evaluation questions
- The complexity of the PBA approach and of the system into which it is being introduced
- The capacity of stakeholders, partners and the evaluation team for different evaluation methods

- The availability of data or feasibility of generating different types of data
- The time and resources available
- Other practical hindrances and limitations.

Chapter 5 - 10 key messages for PBA evaluation

- 1. Working closely with a range of stakeholder and partners is essential for the design and conduct of PBA evaluations.** There will be valid competing perspectives on every aspect of the PBA approach, which need to be surfaced, acknowledged and aligned where possible, although this will never be complete. Ultimately trade-offs, compromises and disagreements need to be made. Capacity needs to be developed for evaluation teams and PBA teams to work well together.
- 2. Understanding the theory of change, theory of place and how the PBA came to be may be aspects of evaluation design and conduct.** Systems mapping is also valuable, to provide contextual understanding and a basis for exploring systems-level change.
- 3. The purpose of the evaluation and the scale of ambition are key drivers of design decisions.** The resources available, key audiences and decision-makers and intended uses will all shape what is feasible and optimal.
- 4. Context needs to be richly understood, at multiple levels, and with a recognition of change over time.** PBAs are shaped by context, and operate through catalysing change within and between dimensions of context. Methods to describe and measure dimensions of local contexts and systems are needed, to support systematic analysis of interactions between context and PBA and changes over time.
- 5. The impacts of many PBAs are not robustly evaluated, and the field needs to evolve and to be more ambitious** to generate better understanding of their effectiveness, and to make the case for the significant investment of funds they involve.
- 6. Evaluation of PBAs needs to combine multiple methods, exploring both impacts (at multiple levels) and processes (at multiple levels).** Methods and perspectives need to be integrated in nuanced interpretation.
- 7. PBAs involve multiple simultaneous strategies and levels.** The interconnectedness between implementation factors, and between the PBA and its context, needs to be a central focus of study, informed by hypotheses about change happens which are tested, refined or rejected throughout the course of evaluation.
- 8. Undertaken well, evaluation can be the most critical ingredient of a PBA.** Evaluations will be most impactful if they generate data that can be used for learning and improvement, and support the strengthening of learning and improvement cultures in the local system. This means making data visible and accessible, and building capacity for its use. PBAs are often concerned with power imbalances in local areas, and this has implications for how evaluations are conducted.

- 9. Alignment between the PBA philosophy and the evaluation approach is an important consideration throughout.** Evaluations are not entirely separate from the system being studied: they become, in a sense, part of the system and an active agent in the change process.
- 10. Evaluation requires approaches that combine rigour, depth, flexibility, learning and agility.** This is likely to be best met by hybrid approaches that combine analysis of implementation and effectiveness, and that use a variety of approaches to measure and assess change, triangulating different forms of data to test and thus provide more confidence in findings.

Appendices

Texts drawn in on developing this report

Aarons, G., Hurlburt, M. & Horwitz, S. (2011). Advancing a Conceptual Model of Evidence-Based Practice Implementation in Public Service Systems. *Adm Policy Ment Health*, 38, 4-23. <https://doi.org/10.1007/s10488-010-0327-7>

Baidawi, S., Valdebenito, S. Smith, S. Irving, M., Wills, E., Mitchell, J., Hall, A., Tan, B., Lewis, J. & Shlonsky, A. (forthcoming, 2023). *How to best deliver and evaluate place-based approaches to tackling local youth violence*. YEF Report.

Barbrook-Johnson, P. & Penn, A. (2021). Participatory systems mapping for complex energy policy evaluation. *Evaluation*, 27(1), 57–79. <https://doi.org/10.1177/1356389020976153>

Bellefontaine, T. & Wisener, R. (2011). *The Evaluation of Place-Based Approaches Questions for Further Research*. Policy Horizons Canada, Government of Canada. https://ccednet-rcdec.ca/sites/ccednet-rcdec.ca/wpcontent/uploads/2022/09/the_evaluation_of_place-based_approaches_questions_for_further_research.pdf

Bicket, M., Christie, I., Gilbert, N., Hills, D., Penn, A. & Wilkinson, H. (2020). *Magenta Book Supplementary Guide: Handling Complexity in Policy Evaluation*. HM Treasury. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879437/Magenta_Book_supplementary_guide_Handling_Complexity_in_policy_evaluation.pdf

Collins, A. & Wringley, J. (2014). Can a neighbourhood approach to loneliness contribute to people's well-being? *Joseph Rowntree Foundation*. <https://www.jrf.org.uk/report/can-neighbourhood-approach-loneliness-contribute-peoples-well-being>

Curran, G., Landes, S., McBain, S., Pyne, J., Smith, J., Fernandez, M., Chambers, D. & Mittman, B. (2022). Reflections on 10 years of effectiveness – implementation hybrid studies. *Front. Health Serv*, 2, 1-10. <https://doi.org/10.3389/frhs.2022.1053496>

Damschroder, L., Reardon, C., Widerquist, M. et al. (2022). The updated Consolidated Framework for Implementation Research based on user feedback. *Implementation Sci*, 17, 75. <https://doi.org/10.1186/s13012-022-01245-0>

Dart, J. (2018). *Place-based Evaluation Framework: A guide for evaluation of place-based approaches in Australia*. Department of Social Services, Australian Government. Plus associated toolkit. https://www.dss.gov.au/sites/default/files/documents/06_2019/place-based-evaluation-framework-final-accessible-version.docx

Davies, R. & Dart, J. (2005). *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. <https://europa.eu/capacity4dev/file/28239/download?token=IWZXyl9R>

Davies, S. (2019). *Place-Based Approaches: Characteristics and Learning*. A New Direction. <https://www.anewdirection.org.uk/asset/4117>

Egan, M., McGill, E., Penney, T., et al. (2019a). *Guidance on Systems Approaches to Local Public Health Evaluation. Part 1: Introducing systems thinking*. London: National Institute for

Health Research School for Public Health Research. https://sphr.nihr.ac.uk/wp-content/uploads/2018/08/NIHR-SPHR-SYSTEM-GUIDANCE-PART-1-FINAL_SBnavy.pdf

Egan, M., McGill, E., Penney, T., et al. (2019b). *NIHR SPHR Guidance on Systems Approaches to Local Public Health Evaluation. Part 2: What to consider when planning a systems evaluation*. London: National Institute for Health Research School for Public Health Research. <https://sphr.nihr.ac.uk/wp-content/uploads/2018/08/NIHR-SPHR-SYSTEM-GUIDANCE-PART-2-v2-FINALSBnavy.pdf>

Fujiwara, D. (n.d.) *A short guide to social impact measurement*. Simetrica. https://uploads-ssl.webflow.com/6274e0c5fb041327b2d5e532/6274e0c5fb04130eed5e6c7_Short-Guide-to-Socail-Impact-Measurement.pdf

Harrison-Evans, P., Kazimirki, A., & McLeod R. (2016). *Balancing Act: A Guide to Proportionate Evaluation*. NPC. <https://www.thinknpc.org/resource-hub/balancing-act-a-guide-to-proportionate-evaluation/>

Hawe, P., Shiell, A. & Riley, T. (2009). Theorising interventions as events in systems. *Am J Community Psychol*, 43 (3-4), 267-76. <https://doi.org/10.1007/s10464-009-9229-9>

Hawkins, J, Catalano, R. & Kuklinski, M. (2014). Communities That Care. In: Bruinsma, G., Weisburd, D. (eds) *Encyclopedia of Criminology and Criminal Justice*. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-5690-2_581

HM Treasury (2020a). *Magenta Book: Central Government guidance on evaluation*. HM Treasury. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/HMT_Magenta_Book.pdf

HM Treasury. (2020b). *Magenta Book Annex A: Analytical methods for use within an evaluation*. HM Treasury. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879418/Magenta_Book_Annex_A_Analytical_methods_for_use_within_an_evaluation.pdf

Holtrop, J., Estabrooks, P., Gaglio, B. et al. (2021). Understanding and applying the RE-AIM framework: Clarifications and resources. *Journal of Clinical and Translational Science*, 5(1), E126. <https://doi.org/10.1017%2Fcts.2021.789>

Kelly, T. (2010). Five Simple Rules for Evaluating Complex Community Initiatives. *Community Investments*, 22(1), 19-22. <https://core.ac.uk/download/pdf/6231642.pdf>

Landes, S., McBain, S. & Curran, G. (2019). An introduction to effectiveness-implementation hybrid designs. *Psychiatry Res*, 280, 112513. <https://doi.org/10.1016/j.psychres.2019.112513>

McEwan, C. (2018). *A guide to evaluating interventions related to violence against women*. ANROWS. https://anrowsdev.wpenginepowered.com/wp-content/uploads/2019/01/ANROWS_IN.18.03_McEwen_Guide-to-Evaluating-Interventions-related-to-VAW.2.pdf

Moullin, J., Dickson, K., Stadnick, N. et al. (2019). Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. *Implementation Sci*, 14, 1. <https://doi.org/10.1186/s13012-018-0842-6>

Nilsen, P. & Bernhardsson, S. (2019). Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research*, 19, 189. <https://doi.org/10.1186/s12913-019-4015-3>

Parolini, A., Tan, W. & Shlonsky, A. (2019). Decision-based models of the implementation of interventions in systems of healthcare: Implementation outcomes and intervention effectiveness in complex service environments. *PLoS ONE*, 14, e0223129. <https://doi.org/10.1371/journal.pone.0223129>

Public Health England (n.d.) *Place-based approaches to reducing health inequalities – Evaluation Report*. Public Health England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1008132/PBA_final_evaluation_report.pdf

Rittel, H. & Webber, M. (1973). "Dilemmas in a General Theory of Planning." *Policy Sciences*, 4(2), 155-169. https://www.sympoetic.net/Managing_Complexity/complexity_files/1973%20Rittel%20and%20Webber%20Wicked%20Problems.pdf

Rogers, L., DeBrun, A. & McAuliffe, E. (2020). Defining and assessing context in healthcare implementation studies: a systematic review. *BMC Health Services Research*, 20, 591. <https://doi.org/10.1186/s12913-020-05212-7>

Skivington, K., Matthews, L., Simpson, S. et al. (2021). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*, 374. <http://dx.doi.org/10.1136/bmj.n2061>

Smith, R. (2011). *How to Evaluate Choice and Promise Neighbourhoods*. The Urban Institute. <https://www.urban.org/sites/default/files/publication/32781/412317-how-to-evaluate-choice-and-promise-neighborhoods.pdf>

Sridharan, S. (2011). *The Top 10 Questions: A Guide to Evaluating Place-Based Initiatives*. Policy Horizons Canada, Government of Canada. https://publications.gc.ca/collections/collection_2012/hpc-phc/PH4-101-2011-eng.pdf

Taylor, M. & Buckley, E. (2017). *Historical review of place based approaches*. Lankelly Chase. <https://lankellychase.org.uk/wp-content/uploads/2017/10/Historical-review-of-place-based-approaches.pdf>

Treasury Board of Canada Secretariat. (n.d.). *Theory-Based Approaches to Evaluation: Concepts and Practices*. Government of Canada. <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/theory-based-approaches-evaluation-concepts-practices.html>

Experts consulted

Dr Bianca Albers, Post-doctoral Researcher, Institute for Implementation Science in Health Care, University of Zurich, Switzerland

Martha Bicket, Senior Research Fellow, CECAN, University of Surrey, UK

Jo Blundell, Co-Lead, Place Matters, UK

Dr Brian Bumbarger, Founding Partner/CEO, Science, Systems & Communities Consulting, LLC; Adjunct Research Associate at the Prevention Research Centers at Colorado State University and Penn State University, US; Research Fellow at the Criminology Institute at Griffith University, Australia.

Professor Jeremy Grimshaw, Senior Scientist, Clinical Epidemiology Program, Ottawa Hospital Research Institute and Professor in the Department of Medicine, University of Ottawa, Canada

Dr Tim Hobbs, CEO, Dartington Service Lab Design, UK

Anne Kazimirski, Evidence and Impact Director, Urban Health, UK

Lily O'Flynn, Principal Consultant for Place-based Evaluation & Learning, Renaisi, UK

Dr Kathryn Skivington, Research Fellow, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, UK



You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at www.youthendowmentfund.org.uk

Registered Charity Number: 1185413