Evaluation Task Force

Geological Disposal Facility Evaluation Plan Technical Appendices

Evaluation Task Force

August 2025

verian

Nuclear Waste Services

Evaluation Task Force

Contents

Appendix A: Theory of Change

Within each phase of the GDF programme Theory of Change, a series of colour-coded boxes represented in the ‘key’ describes inputs, activities, outcomes and impacts.

Inputs are in white boxes and represent the resources dedicated to the programme, including the GDF funding, policy, and initial discussions regarding the initiation of the GDF programme.

Activities are in two shades of dark blue boxes and represent the actions that are being taken to deliver the GDF programme within each phase: for example, the formation of a working group in a community, information gathering and engagement with local authorities by community members, and so forth.

Outcomes are in lighter blue boxes, and cover what is expected to result from the conduct of activities, which includes that the potential host communities have received accurate information about the GDF programme, which they engage with and trust; that sentiment towards the GDF programme is positive; that community projects and the development of the GDF result in jobs creation and skills investment; and that a Test of Public Support is achieved. Some outcomes outlined in the Theory of Change are expected to occur in the longer term: for example, a constructed GDF is resilient to external shocks, and disposal of Most Hazardous Radioactive Waste (MHRW) occurs.

Impacts are in the lightest blue and green boxes and reflect the overall goals and aims of the programme that are expected to result from the achieved outcomes. For example: improved socio-economic wellbeing of the host community/potential host community; that the GDF removes the need for ongoing costs associated with the storage of waste overground; that carbon intensity associated with nuclear power in the UK is reduced; and that the programme overall provides VfM and supports the economic growth of the UK.

Activities, outcomes or impacts relevant to the potential host community or host community only, are distinguished from NWS activities and national-level outputs of impacts. On a few occasions some outlined activities or outcomes are relevant at both the local and national levels, and these are represented by the brown boxes.

Each of the boxes are connected by arrows to describe the theorised ‘next step’ leading to a result. For example, within Phase 2, the activity of providing information about the GDF to the potential host community is expected to result in the outcome of potential host community members receiving accurate information that they need in order to understand the GDF. In some cases arrows are included between different activities - this is where one activity is supposed to, or is only able to, occur after another.

The Theory of Change is visualised in Figures A.1, A.2, A.3 and A.4, each presenting a sub-Theory of Change for each phase of the GDF programme.

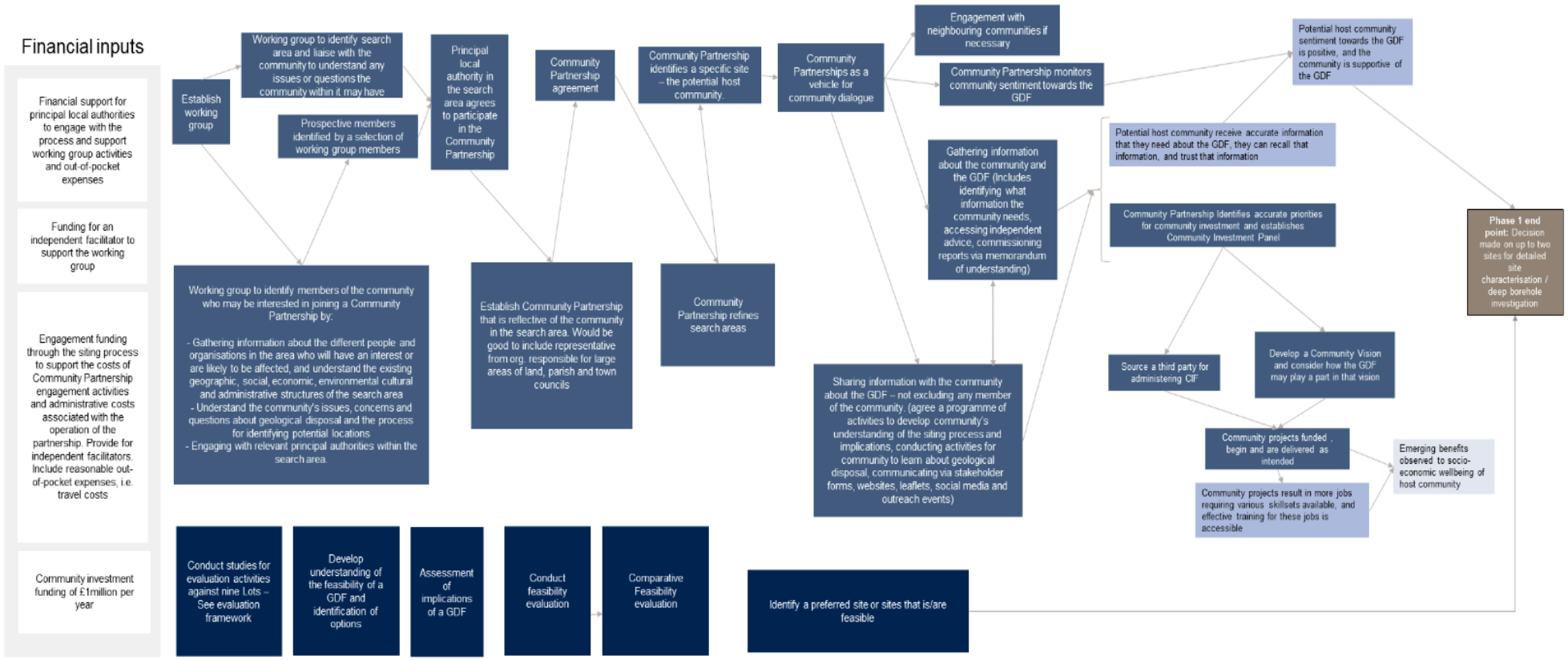
Several assumptions have been made in the development of the Theory of Change, including that:

- necessary delivery and supply chains will have appropriate capabilities to deliver
- safe methods of encapsulation and transportation are approved
- only one GDF will be built
- the GDF will be a final disposal solution
- at least one community is suitable, both geologically and infrastructurally
- construction and waste emplacement operations will run concurrently
- detailed characterisation activities will take place in two communities

A number of risks to the GDF programme have also been identified that could have implications for the theory behind the programme and the level of benefits realised. These include the following:

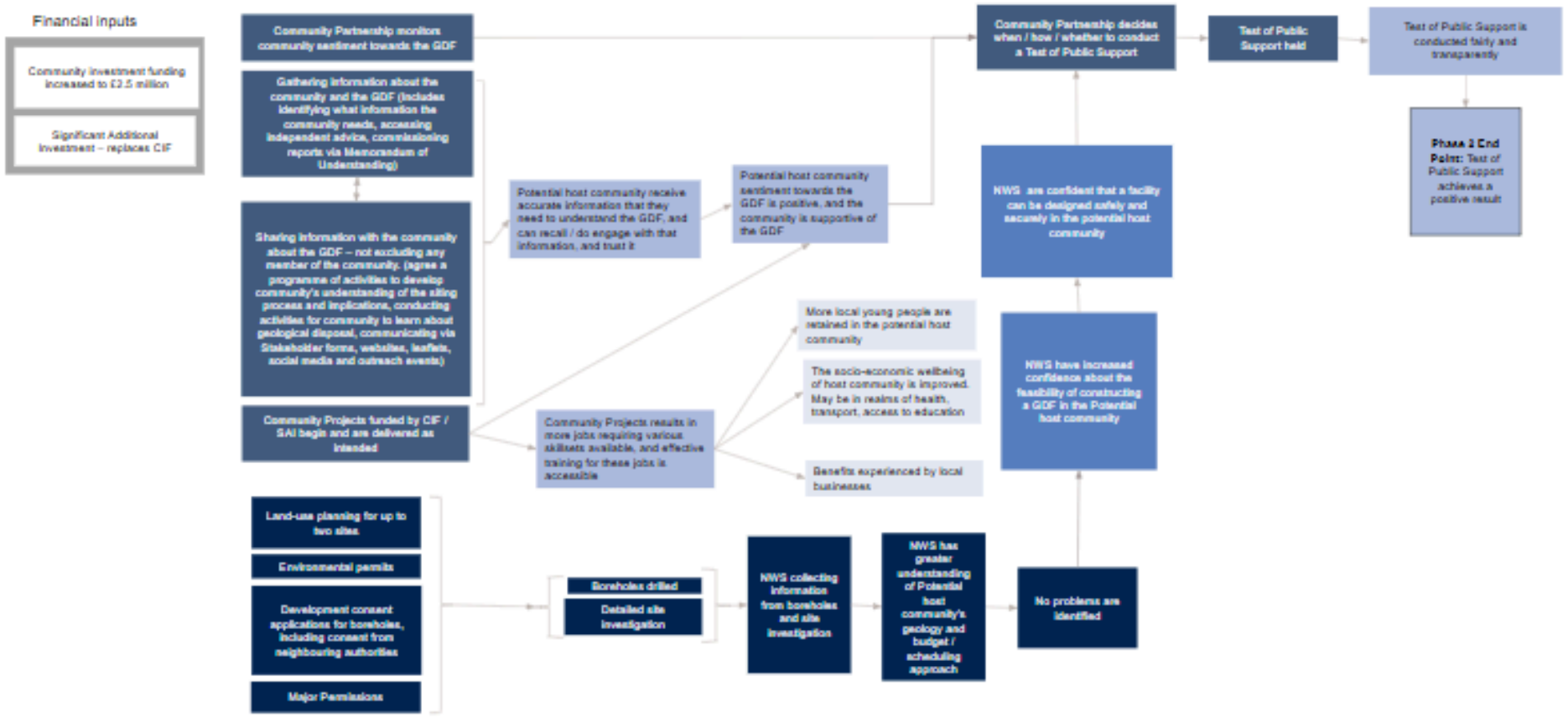
- unexpected loss of one or more communities
- insufficient capability to procure multiple major contracts
- the GDF becomes less of a priority and hence funding for the GDF is reduced
- national and local government political uncertainty
- areas are unable to achieve a positive Test of Public Support
- skills, delivery and supply chain capability shortages

Figure A.1: Phase 1 Community engagement and site evaluation Theory of Change



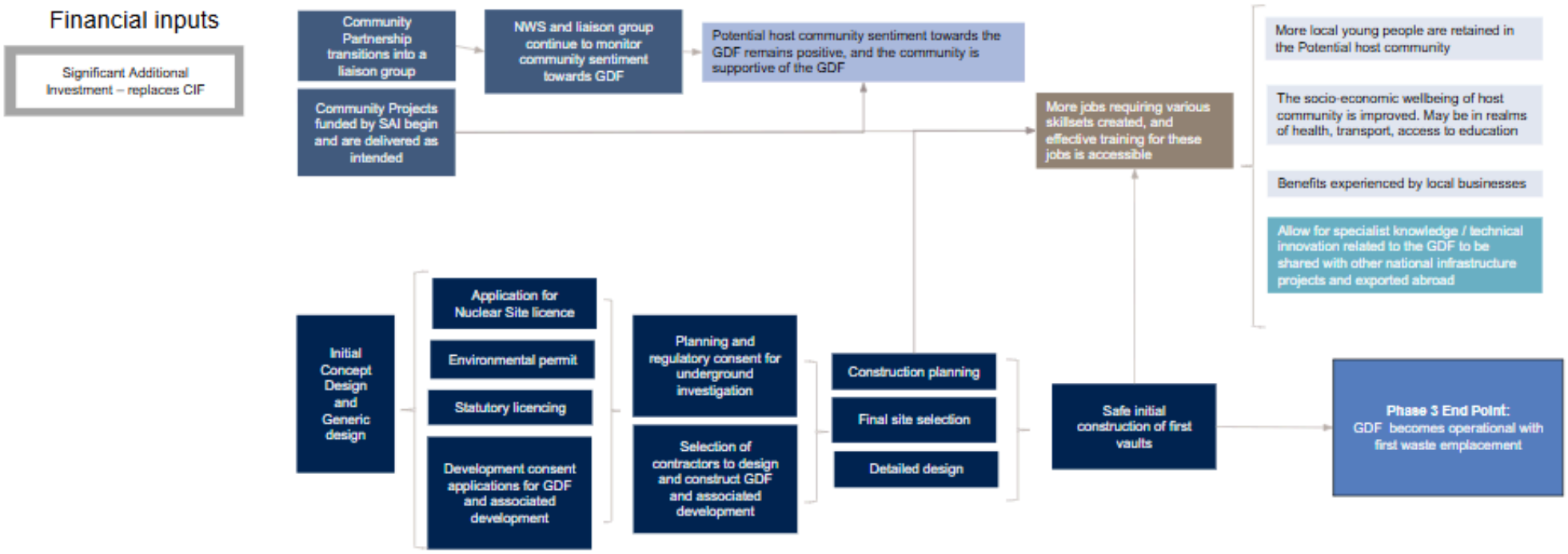
Key	
X	Activities related to community identification and engagement
X	Outcomes related to community identification and engagement
X	Impacts related to community identification and engagement
X	Activities related to Site Evaluation
X	Activities related to both community identification and site evaluation

Figure A.2: Phase 2 site characterisation Theory of Change



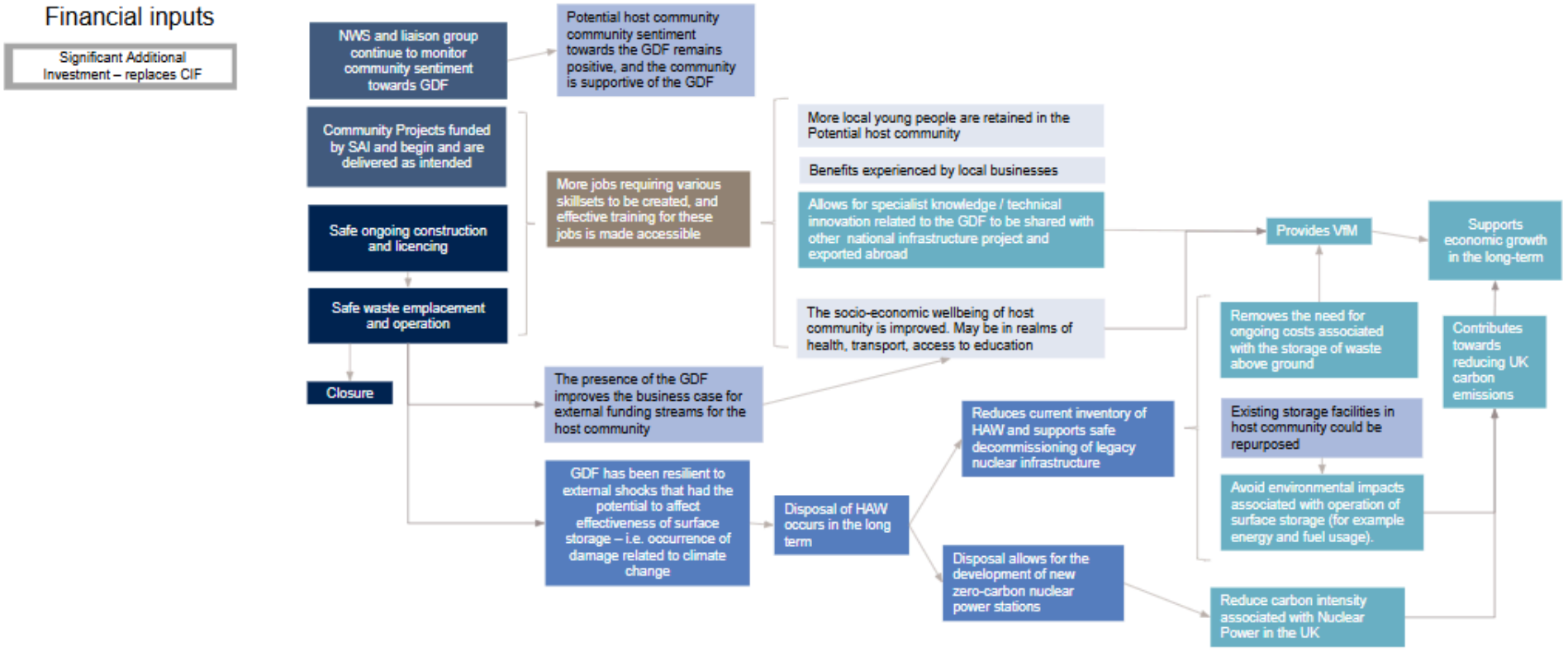
Key	
X	Activities related to community sentiment and Community Projects
X	Outcomes related to community sentiment and Community Projects
X	Impacts related to community sentiment and Community Projects
X	Activities related to Site Characterisation
X	Outcomes related to Site Characterisation

Figure A.3: Phase 3 GDF technical development, design and construction ToC



Key	
X	Activities related to liaison group and Community Projects
X	Outcomes related to liaison group and Community Projects
X	Impacts related to liaison group and Community Projects
X	Activities related to technical development, design or construction
X	Outcomes related to technical development, design or construction
X	Impacts related to technical development, design or construction
X	Outcomes related to both the host community / liaison group and technical development, design and construction

Figure A.4: Phase 4+ Operation ToC



Key	
X	Activities related to the Host Community sentiment and Community Projects
X	Outcomes related to Host Community sentiment and Community Projects
X	Impacts related to Host Community sentiment and Community Projects
X	Activities related to operation of the GDF
X	Outcomes related to operation of the GDF
X	Impacts related to operation of the GDF
X	Outcomes related to both the host community / liaison group and technical development, design and construction

Appendix B: Plan for evaluation periodicity

The green boxes represent the time period during which phase-specific implementation activities will occur. The orange boxes represent the points at which it would be valuable to conduct an IPE, differentiating between scoping, data collection, and results. This would happen within each phase of implementation, until emerging findings and key findings of each phase have been collected and the full findings be presented in the IPE results. The grey box represents the time period for the impact evaluation, which would be continuous across the lifetime of the programme, with re-scoping mapped against the implementation activities and key deliverables.

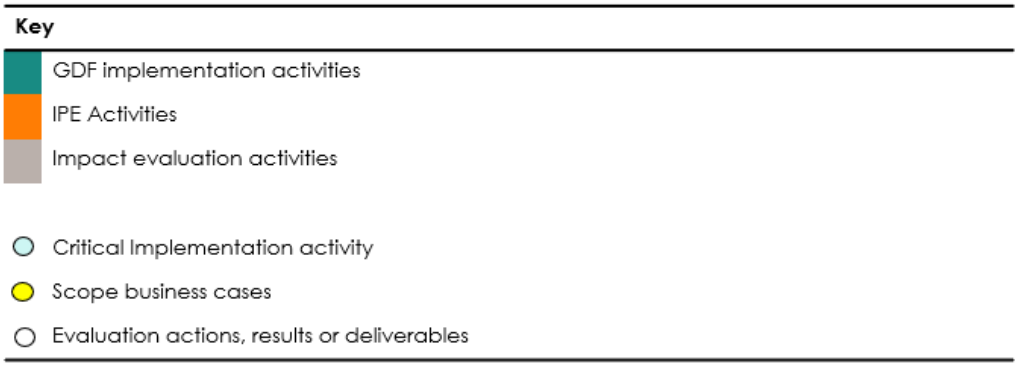
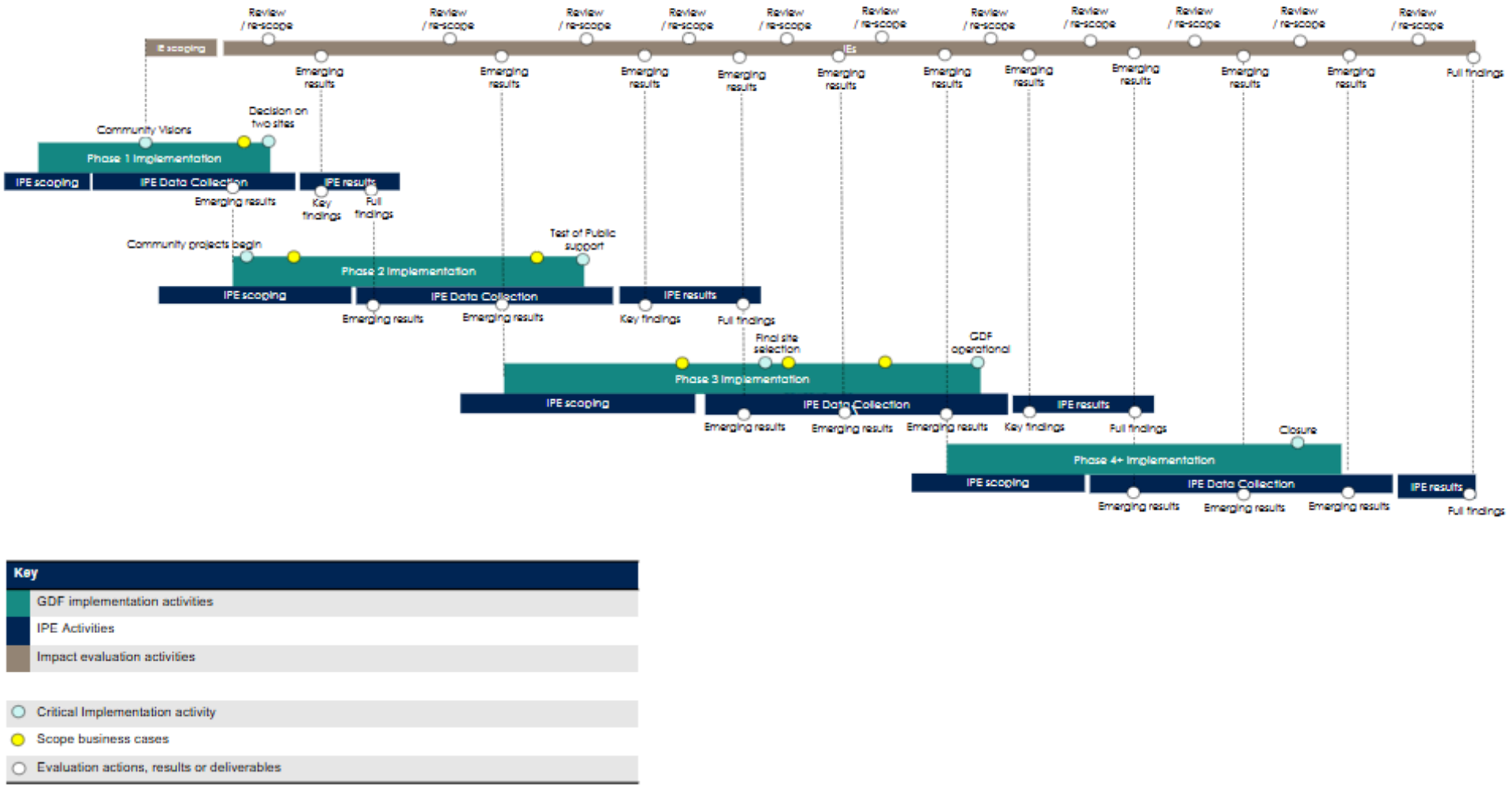


Figure B.1: Mapping implementation against evaluation



Appendix C: Technical report

Spotlight on impact evaluation techniques

Experimental and quasi-experimental approaches to impact evaluation are designed to achieve a robust estimate of the average impact of a policy or intervention. Experimental and quasi-experimental approaches infer the impact of an intervention through statistical comparison with a group or time period that is unaffected by the intervention. This unaffected group acts as a proxy for what would have happened to the affected group in the absence of the policy and is commonly called the counterfactual.

Experimental design

Experimental designs involve providing a robust comparison between one or more groups receiving an intervention (treatment group) and a group that does not receive the same intervention (control group) by randomly assigning participants to each group. This ensures there are no observable or unobservable differences (or bias) between the treatment and control, meaning that any differences in measured outcomes between the two groups can be reliably attributed to the intervention, not an unrelated factor. When implemented correctly, experimental designs are considered to produce robust estimates of impact.

There are some practical limitations of experimental designs but the main issue for this evaluation is that delivery of the GDF programme itself does not lend itself to a scenario where the treatment is ‘randomly’ allocated. First, a GDF cannot be randomly placed within the UK, as there are strict technical, geological and security requirements. However, even among the locations that meet the requirements for a GDF, sites’ local communities have to support the idea of a GDF facility being constructed nearby. Thus, the placement of GDFs is not random as communities will ‘self-select’ into the GDF programme.

Another randomisation option is to identify a list of sites where there is sufficient local support for a GDF, and where the local area meets the technical (e.g. geological) requirements etc. Within this eligible list, sites could then be randomly selected to implement the GDF programme or the timing that they begin constructing a GDF. However, again Verian views this option as impractical in this instance, as this may not result in the most suitable or preferable site being selected for this large investment of public funding. Furthermore, randomising between sites is unlikely to achieve a suitable sample size for performing a robust statistical analysis of outcomes according to effect sizes we can meaningfully detect. There are too few GDFs currently planned to meaningfully implement this evaluation approach. In practice, there would also be a need to start the process for testing local support of the construction of a facility within a much larger sample of areas than the GDF programme is currently doing.

Quasi-experimental designs

Quasi-experimental designs (QED) are more appropriate for this evaluation; they use a counterfactual, but not one achieved through randomisation (hence being quasi-experimental).

Quasi-experimental designs adopt statistical techniques to try to account for known differences between the intervention and control groups (or areas in this case) in the analysis, to increase the likelihood that any differences (impacts) occurring between the treatment and comparison results from the policy/programme of interest. The statistical methods chosen will depend on factors such as the nature of the comparison group (e.g. concurrent or historic); the format of the available data; and the volume of the available data. Relevant quasi-experimental design methods in this case include regression discontinuity design and difference-in-differences design.

Theory-based design

A theory-based approach to impact evaluation is any evaluation that seeks to understand the theory behind why a particular programme is expected to lead to particular changes or benefits, and which then seeks to test that theory and explore why change has or has not occurred. A Theory of Change is a tool used for theory-based evaluations and is a visual representation of the theory behind a programme. A Theory of Change is intended to make explicit the inputs, activities, outputs and outcomes that need to be achieved for the desired impacts and vision to be realised. This is useful from a programming perspective because it helps ensure the programme is well designed and can lead to the intended results. For this reason, assumptions and risks to results being achieved are also included in the Theory of Change so that they can be managed and mitigated as much as possible.

One particular theory-based impact evaluation technique proposed for this evaluation is contribution analysis, which is concerned particularly with investigating the extent to which a particular programme has ‘contributed’ towards its impacts. It is usually used in complex environments: for example where co-current programmes are running. It typically involves collecting multiple different forms of evidence, an explicit exploration of ‘what else’ might have led to change, and validation of results. Overall, contribution analysis seeks to make a conclusion about ‘how confident’ one can be that a programme has led to the changes it expected to cause. Contribution analysis would therefore be a useful approach for an evaluation of the GDF programme, given the length of programme timeframes and various forms of evidence that would be needed to draw conclusions about impact.

Qualitative Impact Protocol studies

The Qualitative Impact Protocol (QuIP) is a qualitative approach to impact evaluation. QuIP studies are designed to collect credible evidence on the impacts of a given intervention. Information is gathered directly from intended beneficiaries across a series of domains related to the project’s Theory of Change and based on their perceptions of what has changed in their lives over a set period. The key principle of QuIP is to try separate the explanations for observed changes in key outcomes from the specific intervention being evaluated – i.e. beneficiaries are asked to talk about the change (i.e. in their wellbeing etc.) and are asked to share the factors which they perceive have caused the change, using open-ended questions, rather than being asked directly about the project. If a range of respondents report (unprompted) that the intervention was a significant factor driving various outcome trends then this is taken to be strong evidence that the intervention did indeed make a contribution to the observed impacts. This is likely to be a useful method to employ within the GDF programme evaluation in order to understand the extent to which the GDF programme has contributed towards local-level benefits.

More information about different evaluation approaches can be found in the HM Treasury guide to evaluation, the Magenta Book.[1]

Impact evaluation approach according to different benefits/phases

Here we provide further descriptions of examples of the evaluation questions and sources of data that could be used to inform the evaluation of the GDF programme according to the different aspects of the programme and the types of impacts it is expected will materialise in the local community.

Evaluation of the impact of GDF CIF on local development (Phases 1 and 2)

There is currently a high-level strategy to use the CIF to target projects which contribute to outcomes within participating communities, such as local economic development opportunities, enhancing the natural or built environment, and enhancing community wellbeing. Which of these issues is prioritised, the specific projects and the way they intend to achieve an impact will directly affect the outcomes of interest and their measurement. The impact evaluation design will therefore need to be iteratively developed and re-scoped as the CIF projects are finalised. However, we envisage that a theory-based evaluation using a contribution analysis approach could investigate questions such as the following:

- Which outcomes or changes did the CIF project contribute to and how?
- What are the perceptions of local beneficiaries and service providers on the impact of the CIF and its value in supporting local services?
- Which other factors also contributed to local development outcomes?
- Have there been any unintended or unexpected contributions of the CIF project? How has the CIF project interacted with other local initiatives?
- To what extent do the CIF projects align with the Community Vision?

The types of data that could be collected to inform a contribution analysis of the CIF could include both qualitative and quantitative methods: for example, a survey of key local stakeholders, depending on the size of the impacted group, to understand how widely shared certain views are, or how common certain experiences are. Stakeholders can also be engaged in in-depth interviews to gather evidence on the extent to which they agree with, or would challenge, the Theory of Change's contribution story. Relevant stakeholders and local experts can be identified widely: for example, from frontline health professionals, service users and their family and friends, the wider community, businesses, organisations, councillors and government, and the CIF managers and wider stakeholders.

[1] HMT (2020). The Magenta Book (PDF, 2.4MB)

Quasi-experimental impact evaluation of impact of CIF on local development outcomes[2]

Some example evaluation questions for a quasi-experimental evaluation of CIF on local development outcomes include:

- What effect has the CIF project had on local development outcomes (e.g. residents' wellbeing), compared to non-beneficiaries within the same community?
- What effects has the CIF project had on local development outcomes (e.g. enhancing the natural or built environment), in comparison to comparable residents in a matched community?

Design

There are two types of quasi-experimental evaluation that would be relevant, depending on the implementation of the CIF project activities:

- difference-in-differences
- regression discontinuity design (RDD)

The first method, difference-in-differences, would estimate the impact of the project by measuring outcomes for two groups of people: a group which was intended to benefit from the project and a comparison group that was not intended to benefit. A difference-in-differences analysis compares the change in outcomes for these two groups. A difference in these trends is attributed to the effects of the project. For example, if a measure of wellbeing increased by 1 point on average in the beneficiary group and by 0.4 points in the comparison group, we would estimate the impact attributable to the project to be 1 – 0.4 = 0.6 points. In other words, this analysis recognises that some of the increase observed for beneficiaries would still have happened without the project.

The nature of the comparison group would be decided based on how the project works. If the project only affects some people in the community, such as access to a children's sports club, and if the likelihood of spillover effects to the rest of the community is low, then the outcomes of project beneficiaries can be compared with non-beneficiaries from the same area. Alternatively, if all people in the community could be affected by the project, such as an improvement in the visual appeal of the neighbourhood, then a different community which is similar could be used as the comparison group.

The second quasi-experimental evaluation method is regression discontinuity design (RDD). RDD assumes that those just below and just above the cut-off for eligibility for an intervention are likely to be otherwise very similar. When comparing the groups who did and did not receive the intervention, any differences in outcomes can therefore be attributed to the intervention. This requires an eligibility criterion with a reasonably fixed cut-off point: for example, being eligible for a skills programme if you have under two years of work experience, but not being eligible if you have more. The suitability of RDD therefore depends on how CIF projects are defined and implemented.

[2] The precise development outcomes and indicators would need to be selected based on the agreed use of the CIF.

Data

Capturing the data required for these evaluation methods would involve the repeated surveying of beneficiaries and non-beneficiaries (in groups determined by the project type). We recommend that 'baseline' data be collected from before the start of a project: that is, the survey would need to be conducted both before and after the implementation of the CIF projects. This is essential for a difference-in-differences analysis, and desirable for an RDD. If outcome data are only available from after the community projects start, there are other quasi-experimental methods (such as matching methods) we could consider.

Some related outcomes could be inferred from secondary data sources, such as NHS data regarding use of mental health services or wellbeing questions recorded in the Annual Population Survey. Analysis of these datasets could supplement analysis of survey data, likely using a similar difference-in-differences approach.

The time period over which the outcomes are measured will depend on the scope of the project and the duration of expected benefits, and their relative size in comparison to other local initiatives. The expected timeframes should be defined once the details of CIF projects are known.

Evaluation of the impact of the GDF programme on local economic outcomes and wellbeing (Phases 1–4)

Economic outcomes from the GDF programme are expected to accrue locally, particularly in the beginning stages of the programme.

Example evaluation questions:

- Has the GDF programme contributed to any unexpected changes or outcomes?
- To what extent has the GDF programme contributed towards achieving the outcomes and changes that it set out to achieve, and how likely is it that it will do so going forwards?
- What are the outcomes that the GDF programme intends to achieve, and how is it intended to achieve these?
- To what extent have the outcomes and changes that the GDF programme set out to achieve occurred, and to what was the GDF programme a contributing factor?
- To what extent does the evidence suggest that the GDF programme will go on to contribute towards the outcomes and changes it intended to achieve?

Types of data collection that would be relevant for this aspect of the programme could include quantitative measures: for example, trends in online jobs postings and other relevant data sources that are dependent on the intended outcomes. Additionally, qualitative focus groups and interviews with residents and local organisations would aid in understanding any changes in economic demand

seen locally, and how this may have translated into outcomes such as hiring decisions. Interviews could be conducted with a wide variety of stakeholders, such as:

- residents and young people from the local area
- local businesses leaders
- charities that support the unemployed and those on low incomes
- representatives from community groups and minority groups
- local Jobcentre staff and users

Quasi-experimental impact evaluation of the impact of the GDF programme on local economic outcomes and wellbeing

One example evaluation question for the quasi-experimental evaluation of the impact of the GDF programme on local economic outcomes and wellbeing could be: what effect has the GDF programme had on local economic outcomes, such as job creation, net migration or demographic change, compared to other comparable communities?

Design

Using a difference-in-differences approach, residents from the GDF host community can be compared to those from another community which are comparable on key features, including demographic and economic variables. It would be necessary to compare the GDF community to an external community as the GDF is expected to accrue effects across the community, meaning there would not be a comparable group from within the community which had not been affected. One limitation of this approach is that it would not be able to separate out the effects of the CIF from the effects of construction and operation of the GDF once these aspects of the programme start. This is because the activities from the construction and operation phase are also expected to affect local economic outcomes.

Data

Existing datasets should be considered for these outcomes. The Census collects almost comprehensive information about various local labour market factors (such as employment), local migration and demographic change. As the Census has national coverage, it should be feasible to compare the change in outcomes in the GDF host community against the change in other comparable areas. The key limitation with the Census is its infrequency, being conducted once every 10 years. However, given the timescales of the evaluation, it can still provide a useful source of evidence when it does become available.

Other existing datasets include administrative data recorded by DWP and HMRC about employment, tax and benefits. With the permission of DWP and HMRC, this data could be used to monitor changes in employment, benefit claims and earnings (for those completing self-assessments or paying tax through PAYE). Again, this data could be compared to a comparison group made up of similar individuals in other areas outside of the GDF host community. Additionally, the Annual Population Survey could provide some evidence about changes in the characteristics of local communities. Similarly, existing secondary data could be captured on outcomes related to housing development and land value or property prices.

For outcomes where widespread secondary data is available at a granular level (e.g. individual, household, street level), the analysis could involve a variation of the difference-in-difference approach described above. This would examine changes in outcomes before and after the programme based on units' proximity to the GDF host community.

In addition to existing data sources, exploration of new data sources should be considered, as relevant to each outcome. Given the significant timeframe of the programme, it is important to

maintain adaptability to new contexts and circumstances around data collection, including the increasing scope for automated approaches.

Evaluation of the impacts of the GDF programme on local levels of trust and support for a GDF (Phases 1 – 4)

The GDF can only go ahead if there is a successful Test of Public Support. How successful different strategies are in generating public support is likely to be of great interest in future public works.

Throughout the siting process, engagement with potential communities has been and will be ongoing. Moving a community from an interested party through to the formation of working groups and subsequent Community Partnerships all rely on building trust and support through various engagement methods, including the CIF. Trust and support for the GDF will continue to be important throughout the phases of the GDF programme, and they will be impacted by the process of delivering the programme and can themselves greatly impact local economic outcomes.

Example valuation questions:

- Is it reasonable to assume that the CIF contributed to generating public support?
- How confident are stakeholders that community projects contribute towards job creation and socio-economic benefits for the host community? What socio-economic benefits do they expect to occur?
- How successful has community engagement been in improving or maintaining awareness / trust / engagement / sentiment of the community towards the GDF?
- How has the delivery of the GDF programme impacted local support?
- What other factors or events affected the level of support?

A Theory of Change can be developed setting out the outcomes, risks and assumptions in regard to understanding public support levels. Multiple forms and sources of data could contribute evidence for evaluating the strength of the Theory of Change.:

- Ethnographic observations in key stakeholder meetings – what key concerns are raised and how are these dealt with?
- Interviews with ‘unsuccessful communities’ – what happens after a town is not selected or drops out? This could offer insights into understanding what ‘went wrong’, and could present a key learning opportunity.
- Sentiment and thematic analysis of media and social media or forums in which the GDF programme is discussed – which topics come up concurrently, alignment with geographies, tracking of sentiment over time?
- Key stakeholders for this set of outcomes could include community organisations, environment organisations, local political groups and local councillors, community figures and those involved in the GDF Community Partnership.

Quasi-experimental impact evaluation of the impacts of the GDF programme on local levels of trust and support for a GDF programme

Example research questions may include:

- Did trust in the GDF programme change in potential host communities relative to other comparable communities that did not receive CIF investment?
- What are the effects of the CIF activities on beneficiaries' trust in the GDF programme relative to non-beneficiaries from the same area? (If comparison is possible.)
- Did trust in the GDF programme change in the host community relative to other comparable communities?

Design

Difference-in-differences. During Phase 1, multiple (currently four) potential host communities will receive the CIF. Potential host communities initially enter the working group phase, before forming a Community Partnership, which entitles them to the CIF funding. The projects undertaken by the communities with the funds can vary greatly, and therefore might be expected to have different effects on public support. Therefore, each project's impact should be measured separately. Outcomes can be compared against a set of comparison locations using a difference-in-differences approach.

Data

A local community survey can be used to measure outcomes such as trust, support, and perceptions of wider support in the area. This can be part of the same survey we have recommended for addressing local development outcomes (such as personal wellbeing). The sample of the survey should be a representative sample of the local community and should also cover residents in the comparison communities. The survey should be repeated at regular intervals, both before and after the project is implemented.

Illustrative design of a local survey

As described above, measuring local effects is likely to rely in part on conducting a survey of the local community. The best approach for conducting this survey will depend on the types of CIF projects which are undertaken and who the intended beneficiaries of these projects are. Here we provide an example of how such a survey could be designed.

The population of interest would be residents of the specific geographic areas intended to benefit from the projects. Depending on the size of the area, either all addresses could be selected, or a random sample of addresses in these areas could be drawn from the Postcode Address File. Data could be collected through either a 'push-to-web' approach or in-person data collection. Push-to-web involves sending sampled addresses a letter, inviting the residents to complete an online survey. Typically, paper versions of the questionnaire are later sent to at least some non-responding households to encourage response, especially among households with less confidence in using the internet. It is possible that other novel survey collection channels could become available across the duration of the GDF programme, and these should be implemented where this is appropriate.

This design can be straightforwardly extended to comparison areas. Here the challenge is identifying a suitable set of comparison areas so that the comparison sample is similar to the sample in the area where CIF projects are delivered. We recommend selecting a comparison location for each potential host community. A shortlist of comparison locations can initially be selected purposively, based on factors such as location, size and local connectivity. Locations on the shortlist can then be compared to the potential host communities according to statistics from the 2021 Census (such as population, demographic profile, employment status and so on), as well as other national data sources, such as the Indices of Multiple Deprivation and Rural/Urban classification. This is to ensure that these locations have similar profiles to the potential host communities. Each comparison location should be associated with a single potential host community. In later phases, as the number of sites under consideration is narrowed down, the set of comparison locations can be similarly reduced.

We recommend that the survey initially be conducted before the start of CIF projects to provide a baseline measure of key outcomes, enabling the difference-in-differences analysis described above. The survey can then be repeated at suitable intervals to record changes in these outcomes. Given the timescales of the evaluation, the sample design (that is, which addresses should be considered in scope) should be reassessed at each time point so that the survey reflects local developments and other changes in the community. It may not be advisable to run the survey at short regular intervals (i.e. every six months) as this may lead to participation fatigue or resistance to the survey, and possibly the GDF programme as well, by extension.

This approach is most suitable for projects which are intended to benefit all people living in the area. This design could also accommodate outcomes aimed at specific sub-groups, as long as these sub-groups are reasonably large. For projects targeted at very specific groups, another approach would likely be more suitable. This would need to be decided once the details of planned community projects are known.