Guidelines on Economic Evaluation in Creative Health

Contents

Contents 1

About this resource 3

Contributors 4

Background 5

Executive Summary 6

How to use these Guidelines 9

Key recommendations for economic evaluation in creative health 12

1. Introduction to economic evaluation 13

1.1. Economic evaluation and the Green Book 13

1.2. Who are these Guidelines for? 15

1.3. When is Economic Evaluation used? 17

1.4. Proportionality 17

1.5. Credibility and rigour 18

2. The HM Treasury Green Book (2026) **19**

2.1. Economic evaluation in the Green Book 19

2.2. Choosing the appropriate form of economic evaluation 20

3. Creative health 22

3.1. What is creative health? 22

3.2. Creative health outcomes 23

3.3. Measurement of outcomes in creative health 24

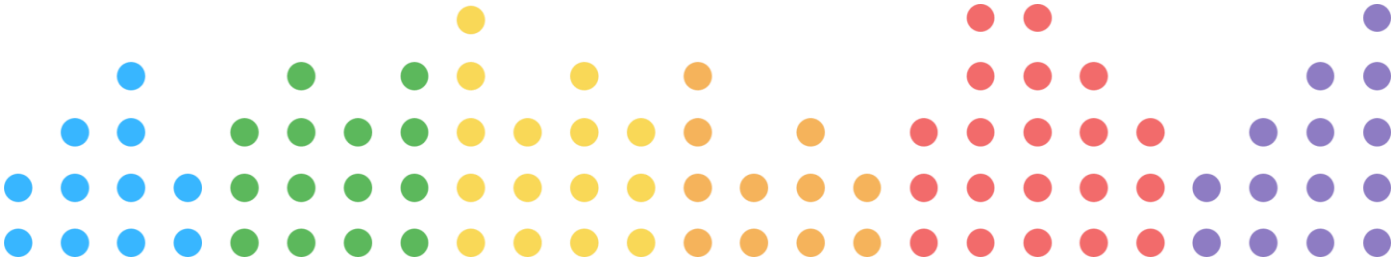
4. Guidelines on economic evaluation for creative health **26**

4.1. Developing a Theory of Change 27

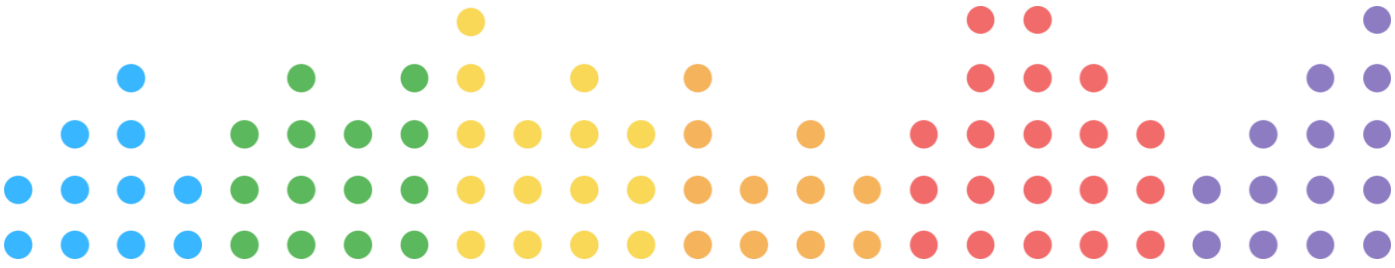
4.2. Measuring impacts 29

4.3. Estimation of costs of creative health interventions 41

5. Cost-Benefit Analysis (CBA) **44**



Step 1. Specify the intervention, identify outcomes and decide groups with standing 44
Step 2. Estimate the impacts of the intervention 44
Step 3. Valuation of costs and benefits of the intervention 45
Step 4. Calculation of CBA results 60
6. Cost-Effectiveness Analysis (CEA) **65**
Step 1. Estimate the impact of the intervention 66
Step 2. Estimate intervention costs 66
Step 3. Calculate cost-effectiveness ratios 66
7. Worked examples of economic evaluations **68**
7.1 Hypothetical worked examples 68
7.2 Applied analyses 73
8. Economic evaluation template **78**
Annexes **82**
Annex A. Examples of experimental and quasi-experimental methods that can be used for creative health evaluations **82**
Annex B. Non-market valuation methods **85**
Annex C. Examples of applied economic evaluation of Flip of the Coin and Odd Arts creative health programmes **88**
Bibliography **121**



About this resource

This template business case has been developed as part of the Mobilising Community Assets to Tackle Health Inequalities research programme to support investment in Creative Health across the UK.

The research is supported by UK Research and Innovation (UKRI), including the Arts and Humanities Research Council (AHRC), Biotechnology and Biological Sciences Research Council (BBSRC), Economic and Social Research Council (ESRC), Medical Research Council (MRC), and Natural Environment Research Council (NERC), in partnership with the National Centre for Creative Health (NCCH). The programme is funded across three phases from 2021 to 2027, led by University College London (UCL)'s Culture Nature Health Research Group (Grant Ref: AH/W006405/1; PI: HJ Chatterjee).

Ethics approval: UCL REC ID: 4525.004 Title: Mobilising Community Assets to Tackle Health Inequalities (Mobilising Community Assets – MCA).

© 2026 University College London. All images © UCL

For further information contact: h.chatterjee@ucl.ac.uk

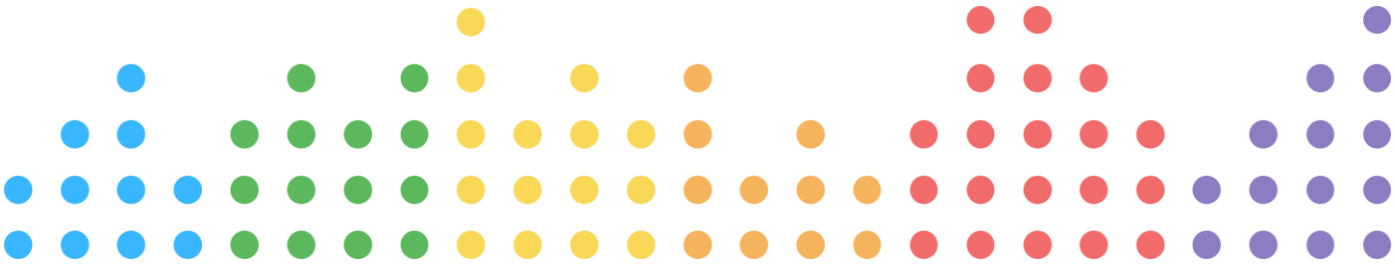
How to Cite

Fujiwara, D., Dolan, P., Muelas, A., Segarra, E., Ruiz, C., Newton, U., McNeill, E., de Andrade, M., Fancourt, D., Mundra, J. and Chatterjee, H.J. (2026). *Guidelines on Economic Evaluation in Creative Health*. London: University College London. Available from: https://ncch.org.uk/uploads/Guidelines-on-Economic-Evaluation-in-Creative-Health.pdf

> *"Creative health can be defined as creative approaches and activities which have benefits for our health and wellbeing. Activities can include visual and performing arts, crafts, film, literature, cooking and creative activities in nature, such as gardening; approaches may involve creative and innovative ways to approach health and care services, coproduction, education and workforce development."*
>
> **The National Centre for Creative Health**

Contributors

Dr Daniel Fujiwara (TSC Impact & Harmon)
Professor Paul Dolan (London School of Economics)
Professor Helen J Chatterjee MBE (University College London)
Alberto Muelas (Harmon)
Elena Segarra (Harmon)
Carlos Ruiz (Harmon)
Uisce Newton (University of Edinburgh)
Marisa de Andrade (University of Edinburgh)
Elaine McNeill (University College London)
Professor Daisy Fancourt (University College London)
Jag Mundra (National Academy for Social Prescribing)
Professor David O'Brien (University of Manchester)
Jayne Howard (National Centre for Creative Health, NCCH)
Professor Martin Marshall CBE (Chair of the Nuffield Trust; Chair, NCCH)
Andrea Sutcliffe, CBE (Trustee, NCCH; former Chief Executive and Registrar, Nursing and Midwifery Council)
Teresa Salami-Oru (FFPH RN Consultant in Public Health East Sussex County Council; Trustee NCCH)
Nancy Hey OBE (Lloyd's Register Foundation; Trustee NCCH)
Ellen Rule (Chief Executive, Herefordshire and Worcestershire NHS Trust)
Harman Sagger (Head Economist for Arts, Heritage and Tourism, Department for Digital, Culture, media and Sport, DCMS)
Matthew Bezzano (Economic Adviser, Culture & Heritage Capital, DCMS)
Anya Harvey (Historic England)
Professor Christopher Smith (Executive Chair, Arts and Humanities Research Council, AHRC, UK Research and Innovation, UKRI)
Dr Jaideep Gupte (Director of Research, AHRC, UKRI)
Lucy Hackett (Head of Health, Environment and Urban Humanities, AHRC, UKRI)
Josie McGregor (Senior Investment Manager for Health, AHRC, UKRI)



Background

Key findings from the Mobilising Community Assets research programme reveal that a significant stumbling block in the closer relationship between the creative health ecosystem, the NHS and public health, is the lack of a defined business case standard that sets out the economic basis for commissioning (Thomson et al., 2021; Thomson et al., 2026). NHS commissioning in England is increasingly focused on using economic evidence to drive "missions" centred on productivity, prevention, and the shift toward community-based care, particularly through Integrated Care. This approach uses data to ensure financial sustainability, reduced health inequalities, and the delivery of high-quality, cost-effective care (NHS Strategic Commissioning Framework).

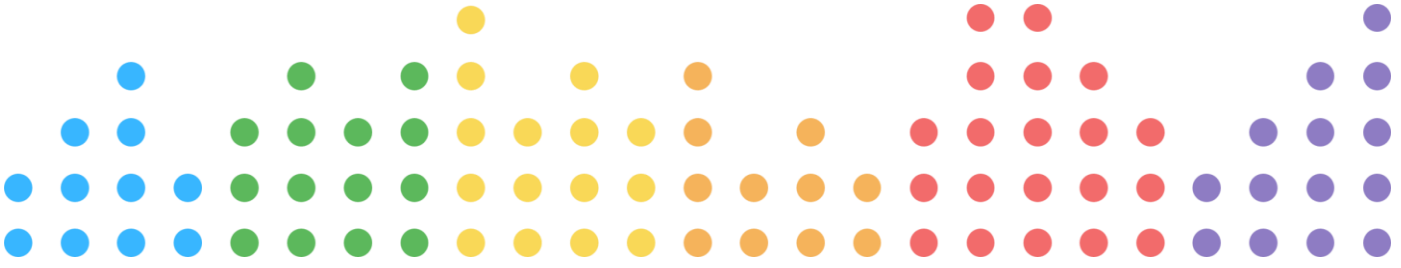
Economic studies in creative health often lack standardised methods and consistent datasets, leading to findings that are not always trusted or valued by decision-makers. In addition, creative health outcomes do not align easily with established NHS performance measures, such as hospital admissions, which limits their integration into existing systems and frameworks.

To address this, there is a need for clearer synthesis and translation of evidence, so that the multiple ways in which creative health can be measured and valued are expressed in terms meaningful to decision-makers (Hearst et al., 2025). These Guidelines are intended to support that process.

These Guidelines have been developed as part of the Mobilising Cultural Assets (MCA) programme, in response to findings from the programme's Mid-Term Review. The Review identified a need to strengthen the economic evidence available to demonstrate the impact and value of creative health and to support its translation into commissioning, funding and policy decisions. In response, additional investment was made in a dedicated economic evaluation workstream to develop a more consistent, credible and practically applicable approach to the economic evaluation of creative health interventions.

The Guidelines are a key output of this workstream and form an important part of the legacy of the MCA programme. They draw on learning generated through the programme, established approaches to public-sector economic evaluation, and consultation with stakeholders from across creative health, health and social care, government, commissioning, funding, research and practice. The purpose is to provide a framework that can continue to support the development, evaluation and commissioning of creative health.

As part of the development of these Guidelines, a Roundtable was convened on 3 June 2026 with key stakeholders from across the health and creative health sectors, including government departments and representatives with expertise in commissioning, evaluation, policy, research, delivery and funding. The discussion explored the practical challenges of evaluating creative health activities, the needs of different audiences for economic evidence, and the importance of proportionate, credible and usable guidance. The comments, insights and recommendations shared during the session have directly informed the structure, emphasis and content of this guidance document.



Executive Summary

Creative health encompasses creative and cultural activities that seek to improve health and wellbeing, ranging from community-based participation and social prescribing to interventions embedded within health and care services. As investment in creative health grows, commissioners, funders, providers and government increasingly need credible evidence about whether interventions improve outcomes and represent good use of resources.

These Guidelines provide a consistent framework for the **economic evaluation of creative health interventions**. Their purpose is to help evaluators produce evidence that is proportionate, credible and useful for decision-making, while aligning creative health evaluation with established principles of public-sector appraisal, particularly the **HM Treasury Green Book**.

What is economic evaluation?

Economic evaluation asks whether the benefits generated by an intervention justify the resources used to deliver it. The appropriate evaluation will depend upon the intervention, the available evidence and, critically, the decision the evaluation is intended to inform.

Before beginning an evaluation, users should therefore identify the intervention and its objectives, the decision to be informed, the intended decision-maker and audience, and the outcomes most relevant to that decision.

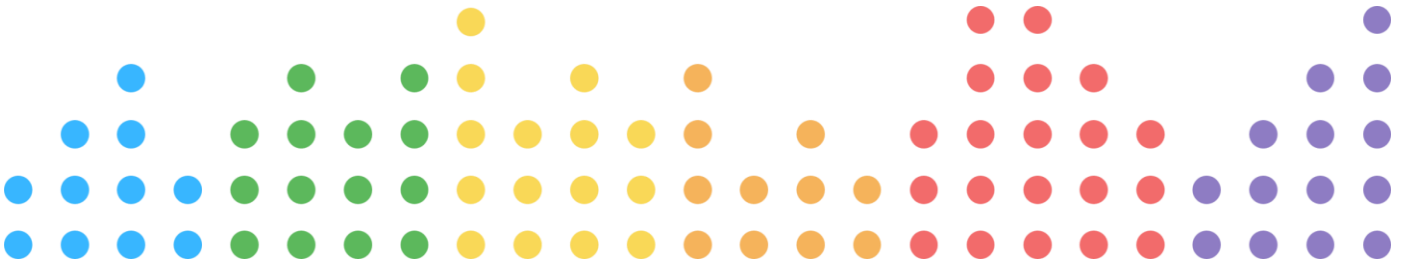
The Guidelines recommend four core steps as part of this process.

1. Develop a Theory of Change. Set out clearly how the creative health intervention is expected to generate change. Identify its inputs, activities and outputs; the people expected to benefit; the outcomes that matter; and the causal pathways connecting participation to those outcomes.

2. Measure outcomes and impact. Collect evidence on the outcomes identified in the Theory of Change. Measuring change alone is not sufficient: wherever possible, evaluators should consider what would have happened without the intervention — the counterfactual — so that the additional impact attributable to creative health can be estimated. The strength of the evaluation design should be proportionate to the scale, resources and intended use of the intervention.

3. Measure the costs of delivering the intervention. Identify the resources required to deliver the creative health intervention, including relevant financial expenditure and other resource costs.

4. Undertake an appropriate economic evaluation. Where outcomes can credibly be valued in monetary terms, **Cost-Benefit Analysis (CBA)** should be used to compare the social benefits of an intervention with its costs. Where the principal question concerns the cost of achieving a particular outcome, **Cost-Effectiveness Analysis (CEA)** may be



more appropriate. The method chosen should reflect the decision being made and the information needs of the decision-maker and audience.

Outcomes and evidence

Creative health interventions can generate multiple forms of value, including improvements in health, subjective wellbeing, loneliness and social connection, confidence and self-efficacy, cultural participation, employability, and healthcare utilisation. Evaluators should begin by identifying the outcomes that matter for the intervention and the people participating in it, rather than selecting outcomes simply because they can readily be given a monetary value.

These Guidelines distinguish clearly between measurement and valuation. The first task is to identify and measure the health, wellbeing, social and experiential outcomes that matter and to establish whether they are attributable to the intervention. Economic valuation is undertaken only after this, where it is appropriate to the decision being informed.

To improve consistency and comparability across the sector, these Guidelines propose a **Core Outcome Framework**: a small set of common outcomes and measures that creative health evaluations should collect where relevant and feasible. These cover health utility (Quality Adjusted Life Year); subjective wellbeing (Life satisfaction); and healthcare utilisation (GP visits and A&E attendances). Wider outcomes that could be included such as social connection, self-efficacy, mental health, participation and other outcomes where supported by the Theory of Change are also discussed.

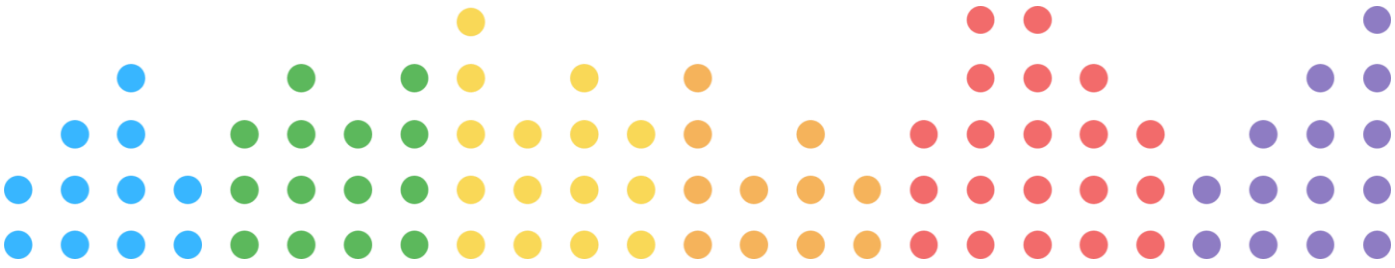
The Framework is not intended to imply that all creative health interventions have the same objectives, or to replace intervention-specific measures. Rather, it provides a common measurement foundation that can be supplemented by additional quantitative and qualitative evidence reflecting the particular intervention, population and Theory of Change.

Greater consistency in collecting these outcomes should, over time, make it easier to compare findings across creative health interventions and strengthen the evidence base available to commissioners, funders and practitioners.

Proportionality and credibility

A central principle of these Guidelines is that the strength of impact evaluation required should be proportionate to the scale of the intervention and the decision the evidence is intended to inform. Economic evaluation needs to consider the counterfactual: what would have happened to participants in the absence of the intervention.

For larger-scale interventions seeking substantial public investment, or evaluations intended to support significant commissioning decisions, the Guidelines recommend robust impact evaluation that explicitly estimates the counterfactual. Wherever



feasible, this should involve an appropriate control or comparison group in order to produce credible estimates of the additional impact attributable to the intervention.

For smaller creative health interventions, particularly community and voluntary-sector programmes with limited resources, constructing a robust control group may be neither feasible nor proportionate. The Guidelines therefore set out pragmatic approaches that allow these programmes to estimate impact without undertaking a full counterfactual evaluation, while making the greater uncertainty explicit. This includes conservative approaches to deadweight and attribution and transparent reporting of assumptions and limitations.

The principle is therefore **proportionate rigour rather than a single evidential standard for every intervention**. A small community arts programme should not be expected to undertake the same evaluation as a large regional or national programme. However, all evaluations should distinguish as clearly as possible between observed change and impact attributable to the intervention and should avoid making claims that are stronger than the underlying evidence supports. Credible estimates are more valuable to decision-makers than large but weakly supported claims of impact, value or financial savings.

Using the Guidelines

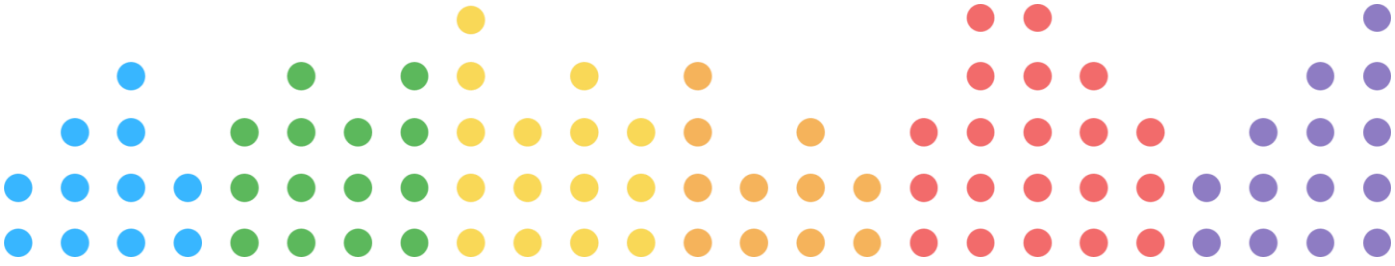
Different users will require different parts of this document. Providers may principally need guidance on Theory of Change, outcome measurement and proportionate data collection. Evaluators and analysts will require the more detailed sections on impact estimation, monetary valuation, CBA and CEA. Commissioners and funders may focus particularly on how evidence should be interpreted and used to support investment and commissioning decisions.

Readers do not therefore need to read the document sequentially. The **"How to use these Guidelines"** section that follows provides recommended routes through the guidance for different users.

The overall objective is that: **creative health should be evaluated using evidence that is proportionate to the intervention, credible about what can be attributed to it, consistent enough to support comparison, and designed from the outset to inform real decisions.**

Economic evaluation template

The Economic Evaluation Template in Section 8 provides a standardised, practical framework for applying these Guidelines. It guides users through defining the intervention and decision context, developing a Theory of Change, selecting CBA or CEA, assessing impact, measuring costs and outcomes, reporting economic results, addressing uncertainty, and presenting findings for decision-making.



How to use these Guidelines

The purpose of these Guidelines is to improve the quality, consistency and decision-usefulness of economic evaluation in creative health. They provide a common methodological framework for assessing the impacts, costs and value of creative health interventions, with the aim of producing economic evidence that is sufficiently credible, transparent and comparable to inform decisions about commissioning, investment, adoption and scaling.

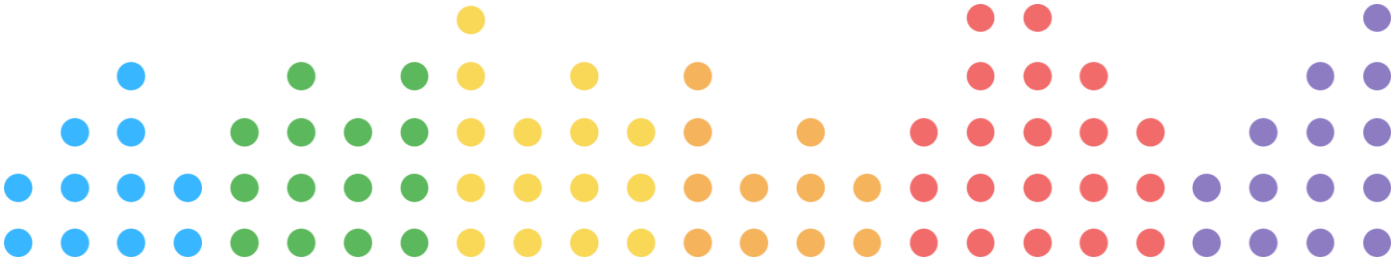
The Guidelines are principally a technical and methodological resource. They are intended to support those designing, undertaking, commissioning or interpreting economic evaluations of creative health interventions. They set out how to develop a Theory of Change, measure outcomes and attributable impact, estimate costs and benefits, and undertake Cost-Benefit Analysis (CBA) or Cost-Effectiveness Analysis (CEA), using methods that are proportionate to the scale of the intervention and the decision being informed.

These Guidelines are intended for a wide range of users, including creative health practitioners and providers, commissioners, funders, policy makers, researchers and economic evaluators. These users have different reasons for undertaking or using economic evaluation and will require different levels of methodological detail.

Readers should, therefore, use the sections that are most relevant to the intervention they are evaluating, the decision the evaluation is intended to inform, and the level of evaluation that is proportionate to the scale and resources of the intervention. For example, a small community organisation seeking to understand and demonstrate the outcomes of its programme will have different evaluation requirements from a large programme seeking substantial public investment. Similarly, an NHS commissioner considering whether to commission a service may require different economic evidence from a government department or funder interested in the wider social value generated by creative health.

Users of these Guidelines need to follow a process covering two core decisions. These decisions determine both the type of economic evaluation that should be undertaken and the level of statistical rigour that is proportionate to the intervention in terms of measuring impact. These terms are described below and a decision flowchart is presented in Figure 1 (page 10).

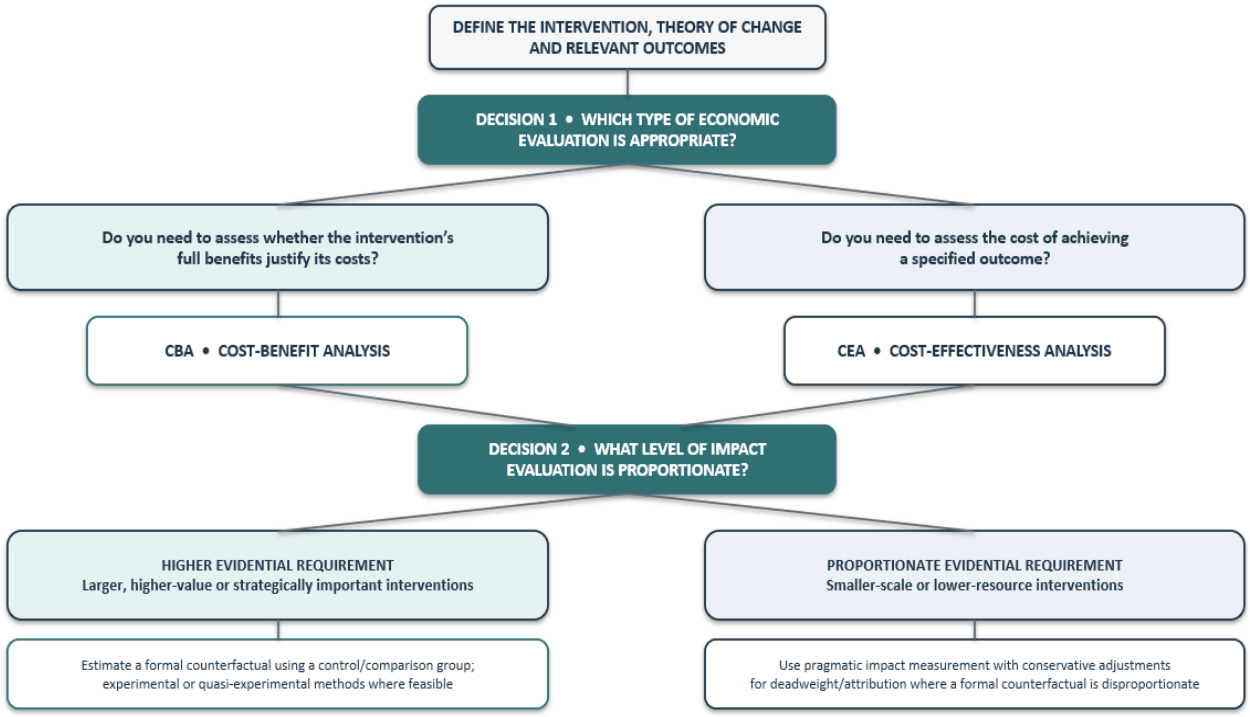
i. **Cost-Benefit Analysis (CBA) or Cost-Effectiveness Analysis (CEA) as the appropriate form of economic evaluation**. This choice should be driven by the decision that the evaluation is intended to inform. CBA is appropriate where the aim is to assess whether the full set of benefits generated by a creative health intervention justify its costs. This goes beyond just the healthcare cost savings to cover impact areas such as wellbeing and employability. This is suited to decisions that require consideration of multiple outcomes and, where possible,



their monetary valuation. CEA is more appropriate where the decision focuses on the cost of achieving a particular specified outcome, such as an improvement in health-related quality of life or a reduction in healthcare utilisation. The relevant method should therefore follow from the needs of the decision-maker.

ii. **Level of rigour required for estimating impact**. The Guidelines take a proportionate approach: larger, higher-value or strategically important interventions should provide stronger evidence that observed changes were caused by the intervention rather than by other factors. Where feasible, this means explicitly estimating the counterfactual using an appropriate control or comparison group and employing appropriate statistical methods. Smaller or lower-resource interventions should not be expected to meet the same evidential standard. In these cases, more pragmatic approaches to impact measurement can be used, provided assumptions about deadweight, attribution and uncertainty are made explicit and reported conservatively. The appropriate level of rigour therefore depends on factors such as scale, investment, participant numbers, strategic importance and the intended use of the evidence.

Figure 1. Decision flowchart to determine which economic evaluation methods to use



For users of these Guidelines interested in conducting a full assessment of the costs and benefits of creative health interventions under CBA Chapter 4 (how to create a Theory of Change and statistical methods for impact measurement) and Chapter 5 (how to conduct a Cost-Benefit Analysis) will be of most relevance.

For those interested in understanding the impact of creative health interventions on public health resources and costs and on health outcomes using CEA Chapter 4 (how to create a Theory of Change and statistical methods for impact measurement) and Chapter 6 (how to conduct a Cost-Effectiveness Analysis) will be of most relevance.

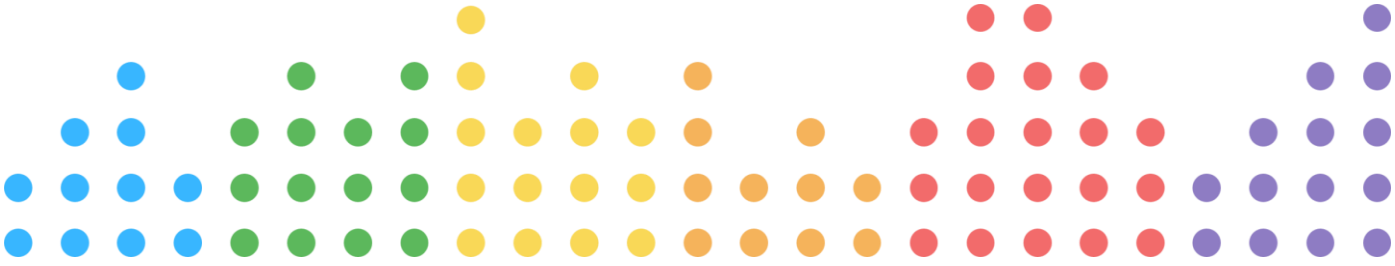
The **Economic Evaluation Template in Section 8** provides a practical tool for applying these Guidelines. It takes users through the key stages of an evaluation, including defining the intervention and decision context, selecting CBA or CEA, determining a proportionate impact evaluation approach, measuring outcomes and costs, and reporting results and uncertainty. It provides a minimum framework that can be supplemented with additional analysis where appropriate and proportionate.

These Guidelines form part of a wider suite of resources designed to strengthen the evidence and decision-making infrastructure for creative health. Each resource has a distinct purpose and primary audience, including specific guidance for health commissioners and creative health providers. For more information visit: https://ncch.org.uk/resources/creative-health-economics-the-economic-business-case-for-creative-health.

Main findings: McNeill, E., Newton, U., de Andrade, M., Fancourt, D., Fujiwara, D., Mudra, J., et al & Chatterjee, H.J. (2026) **Creative Health Economics: Building the foundations for a commission-ready sector.** London: University College London. Available at: https://ncch.org.uk/uploads/Creative-Health-Economics-Building-the-foundations-for-a-commission-ready-sector.pdf

For providers of Creative Health: McNeill, E., Newton, U., de Andrade, M., Fancourt, D., Fujiwara, D., Segarra, E., Muelas, A., Ruiz, C., Frier, R., Chapman, L., Thomson, L.J. et al. & Chatterjee, H.J. (2026) **Delivering Commission-ready Creative Health: The Economic Business Case for Investment.** London: University College London. Available at: https://ncch.org.uk/uploads/Delivering-Commission-ready-Creative-Health-The-Economic-Business-Case-for-Investment.pdf

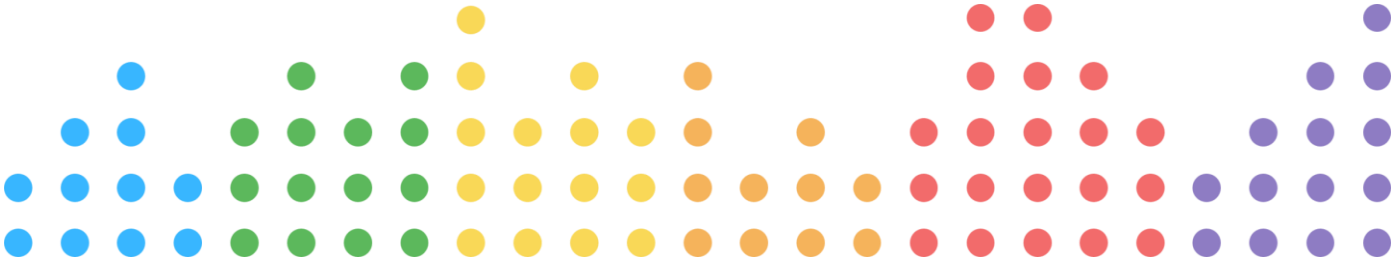
For those commissioning Creative Health: McNeill, E., Newton, U., de Andrade, M., Fancourt, D., Fujiwara, D., Mundra, J., Moss, D., Rule, E., Hobart, E., Sutcliffe, A. et al. & Chatterjee, H.J. (2026). **Creative Health Commissioning: An economic business case for improving population health and reducing health inequalities.** London: University College London. Available at: https://ncch.org.uk/uploads/Creative-Health-Commissioning_An-Economic-Business-Case.pdf



Key recommendations for economic evaluation in creative health

The principal contribution of these Guidelines is to translate the economic evaluation methodology into a consistent and proportionate framework for creative health. The recommendations below summarise the creative health-specific standards proposed in these Guidelines. Detailed explanations, evidence and worked examples are provided in the subsequent sections.

Area	Recommendation
Choosing the economic evaluation method	Select Cost-Benefit Analysis where the decision is whether the intervention's full benefits justify its costs. Select Cost-Effectiveness Analysis where the decision concerns the cost of achieving a specified outcome. The method should follow from the aims of the decision at hand.
Proportionality	Larger, higher-value or strategically important interventions should use stronger statistical methods; smaller interventions can use proportionate approaches where full control-group evaluation would be unreasonable.
Minimum impact standard	Where a robust control group is not feasible, use the recommended minimum approach: preferably Interrupted Time Series where appropriate, otherwise structured pre-post measurement with counterfactual adjustment.
Core Outcome Framework	Collect data on the core outcomes relevant to the intervention and supplement these with intervention-specific outcomes identified through the Theory of Change.
Recommended economic measures	Use the Quality-Adjusted Life Year (QALY) where the primary focus is health; life satisfaction where broader wellbeing is central; and GP/A&E utilisation where healthcare resource impacts are relevant. Avoid double counting overlapping benefits.
Reporting	Report assumptions, uncertainty, limitations and sensitivity analysis transparently.
Application	Use the worked examples and templates in Sections 7 and 8 to apply the framework in practice.



1. Introduction to economic evaluation

1.1. Economic evaluation and the Green Book

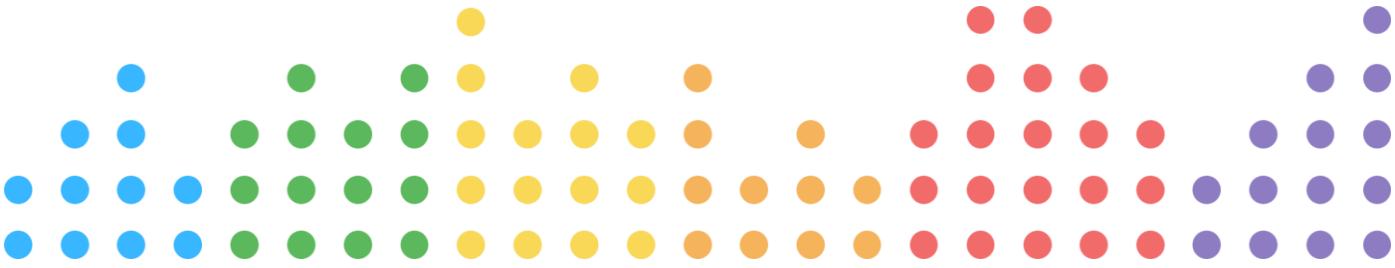
Economic evaluation is the systematic comparison of alternative uses of resources in order to inform decisions about how to maximise **benefits and value to society**. At its core, economic evaluation recognises that resources are scarce and that choosing one intervention means forgoing others. It, therefore, requires explicit identification of a counterfactual (often "business as usual"), clear specification of objectives, and measurement of costs and benefits across options.

The **HM Treasury Green Book** (2026) is the UK Government's authoritative guidance on how public sector bodies should appraise policies, programmes and projects that use public resources and it establishes economic evaluation as the core methodology for making these appraisals. The Green Book sets out the principles, standards and analytical framework for assessing **costs**, **benefits** and **risks** before decisions are made and its core purpose is to ensure that advice to decision makers is evidence-based, objective and focused on maximising **social value** for the United Kingdom as a whole.

The Green Book defines economic evaluation as the **structured comparison of options** to determine which delivers best **value for money** for society, where "value for money" includes a comprehensive assessment of all **financial impacts** such as public-sector costs and tax receipts for government and income effects for individuals, and **non-financial impacts** such as health, justice, arts and cultural participation, security, biodiversity and the environment (2026, p.4).

This expansive conception of value in the Green Book is particularly relevant for creative health interventions, whose benefits often extend beyond conventional healthcare metrics.

Creative health is an approach to improving health and wellbeing through participation in creative and cultural activities such as music, visual arts, dance, theatre, crafts, literature and engagement with heritage or nature. Creative health recognises that health outcomes are influenced not only by medical treatment but also by social, cultural and environmental determinants, and it seeks to embed creative activities within health and social care services, communities and social prescribing pathways so that cultural participation complements conventional medical and social care, contributes to prevention and supports health, independence and wellbeing. Creative health initiatives, therefore, aim to mobilise cultural organisations, community groups and creative practitioners as part of wider health and care systems.



Creative health brings together a wide range of practices across a continuum of engagement, from clinically embedded work within healthcare settings, creative activity within social care and supported living settings, and targeted health and wellbeing programmes, through to broader social prescribing activity and everyday creative participation. Across this spectrum, cultural organisations, community groups and creative practitioners influence and interact with the social determinants of health to improve health and wellbeing.

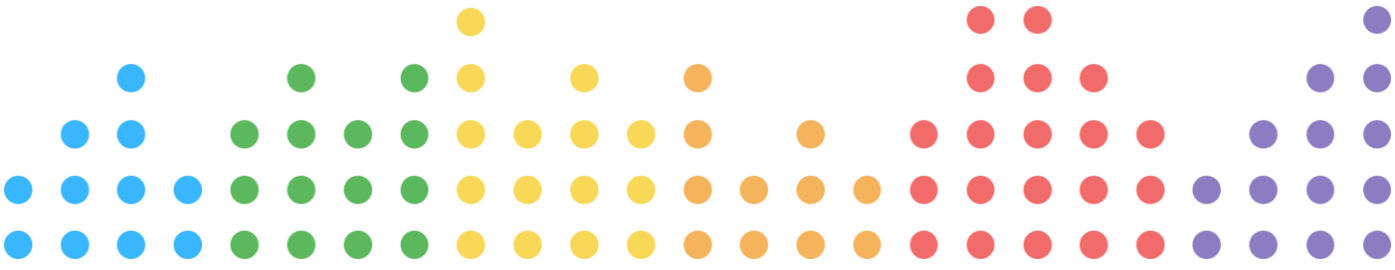
Creative health typically operates through a variety of mechanisms. These include clinical interventions delivered in partnership with NHS services as part of treatment and recovery; creative activities delivered in or alongside social care services to support the wellbeing, independence, social connection and quality of life of older people, people with physical or learning disabilities, and others drawing on care and support; and structured and facilitated programmes designed to support specific groups or health and wellbeing needs, shaped through collaboration between cultural, health and social care partners. Beyond this, creative health encompasses a wide range of open, community-led and place-based opportunities that build social capital, reduce isolation and support wellbeing.

Throughout these Guidelines, the term "intervention" is used as a generic term to refer to any creative health programme, project, activity, initiative, service, or model of delivery. This terminology is adopted purely as a simplification for readability and does not imply any loss of nuance regarding the diversity of creative health practice.

The Green Book (2026) and economic evaluation are mandatory for government departments and arm's-length bodies when developing proposals that involve public spending. The Green Book is also widely adopted across devolved administrations and local authorities in the UK (2026, p.1). Its purpose is to ensure that decisions about the use of public resources are based on consistent, evidence-based analysis of costs, benefits, risks and distributional impacts, with the aim of maximising social value and achieving value for money for UK society as a whole (HM Treasury, 2026).

The decision context for creative health economic evaluation varies considerably, and different decision-makers may place different emphasis on particular outcomes and resource impacts. For illustration, the priorities of three important stakeholder groups may include:

i. **Central government and funders**, who may be interested in the full range of financial and non-financial impacts generated by creative health;
ii. **Health, social care and integrated care commissioners**, including NHS bodies and local authorities, who may be concerned with health and wellbeing



outcomes, independence and quality of life, demand for health and social care services, prevention and the costs falling on different parts of the health and care system; and

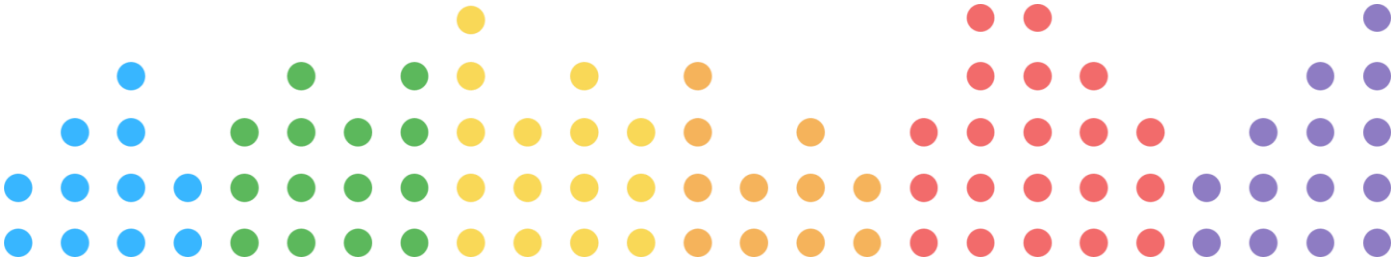
iii. **Providers and practitioners**, who require proportionate evidence to understand whether interventions are achieving their intended outcomes, improve delivery and demonstrate value to commissioners and funders. They tend to be interested in wellbeing, participation, inclusion and social value. That is, the non-financial impacts of creative health.

Increasingly, commissioning decisions are made across organisational and sectoral boundaries. Creative health interventions may generate benefits for both health and social care, while their costs and savings may fall on different organisations. Economic evaluation should therefore identify the relevant commissioning context and consider the distribution of costs and benefits across the health and care system where this is material to the decision. These examples describe typical priorities rather than determining the appropriate form of economic evaluation, which should be selected according to the decision being informed.

These Guidelines have been developed through the **MCA programme**, following its Mid-Term Review and subsequent additional investment in a dedicated economic evaluation workstream. This work recognised that the continued development and scaling of creative health requires economic evidence that is credible to decision-makers while remaining proportionate and practical for the diverse organisations delivering creative health interventions. The Guidelines therefore represent an important legacy of the MCA programme, providing a common framework intended to strengthen the consistency, quality and application of economic evaluation across the creative health sector.

1.2. Who are these Guidelines for?

These Guidelines are intended for people who produce, commission, interpret or use economic evidence about creative health interventions. This includes economists, analysts and researchers undertaking economic evaluation; commissioners, funders and policy makers who need to assess the quality and relevance of economic evidence; and creative health providers and practitioners who need to understand the evidence requirements associated with evaluation and future commissioning or investment. Different users will require different levels of methodological detail. The Guidelines therefore establish a common methodological standard and language through which providers, evaluators, commissioners and funders can understand what constitutes credible and proportionate economic evidence.



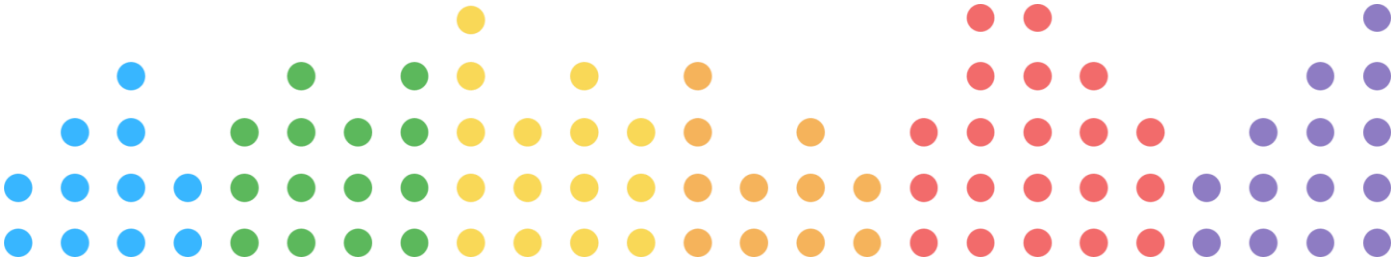
The appropriate economic evaluation will depend on the decision being informed, the perspective adopted and the information needs of the decision-maker. For example, central government and some funders may require evidence about the full financial and non-financial impacts of creative health, while health, social care and integrated care commissioners may place particular emphasis on outcomes and health resource impacts relevant to the populations and services for which they are responsible.

For the purposes of these Guidelines, we distinguish between **two broad decision contexts**. The first is where the decision-maker wishes to understand whether the **full benefits of a creative health intervention justify its costs**, taking account of relevant financial and non-financial impacts across society. The second is where the decision-maker is principally interested in the **cost of achieving a specified outcome**, such as an improvement in health-related quality of life or a reduction in the use of health or care services. These different decision contexts have implications for the outcomes and measures required and, importantly, for the appropriate form of economic evaluation. The Guidelines therefore explain which types of outcomes, measures and economic evidence are most relevant to each.

These decision contexts are intended to provide broad guidance rather than prescribe a fixed evaluation approach for particular organisations. In practice, the evidence needs of decision-makers overlap. NHS commissioners, social care commissioners and other funders may place considerable importance on healthcare utilisation, social care outcomes and public service costs, whilst central government and local authorities may be interested in wider health, wellbeing and social outcomes, particularly where these support prevention or the reduction of health inequalities. Evaluators should therefore begin by identifying the **decision to be informed and the needs of the decision-maker**, rather than assuming that a particular economic evaluation method should be used solely because of the type of organisation commissioning or using the evaluation.

These two decision contexts broadly correspond to the two principal forms of economic evaluation considered in these Guidelines: **Cost-Benefit Analysis (CBA)** is appropriate where the objective is to assess whether the full benefits generated by an intervention justify its costs, requiring relevant financial and non-financial benefits to be valued in monetary terms where this can be done credibly. **Cost-Effectiveness Analysis (CEA)** is appropriate where the objective is to assess the cost of achieving a specified outcome, such as health utilisation rates.

The Green Book provides the institutional framework and principles through which creative health interventions can be compared with alternative uses of public resources using CBA or CEA.



1.3. When is Economic Evaluation used?

Economic evaluation can be undertaken both **before and after an intervention is implemented**. In **Green Book terminology, appraisal** is undertaken before a decision is made (ex-ante), to assess options and inform the choice between them, while evaluation is undertaken during or after implementation (ex-post), to assess what happened in practice and generate learning for future decisions. Throughout these Guidelines, "economic evaluation" is used as an overarching term where the principles and methods apply to both contexts.

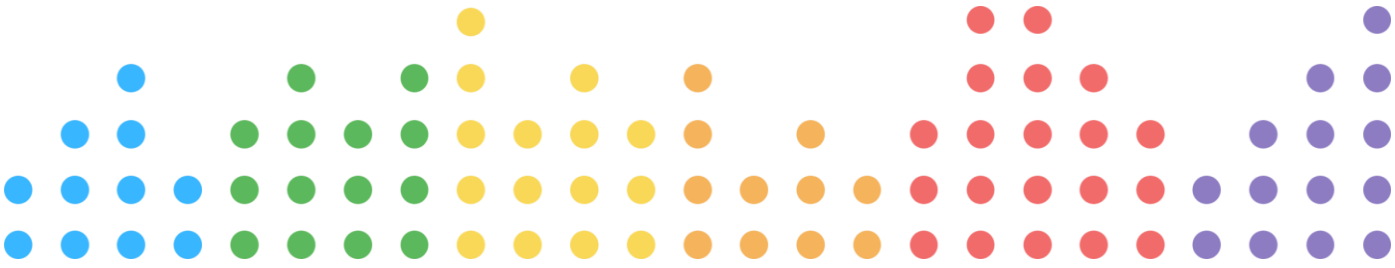
In creative health, ex-post economic evaluations are used to estimate the impacts generated by an intervention that has already been delivered (or is being delivered). Ex-ante economic evaluations, by contrast, seek to estimate the likely impacts of a future intervention before implementation. Although ex-ante evaluations concern future programmes, they remain dependent on robust evidence regarding programme effectiveness and in practice, predictions of future impacts are typically based on evidence from previous similar interventions, pilot studies, or comparable programmes.

Consequently, the principles concerning evidence quality, causal inference and impact measurement set out in these Guidelines are relevant to both ex-ante appraisal and ex-post evaluation, although the evidence available and its application will differ between the two.

1.4. Proportionality

These Guidelines take a **proportionate** approach, as advised by the Green Book (p.2), whereby the level of evidence and its standards should reflect the scale and cost of an intervention. Therefore, a small community arts project should not be expected to undertake the same level of evaluation as a national creative health intervention seeking major public investment. We provide guidance on the level of evaluation required in these different types of scenarios.

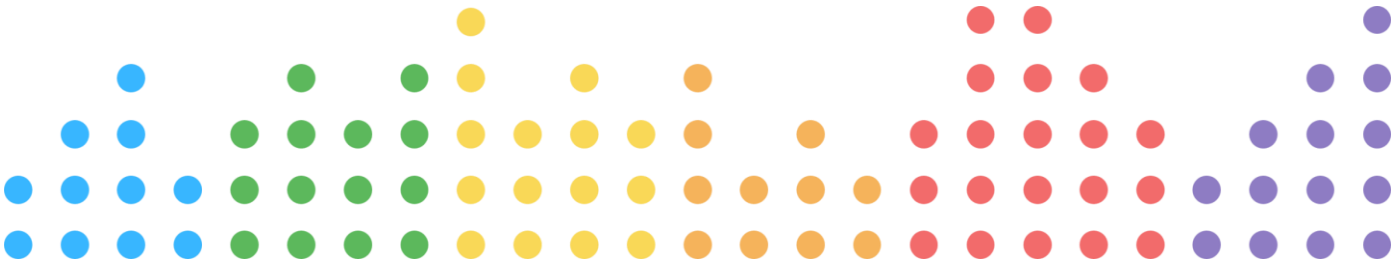
Proportionality also applies to the timing and purpose of evaluation. Economic evaluation should not be viewed solely as a retrospective exercise undertaken at the end of an intervention to support external funding or commissioning decisions. Proportionate measurement and analysis during delivery can support rapid learning and continuous improvement by identifying at an early stage whether an intervention is achieving its intended outcomes. This allows providers to adapt delivery where necessary and reduces the risk of reaching the end of an intervention before identifying limited impact.



1.5. Credibility and rigour

A final important principle of these Guidelines is that **credibility** is the core foundation of economic evaluation, rather than the ability to show large and persuasive headline figures. Many stakeholders in the creative health sector feel that unrealistic claims can undermine confidence in creative health evidence and they warn against excessive extrapolation; inflated value estimates (including overstated healthcare savings); and opaque assumptions.

These Guidelines, therefore, place emphasis on conservative modelling; transparent assumptions; sensitivity analysis; explicit treatment of uncertainty; and robust valuation methods for creative health outcomes.



2. The HM Treasury Green Book (2026)

The Green Book process is a multi-faceted one with the core aim of providing policy makers with an evidence-based decision-making process. It does this through a structured framework for the development and appraisal of public spending proposals through the **Business Case process**, which structures public spending proposals around five perspectives: the **strategic case** (the rationale and objectives); the **economic case** (assessment of options and value for money); the **commercial case** (procurement and delivery arrangements); the **financial case** (affordability), and the **management case** (implementation, governance and evaluation).

Economic evaluation principally contributes to the **economic case**. The strategic, commercial, financial and management cases are important components of a full business case but fall outside the primary scope of these Guidelines. Readers should refer to the Green Book and Business Case guidance for these wider requirements.

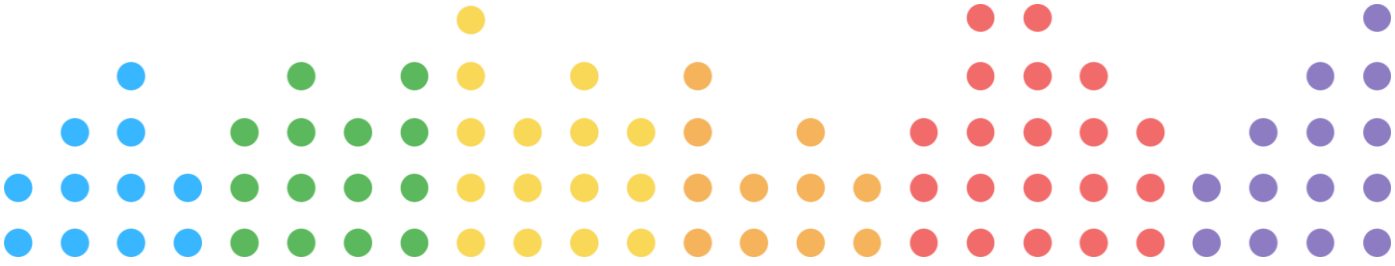
2.1. Economic evaluation in the Green Book

The aim of the **economic case** is to compare alternative ways of achieving the policy objectives and ultimately to identify the preferred option: the intervention that optimises value for money, given the objectives and constraints of the proposal.

Economic evaluation is the analytical method used within the economic case to perform this assessment. It involves identifying and quantifying the social costs and benefits of each option over time and, where possible, expressing them in monetary terms. There are two endorsed methods for economic evaluation in the Green Book:

1. **Social Cost–Benefit Analysis (CBA)**. CBA is the preferred method for economic evaluation in the Green Book. It involves identifying, measuring and, to the greatest extent possible, monetising the social costs and benefits of alternative policy options. This should comprehensively identify and value all relevant impacts of an intervention, capturing not only economic effects but also wider social and environmental impacts. These impacts are assessed over the lifetime of the intervention and adjusted for inflation and discounting to reflect time preferences.

   By expressing costs and benefits in monetary terms, CBA allows them to be compared on a common basis. The results are summarised as the **net present social value (NPSV)**, which is the total benefits minus the total costs, and the **benefit–cost ratio (BCR)**, which is the total benefits divided by the total costs. This enables decision makers to assess which option generates the greatest overall value for money for society.



2. **Cost-Effectiveness Analysis (CEA).** The Green Book states that CEA can be used when the benefits of an intervention cannot be reliably expressed in monetary terms. Instead of monetising benefits, CEA compares the costs of alternative options relative to their effectiveness in achieving a specified outcome, such as reduced hospital waiting times, or reductions in healthcare spending. Outcomes are measured in natural units (for example, average waiting times) while costs are expressed in monetary terms.

   Results are expressed as the **cost-effectiveness ratio (CER)**, which is calculated as the cost of the intervention divided by the number of outcomes achieved. This shows the marginal cost of generating one additional outcome. This allows decision makers to identify the option that delivers the desired outcome at the lowest cost.

CBA is the preferred method for economic evaluation in the Green Book because it provides the most comprehensive way of assessing the overall social value of policy options. CBA monetises the social costs and benefits of alternative interventions, which allows them to be aggregated and compared on a common monetary basis, summarised as the NPSV or BCR.

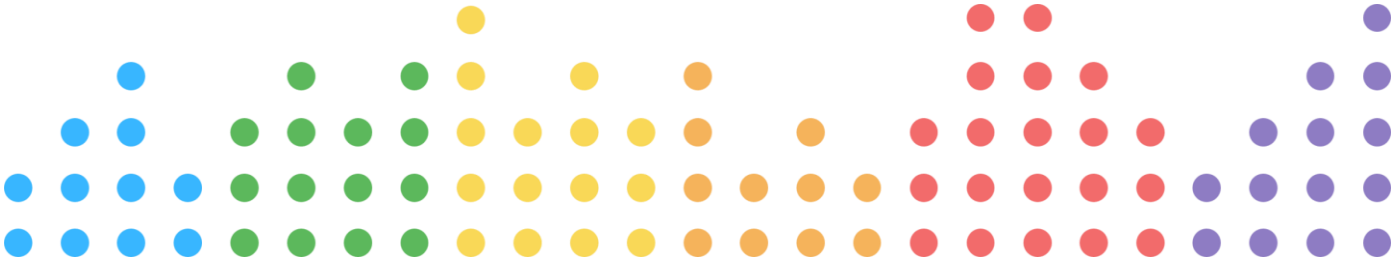
A key advantage of CBA is that it supports comparison not only between options within a single policy area (e.g. different interventions that all have the aim of improving mental health), but also across different types of interventions (e.g. interventions in crime prevention vs interventions in education) competing for public resources. Because all impacts are expressed in monetary terms, decision makers can assess trade-offs between very different outcomes and determine which investments deliver the greatest social value.

By contrast, CEA can only be deployed when assessing options within a single policy area. CEA can identify the least costly way of achieving a given outcome, but it cannot measure the overall value of benefits relative to costs or compare interventions that produce different types of outcomes (e.g. crime reduction vs educational attainment). For this reason, the Green Book recommends using CBA wherever credible monetary valuation is possible, and reserves CEA for cases where benefits are largely unmonetizable (2026, p.32).

2.2. Choosing the appropriate form of economic evaluation

The choice between CBA and CEA should be determined primarily by the decision that the economic evaluation is intended to inform:

CBA is appropriate where the decision requires an assessment of whether the full benefits generated by a creative health intervention justify its costs. This requires



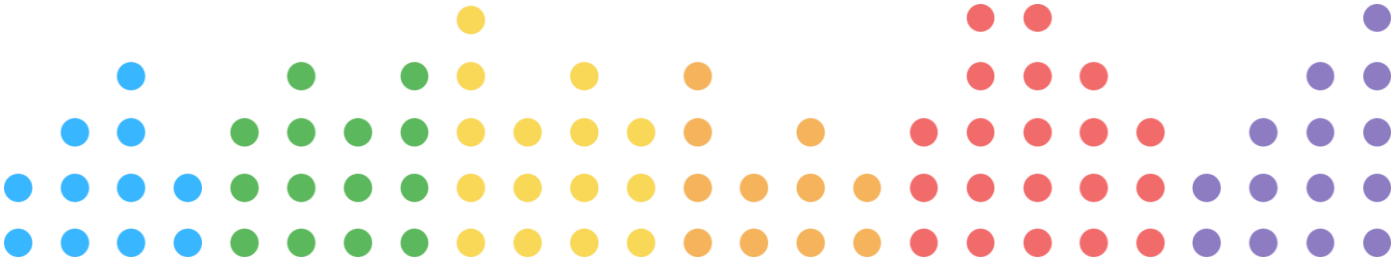
evaluators to identify the relevant financial and non-financial outcomes generated by the intervention and, where credible monetary values are available, express these benefits in monetary terms so that they can be compared with the costs of delivery.

CEA is appropriate where the decision concerns the cost of achieving a specified outcome. For example, the decision-maker may wish to compare the cost of alternative interventions for improving health-related quality of life, reducing healthcare utilisation or achieving another clearly defined outcome. The effectiveness measure should reflect the objective of the intervention and the decision being informed.

For central government and funders the focus may be on assessing and including a comprehensive set of outcomes related to creative health interventions which will include not only impacts on personal health and healthcare spending, but also on wider outcomes such as loneliness, cultural participation and wellbeing. For NHS commissioners, the focus may be on reductions in healthcare utilisation, such as GP visits and A&E attendances, or on improvements in health-related quality of life. The appropriate effectiveness measure should reflect the primary objective of the intervention and the decision being informed.

Local authorities and social care commissioners may have different but overlapping evidence requirements. Creative health interventions delivered through or alongside social care may seek to improve wellbeing, independence, social connection and quality of life for people drawing on care and support, including older people and people with physical or learning disabilities.

Where commissioning is integrated across health and social care, evaluation should consider impacts across organisational boundaries. An intervention may, for example, require investment from one part of the system while generating benefits or avoided costs elsewhere. The appropriate economic perspective should therefore reflect the decision being made rather than the budget of a single organisation.



3. Creative health

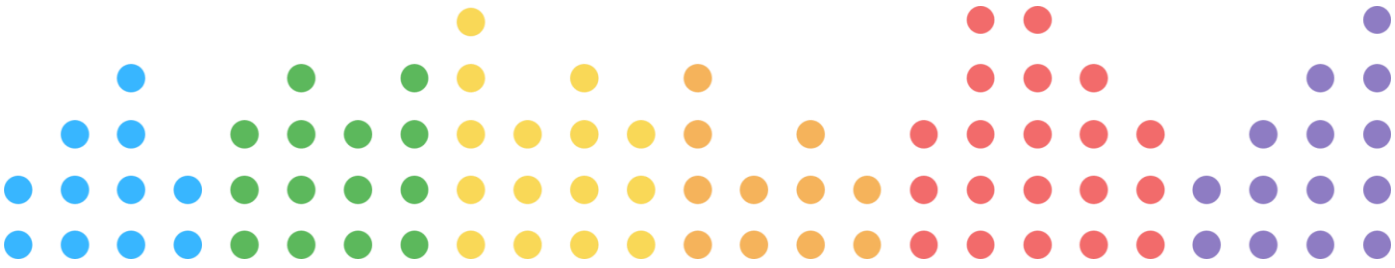
3.1. What is creative health?

Creative health refers to the use of creative, cultural and participatory activities across the arts, culture, heritage and nature to improve health and wellbeing across the life course. It recognises that health is shaped not only by medical and nursing treatment and care, but also by social, cultural and environmental determinants, positioning creative activity as a core component of prevention, community wellbeing, and health and care delivery (Marmot et al., 2010; Marmot et al., 2020; APPG, 2017). Creative health activities are characterised by intentional creative engagement, whether participatory or receptive, distinguishing them from general recreational or lifestyle activities.

Encompassing a continuum of activities, approaches and interventions, creative health responds to clinical, health and social needs across different levels of intensity and formality. At one end, this includes clinically embedded services within the NHS, where creative interventions are integrated into healthcare pathways to support treatment, rehabilitation and recovery (including social prescribing). Alongside these are targeted creative health activities designed to respond to specific health conditions or population groups, such as respiratory choirs or mental health recovery drama groups, which may be accessed both through referral pathways and through open, self-directed participation.

Creative health is also closely aligned with clinical arts therapies, which require a trained therapist and a psychotherapeutic relationship. Notwithstanding this, many creative health interventions operate through participatory and community-based mechanisms that emphasise social connection, agency and meaning-making (Fancourt & Finn, 2019; Jensen et al., 2024). Central to this approach is the recognition of **lived experience** as a form of expertise: people with direct experience of ill health, disability or social exclusion are understood as active agents in the co-design and co-delivery of creative health activities rather than being passive recipients of care (NCCH, 2023; APPG/NCCH, 2023).

A significant proportion of creative health activity exists beyond formal referral systems, embedded within communities, cultural institutions and the voluntary, community, faith and social enterprise (VCFSE) sector. This includes open-access, community-led and culturally rooted activity such as arts participation, heritage engagement and nature-based practices. These forms of provision play a vital role in supporting prevention, social connection and wellbeing at a population level. Together, these interconnected modes of delivery enable creative health to operate across both treatment-focused and preventative models. Furthermore, emerging evidence highlights the role of creative



health in addressing health inequalities through integrated care and neighbourhood-based approaches, targeting the wider determinants of health (Thomson and Chatterjee, 2024; Thomson et al., 2025; Thomson et al., 2026).
These Guidelines adopt a creative health framework aligned with the WHO classification of arts and health interventions across prevention, promotion, management and treatment (Fancourt and Finn, 2019), while articulating how these interventions function within an integrated health system. Creative health should therefore be understood as a continuum of health-promoting behaviour spanning everyday participation through to clinically embedded interventions (Fancourt et al, 2026), the implications for how it is understood, measured and integrated within health systems, needs to reflect how services are organised, resources allocated and outcomes evaluated (NHS England, 2025).

3.2. Creative health outcomes

These Guidelines start from the premise that the fundamental purpose of creative health is to improve people's lives and lived day-to-day experience. Creative health may achieve this through many different pathways, including improvements in physical and mental health, reduced loneliness, stronger social connection, increased confidence and agency, greater participation, purpose and meaning, and improved quality of life.

These outcomes should be understood as different dimensions of, or pathways towards, improvements in people's lived experience and wellbeing and they should not be treated simply as inputs into an economic valuation exercise. The first task of evaluation is to identify and measure the outcomes that matter to participants and that are plausibly generated by the intervention. Valuation is then a subsequent step. Once credible evidence of change and impact has been established, economic values can help decision-makers compare benefits with costs, but they do not determine which outcomes are important in the first place.

This approach allows the framework to evolve as the measurement of lived experience and wellbeing develops in creative health. Table 1 provides example health and non-health outcomes associated with creative health interventions, as documented in the literature. They will form the core of our economic evaluation framework as they represent the best currently practical approaches for economic evaluation, but they are not intended to imply that current measures provide a complete account of people's experiences. As stronger measures emerge, they can be incorporated within the framework.

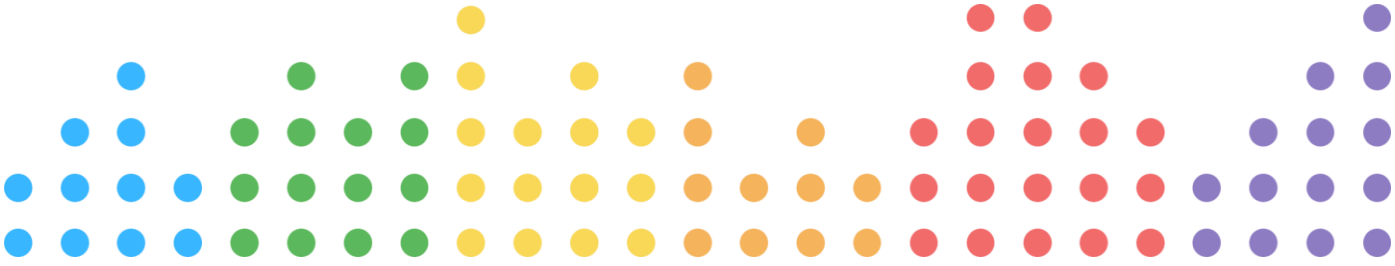


Table 1. Example outcomes associated with creative health interventions

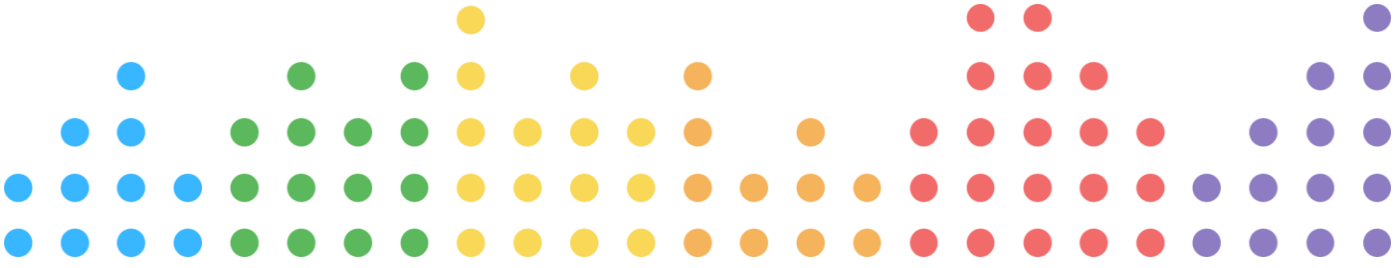
Health-related outcomes	Non-health-related outcomes
Mental health	Social outcomes
Reduced symptoms of depression and anxiety	Reduced social isolation and loneliness
Improved emotional regulation	Increased social connectedness
	Improved social cohesion
Physical health	Psychological & personal development
Improved physical activity levels	Increased self-confidence
Improved mobility and functional ability	Increased self-efficacy
Reduced risk factors for chronic disease	Increased sense of purpose and meaning
Clinical outcomes	Community outcomes
Improved management of long-term conditions	Strengthened community engagement
Reduced pain and fatigue	Increased social capital
Improved cognitive functioning (e.g. dementia support)	Improved community resilience
Quality of life	Skills & participation
Improved health-related quality of life	Development of creative skills
Improved subjective wellbeing	Increased cultural participation
	Increased volunteering and civic engagement
Health service outcomes	Economic & societal outcomes
Reduced GP visits	Increased productivity and employability
Reduced hospital admissions	Reduced absenteeism
Reduced healthcare utilisation	Contribution to local cultural economies

Sources: Adapted from APPG (2017), WHO (2019), Fancourt & Finn (2019), Fancourt, et al. (2020), and NCCH (2023).

3.3. Measurement of outcomes in creative health

A substantial and growing evidence base has examined the relationship between creative and cultural participation and health and wellbeing. Across this literature, creative health has been associated with outcomes including improved subjective wellbeing, reduced loneliness and social isolation, improved mental health, greater confidence and self-efficacy, stronger social connection, increased participation and a greater sense of purpose and meaning [1].

[1] Evidence can be found in a range of reviews including:
» Rapid review of wellbeing evaluation research using the WEMWBS scales
» What works to improve ONS4 personal subjective wellbeing?
» Creativity and wellbeing
» Arts and wellbeing: review of the social value of place-based arts interventions
» Music, Singing and Wellbeing in adults with diagnosed conditions or dementia summary insights



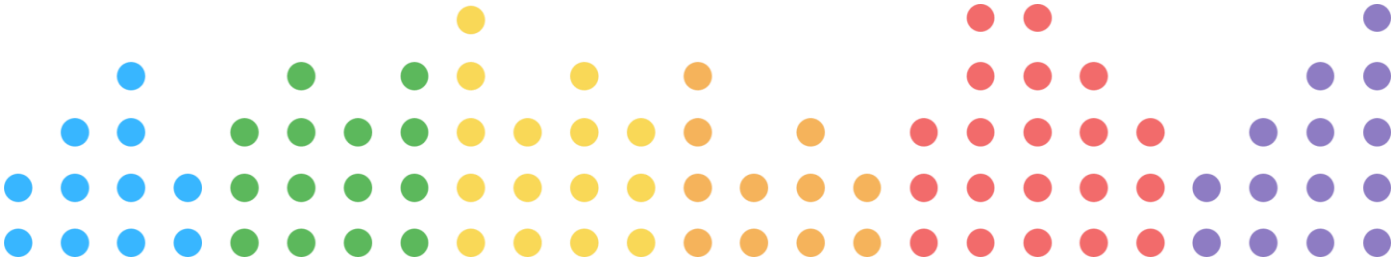
Evidence comes from a range of settings and creative activities, including music and singing, dance, visual arts, social prescribing and place-based cultural programmes. The evidence also suggests that creative health may generate these outcomes through multiple and interacting pathways. These include opportunities for social connection and belonging, creative expression, learning and mastery, increased confidence and agency, meaningful participation, physical activity, routine and purpose. The importance of these pathways will vary according to the intervention, population and context, and they should therefore be reflected explicitly rather than assumed to operate uniformly across creative health.

Existing evidence has an important role in economic evaluation. It should inform the development of the Theory of Change, the selection of outcomes and measures, and judgements about the plausibility of expected impacts. It is particularly important for ex-ante evaluation, where estimates of future benefits necessarily depend on evidence from previous studies, evaluations or comparable programmes. Where robust existing evidence is unavailable, evaluation should be designed not only to inform the immediate decision but, where feasible, to strengthen the wider evidence base for creative health.

A range of attempts have been made to measure the types of outcomes in Table 1. Empirical studies of arts participation and community arts programmes have highlighted improvements across wellbeing, social inclusion, and personal development outcomes, illustrating the need for measurement frameworks that extend beyond traditional clinical indicators (Jensen et al., 2024).

Some studies have attempted to assess the economic value of creative health interventions using frameworks commonly applied in health economics. A systematic review of economic evaluations of arts and creativity interventions for older adults identified a small number of studies employing quality-adjusted life years (QALYs) derived from instruments such as the EQ-5D, alongside Social Return on Investment (SROI) approaches (Crealey et al., 2023). In part, this limited use reflects concerns that QALYs are not well suited for capturing the social and psychological outcomes central to these interventions and that the overall evidence base is methodologically heterogeneous, making cross-study comparisons difficult (Crealey et al., 2023).

To better reflect these broader outcomes, reviews of arts and health research document that evaluations commonly employ validated wellbeing scales alongside measures of health status, quality of life and healthcare utilisation in order to reflect the multi-dimensional nature of these interventions (Fancourt & Finn, 2019; Fancourt, et al., 2020). However, although wellbeing measures are generally better aligned with creative health's outcomes than generic health measures, their standardised nature means they



may not be sufficiently sensitive to the context-specific outcomes associated with different activities, populations, and places.

One of the most widely used instruments in creative health research is the Warwick–Edinburgh Mental Wellbeing Scale (WEMWBS), which assesses subjective mental wellbeing across domains such as positive affect, social functioning, and sense of purpose. Despite its prevalence, the sector reports that its standardised nature may misrepresent the changes experienced by participants (de Andrade and Angelova, 2018) and that the scale is challenging to implement for certain health conditions and with certain population groups, particularly where cognitive and communicative barriers are present. Furthermore, for communities facing structural constraints, decontextualised measures may not adequately capture the social and material conditions shaping their responses.

These limitations identified by the sector are accompanied by methodological challenges for economic evaluation. WEMWBS is not preference-based and does not readily translate into the economic units required for CBA, prompting calls for more complex, multi-dimensional wellbeing models (Jensen et al., 2024).

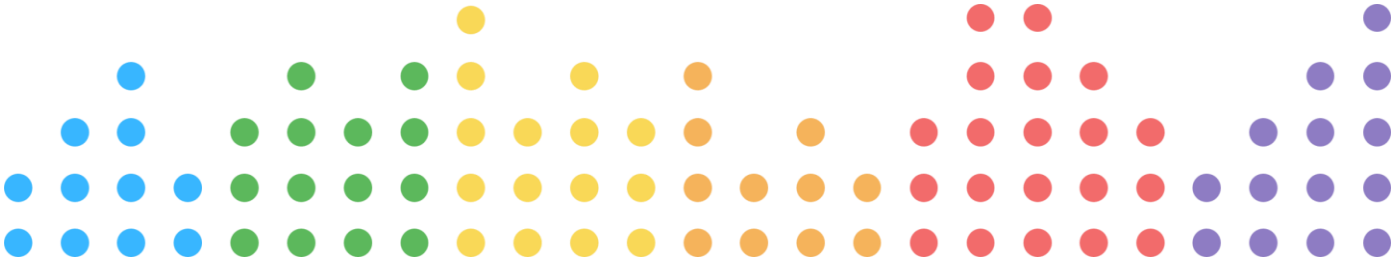
One response has been the development of approaches that monetise wellbeing outcomes directly, using scales such as the WELLBY framework to generate estimates comparable across health and non-health domains. This approach has been applied in assessments of the economic impact of arts and cultural engagement on NHS costs and wider population health outcomes in both Wales (Centre for Health Economics and Medicines Evaluation & Research Centre for Arts and Wellbeing, 2026) and England (Frontier Economics, 2024), representing a methodological step towards more comprehensive economic appraisal of creative health interventions.

4. Guidelines on economic evaluation for creative health

The aim of these guidelines is to align the appraisal of creative health with best-practice principles of economic evaluation set out in the Green Book and its supplementary guidance to ensure that creative health proposals and programmes can be assessed on the same analytical footing as other public investments.

The process and steps for conducting an economic evaluation in creative health are:

i. Develop a **Theory of Change** model that explains how the inputs and outputs lead to desired outcomes in creative health and who the main beneficiaries are.
ii. Measure the **impacts** of the intervention on the outcomes.
iii. Estimate the **cost and benefits** required to conduct **Cost-Benefit Analysis (CBA)**, or **Cost-Effectiveness Analysis (CEA)**. This includes key topics such as



the valuation of creative health outcomes; accounting for risk and uncertainty; distributional analysis; discounting; and optimism bias.

iv. **Presentation** of key economic evaluation **results** such the **NPSV**, **BCR**, and **CER**.

We have developed a **Template for Economic Evaluation in Creative Health** that sets out a standardised reporting format and guides analysts, practitioners and other stakeholders through this process. The Template is set out in Section 8 and comes with a set of standardised unit value metrics for key outcomes related to creative health.

The guidelines are intended for a wide audience covering creative health practitioners and providers, health care practitioners, policy analysts, economists, government departments and arms-length bodies and hence are non-technical in nature. Technical aspects will be covered in Annexes or referenced in external sources.

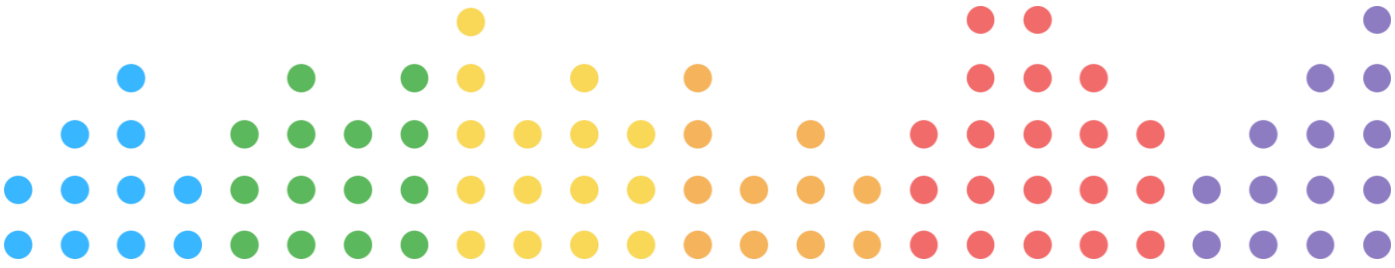
4.1. Developing a Theory of Change

The development of a Theory of Change (ToC) should begin by clearly identifying the problem or need that the intervention is intended to address and the wider strategic priorities to which it contributes. For commissioned creative health interventions, this may include priorities such as reducing urgent care demand, supporting hospital discharge, addressing health inequalities, improving the management of long-term conditions, or meeting priorities for specific population groups. Evaluators should also consider how the intervention fits within existing health, care and community pathways and interacts with other services or interventions. For example, a creative health intervention may form part of a social prescribing pathway, neighbourhood health model, or wider programme of support rather than operating as a standalone activity. The ToC model should make these relationships explicit and demonstrate how the creative health intervention contributes to the wider system objectives.

Having identified the problem, strategic context and role of the intervention within the wider system, the next step is to set out the **intended outcomes** of the intervention. Key outcomes with regards to creative health have been summarised in Table 1 above.

A **Theory of Change (ToC)** model sets out the relationship between the **inputs**, **outputs** and **outcomes** of an intervention, defined as:

Inputs are the resources required to deliver the intervention. Examples in creative health include funding for intervention; creative practitioners such as artists, musicians or facilitators; healthcare or community staff; venues such as community centres, museums or healthcare settings; materials and equipment for creative activities; and partnerships with cultural organisations, charities or health services.



Outputs are the direct activities delivered using these inputs. In creative health interventions, outputs may include structured creative sessions (for example music groups, dance classes, art workshops or theatre projects); the number of sessions delivered; the number of participants engaged; social prescribing referrals to creative interventions; and opportunities for participants to produce or exhibit creative work.

Outcomes are the changes that occur as a result of participation. Where these outcomes are attributable to the intervention and generate positive social value, they constitute benefits for the purposes of economic appraisal or evaluation. For CBA, the relevant benefits should encompass the financial and non-financial impacts arising from the intervention. For CEA, the analysis focuses on a specified measure of effectiveness rather than monetising the full range of benefits. For NHS commissioners, this may be a measure of healthcare utilisation or health-related quality of life. The ToC should therefore clearly identify the pathway through which the creative health intervention is expected to affect the selected primary outcome.

Throughout these Guidelines, 'outcomes' therefore refers to the changes experienced by participants or others, while 'benefits' refers to the positive social value arising from attributable outcomes.

A ToC should be developed based on **good evidence** which can come from a range of possible sources such as evaluations of previous interventions; academic research; commissioned research; and international comparisons.

Creative health interventions may have an important role in reducing health and wider socioeconomic **inequalities**, particularly where they engage groups experiencing poor health, social isolation, deprivation or barriers to accessing conventional health and cultural services. Evaluators should therefore consider not only the average impact of an intervention, but also who participates and who experiences the benefits.

The ToC should identify whether reducing inequalities is an explicit objective of the intervention and specify the population groups expected to benefit. Where sample sizes and data permit, evaluators should report participation, retention and outcomes by relevant population characteristics, such as socioeconomic disadvantage, age, disability or geographic deprivation. This can help identify whether an intervention is reaching underserved groups and whether impacts differ across the population.

Where reducing socioeconomic inequalities is a material objective, analysts may also consider applying the Green Book distributional weighting approach described in Section 5.

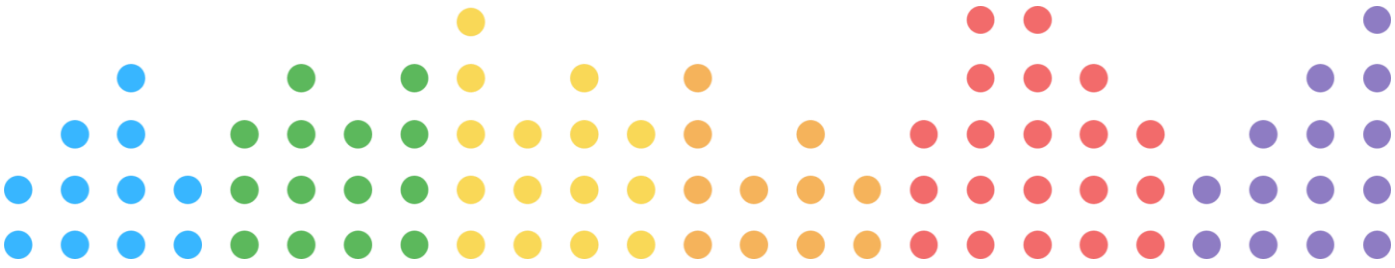
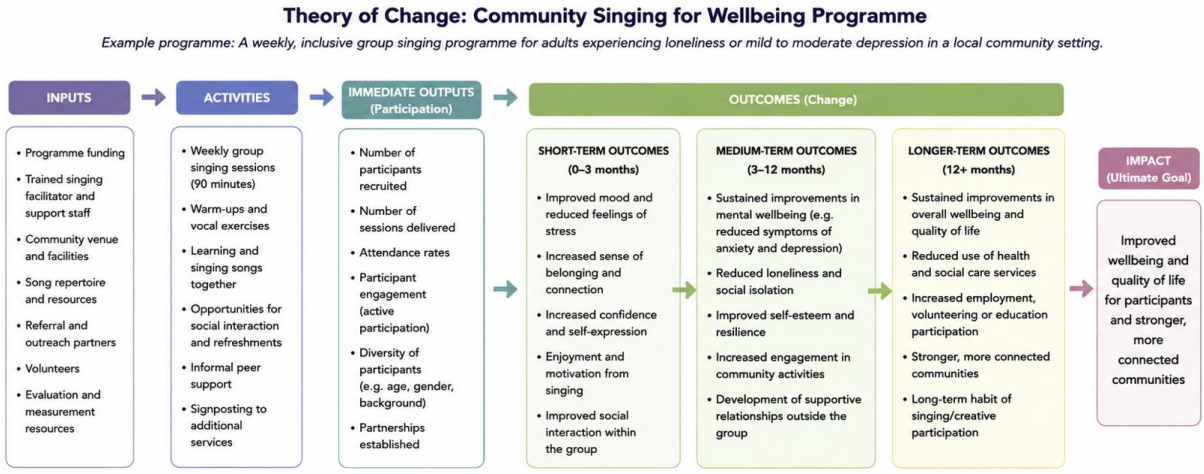


Figure 2 provides an illustrative ToC model for a hypothetical community singing programme. It is intended to demonstrate how the inputs, activities, outputs and outcomes of a specific creative health intervention can be mapped and linked through a clear causal pathway. Evaluators should develop a ToC that reflects the specific activities, participants, objectives and expected outcomes of their own intervention.

Figure 2. Example Theory of Change model for a Community Singing Programme



Alongside outcome measurement, evaluators should collect and report a small set of **implementation and delivery metrics**. These should include the **number of people referred or eligible to participate** where relevant, **programme uptake**, **attendance** or **engagement**, **programme completion** and **dropout**. These measures are not outcomes and should not be monetised within the economic evaluation, but they are important for interpreting impact estimates and understanding how an intervention operates in practice. For example, a programme may generate substantial benefits for participants who complete it but have low uptake or high dropout, which may limit its overall population impact and scalability. Reporting these metrics alongside outcome and economic results provides important context for decision-makers and can identify where changes to programme delivery may improve effectiveness.

4.2. Measuring impacts

Measuring impact refers to understanding the effect that is attributable to the creative health intervention. The key to this is to measure what would have happened without the intervention, which is known as the **counterfactual**. We distinguish between **deadweight** (outcomes that would have occurred without the intervention) and **additionality** (outcomes causally created by the intervention).

Understanding the counterfactual to arrive at an estimate of additionality is a critical part of the process of economic evaluation in creative health and, as discussed, it is equally relevant and important to both ex-ante and ex-post evaluations.

Whilst the intention of creative health is to improve outcomes such as health, wellbeing and loneliness, it cannot be deduced that any observed changes in such outcomes amongst participants are *wholly* due to creative health interventions. This is because a range of other factors could plausibly explain some of the observed changes, including, for example, that some of the participants may have experienced improvements in these outcomes anyway (this is especially the case for more motivated participants); their health may have been on an improving trend before starting the creative health intervention; and some participants may also have been involved in other health initiatives at the same time.

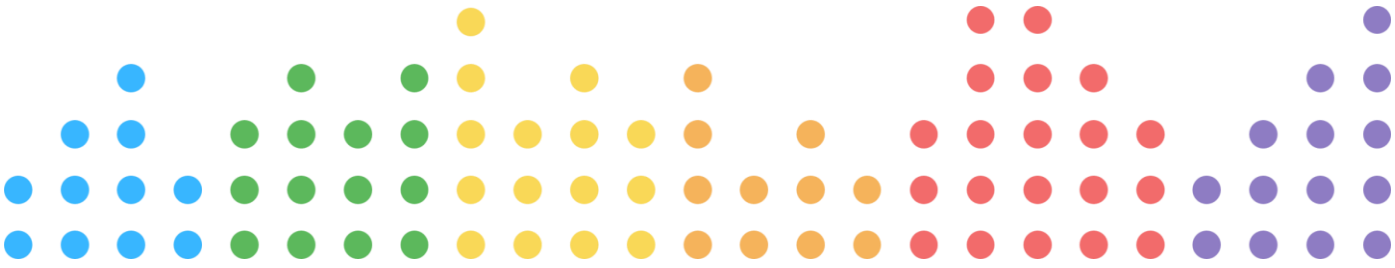
The key question for creative health is to what extent does an Output such as dance classes lead to an Outcome such as improved mobility and mental health, or reduced health resource utilisation?

To assess causality, evaluators must consider the counterfactual, but this scenario cannot be directly observed as the intervention either takes place or it doesn't. Therefore, the counterfactual must be approximated using a suitable **control group**. The strongest evidence of causality is generated when the outcomes for participants receiving an intervention can be compared with those from a group of very similar individuals who do not receive it.

The **Magenta Book** (2019) provides HM Treasury's central guidance on how to design and implement robust evaluations of government interventions using control group analysis. It stresses that estimating the counterfactual through control groups is essential for determining additionality. The guidance distinguishes between process, impact, and value-for-money evaluations, highlighting that economic evaluation depends critically on impact evaluation.

The credibility of CBA and CEA depends directly on the quality of the underlying impact evidence. Stronger experimental and quasi-experimental methods provide greater confidence that estimated outcomes are causally attributable to the intervention, whilst weaker evaluation designs produce correspondingly more uncertain economic evaluation results.

The Magenta Book has developed **Quality in Policy Impact Evaluation (QPIE)**, a framework which focuses on the **strength of evaluation designs in establishing causal impact**. QPIE presents a **hierarchy of evaluation designs** based on their ability to estimate the counterfactual and produce robust causal impact estimates.



At the top of the hierarchy are **Experimental designs**, particularly **Randomised Controlled Trials (RCTS)**, which minimise bias by ensuring that intervention and control groups are statistically equivalent.

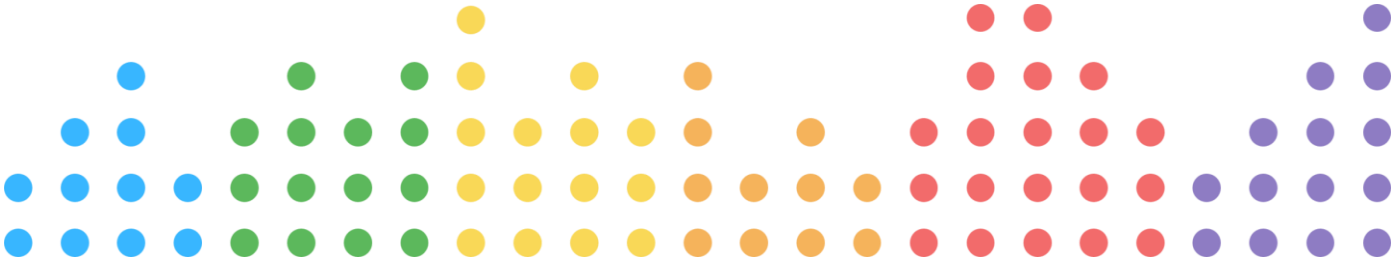
Where randomisation is not feasible, well-designed **Quasi-experimental methods**, such as **Matching**, **Difference-in-Difference** designs, or **Regression Discontinuity** approaches, can also provide causal estimates. Quasi-experimental methods attempt to mimic what happens in an RCT by controlling for as many differences as possible between the intervention and control groups. This can take the form of statistical control methods such as regression analysis or matching whereby the control group is designed to closely match the intervention. However, the downside of quasi-experimental methods in relation to RCTs is that it is only possible to match people based on observable factors such as age, education and income whilst important difficult-to-measure factors, such as ability or motivation often cannot be observed and hence there is a risk that the two groups are not closely matched on all key aspects.

While generally less robust than RCTs, well-designed quasi-experimental methods can provide strong evidence of causality and are often the most practical approach for evaluating real-world policies and programmes.

Lower down the QPIE hierarchy are designs that do not explicitly include counterfactual analysis, such as **unmatched comparison groups** (whereby the intervention group's outcomes are compared against the outcomes of another group that has not been statistically matched. This method is known as the 'Naïve Estimator' in statistics) and **simple before-and-after comparisons** (whereby the change in outcomes observed pre and post intervention is taken as the impact of the intervention). These approaches have been shown to be highly problematic and provide little confidence in causal attribution because they do not adequately account for the counterfactual, often overstating the impact.

Robust causal assessment therefore requires a clearly defined control group and the greater the confidence that the control group represents what would have happened in the absence of the intervention, the greater the confidence we can have that the difference in observed outcomes between the intervention and control groups can be attributed to the intervention itself.

In creative health, impact measurement should prioritise the ability to generate **credible and quantifiable impact estimates** whilst recognising the **proportionality** issue of size, scale and budget of the interventions.

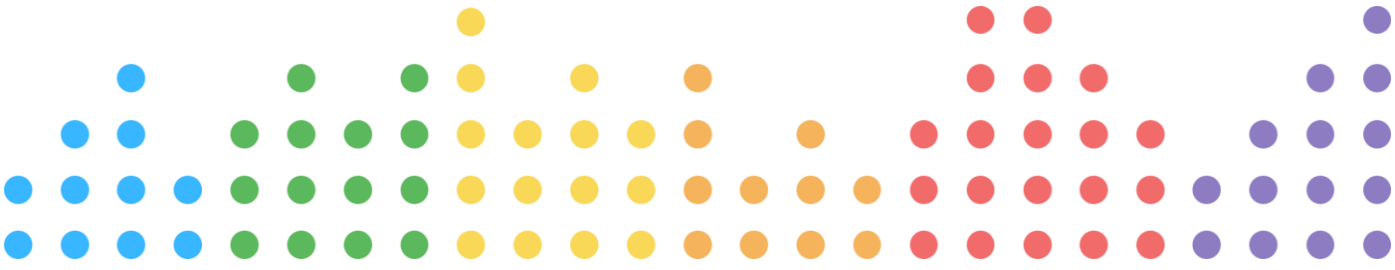


Our recommendations on impact measurement depend on the level of resource available and the size of the intervention, broadly differentiated by large- and small-scale interventions. The definition of the scale of the intervention will depend on a range of factors such as participant numbers; level of investment; geographic reach; maturity of the intervention; strategic importance; and intended use of the evidence.

In these Guidelines, therefore, the minimum evidence standard depends on the scale, strategic importance and intended use of the evaluation: larger interventions intended to inform significant commissioning, investment or policy decisions should normally use an experimental or appropriate quasi-experimental design incorporating a credible counterfactual. Where this would be disproportionate for smaller interventions, the minimum approaches set out below should be used. These provide a pragmatic basis for economic evaluation but should not be interpreted as providing equivalent causal evidence.

Evaluators should also consider the potential for **displacement**. This occurs when benefits generated by an intervention are offset by reductions in activity or benefits elsewhere. For example, a participant may attend a creative health programme instead of participating in another community activity that would have generated similar benefits. In many creative health interventions displacement is likely to be relatively small, particularly where programmes engage individuals who would not otherwise participate in alternative activities. Nevertheless, evaluators should consider whether there is evidence that participation in the intervention has displaced engagement in alternative activities that would have generated similar benefits. Where displacement is considered likely to be material, evaluators should estimate its scale using participant surveys, qualitative evidence, or evidence from previous evaluations and deduct the displaced proportion from the estimated impact. Any displacement assumptions should be clearly reported and justified. In the absence of evidence suggesting otherwise, a displacement rate of zero may be assumed.

Furthermore, creative health interventions may be delivered alongside other forms of support, including social prescribing, healthcare services, community support programmes, employment services and voluntary sector activities. Consequently, observed outcomes may sometimes be influenced by multiple interventions operating simultaneously. Where evaluators believe that other interventions have contributed materially to the outcomes achieved, they should consider **attribution** within the analysis and avoid assuming that all observed impacts are solely due to the creative health intervention. Attribution may be assessed through evaluation design, statistical analysis, or participant surveys. Any assumptions regarding attribution should be clearly documented and justified. Where no evidence on attribution is available, evaluators should adopt a cautious approach and explicitly acknowledge the possibility



that some observed outcomes may have been influenced by other interventions or services operating at the same time.

Practical considerations for data collection

Data collection should, wherever possible, be integrated into the routine delivery of creative health interventions rather than undertaken as a separate evaluation exercise. Evaluators should seek to collect the outcome measures during normal contact with participants, for example at enrolment, during programme sessions and at the end of the intervention. In-person data collection during existing contacts is likely to reduce participant burden and improve completion compared with relying solely on separate follow-up surveys. Data collection should be kept as short and simple as possible, particularly where participants may experience poor health, low confidence or other barriers to survey completion.

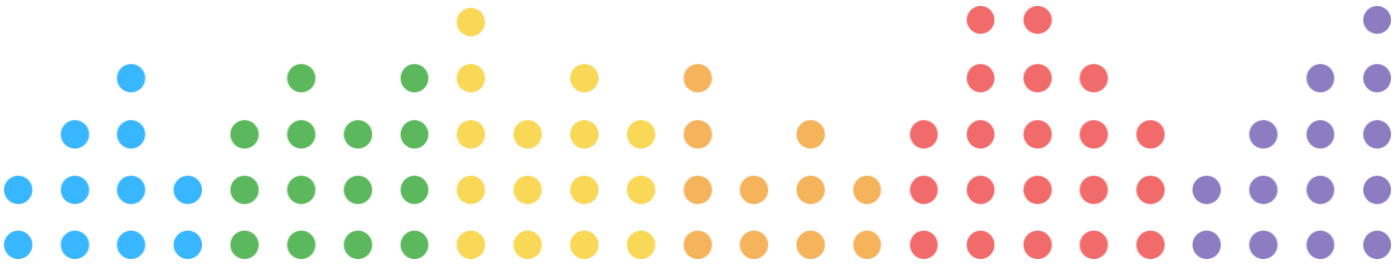
Evaluators should monitor response and follow-up rates and report these transparently. Particular attention should be paid to whether participants who complete follow-up measures differ systematically from those who do not, as this may introduce attrition bias and reduce the representativeness of the findings. Where feasible, evaluators should compare the characteristics of respondents and non-respondents and consider the implications for the interpretation of results.

Where creative health interventions are delivered in partnership with primary care or other NHS services, commissioners and providers should also consider whether participation in the intervention can be recorded within routine data systems, subject to appropriate information governance and data protection requirements. This may create opportunities for future linkage to healthcare utilisation data and enable more robust evaluations.

i. Large-scale creative health interventions

Large-scale creative health interventions are typically established interventions that operate across multiple sites, organisations, local authorities, Integrated Care Systems (ICSs), or regions and are intended to influence commissioning, service delivery, or public policy. These interventions typically involve substantial investment from NHS organisations, local authorities, national funders, or government and often serve large numbers of participants across multiple communities. Their purpose extends beyond local service delivery to informing strategic decisions about future investment, scaling, or commissioning of creative health services.

Examples include region-wide Arts on Prescription programmes operating across an Integrated Care System, national social prescribing initiatives incorporating creative



health pathways, large-scale museum or heritage programmes supporting health and wellbeing across multiple locations, and major place-based creative health programmes funded by organisations such as Arts Council England, AHRC, NHS England, or UKRI.

Since the evidence generated may influence significant resource allocation decisions, these interventions should be expected to meet higher standards of evaluation, including stronger approaches to control group analysis and transparency regarding assumptions and uncertainty.

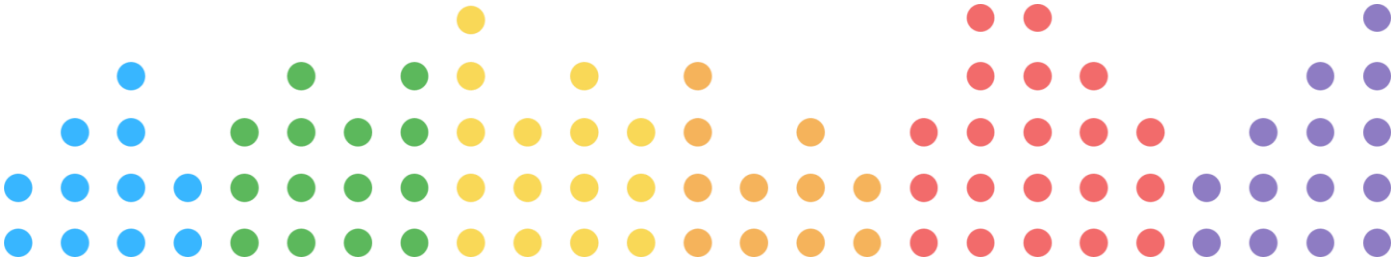
Large-scale creative health interventions and activities should use **statistical methods that incorporate counterfactual analysis through a well-designed control group.** The closer the match with the control group, the more robust the impact estimates are.

To deliver these methods will require analysts trained in statistics and econometrics and primary data collection.

It should be noted that for larger-scale interventions, the use of a robust experimental or quasi-experimental evaluation does not remove the need for proportionate and timely measurement during programme delivery. These Guidelines recommend that larger interventions also apply the minimum outcome measurement framework set out below for smaller-scale interventions, alongside the more robust evaluation design. This can provide earlier evidence on changes in the core outcomes and support learning, course correction and decisions regarding continuation or funding before the final evaluation results are available. Applying a consistent minimum set of measures across interventions may also improve comparability of results across the creative health sector.

Results from this interim measurement should be interpreted as monitoring and early learning evidence and should not be presented as equivalent to the impact estimates generated by the robust evaluation. Where the two approaches produce different findings, the results of the more robust experimental or quasi-experimental evaluation should be used as the primary basis for estimating impact and undertaking the final economic evaluation.

Annex A sets out descriptions of statistical methods that can be applied to measure impact incorporating control group analysis. As mentioned above these are known as **Experimental and Quasi-experimental statistical methods**. Further details on these methods can be found in the Magenta Book.



Reporting standards and robustness checks
When using quasi-experimental methods, evaluators should provide sufficient information to enable readers to assess the credibility of the counterfactual and the robustness of the resulting causal estimates. The evaluation report should clearly describe the intervention, target population, outcomes measured, data sources, analytical methods, and rationale for selecting the chosen evaluation design.

A key requirement is **transparency** regarding the **assumptions**. Evaluators should explicitly state the assumptions required by the chosen method and discuss their plausibility. For example, DiD analyses require the assumption that intervention and control groups would have followed parallel trends in the absence of the intervention. PSM and regression assume that all important factors influencing participation and outcomes have been observed and included in the analysis.

Sensitivity analysis should be undertaken wherever feasible to assess whether conclusions are robust to alternative analytical choices. Examples include:

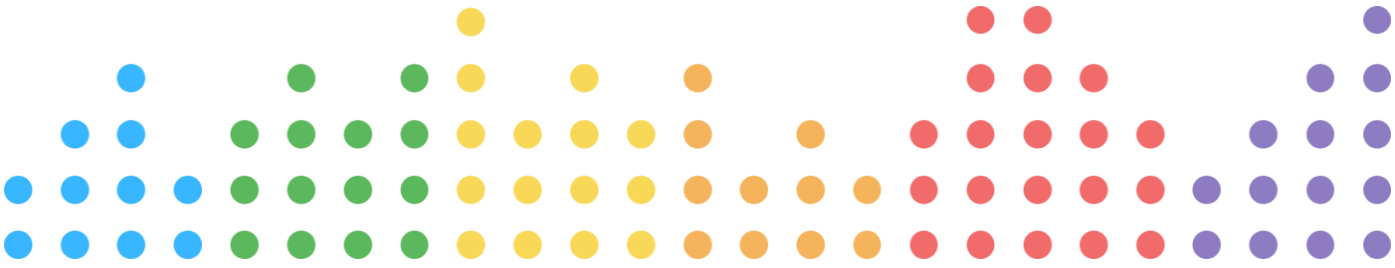
- Using alternative matching algorithms or matching specifications.
- Testing alternative regression model specifications.
- Using different control groups.
- Examining the influence of outliers or extreme observations in the data.
- Assessing sensitivity to missing data assumptions.

Results should be reported with appropriate measures of **statistical uncertainty**, including **confidence intervals** and **standard errors**. The objective is not to eliminate uncertainty (as that is not possible), but to ensure that decision-makers understand its implications.

Given the uncertainty that may arise at multiple stages of creative health evaluation, evaluators should also present a central estimate together with plausible **lower and upper bound estimates** of the results wherever feasible. These scenarios should reflect the combined uncertainty in the assumptions most likely to drive the results, including impact estimates, deadweight or counterfactual assumptions, the duration and persistence of benefits, and healthcare utilisation impacts. The lower bound should apply credible conservative assumptions and the upper bound should represent a plausible favourable scenario rather than a theoretical maximum. The three estimates should be reported together so that decision-makers can assess the reasonable range of potential results.

ii. Smaller-scale creative health interventions

Smaller-scale creative health interventions are typically local, community-based interventions delivered by voluntary, community, cultural, or charitable organisations to



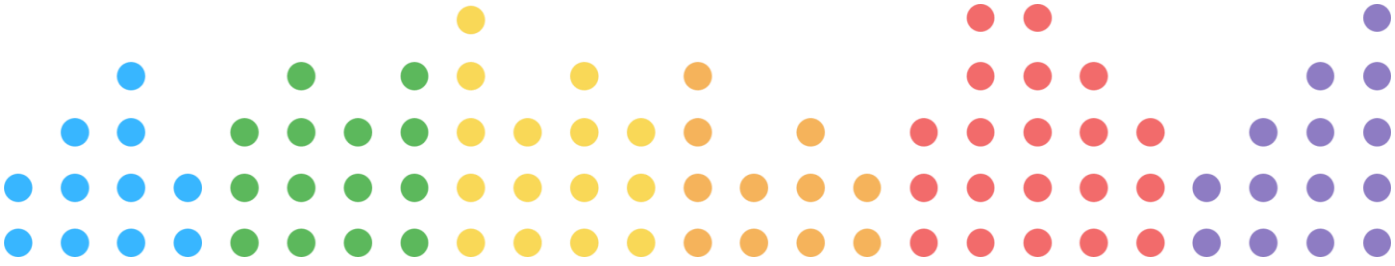
support health and wellbeing outcomes. These interventions are often delivered within a single locality, healthcare setting, cultural venue, or community organisation and are frequently funded through short-term grants, local authority funding, charitable foundations, social prescribing budgets, or small NHS contracts. Examples may include a community choir for people experiencing loneliness, an arts-on-prescription programme operating within a single primary care network, a museum-based programme for people living with dementia, creative workshops for young people experiencing anxiety, or an arts programme supporting people with long-term health conditions.

Consistent with the principle of **proportionality** set out in the Green Book, small-scale interventions should not be expected to undertake the same level of evaluation as large-scale interventions intended to inform major commissioning, investment, or policy decisions. In such cases, a combination of outcome measurement, qualitative evidence, theory-based evaluation, and basic economic analysis may provide sufficient evidence to support learning, service improvement, and local decision-making. More resource-intensive approaches, such as experimental or advanced quasi-experimental methods, should generally be reserved for larger interventions where the resulting evidence is expected to influence substantial funding, commissioning, or policy decisions.

The principle of **proportionality** implies that the scale and complexity of the evaluation should be proportionate to the size, maturity, and strategic importance of the intervention.

Wherever feasible, evaluators should seek to use experimental or quasi-experimental approaches to generate robust estimates of impact even for smaller-scale creative health interventions, but for many small-scale interventions, the costs, data requirements, and sample sizes associated with these methods may be disproportionate to the scale of the intervention and the decisions being informed, which may mean that these methods are not practical or proportionate.

The issue that we have observed in the creative health sector is that data on control groups is almost never collected and often it may not be feasible to do so for these types of interventions. We have therefore produced guidance for impact measurement under these circumstances, which is set out below. These can be taken as the **minimum standards of analysis** required to measure outcomes in creative health for the purposes of economic evaluation. **A fundamental requirement in these minimum standards is that before and after data on outcomes for participants is collected**.



The approaches set out below are intended as **pragmatic solutions** for resource-constrained evaluation settings where a robust control group or more rigorous counterfactual design is not feasible.

These methods should not be interpreted as providing the same strength of causal evidence as experimental or quasi-experimental methods.

In particular, the pre-post approach with counterfactual adjustment provides a relatively crude approximation of the outcomes that may have occurred without the intervention. However, it provides a more structured and conservative approach than assuming that all observed pre-post changes are attributable to the intervention or making no adjustment for the counterfactual at all.

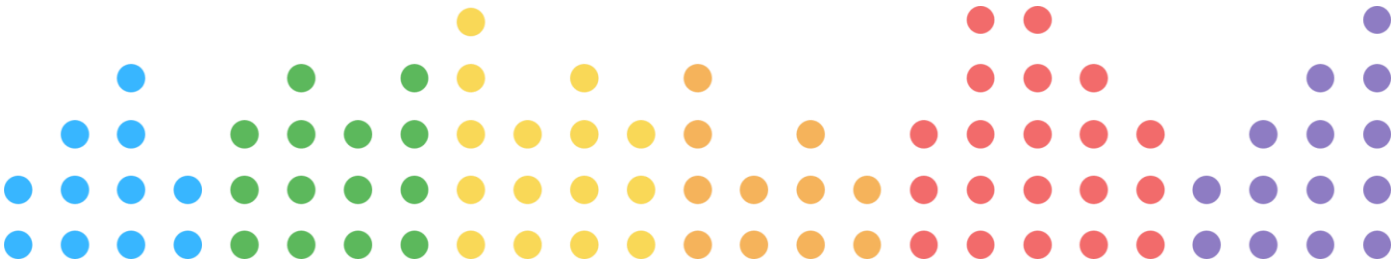
Minimum standards for impact measurement

Where experimental or quasi-experimental methods involving control group analysis are not feasible, these Guidelines recommend two minimum approaches for estimating programme impact.

1. Preferred Approach: Interrupted Time Series Analysis
Where interventions are delivered repeatedly over time, evaluators should use Interrupted Time Series (ITS) analysis (see Annex A for a description of the ITS methodology). This approach is particularly suitable for interventions that are introduced and withdrawn at regular intervals, such as holiday-based programmes for young people, seasonal arts interventions, or programmes delivered in repeated annual cycles. By examining whether outcomes consistently improve when the intervention is introduced and subsequently deteriorate or stabilise when the intervention is withdrawn, evaluators can generate evidence regarding whether observed changes are attributable to the intervention.

ITS evaluations should seek to include at least three intervention cycles, with outcome measurements collected before, during, and after each cycle. This is to ensure that the impacts of the intervention can be isolated from other impacts and trends. Data may be collected from the same individuals over time or from repeated samples drawn from the same underlying population. Where repeated samples are used, the samples should be broadly similar at the start of each intervention cycle and random sampling would provide the best method for doing so wherever practical.

The analysis should test for statistical changes in the outcomes before and after each intervention cycle and should present estimated effect sizes, confidence intervals, p-values, and the level of statistical significance. Sample sizes will need to be sufficient to



enable this type of statistical analysis. Evaluators should also discuss whether other events occurring at the same time could plausibly explain the observed changes.

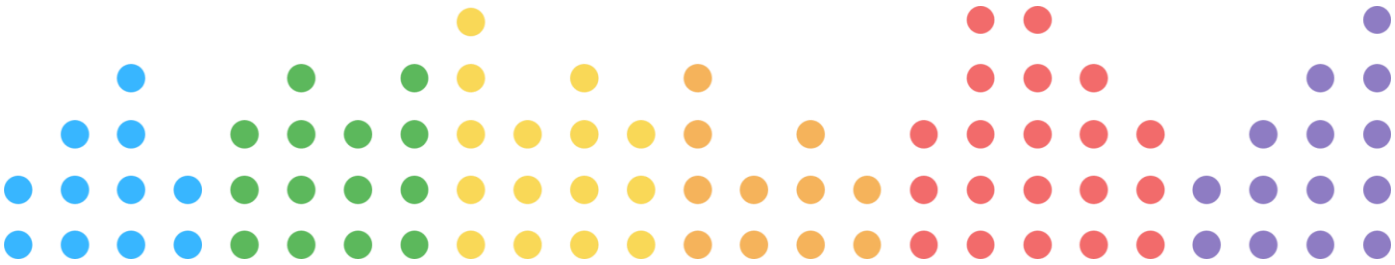
2. Alternative Approach: Pre-post measurement with counterfactual adjustment
Where interrupted time series analysis is not feasible, evaluators should undertake a structured pre-post assessment of outcomes. This should involve collecting outcome data immediately before and after participation and testing whether observed changes are statistically significant. Effect sizes and confidence intervals should also be reported.

Because this approach does not directly observe a counterfactual, evaluators must incorporate explicit adjustments to account for the possibility that some of the observed changes would have occurred in the absence of the intervention. This is a two-step process.

i. **Participant attribution.** Firstly, participants should be asked to what extent they believe their reported changes were attributable to the intervention. Although this type of method is susceptible to a range of recall biases, there is some empirical support for participant-attribution approaches. Studies of counterfactual self-estimation have found participant-derived estimates to be reasonably close to quasi-experimental estimates for attitudes and behavioural intentions (Mueller & Gaus, 2015), with similar findings across multiple experiments (Mueller & Gaus, 2018). Evidence from long-term care also showed broad consistency with instrumental-variable estimates (Malley et al., 2019). Therefore, the following type of question should be asked of all participants experiencing an improvement: "There are many reasons why you may have experienced an improvement in [outcome X]. Taking account of any other support or changes in your life, do you think you would have experienced this improvement without taking part in the programme?"

   If, for example 100 participants report an improvement, but 20 say they believe the improvement would have occurred anyway, then deadweight is 20% and additionality is 80 people.

ii. **Deadweight factors comparison.** Participant attribution assessments cannot be relied upon as the sole estimate of additionality because participants may not accurately know what would have happened in the absence of the intervention. Furthermore, responses may be influenced by social desirability bias (participants may overstate programme impacts because they want to provide favourable responses or please programme providers and survey administrators) and cognitive dissonance (participants may exaggerate programme benefits to justify the time, effort, or commitment they invested in participation in order to look like a 'rational'



person to the programme providers and survey administrators). This tends to create a downward bias in participant estimates of deadweight (i.e. they are understated).

The Homes and Communities Agency (2014) *Additionality Guidelines* was developed to support appraisal and evaluation in line with HM Treasury's Green Book. Its primary purpose is to help evaluators estimate the extent to which observed outcomes are genuinely attributable to an intervention and would not have occurred in its absence. The guide provides a set of benchmark deadweight factors derived from previous evaluations and empirical research. These estimates were compiled from large-scale evaluations of regeneration, economic development, housing, health, community, and social programmes, including the City Challenge evaluation and the Neighbourhood Renewal Fund evaluation. Reported deadweight rates varied substantially by intervention type, reflecting differences in context and programme design.

The guide provides several benchmark deadweight estimates that are relevant to creative health interventions. Evidence from the **City Challenge evaluation** suggests that health interventions have an average deadweight rate of approximately **27%**, meaning that around one-quarter of observed outcomes may have occurred in the absence of the intervention. **Community and social interventions** have a lower estimated deadweight rate of around **19%**, while the **Neighbourhood Renewal Fund** evaluation reports a deadweight estimate for health interventions averaging **18%.** Taken together, these findings suggest that deadweight rates for interventions most closely aligned with creative health programmes typically fall within the range of 18–27%, with a **mid-point of 22.5%**.

To derive the **final deadweight figure to be used in the evaluation** the deadweight figure estimated from the participant attribution approach in (i) should be compared with the benchmark deadweight factor of 22.5% (obtained from previous evaluations in (ii)) and to ensure a **conservative assessment of impact**, the **higher of the two estimates** should be used.

The **selected deadweight factor should then be deducted from observed outcomes to derive an adjusted estimate of programme impact**. Evaluators should clearly report the participant attribution estimate, the benchmark estimate and the final deadweight assumption.

Developing the example above - where through participant attribution we estimated that deadweight is 20% - the final deadweight figure used would be 22.5% (as it is higher than the participant attribution figure) and the final impact would be that 77.5 people experienced an improvement in the outcome.

The same approach can be applied to continuous outcome measures, such as mental health or social connectedness scales. In these cases, the evaluator would first estimate the average change in the outcome between baseline and follow-up. The selected deadweight estimate would subsequently be applied to the observed change in the outcome measure to derive an estimate of additional impact.

For example, if participants' average mental wellbeing score increased by 12 points and the final deadweight estimate was 36%, the estimated additional impact attributable to the intervention would be 7.68 points per person. This approach enables the methodology to be applied consistently across both binary outcomes (e.g., whether an improvement occurred) and continuous outcome measures.

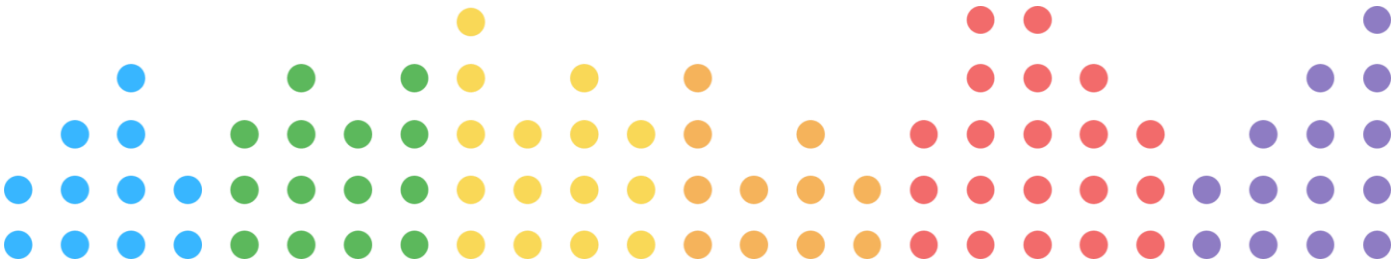
Where possible (i.e. where sample sizes are large enough), evaluators should test whether the observed change in outcomes between baseline and follow-up is statistically significant. **Statistical significance testing** should be undertaken on the observed outcome change prior to the application of any deadweight adjustments. Evaluators should report the mean change in outcomes, the relevant statistical test used, p-values, confidence intervals, and effect sizes where appropriate. However, it is recognised that in some cases participant numbers will not be sufficient to measure statistical significance. Failure to achieve statistical significance does not necessarily imply absence of impact, particularly where sample sizes are small.

As the evidence base for creative health continues to grow, benchmark deadweight estimates should increasingly be derived from previous creative health evaluations rather than from broader health, community, or social intervention evidence. Over time, the accumulation of robust evaluations will enable the development of sector-specific deadweight benchmarks that better reflect the characteristics, populations, outcomes, and delivery models associated with creative health interventions. This will improve the accuracy and credibility of additionality estimates. Evaluators should therefore seek to contribute to the development of this evidence base by transparently reporting deadweight assumptions, participant counterfactual estimates, and evaluation findings wherever possible.

Examples of evaluation studies using quasi-experimental methods and the pre-post measurement with the counterfactual adjustment method for small-scale interventions are provided below.

Use of existing evidence

Where these minimum standards for impact measurement cannot reasonably be achieved, **existing evidence** may provide a pragmatic basis for estimating likely programme impacts. This should be regarded as a fallback approach rather than an



equivalent substitute for primary impact analysis, and the resulting estimates should therefore be interpreted with appropriate caution.

In these circumstances, where an appropriate impact estimate is available from existing research, it may be used provided that the underlying evidence meets the relevant standards of causal inference set out in these Guidelines and is sufficiently applicable to the intervention being appraised. Evaluators should assess both the **quality and transferability** of the evidence, including the robustness of the original impact estimate and the comparability of the intervention, population, setting, dosage and frequency of engagement. Any assumptions required to transfer the evidence to the intervention being appraised - including assumptions about participation, duration and persistency of effects - should be stated explicitly and, where material, tested through sensitivity analysis. Where these requirements are not met, existing evidence should not be used as a basis for estimating programme impacts.

The strength of the conclusions drawn from an economic evaluation should reflect the strength of the underlying impact evaluation design. Experimental and well-designed quasi-experimental methods can support causal conclusions regarding the impact of an intervention. By contrast, pre-post analysis with counterfactual adjustments does not establish causality to the same standard because the counterfactual is not directly estimated using a control group. Results generated using the minimum pragmatic approach should therefore be described as indicative or adjusted estimates of programme impact and should not be presented as definitive causal effects and economic evaluation results based on these estimates should be interpreted with corresponding caution, particularly where they are used to inform decisions regarding scaling or substantial public investment.

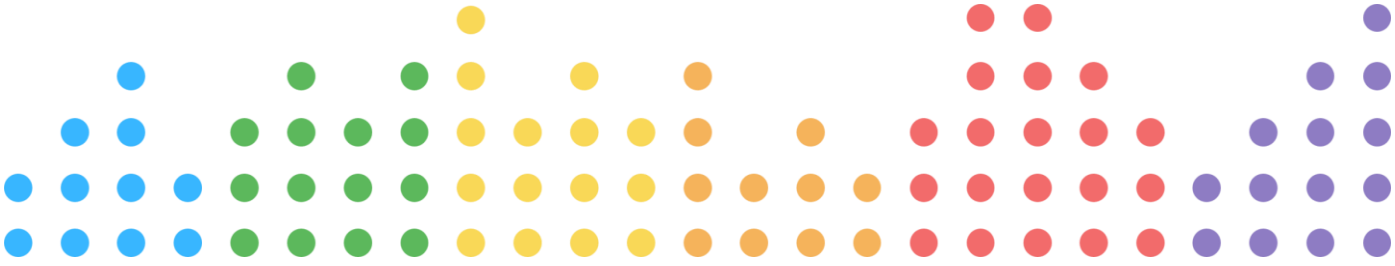
4.3. Estimation of costs of creative health interventions

The measurement of costs is an integral part of both CBA and CEA and is therefore relevant to both approaches to economic evaluation. This is covered here first before the Guidelines branch into two distinct sets of recommendations for CBA and CEA.

There are two types of costs to consider in economic evaluation:

i. The **costs of running the intervention**; and
ii. Any **potential negative effects or outcomes** of the intervention.

Negative outcomes in policy evaluations include impacts such as noise, congestion, pollution, negative health impacts, crime, littering and destruction of property. By and large these types of impacts are not relevant for creative health interventions, although there could be instances of accidental damage or wear and tear of cultural and heritage assets and disruptions to local communities such as noise or increased traffic for larger



creative health interventions. **Since these types of impacts are likely to be rare they are not included in these Guidelines**. However, analysts should report any instances of negative impacts and their magnitude in economic evaluations.

The **running costs** of creative health interventions often involve a **combination of public, voluntary and community resources**. A comprehensive assessment of programme running costs should therefore consider the following cost categories:

Programme delivery costs

- Salaries for programme staff
- Payments to artists, facilitators or creative practitioners
- Administrative and coordination staff

Venue and facility costs

- Rental or hire of community spaces, studios or halls
- Use of museums, galleries or cultural venues
- Utilities and facility overheads

Materials and equipment

- Art materials (paint, clay, textiles, etc.)
- Musical instruments
- Craft materials and tools
- Digital equipment and software

Training and development

- Training for facilitators or volunteers
- Professional development
- Safeguarding or clinical training where required

Programme development costs

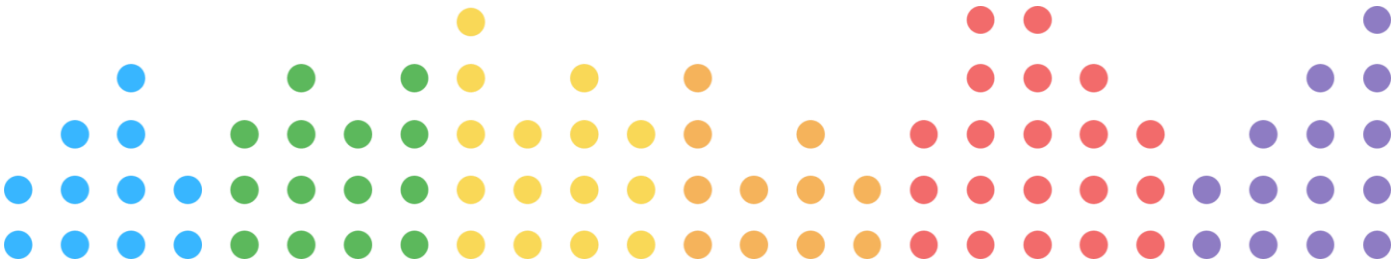
- Design of the intervention
- Pilot programme development
- Partnership development and planning

Volunteer time

- Time contributed by volunteers or community members valued using appropriate shadow wage estimates or wellbeing values

Healthcare system costs

- GP or social prescribing link worker time
- Administrative costs of referrals
- Monitoring and evaluation systems



Management and overheads

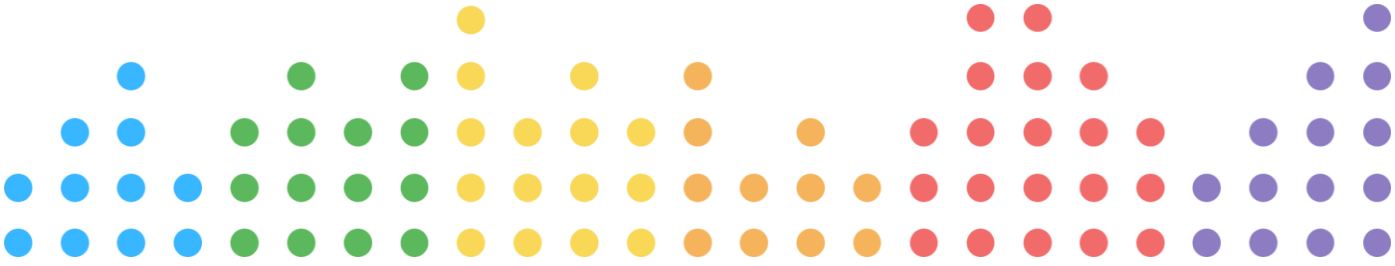
- Organisational overhead costs
- Programme management and governance
- Insurance and compliance costs

When assigning values to these types of programme costs, the objective is to measure the real economic resources used, rather than simply the accounting or budgetary expenditure incurred by the organisation delivering the programme. This ensures that economic evaluation reflects the full social cost of using labour, capital and other inputs for a particular intervention instead of for alternative activities, known as the **opportunity costs**.

In practice, the Green Book (2026, p.53) recognises that **market prices** are often a good proxy for opportunity costs; where markets are competitive and functioning well, the price paid for labour, materials or services will generally reflect the value of those resources in their next best use.

Therefore, for many creative health intervention inputs, such as staff salaries, venue hire, equipment and materials, analysts can use observed market prices as the basis for cost estimates.

Adjustments, however, may sometimes be required - for example, where resources are **donated** or **subsidised**, such as volunteer labour or free venue space. Here analysts should estimate the shadow price of the resource, reflecting what it would cost if it were obtained in the market.



5. Cost-Benefit Analysis (CBA)

This part of the Guidelines sets out the process required for performing Cost-Benefit Analyses (CBA) of creative health interventions. It is intended for evaluations where the decision requires an assessment of whether the full benefits of a creative health intervention justify its costs, including health and non-health outcomes such as loneliness, wellbeing and cultural participation.

The steps below provide the core framework for undertaking CBA in creative health. They should be read alongside the guidance developed through the Culture and Heritage Capital programme, including the work undertaken by Frontier Economics (2024), which provides more detailed guidance on applying CBA to cultural and heritage interventions. In particular, this guidance illustrates how existing evidence and economic values can be adapted to different use cases, including the assumptions required around factors such as participation, dosage, frequency of engagement and persistence of effects. This provides a useful practical complement to the worked examples presented in these Guidelines.

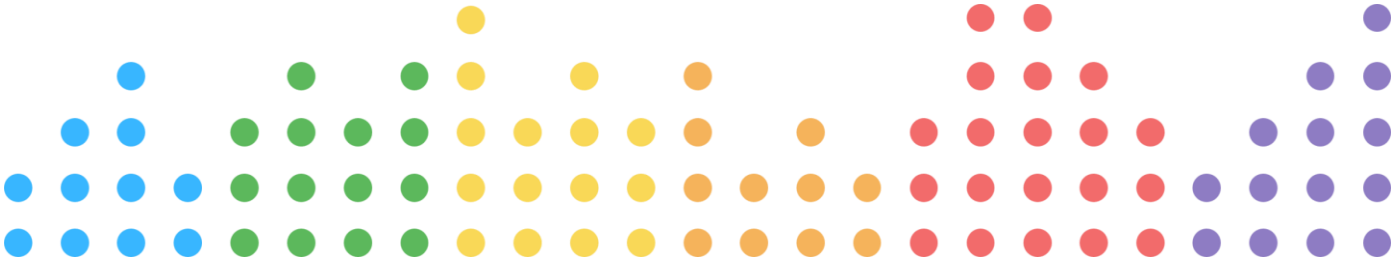
Step 1. Specify the intervention, identify outcomes and decide groups with standing

Using the ToC model, the first step is to clearly define the intervention being analysed, which involves specifying the scope, objectives and main components of the project. Analysts must also determine which individuals or groups are considered to have **standing** in the analysis – these are the people who will be impacted by the intervention.

Then the next step is to identify the full range of outcomes associated with the intervention. All significant costs and benefits affecting individuals, households, businesses, public services and communities should be considered including economic, social, and environmental effects. At this stage analysts should distinguish between **outputs** (activities delivered) and **outcomes** (changes experienced by participants or society), focusing the analysis on the latter.

Step 2. Estimate the impacts of the intervention

Once outcomes have been identified, analysts should estimate the intervention's impacts on the outcomes using the methods set out in Section 4.2 (Measuring impact) to account for deadweight and additionality.



Analysts must consider not only the magnitude of the outcome achieved but also the **length of time** for which that outcome is expected to persist. The **duration** of benefits is likely to be one of the most important drivers of estimated value and should therefore be treated as a key analytical input.

Where follow-up data are available, the observed pattern of outcomes over time should be used to estimate persistence. Where benefits decline following the intervention, analysts should model this decline rather than assume that the full measured impact remains constant. This may involve applying a decay rate so that the estimated impact reduces progressively over time. Evidence on persistence and decay may also be drawn from robust evaluations of sufficiently similar interventions and populations.

In the absence of follow-up evidence or relevant external evidence, evaluators should not assume that benefits persist beyond the period over which they have been observed. Longer-term persistence may be modelled as an upper bound or sensitivity scenario, but it should not form the central estimate without supporting evidence. Benefits should not be assumed to continue indefinitely unless there is robust evidence to support this assumption.

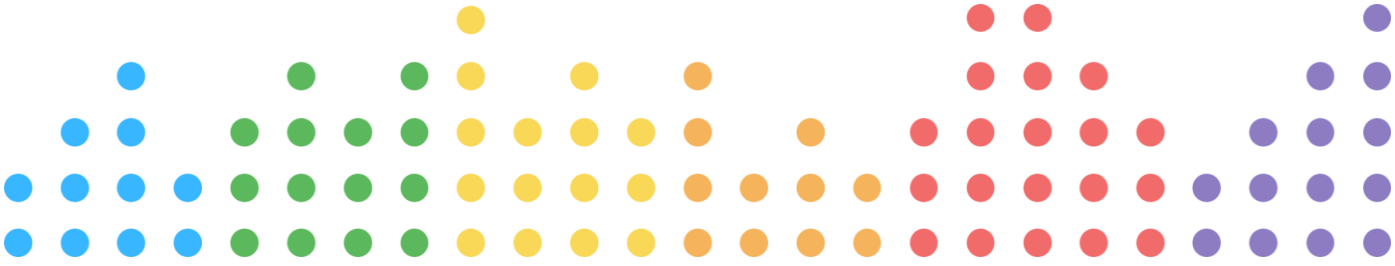
Step 3. Valuation of costs and benefits of the intervention

The costs of running the intervention should be measured as described in Section 4.3 (Estimation of costs of creative health interventions). The focus of this section is on the valuation of the benefits of creative health interventions.

This section strikes a **balance** between the technical rigour required by CBA - which seeks to measure the costs and benefits of all outcomes related to a project - and the real-world application, resources and realities in the creative health sector, which mean that a narrower core set of metrics and questions is more likely to be measured and used correctly given resource constraints.

The Roundtable on Creative Health highlighted that **wellbeing** is the outcome that most closely reflects the **lived experience of participants**. Stakeholders emphasised that creative health generates benefits that extend beyond clinical outcomes and healthcare utilisation, including meaning and purpose; confidence; mental and physical health; identity; agency; social connection; and participation.

There was consensus, therefore, that the outcomes framework for creative health should be structured around quality of life rather than treating it as one outcome among many. And that outcomes such as healthcare utilisation; economic participation; and community and cultural outcomes are more secondary in importance.



Types of values

There are two broad categories of value in CBA that require different valuation techniques.

1. **Primary value impacts** refer to outcomes experienced directly by individuals as a result of participating in the intervention. In creative health programmes these impacts typically occur through changes in wellbeing, health, behaviour, and economic circumstances. They can be **financial** (e.g. increases in household income from getting a new job) or **non-financial** (e.g. improved confidence and reduced loneliness) in nature. They represent the core benefits that the intervention aims to achieve and should therefore form the foundation of the valuation exercise.

2. **Secondary value impacts** are additional **spillover outcomes** from primary impacts experienced by participants. These are usually **financial** in nature and accrue to the government (i.e. taxpayer) and businesses. In creative health interventions secondary impacts include for example reductions in benefit payments; increases in productivity as someone moves from unemployment into employment; and reduced healthcare costs. These spillover effects arise indirectly as a consequence of changes in behaviour, health or wellbeing among participants.

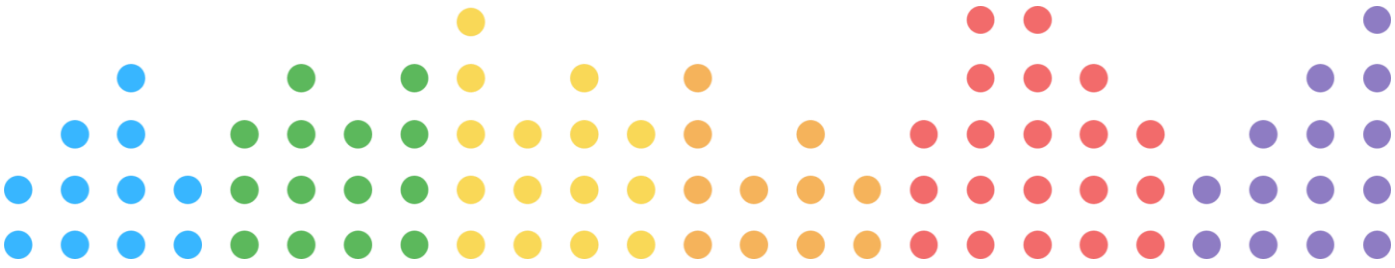
Secondary impacts are already denominated in financial terms and therefore do not need to be valued separately; they can enter into the CBA calculations directly once the relationship between the outcome achieved (e.g. improved health) and the secondary impact (e.g. healthcare usage such as reduced nights in hospital) has been estimated.

The focus of this part of the Guidelines is on the methods for valuing non-financial primary impacts as these are more technical in nature requiring specialised non-market valuation methods.

Programme-based versus outcomes-based approaches to valuation

There are two approaches for valuing the primary benefits generated by creative health interventions: an **outcomes-based approach** and a **programme-based approach**. **These Guidelines adopt the outcomes-based approach as the recommended method for economic evaluation**.

Under the outcomes-based approach, outcomes are first measured using the impact evaluation methods described in Section 4.2 and are then assigned monetary values using recognised valuation techniques. These values are subsequently aggregated to



estimate the overall value generated by the intervention. This approach enables a transparent link between observed outcomes and estimated value.

A key advantage of this approach is that it allows evaluators to identify which outcomes are driving value and to assess impacts across a broad range of domains. Unlike traditional health economic evaluations that focus primarily on health outcomes, such as quality-adjusted life years (QALYs), creative health interventions often generate wider benefits, including improvements in wellbeing, social connection, cultural participation, confidence, skills, learning, and community engagement.

An alternative approach involves directly valuing interventions through stated preference methods, such as contingent valuation surveys that ask participants about their willingness to pay for programme benefits. While this approach can capture a broad range of benefits and may be useful in some circumstances, it requires additional data collection, bespoke surveys and specialist expertise. Furthermore, asking people their willingness to pay for a good or service is not always appropriate and would be difficult in creative health since the model is free provision. For these reasons, the outcomes-based approach is recommended as the primary valuation method within these Guidelines.

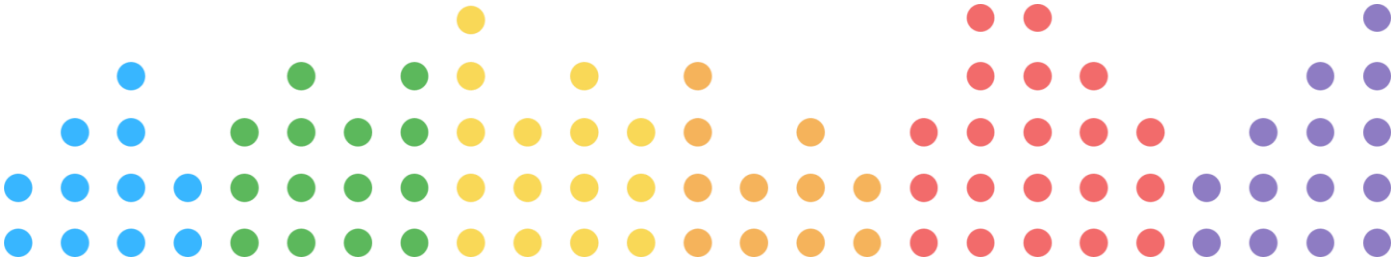
Valuation methods

Many of the benefits generated by creative health interventions, such as improvements in wellbeing, reduced loneliness, better mental health and stronger social connection, do not have observable market prices. The Green Book therefore recommends the use of **non-market valuation methods** where relevant outcomes cannot be valued directly through market data.

These approaches seek to estimate the monetary value of outcomes that affect individual wellbeing. In broad terms, the Green Book recognises three principal approaches:
i. Revealed Preference valuation methods, which infer value from observed behaviour in related markets;

ii. Stated Preference valuation methods, which ask individuals directly about their willingness to pay for benefits or willingness to accept compensation for losses; and

iii. Wellbeing Valuation, which estimates the monetary value of outcomes by examining how changes in life circumstances are associated with changes in life satisfaction.



For creative health, revealed and stated preference methods may be useful in some contexts, but their practical application can be limited because many important outcomes, such as reduced loneliness or improved wellbeing, are not directly reflected in market behaviour and robust bespoke valuation studies can be resource-intensive. Wellbeing valuation may therefore be particularly relevant where interventions generate broad improvements in subjective wellbeing, social connection, confidence or quality of life.

Health outcomes are treated separately within established health economic guidance. **Quality-Adjusted Life Years (QALYs)** combine changes in health-related quality of life and duration of life into a single measure and can be used directly in Cost-Effectiveness Analysis or assigned a monetary value for Cost-Benefit Analysis. Health-related quality of life is commonly measured using instruments such as the EQ-5D.

The detailed choice of valuation method should reflect the outcome being assessed, the quality of the available evidence and the decision being informed. These Guidelines focus primarily on the valuation approaches considered most practical and relevant for creative health, particularly **health-related quality of life, life satisfaction and healthcare utilisation**. Further detail on non-market valuation methods, their assumptions and limitations is provided in Annex B, with additional methodological guidance available in the Green Book and relevant health economic guidance.

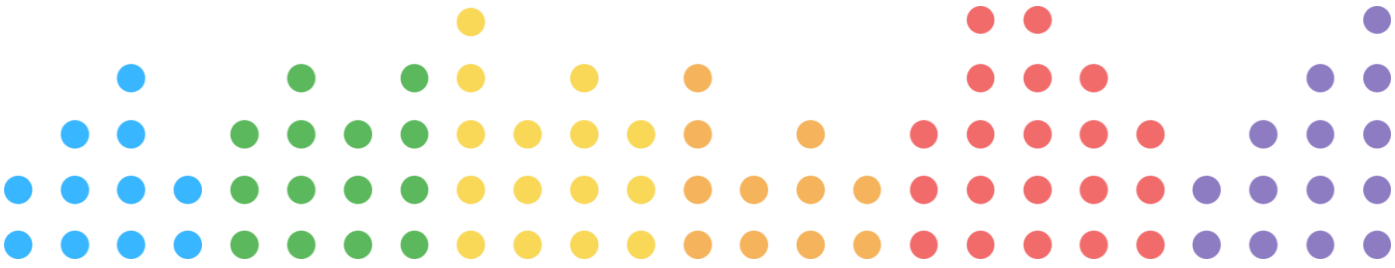
Creative health outcomes to be measured and valued for CBA

This section sets out a recommended set of outcome metrics to be valued and included in CBA of creative health interventions. Participants in the Roundtable on Creative Health recognised the need for greater consistency across the sector but warned against imposing overly complex or burdensome requirements.

There was agreement that the Guidelines should, therefore, introduce a core set of outcomes that is easy to apply and is measured consistently in creative health interventions to ensure consistency across the sector. This is supplemented by a broader set of outcomes that can be captured in addition, if budgets and resources allow.

Core Outcome Framework (COF)

The COF is a minimum set of four outcomes for creative health interventions that all providers and delivery organisations are encouraged to collect. These represent a small number of common measures covering wellbeing and health utilisation that can be used in CBA. The COF is based on feedback and discussions within the sector and is



designed to be easy to apply and collect, whilst at the same time being of sufficient rigour and depth to permit valuation of outcomes for CBA as per the Green Book.

The **focus of the COF** is on the **wellbeing of participants** and **health utilisation** – the key metrics for different stakeholders in creative health. They cover the primary and secondary benefits of creative health interventions.

Primary and secondary outcomes should be understood as part of an integrated pathway of impact rather than as separate categories of value to be automatically combined. Primary outcomes capture the principal changes that an intervention is intended to generate, while secondary outcomes may capture wider or downstream effects arising through those changes. The relationship between them should be clearly specified in the intervention's Theory of Change and supported by appropriate evidence. Where outcomes are subsequently valued or aggregated within an economic evaluation, evaluators should demonstrate the pathway through which each benefit arises and consider whether outcomes overlap. Benefits should not be combined where this would result in double counting, or where the evidential basis for linking a primary outcome to a secondary outcome is insufficient.

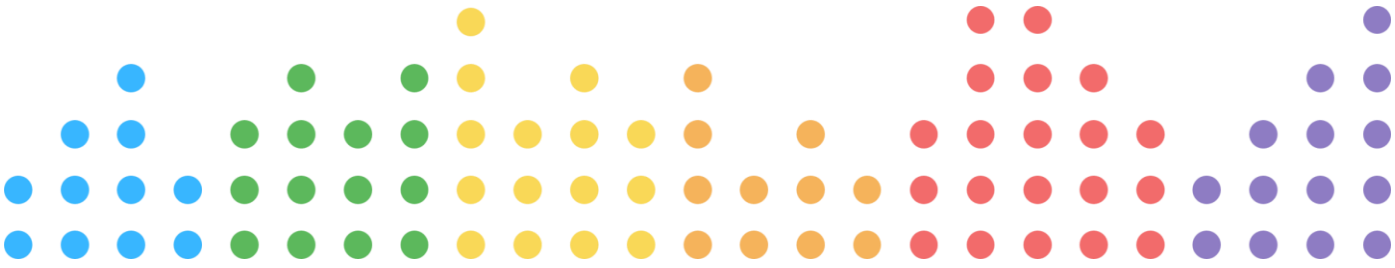
COF metrics

1. Health utility

Health utility is a measure of the quality of an individual's physical and mental health state expressed on a scale where 0 represents being dead and 1 represents full or perfect health. It reflects the value that individuals or society place on different health states and is used in health economics to compare the health benefits generated by different interventions.

A **Quality-Adjusted Life Year (QALY)** is a health utility measure that combines both the quantity and quality of life into a single metric. One QALY represents one year of life lived in perfect health. QALYs are calculated by multiplying the number of years lived by a health utility score (on a scale of 0 to 1). For example, 10 years lived with a health utility of 0.3 would generate 3 QALYs, whilst 10 years lived in full health would generate 10 QALYs. Improvements in health utility, therefore, translate directly into QALY gains.

The **EQ-5D** survey is a standardised instrument used to measure health-related quality of life and derive health utility values. It comprises five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Respondents select the level that best describes their health state in each dimension, and the resulting health profile is converted into a single utility value (on a scale of 0 - 1) using population preference



weights [2]. These utility values are widely used in economic evaluation to calculate Quality-Adjusted Life Years (QALYs) and assess the value for money of healthcare interventions.

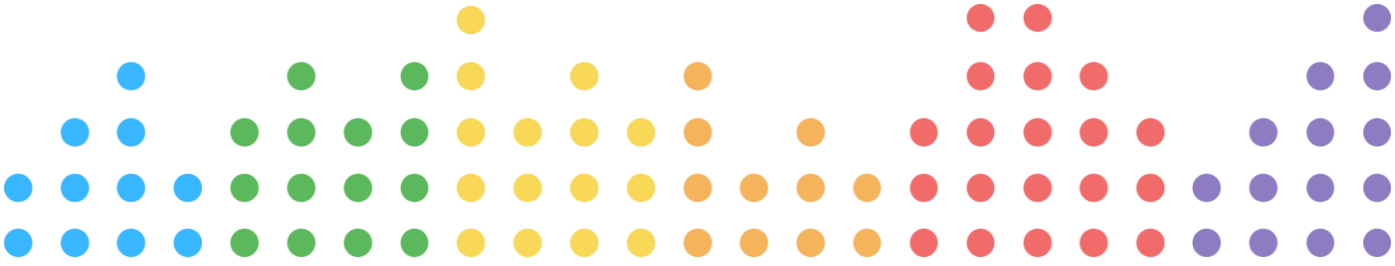
EQ-VAS (EuroQol Visual Analogue Scale) is a complementary measure of health-related quality of life that asks respondents to rate their overall health on a scale from 0 to 100, where 0 represents the worst imaginable health and 100 represents the best imaginable health. Unlike EQ-5D, EQ-VAS does not directly generate a preference-based health utility value, but it can be converted into an approximate health utility measure by rescaling scores from a 0–100 scale to a 0–1 scale. In practice, many applied evaluations and some published studies estimate QALYs directly from EQ-VAS by using this rescaling technique and given that this approach is simpler to administer than EQ-5D it may be particularly appropriate for creative health evaluations where minimising respondent burden is an important consideration.

For creative health, health utilities will capture improvements in mental health and physical health, but it is unlikely to fully capture the wider social, cultural, and community outcomes generated by creative health interventions such as cultural participation; creative expression; community cohesion; loneliness and social capital; and skills and learning. As such, health utilities should be used alongside broader outcome measures that can pick up these wider outcomes.

Health utility scores can be valued in monetary terms using a range of indicators such as (i) the NHS **opportunity cost** of around £13,000 per QALY in 2008 prices (Klaxton et al., 2015)[3], which reflects the estimated health forgone elsewhere in the NHS when resources are reallocated to a new intervention (approx. **£21,000 in 2026 prices**), (ii) the National Institute for Health and Care Excellence's (NICE) (2025) threshold of around **£30,000** per QALY gained [4]. This represents a **cost-effectiveness threshold** that represents the maximum amount the NHS is prepared to spend to generate one additional QALY, or (iii) the Green Book's (2002) £70,000 per QALY value based on **stated preference valuation** methods (approx. **£83,000 in 2026 prices**).

Which value to use in economic evaluation will depend on the audience and decision context. For evaluations undertaken primarily to inform NHS commissioning and resource allocation decisions, these Guidelines recommend the use of the estimated NHS opportunity cost of approximately £21,000 per QALY in 2026 prices. This value

[2] Health utility values are anchored at 1, representing full health, and 0, representing a health state equivalent to being dead. However, values below zero are possible and represent health states that are valued, based on general population preferences, as worse than being dead.
[3] Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold - PubMed
[4] Changes to NICE's cost-effectiveness thresholds confirmed | NICE



reflects the estimated health forgone elsewhere in the NHS when resources are reallocated to fund a new intervention and is therefore directly relevant to assessing whether an intervention is expected to generate a net increase in health within a fixed NHS budget.

The NICE cost-effectiveness threshold provides an alternative benchmark and may be reported as part of sensitivity analysis. However, an intervention falling below the NICE threshold should not necessarily be interpreted as demonstrating that the NHS would fund the intervention or that reallocating resources towards it would increase overall population health. For this reason, the NHS opportunity-cost value is recommended as the primary value for NHS-focused economic evaluation in these Guidelines.

NHS commissioner QALY value recommendation: £21,000 per QALY (for use in CEA)

Evaluations undertaken for central government departments, local authorities, combined authorities, and other public sector organisations should use the Green Book value, as this is designed for CBA and reflects the wider societal value of health improvements rather than NHS budget constraints alone.

Central government QALY value recommendation: £83,000 per QALY (for use in CBA)

Wherever possible the **EQ-5D survey instrument should be used in economic evaluations** of creative health interventions. Responses to the EQ-5D should be converted into health utility values using the appropriate EQ-5D value set (also referred to as a tariff).[5] Evaluators should refer to the EuroQol EQ-5D value set and scoring guidance for the appropriate value set and tools for converting EQ-5D health profiles into health utility values. The value set used should be clearly reported in the evaluation.

If EQ-5D cannot be used then the **EQ-VAS should be used instead** using the following question:

"We would like you to indicate on this scale how good or bad your health is today?"

[5] A value set (also referred to as a tariff) assigns a health utility value to each possible EQ-5D health profile based on preferences elicited from the general population. A respondent's answers across the five EQ-5D dimensions generate a health profile (for example, 1-2-1-3-1), which is converted into a single health utility value by applying the relevant country-specific value set. EuroQol provides scoring tools that apply the selected value set to EQ-5D data and calculate the corresponding utility values.

(0 = "worst imaginable health state" to 100 = "best imaginable health state")

The EQ-VAS score can be converted into a health utility by dividing the score by 100.

2. Office for National Statistics (ONS) Life Satisfaction

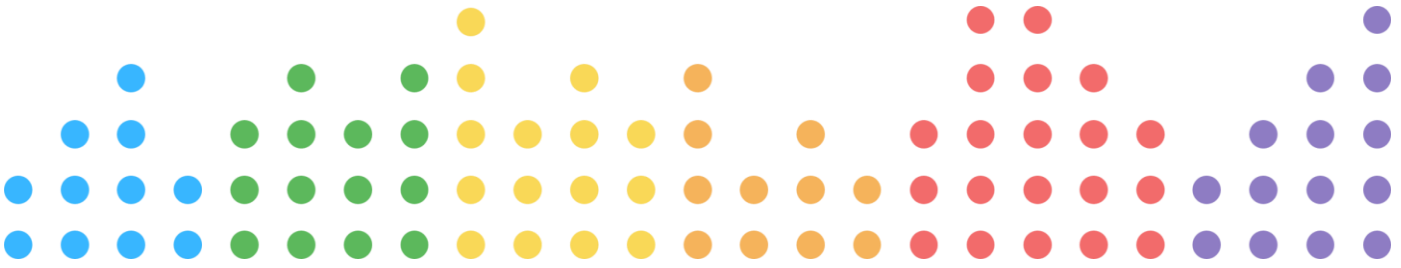
Life satisfaction is one of the personal wellbeing measures used extensively in UK surveys. Life satisfaction can be used to estimate how a creative health intervention affects individuals' overall quality of life.

Many creative health interventions generate benefits that extend beyond health alone. These may include improvements in social connection, reduced loneliness, increased confidence, cultural participation, sense of purpose, personal development, community engagement, and so on. Such outcomes may not be fully reflected in health-focused measures and therefore may be underestimated when using EQ-VAS alone.

The ONS life satisfaction measure provides a broader assessment of wellbeing by asking individuals to evaluate their life as a whole. Life satisfaction can, therefore, capture the combined effect of many of the impact pathways through which creative health interventions may improve people's lives. These may include improvements in physical and mental health, reduced loneliness, increased social connection, greater confidence and self-efficacy, and an increased sense of purpose or meaning. Rather than valuing each of these outcomes separately, changes in life satisfaction can provide an overarching measure of their combined impact on an individual's overall wellbeing. Furthermore, another benefit of the life satisfaction measure is that it provides a net impact estimate over and above any **personal costs** associated with participating in creative health interventions, which include time used to travel to/from the activity and out-of-pocket costs incurred to attend the activity, such as petrol, parking fees, train tickets, etc.

Used alongside EQ-VAS, life satisfaction can therefore provide a more comprehensive assessment of the benefits generated by creative health interventions. Together, the two measures enable evaluators to capture both health-related outcomes and wider social and wellbeing impacts, reflecting the broad range of outcomes that creative health programmes are designed to achieve.

Changes in life satisfaction can be converted into monetary values using the Green Book WELLBY method, which is a standardised value of a unit change in life satisfaction for one year. This was £13,000 in 2019 (approx. £17,000 in 2026 prices). This approach is widely used in UK policy appraisal and social value analysis and is particularly



relevant for creative health interventions that generate improvements in wellbeing, social connection, confidence, and quality of life.

WELLBY value recommendation: £17,000 per WELLBY (for use in CBA)

The ONS recommends the following question:

"Overall, how satisfied are you with your life nowadays?"

(0 = "Not at all satisfied" to 10 = "Completely satisfied")

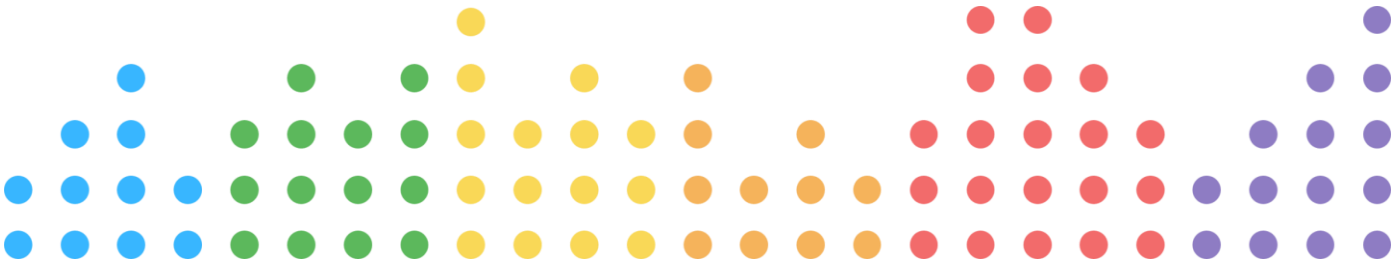
Life satisfaction also has limitations. It is a broad evaluative measure that requires individuals to make an overall assessment of their life and respondents often provide an answer relatively quickly. As a result, it may not distinguish clearly between the different pathways through which creative health interventions generate value. For example, equivalent changes in life satisfaction may arise through improvements in health, reduced loneliness, greater confidence, social connection or an increased sense of purpose.

Life satisfaction should, therefore, be interpreted as an overarching measure of overall wellbeing rather than a diagnostic measure of the specific mechanisms through which an intervention has generated impact. Supplementary outcome measures and qualitative approaches can be used alongside life satisfaction to provide greater insight into these impact pathways.

These Guidelines therefore recommend life satisfaction as one important broad measure of wellbeing for economic evaluation, whilst recognising wellbeing measurement as an important area for future research in creative health.

Health utility scores and life satisfaction capture different but overlapping dimensions of impact. Health utility scores provide a health-focused measure that can be converted into QALYs and are particularly relevant where the primary objective is to improve health outcomes or inform NHS resource allocation decisions. Life satisfaction is a broader measure of overall wellbeing that can capture the combined effects of improvements in health alongside wider outcomes such as social connection, purpose and skills attainment, and can be converted into WELLBYs for wider social value analysis.

Wherever feasible, evaluators should collect and report both measures. This can provide complementary evidence on the impacts of creative health interventions,



improve understanding of the pathways through which benefits are generated, and support different decision contexts and economic evaluation questions. The choice of primary measure for economic evaluation should reflect the objectives of the intervention and the decision context. QALYs should generally be prioritised where the primary focus is health outcomes and NHS resource allocation, while WELLBYs may provide a more comprehensive basis for CBA where interventions are expected to generate broader social and wellbeing benefits.

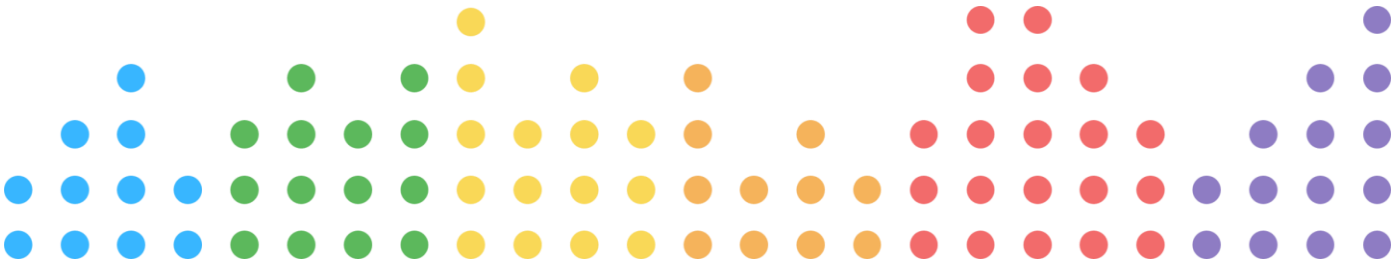
However, the monetary values derived from QALYs and WELLBYs should not be added together within the same CBA. As improvements in health may also contribute to changes in life satisfaction, combining the two valuations risks double counting overlapping benefits. Evaluators should therefore report valuations of QALY- and WELLBY-based results separately and select the most appropriate approach as the primary basis for the economic evaluation.

3. Health utilisation

The focus of the health utilisation metrics is on measures that serve as the primary care touchpoints most likely to be influenced by creative health interventions. Creative health interventions are unlikely to immediately impact complex hospital admissions and other secondary or tertiary healthcare systems.

In contrast, **general practitioner (GP) visits** often address milder health concerns, stress, or mental health, and **accident and emergency (A&E) visits** may reflect urgent but non-complex issues. Both metrics provide a clearer link to wellbeing and behaviour changes that creative health interventions could feasibly influence. Importantly, these data are relatively straightforward to collect: participants can simply report how many times they visited a GP or A&E in a defined period; they are well understood; they are readily linked to NHS cost data; and they provide tangible evidence of potential healthcare resource shifts. Focusing on primary care utilisation provides a proportionate and realistic way to track potential cost offsets within the healthcare system.

GP visits and A&E attendances are, therefore, recommended as the core healthcare utilisation measures, but these measures are not intended to represent the full range of potential healthcare resource impacts. Where the ToC identifies a credible pathway to other forms of healthcare utilisation, evaluators may additionally measure outcomes such as non-elective hospital admissions, outpatient care or medication use. Given the greater complexity, lower frequency and potential difficulties in attributing changes in these outcomes to creative health interventions, their inclusion should be supported by a clear causal rationale and, where possible, robust evaluation design or linked administrative healthcare data.



The recommended health utilisation questions are:

GP visits: *"In the past [three months], how many times have you visited or spoken to a GP (including face-to-face, phone, or video consultations)?"*

A&E visits: *"In the past [three months], how many times have you attended an Accident & Emergency (A&E) department for your own care?"*

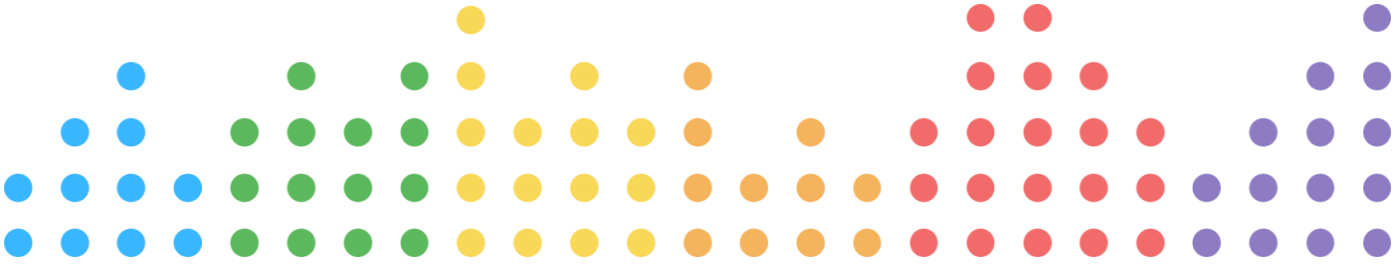
Self-reported GP and A&E utilisation provides a pragmatic approach where routine data are not available but may be subject to recall error and measurement bias. Where feasible and proportionate, evaluators should seek to validate or replace self-reported health usage measures through linkage to routinely collected NHS data.

As set out in section 4.2 (Measuring impact) these questions should be asked both before the intervention begins and after it ends. When collecting pre- and post-intervention data on GP and A&E visits, the recall period - in [brackets] in the question text - should match the intervention's duration. For shorter interventions, use shorter recall windows (e.g., past month); for longer interventions, use longer windows (e.g., past three or six months). It is essential that the pre- and post-survey recall periods do not overlap. For example, if the intervention lasts one month, do not use a three-month recall period after the intervention, as it would include the pre-intervention period.

There is an emerging area of research exploring how self-reported measures of health, such as confidence in managing one's own health, can be mapped onto healthcare utilisation and associated costs savings. This approach can enable estimation of cost savings without directly asking participants to recall service use, while capturing a broader range of healthcare costs.

An example of this is the Patient Activation Measure (PAM), which is a validated 13-item measure of a person's knowledge, skills and confidence in managing their health. Barker et al. (2018) found that patients with higher PAM scores of self-management used substantially fewer healthcare services (made up of GP contacts, A&E visits, hospital admissions and outpatient appointments). Although the study was observational and cannot prove causality, it provides evidence that measures of health confidence and self-management capability may be useful predictors of future healthcare utilisation.

Although not currently included in the Green Book or DHSC guidelines, this is a promising area for future research and development. As this evidence base is still evolving, it is recommended to continue using direct service-use questions such as GP



and A&E visits for now, while monitoring advancements in mapping self-reported measures to healthcare costs.

Below are standardised NHS cost metrics for GP visits [6] and A&E attendances [7].

Average NHS cost of a GP visit is £48

Average NHS cost of an A&E attendance is £220

Supplementary outcome metrics

In addition to the outcome measures in the COF, evaluators may select supplementary outcome measures that reflect the primary objectives of the intervention. Where possible, these should be validated and widely used measures to facilitate comparability. These should align with the key outcomes of creative health identified in the literature as set out in Table 1.

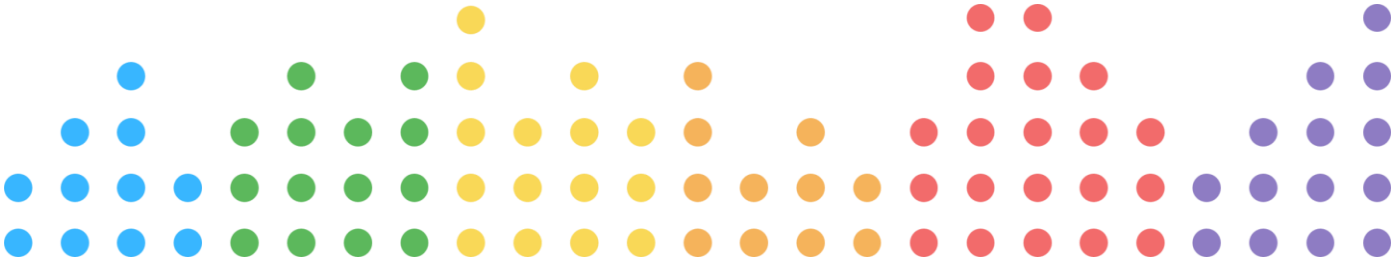
Examples of some supplementary measures are presented below. They cover the key areas in Table 1 and have been selected because they reflect outcomes commonly reported within the creative health evidence base and are supported by validated or frequently deployed measurement instruments. Their use can help improve consistency and comparability across evaluations while contributing to the development of a stronger evidence base for the sector.

However, evaluators are free to collect additional outcome data that are relevant to the objectives, theory of change, and target population of their intervention. Indeed, intervention-specific outcomes may often be necessary to capture important impacts that are not adequately reflected in the core or supplementary measures recommended here. The choice of outcome measures should therefore be guided by the intervention's intended outcomes, whilst balancing considerations of data quality, respondent burden, proportionality, and comparability with other evaluations.

1. Mental wellbeing. Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS)

Rationale: Captures positive mental health, resilience, optimism, and functioning. Widely used across the creative health sector and supported by a strong evidence base.

[6] PSSRU Table 9.4.2: 'Unit costs for a GP'.
[7] NHS National Cost Collection.



2. Experience-based wellbeing. Measures of individuals' emotions and experiences during specific activities or periods of the day. Approaches such as the Day Reconstruction Method (DRM) assess how people feel during particular episodes rather than asking them to evaluate their life as a whole.

Rationale: May be particularly relevant to creative health interventions by capturing the immediate experiential effects of participation in creative activities. The reflective, diary-based format of the DRM surveying method may also be well suited to creative health settings, where creative writing, visual expression and reflective approaches are already used to help participants communicate their experiences. DRM can complement life satisfaction by providing greater insight into how and when wellbeing benefits are generated, although experience-based wellbeing measures are not currently as readily incorporated into standardised economic evaluation.

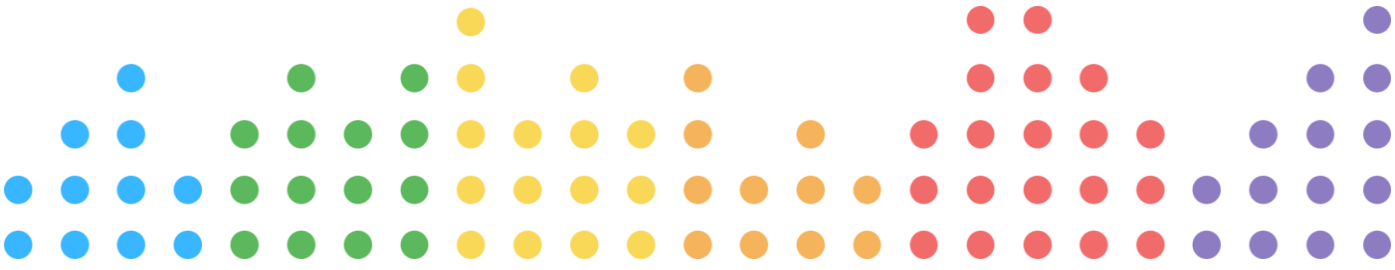
3. Mental health (anxiety and depression). The Generalised Anxiety Disorder-2 (GAD-2) and Patient Health Questionnaire-2 (PHQ-2) are brief, validated two-item screening measures of anxiety and depressive symptoms respectively. Each asks about the frequency of key symptoms over the previous two weeks and generates a score from 0 to 6, with higher scores indicating greater symptom severity. Where a creative health intervention is specifically intended to improve mental health, evaluators may consider collecting the GAD-2 and/or PHQ-2 alongside the core outcome metrics.

Rationale: These measures provide short, low-burden indicators of changes in anxiety and depression and can support comparisons with established mental health services and interventions using similar clinical outcome measures. Changes in scores, or the proportion of participants achieving a clinically meaningful improvement, may be used in CEA to estimate the cost-per-improvement achieved and benchmark the intervention against comparable NHS mental health provision. This may provide useful evidence for commissioners on the relative cost-effectiveness of creative health interventions.

4. Social connectedness and loneliness. "How often do you feel lonely?" (Often/Some of the time/Occasionally/Hardly ever/Never)

Rationale: Reducing loneliness and strengthening social connections are among the most frequently reported outcomes of creative health interventions. This single-item question is used in a number of national datasets in the UK including Understanding Society and the Annual Population Survey.

5. Self-efficacy. The General Self-Efficacy Scale measures an individual's confidence in their ability to cope with difficult situations and achieve desired goals. The short form asks respondents to rate a small number of statements concerning their perceived ability to deal with challenges. Evaluators wishing to use the measure should use the



validated wording and scoring procedure for the selected version rather than adapting individual questions.

Rationale: Captures individuals' confidence in their ability to manage challenges and achieve goals. Relevant across many creative health interventions and linked to long-term behaviour change.

6. Community engagement. Frequency of participation in community activities, groups, events, volunteering, or local organisations. For example the following question is used in a number of national surveys: "How often do you take part in activities organised by groups, clubs or organisations in your community?"

Rationale: Captures wider social and community benefits that extend beyond individual participants and are often central to place-based creative health programmes.

7. Cultural participation. Frequency of participation in creative, artistic, or cultural activities. For example, the following question is used in national surveys: "During the last [12 months], have you taken part in any creative, artistic, or cultural activities?"

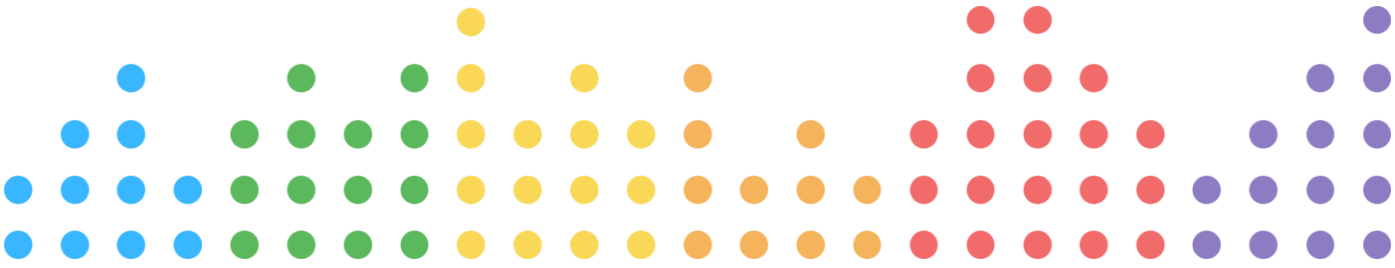
Rationale: Measures engagement with creative activities and helps identify whether interventions lead to sustained participation in arts and culture beyond the intervention.

8. Work readiness. For creative health interventions involving working-age participants, evaluators should consider measuring work ability where there is a credible pathway from the intervention to improved employment outcomes. This covers an individual's perceived ability to work. For example, the following single-item score can be used: *"Assume that your work ability at its best has a value of 10 points. How many points would you give your current work ability?"* (0 = completely unable to work; 10 = work ability at its best).

Rationale: Captures potential impacts on an individual's ability to remain in, return to, or participate more fully in work, with potential wider impacts on employment, welfare costs, tax revenues, absenteeism and productivity.

9. Physical health. Intervention-specific measures of physical health, such as physical activity, pain, body weight, Body Mass Index (BMI) or diet. The specific measure selected should reflect the outcomes and causal pathways identified in the intervention's ToC.

Rationale: Captures physical health impacts that may not be fully reflected in the core outcome measures and, where comparable outcome metrics are available, may allow



cost-effectiveness results to be benchmarked against established NHS or public health interventions.

Qualitative analysis

There is a general concern in the creative health sector that economic evaluation frameworks often overlook the experiences of participants. The sector highlights the importance of:

- Personal narratives
- Participant voice
- Co-production
- Understanding the mechanisms of change
- Family and community impacts

Quantitative analysis should be complemented by qualitative evidence wherever possible. Whilst quantitative methods provide estimates of the scale and value of outcomes, qualitative approaches can provide important insights into how and why outcomes occur, the mechanisms through which value is created, participants' experiences of the intervention, and impacts that may not be fully captured through standard outcome measures.

Qualitative analysis can be particularly valuable in creative health, where interventions often generate complex changes relating to identity, confidence, creativity, social connection, belonging, empowerment, and community participation. These outcomes may be difficult to quantify fully but may nevertheless represent important benefits for participants and communities.

Evaluators should, therefore, consider collecting qualitative evidence through a range of methods set out in Table 2.

Table 2. Qualitative methods for assessing creative health interventions

Method	Approach	Application and use cases
Semi-structured interviews	Guided one-to-one conversations using a common set of open questions while allowing participants to describe experiences in their own words.	Understanding individual experiences, mechanisms of change and unintended outcomes.
Focus groups	Facilitated discussions with a small group of participants or stakeholders.	Exploring shared experiences, differences in perspectives, accessibility and acceptability.

Method	Approach	Application and use cases
Reflective participant feedback	Short written, recorded or creative reflections collected during or after participation.	Lower-burden evaluation and capturing experiences close to the point of delivery.
Observation	Systematic observation of activities, participation and interaction.	Understanding engagement, relationships and changes that participants may find difficult to articulate.
Case studies	Combining quantitative and qualitative evidence to examine particular participants, settings or interventions in depth.	Showing how outcomes arise in context and illustrating pathways identified in the Theory of Change.
Practitioner interviews	Interviews with those delivering or referring into the intervention.	Understanding implementation, barriers, mechanisms and differences between intended and actual delivery.

The choice of method should reflect the nature and scale of the intervention and the evaluation objectives.

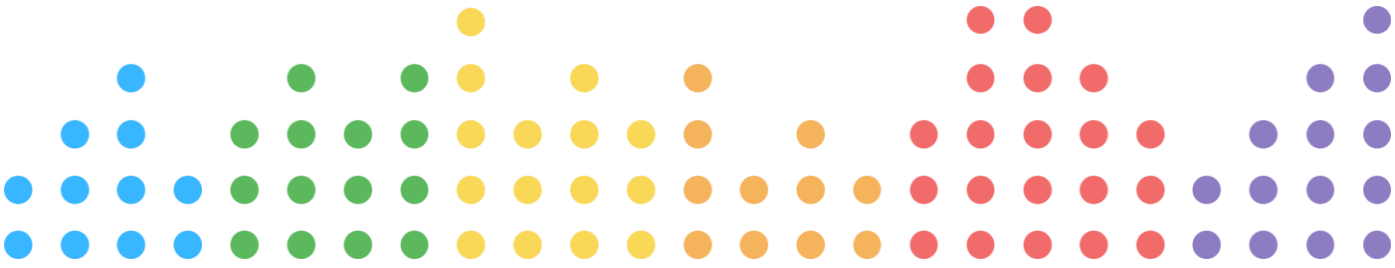
Step 4. Calculation of CBA results

Once the outcomes of the creative health intervention have been measured and the costs and benefits associated with these outcomes have been estimated using the recommendations and standardised metrics set out in these Guidelines, the final step in the CBA is to present the results.

How this final step is performed will depend on whether it is a CBA of a previous creative health intervention (ex-post analysis) or of a future creative health intervention (ex-ante analysis).

Ex-post CBA
The purpose of an ex-post CBA is to estimate the social value generated by a creative health intervention that has already been delivered. The starting point for this analysis is the estimate of impact derived using the methods described in Section 4.2. This will provide an estimate of the additional outcomes attributable to the intervention after accounting for deadweight.

Under these Guidelines, the primary benefits of creative health interventions should be estimated using one of two approaches. The first is a health-focused approach based on changes in EQ-VAS scores. Changes in EQ-VAS can be converted into QALYs and



valued using the standard health values recommended in in these Guidelines. This approach is most appropriate where the primary objective of the intervention is to improve health outcomes.

The second is a wellbeing-focused approach based on changes in life satisfaction. Changes in life satisfaction should be valued using the Wellby value provided in these Guidelines. This approach is most appropriate where interventions generate broad impacts on wellbeing covering health; social connection; confidence; purpose; community engagement, and so on.

EQ-VAS and life satisfaction should not be valued simultaneously within the same CBA because improvements in health contribute directly to overall life satisfaction and wellbeing. Monetising both measures would therefore risk double counting the benefits of health improvements. Evaluators should select the measure that best reflects the primary objectives and expected outcomes of the intervention.

In addition to these primary benefits, changes in healthcare utilisation should be valued separately using the recommended unit costs for GP visits and A&E attendances set out in these Guidelines. These represent financial savings to the NHS and should be added to the primary benefits estimated using either the EQ-VAS or life satisfaction approach.

The **total benefits** of the intervention are therefore calculated as: Total Benefits = Primary Outcome Value + Healthcare Cost Savings.

The **total costs** of delivering the intervention should then be calculated.

Net Present Social Value (NPSV)

NPSV represents the total value generated by the intervention after all costs have been deducted:

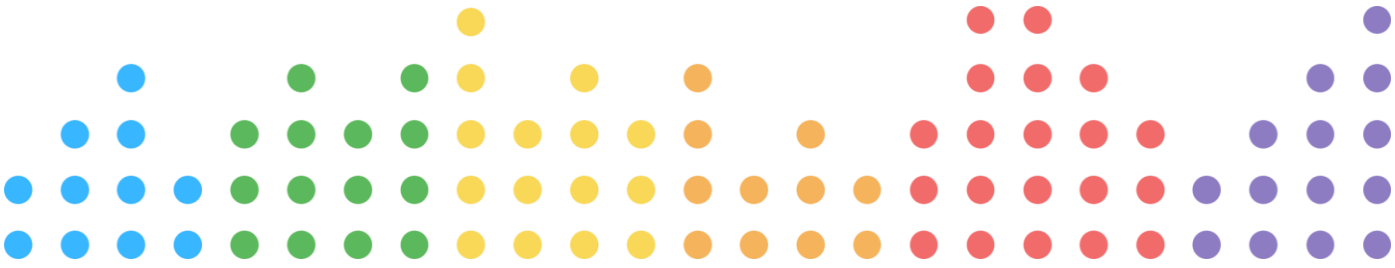
NPSV = Total Benefits – Total Costs

A positive NPSV indicates that the intervention generated more value than it consumed in resources.

Benefit-Cost Ratio (BCR)

The BCR expresses the value generated for every £1 invested:

BCR = Total Benefits ÷ Total Costs



For example, a BCR of 3.5 indicates that every £1 invested generated £3.50 of social value.

Ex-ante CBA

The purpose of an ex-ante CBA is to estimate the likely value generated by a proposed future creative health intervention. The calculation process is identical to that used in ex-post CBA, except that future outcomes must be forecast rather than directly observed.

Forecasts should be based on the strongest available evidence, including previous evaluations of similar creative health interventions, pilot studies, published research, meta-analyses, or evidence from comparable programmes. The evidence used from previous studies must be of the level of rigour set out in Section 4.2 of these Guidelines. Assumptions regarding participation rates, outcome achievement, persistence of benefits, and healthcare impacts should be clearly stated and justified.

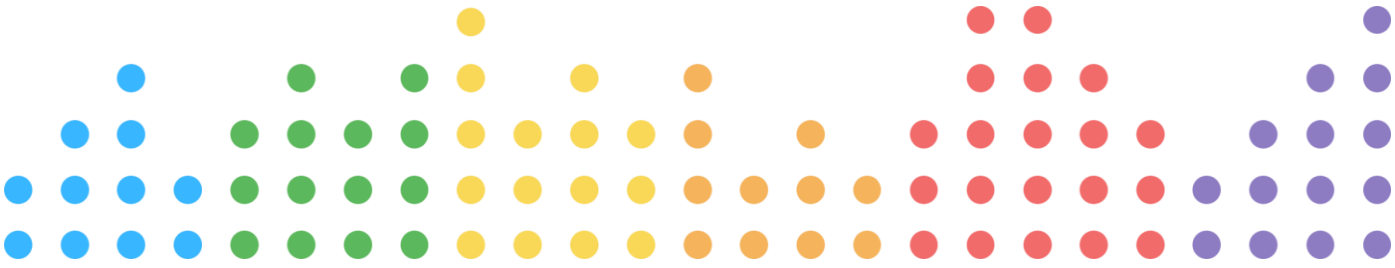
Given the greater uncertainty associated with forecasting future outcomes, ex-ante analyses require a number of additional considerations.

Ex-ante CBA involves forecasting the impacts, costs and benefits of interventions that have not yet been implemented. Consequently, all ex-ante evaluations are subject to **uncertainty**. Uncertainty may arise from a range of sources, including uncertainty regarding participation rates, programme delivery, the magnitude of outcomes achieved, the persistence of benefits over time, healthcare utilisation impacts, and the monetary values assigned to outcomes.

The purpose of economic evaluation is not to eliminate uncertainty, which is impossible, but rather to identify, assess and transparently communicate its implications for decision-making. Evaluators should therefore clearly **document all key assumptions** used within the analysis and provide justification for their selection.

Sensitivity analysis should be undertaken to assess how robust the results are to changes in these assumptions. As a minimum, evaluators should present a central estimate together with plausible lower and upper bound scenarios. Sensitivity testing should focus on the assumptions most likely to influence results, including impact estimates, participation rates, duration of benefits and healthcare utilisation rates. Where possible, evaluators should identify the assumptions that have the greatest influence on the NPSV and BCR.

The Green Book also highlights the importance of **optimism bias**. Evidence from previous public sector projects demonstrates that appraisals frequently overestimate benefits and underestimate costs. This can occur because of limited information,



excessive confidence in programme delivery, or a tendency to focus on best-case scenarios during project development. Evaluators should therefore adopt a cautious and evidence-based approach when forecasting future impacts, especially the costs associated with delivery. Assumptions should be informed by robust evidence from previous evaluations, pilot studies, academic research, or comparable interventions wherever possible. Where evidence is limited, conservative assumptions should be preferred.

Finally, future costs and benefits should be **discounted** and expressed in **present value terms** before calculating summary metrics such as NPSV and BCR. Discounting reflects social time preference: society generally places a higher value on benefits received today than on identical benefits received in the future. For many creative health interventions, where most costs and benefits occur within a relatively short period, discounting may have only a modest effect on results. However, where benefits are expected to persist over multiple years, appropriate Green Book discount rates should be applied to ensure that future costs and benefits are valued consistently and comparably across alternative options.

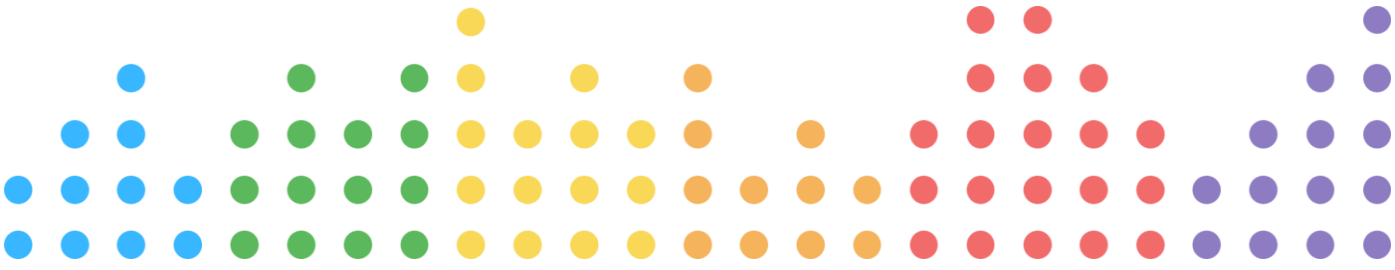
Discounting

Many of the costs and benefits associated with creative health interventions occur at different points in time. Programme costs are often incurred immediately, whereas benefits such as improved health, enhanced wellbeing, reduced healthcare utilisation, and greater community participation may continue for months or years after the intervention has ended. The Green Book therefore requires future costs and benefits to be discounted and expressed in present value terms before they are included in a cost-benefit analysis.

Discounting reflects the principle of **social time preference**: society generally prefers to receive benefits sooner rather than later and to defer costs into the future. A benefit received today is therefore valued more highly than an identical benefit received many years from now. Similarly, costs incurred in the future are valued less than costs incurred today. The Green Book refers to this as the Social Time Preference Rate (STPR).

For most creative health interventions, discounting is unlikely to have a substantial effect where costs and benefits occur within a relatively short timeframe. However, where benefits are expected to persist for several years, discounting becomes increasingly important. Analysts should therefore estimate the duration of benefits as realistically as possible and apply the Green Book discount rates to future costs and benefits before calculating the final CBA results.

The current standard discount rate in the Green Book is **3.5% per annum** for both costs and benefits over the first 30 years of an appraisal.



Distributional analysis in CBA

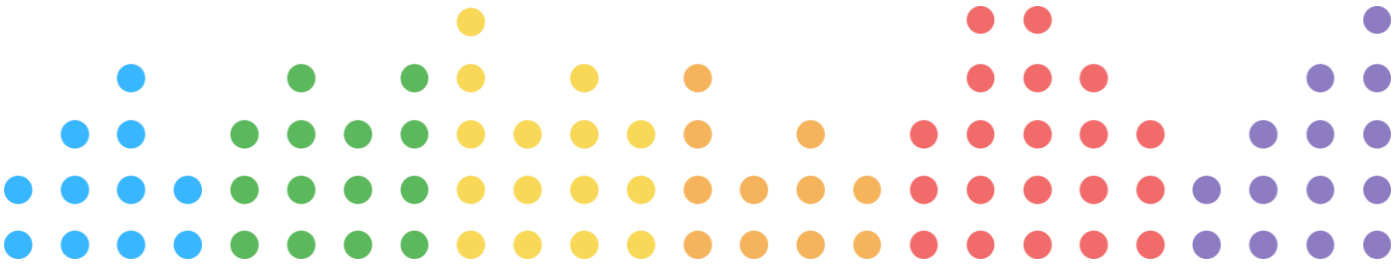
The Green Book emphasises that economic evaluation should consider not only the total value generated by an intervention, but also how that value is distributed across society. Some interventions primarily benefit individuals who are already relatively advantaged, whilst others generate benefits for groups experiencing socioeconomic disadvantage, poor health, social exclusion, or low levels of wellbeing. Distributional analysis is intended to make these differences transparent and support decision-makers in understanding who benefits from public spending.

A key concept underpinning distributional analysis is the principle of **diminishing marginal utility of income**. An additional £1 of benefit is generally considered to generate greater social value for a low-income household than for a high-income household because the additional resources represent a larger improvement in living standards and wellbeing. The Green Book therefore permits the use of **distributional weights**, often referred to as welfare weights, to reflect differences in the social value of benefits received by different groups.

Creative health interventions are often targeted towards populations experiencing health inequalities, deprivation, social isolation, disability, long-term health conditions, unemployment, or other forms of disadvantage. In these circumstances, distributional analysis may be particularly relevant because the benefits generated by an intervention may be concentrated amongst groups for whom improvements in health and wellbeing have greater social value.

The calculation of distributional weights requires information on the average income levels of the beneficiary population and applies weighting factors derived from the Green Book's recommended methodology [8]. Given the technical nature of this analysis and the additional data requirements involved, distributional weighting is likely to be most relevant and proportionate for larger evaluations or where reducing socioeconomic inequalities is a primary policy objective.

[8] The Green Book calculates the distributional weight by dividing the **median equivalised income of the national population** by the **median equivalised income of the beneficiary group**, and raising this ratio to the power of **1.3**. This gives higher weights to benefits received by lower-income groups and lower weights to benefits received by higher-income groups. Analysts should refer to Chapter 7, page 51 of the Green Book (2026) for the full methodology. Where welfare weights are applied, evaluators should **present both weighted and unweighted results** as per Green Book recommendations so that decision-makers can clearly see the impact of applying distributional weights. For most creative health evaluations, distributional analysis should be viewed as a **supplementary analytical tool** rather than a mandatory requirement.



6. Cost-Effectiveness Analysis (CEA)

This part of the Guidelines sets out the process for undertaking Cost-Effectiveness Analysis (CEA) of creative health interventions. CEA is appropriate where the decision-maker wishes to understand the cost of achieving a specified outcome and to compare the relative cost-effectiveness of alternative approaches to achieving that outcome.

Depending on the objectives of the intervention and the decision being informed, effectiveness might be measured through health-related quality of life, healthcare or social care utilisation, or another clearly defined outcome.

CEA may be particularly useful for NHS health and care commissioning decisions where there is a clearly specified outcome of interest, but its use is not restricted to a particular type of commissioner or organisation. The appropriate outcome measure should follow from the decision being informed.

Unlike CBA, CEA does not seek to value all outcomes generated by creative health interventions. Instead, it focuses on a single outcome measure and assesses the cost of achieving that outcome. This approach is particularly useful where the primary objective is to reduce healthcare utilisation and associated pressures on NHS services.

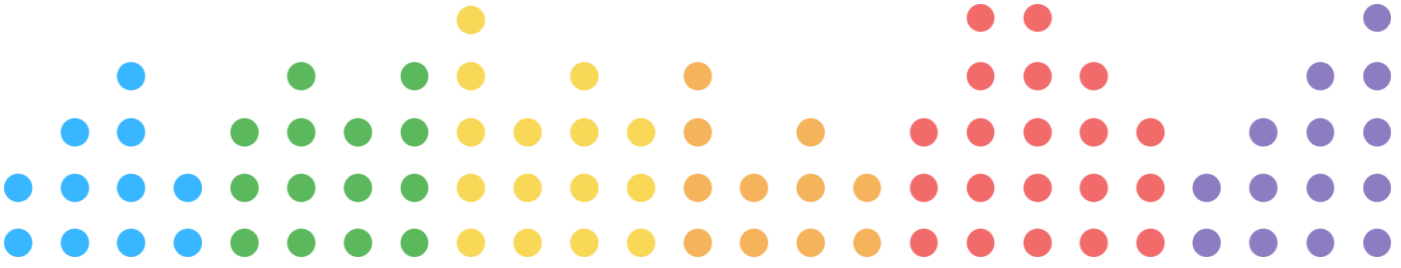
Consistent with the Core Outcome Framework (COF) set out in these Guidelines, the recommended effectiveness measures for CEA are **Number of GP visits** and **Number of A&E attendances**.

As discussed, these outcomes have been selected because they are among the healthcare utilisation measures most likely to be influenced by creative health interventions, can be collected relatively easily through participant surveys and are particularly relevant where the decision concerns healthcare utilisation and NHS resource use.

Where the primary objective of a creative health intervention is to improve health-related quality of life, evaluators may alternatively undertake a CEA using the **QALY** to produce a cost per QALY analysis. QALYs should be estimated using the health utility measures described in the Core Outcome Framework, preferably using EQ-5D. The estimated additional QALY gain attributable to the intervention should then be compared with the costs of delivering the intervention.

The resulting cost per QALY can be compared with the NICE cost-effectiveness threshold set out in these Guidelines. This approach may be particularly relevant for NHS commissioners because it enables creative health interventions to be assessed using the same broad health economic framework applied to other healthcare interventions.

Where the decision concerns a different specified outcome, evaluators should select an effectiveness measure that is appropriate to that decision and supported by the intervention's Theory of Change.



Step 1. Estimate the impact of the intervention

The first step is to estimate the impact of the intervention on healthcare utilisation (GP visits and A&E attendance) using the methods described in Section 4.2. The objective is to estimate the additional reduction in GP visits and A&E attendances attributable to the creative health intervention after accounting for deadweight.

Step 2. Estimate intervention costs

The costs of delivering the intervention should be measured using the approach set out in Section 4.3. All relevant programme delivery costs should be included, including staffing, venues, materials, management, administration, volunteer inputs and any other resources required to deliver the intervention.

Step 3. Calculate cost-effectiveness ratios

The primary output of a CEA is the Cost-Effectiveness Ratio (CER).

For GP visits this is calculated as: CER = Total Intervention Cost ÷ Reduction in GP Visits.

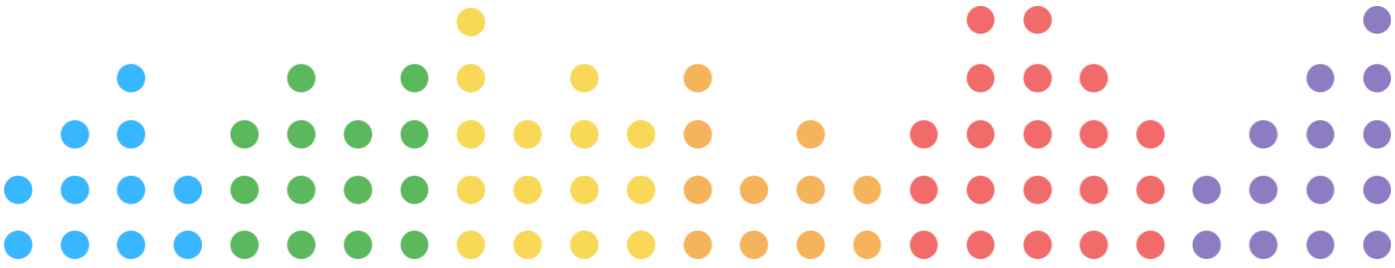
For A&E attendances this is calculated as: CER = Total Intervention Cost ÷ Reduction in A&E Attendances.

The resulting figures show the cost required to avoid one additional GP visit or one additional A&E attendance.

Where QALYs are used as the effectiveness measure, the CER should be reported as the cost per additional QALY gained: CER = Total Intervention Cost ÷ Additional QALYs gained.

The estimated additional QALY gain should be derived from the impact analysis and should reflect the duration of the health improvement. The cost per QALY can then be compared with the NICE cost-effectiveness threshold recommended in these Guidelines.

In addition to reporting cost-effectiveness ratios for GP visits and A&E attendances separately, evaluators may also estimate the overall impact of the intervention on NHS expenditure. This can be achieved by multiplying the estimated reductions in GP visits and A&E attendances by the relevant NHS unit costs and summing the resulting cost savings. The resulting figure represents the estimated NHS cost savings generated by the intervention.



Where this approach is adopted, evaluators should report the total NHS cost savings alongside the intervention costs and calculate the net cost to the NHS as:

Net NHS Cost = Intervention Cost – NHS Cost Savings

This analysis provides a useful indication of the extent to which the intervention offsets its own costs through reduced demand on NHS services. It may be particularly useful for commissioners and healthcare decision-makers whose primary objective is to understand the likely financial implications of investing in creative health interventions.

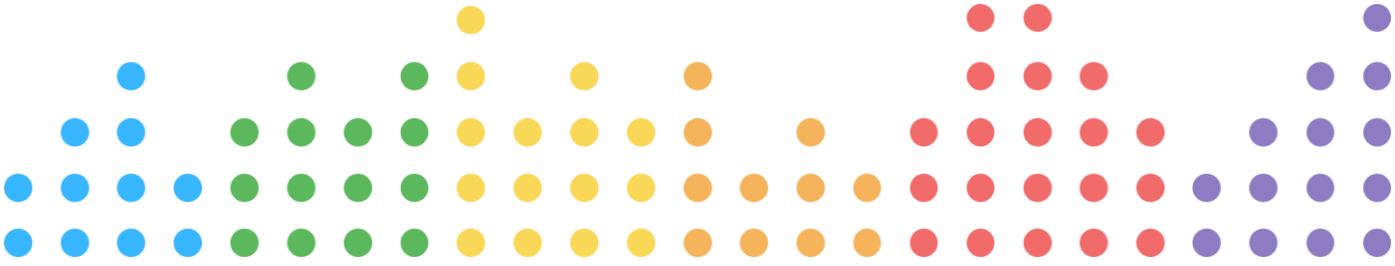
However, evaluators should note that this represents a budget impact assessment focused solely on healthcare utilisation and should not be interpreted as a measure of the full social value generated by the intervention.

Ex-post CEA
The purpose of an ex-post CEA is to assess the cost-effectiveness of a creative health intervention that has already been delivered. The observed reductions in GP visits and A&E attendances should be used to calculate the CERs. Results should be accompanied by confidence intervals and sensitivity analysis where possible.

Ex-ante CEA
The purpose of an ex-ante CEA is to estimate the likely cost-effectiveness of a proposed future intervention. Forecasts of reductions in GP visits and A&E attendances should be based on the strongest available evidence from previous evaluations, pilot studies, academic research, or comparable programmes. The evidence used should meet the standards of rigour set out in Section 4.2 of these Guidelines.
Given the **uncertainty** associated with forecasting future outcomes, evaluators should present central, optimistic and conservative scenarios.

Sensitivity analysis should focus on assumptions regarding participation rates, intervention costs, healthcare utilisation impacts, and the persistence of benefits over time.

As with CBA, all assumptions should be clearly documented and justified to support transparent and evidence-based decision-making.



7. Worked examples of economic evaluations

This section provides worked examples illustrating how the economic evaluation approaches set out in these Guidelines can be applied in practice. It includes four hypothetical examples designed to demonstrate the application of these guidelines in different contexts featuring different impact measurement methods and outcome metrics presented in CBA and CEA contexts. We also present applied analyses of two creative health programmes in the UK.

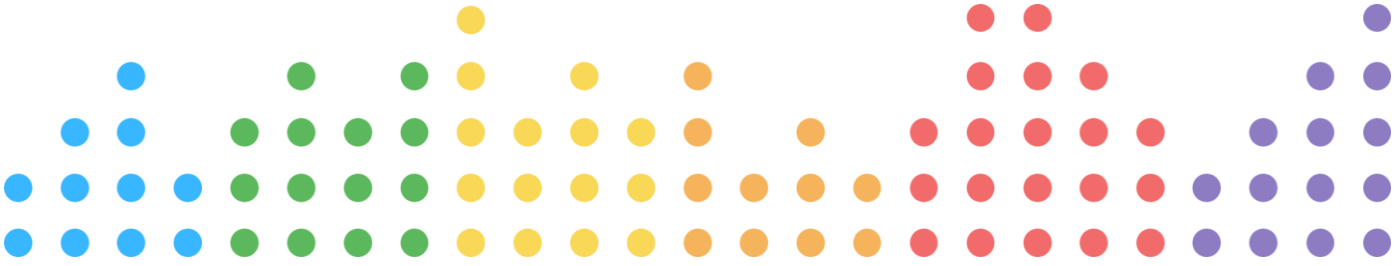
7.1 Hypothetical worked examples

The following hypothetical worked examples illustrate how the economic evaluation methods and principles set out in these Guidelines can be applied to different types of creative health interventions. The examples cover both Cost-Benefit Analysis (CBA) and Cost-Effectiveness Analysis (CEA) and use different approaches to impact measurement and different outcome metrics. They are designed to demonstrate the main analytical steps involved, from estimating the additional impact of an intervention and identifying relevant costs and benefits through to calculating and interpreting the resulting economic evaluation metrics.

The examples are illustrative rather than prescriptive. The appropriate evaluation design, outcome measures and analytical approach will depend on the objectives, scale and characteristics of the intervention, the decision that the evaluation is intended to inform, and the quality and availability of data. In practice, evaluators should apply the principles of proportionality and rigour described in these Guidelines when determining the appropriate approach.
The results presented in the hypothetical examples can be interpreted as follows:

- The **positive Net Present Social Value** indicates that, on the assumptions used, the estimated social benefits of the programme exceed its costs. This provides evidence in support of continued or expanded investment, although the decision-maker should also consider the robustness of the impact estimates, uncertainty around the valuation assumptions, affordability and the distribution of costs and benefits.

- **The cost-effectiveness ratios** provide the decision-maker with an estimate of the cost of achieving the specified outcome. This can be compared with relevant benchmarks, alternative interventions and available budgets when considering whether the programme represents an appropriate use of resources.



Worked Example 1: Cost-Effectiveness Analysis estimating a cost per QALY of a Creative Arts Programme

This is a six-month creative arts programme for adults experiencing mild-to-moderate depression and anxiety. The programme involves weekly group-based visual arts and creative activities delivered by trained creative practitioners and is attended by 500 participants.

The total cost of delivering the programme, including creative practitioners, venue hire, referral coordination, programme management, administration and materials, is £600,000.

As this is a large-scale intervention, the evaluation uses a quasi-experimental statistical method using propensity score matching to estimate impact. Participants are matched to similar non-participants using observable characteristics such as age, gender, baseline health, baseline EQ-5D score, deprivation, mental health status and prior healthcare utilisation.

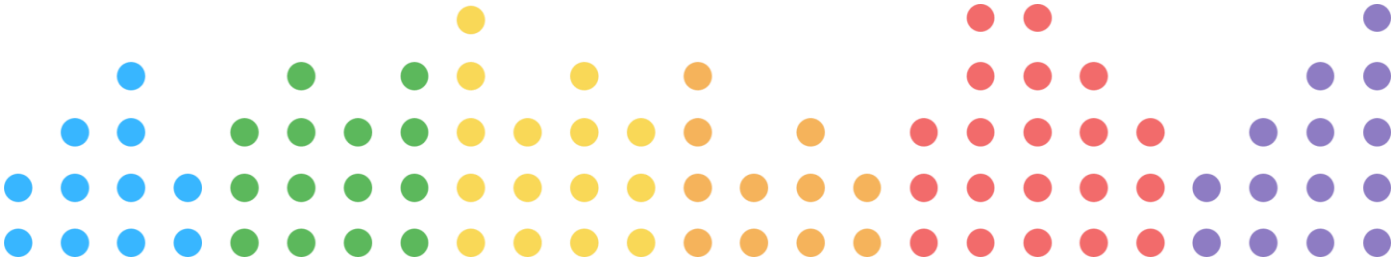
The primary outcome for the CEA is health-related quality of life, measured using EQ-5D. Responses to the EQ-5D are converted into health utility values, where 0 represents being dead and 1 represents full health.
The analysis finds that participants experienced an average additional improvement in health utility of 0.10 compared with the matched control group. The improvement is estimated to persist for six months.

The QALY gain per participant is therefore calculated as: 0.10 × 0.5 years = 0.05 QALYs. Across 500 participants, the total additional QALY gain is 0.05 × 500 = 25 QALYs.

The CER is calculated as £600,000 ÷ 25 = £24,000 per QALY. This is the cost per QALY gained and represents the estimated cost of generating one additional QALY.

This is below the £30,000 per QALY NICE cost-effectiveness threshold.

The analysis should be accompanied by sensitivity analysis examining uncertainty around the estimated health utility impact, the duration of the health improvement and programme costs. For example, if the health improvement persisted for only three months rather than six months, the programme would generate 12.5 QALYs and the estimated cost per QALY would increase to £48,000, exceeding the £30,000 threshold.



Worked Example 2: Cost-Benefit Analysis of a Regional Arts-on-Prescription Programme

An Integrated Care System commissions a regional arts-on-prescription programme for adults experiencing mild-to-moderate anxiety, depression, loneliness, or long-term health conditions. The programme is delivered across community arts centres, libraries, museums, and voluntary-sector venues, with referrals made through social prescribing link workers and primary care networks. The programme runs for 6 months and is attended by 1,000 participants.

The total cost of delivering the programme, including creative practitioners, venues, referral coordination, programme management, administration, training and materials, is £900,000.

As this is a large-scale intervention, the evaluation uses a quasi-experimental statistical method using propensity score matching to estimate impact. Participants are matched to similar non-participants using observable characteristics such as age, gender, baseline health, baseline wellbeing, deprivation, prior GP visits, prior A&E attendances and long-term condition status.

Since the programme goes beyond just health impacts the focus of the CBA is on the impact on life satisfaction. The analysis finds that participants experienced an average improvement of 0.15 points in life satisfaction for 6 months compared to the control group. Using an illustrative wellbeing value of £17,000 per one-point annual improvement in life satisfaction, the estimated wellbeing benefit per participant is:

Life satisfaction benefit per participant = 0.15 × (£17,000 x 0.5) = £1,275.

The estimated total wellbeing benefit is: £1,275 x 1,000 = £1,275,000.

The evaluation also measures healthcare utilisation. Compared with the matched control group, the programme reduces GP visits by 1.2 per participant and A&E attendances by 0.4 per participant.

Across 1,000 participants, this equals 1,200 fewer GP visits and 400 fewer A&E attendances. Using illustrative NHS unit costs of £45 per GP visit and £220 per A&E attendance, the estimated NHS savings are:

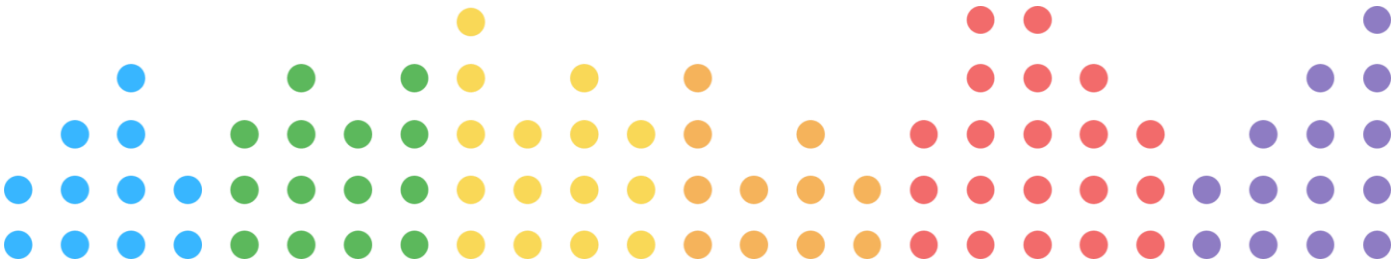
GP savings = 1,200 × £48 = £57,600

A&E savings = 40 × £220 = £88,000

Total NHS savings = £145,600

Total benefits = £1,275,000 + £145,600 = £1,420,600

Net Present Social Value = £1,420,600 − £900,000 = £520,600.



Worked Example 3: Cost-Benefit Analysis of a Dance and Movement Programme using an RCT and the EQ-5D health utility measure

An Integrated Care System commissions a 12-month dance and movement programme for adults living with long-term musculoskeletal conditions, mild depression, or reduced mobility. The programme is delivered across leisure centres, community halls, cultural venues, and voluntary-sector settings and is attended by 1,500 participants.

The total cost of delivering the programme, including dance practitioners, venue hire, referral coordination, programme management, administration, training and materials, is £1,800,000.

As this is a large-scale intervention, the evaluation uses a Randomised Controlled Trial (RCT) to estimate impact. Eligible participants are randomly allocated either to receive the dance and movement programme immediately or to a control group receiving usual care. Randomisation ensures that, on average, the intervention and control groups are similar in both observed and unobserved characteristics before the programme begins.

The primary outcome for the CBA is health-related quality of life, measured using EQ-5D. Responses to the EQ-5D are converted into health utility values, where 0 represents being dead and 1 represents full health. The analysis finds that, compared with the control group, participants experience an average health utility improvement of 0.03 for 12 months.

The QALY gain per participant is therefore:

QALY gain per participant = 0.03 × 1 = 0.03 QALYs.

Across 1,500 participants, the total QALY gain is:

Total QALYs = 0.03 × 1,500 = 45 QALYs.

Using the value of £83,000 per QALY, the estimated health benefit is:

Health benefit = 45 × £83,000 = £3,735,000.

The evaluation also measures healthcare utilisation. Compared with the control group, the programme reduces GP visits by 1.6 per participant and A&E attendances by 0.40 per participant.

Across 1,500 participants, this equals 2,400 fewer GP visits and 600 fewer A&E attendances. Using the NHS unit costs of £48 per GP visit and £220 per A&E attendance, the estimated NHS savings are:

GP savings = 2,400× £48 = £115,200.

A&E savings = 600 × £220 = £132,000.

Total NHS savings = £247,200.

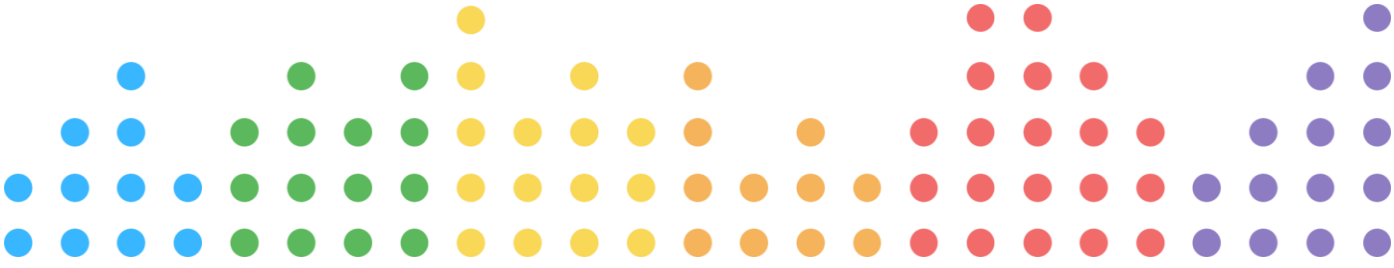
Total benefits are therefore:

Total benefits = £3,735,000 + £247,200 = £3,982,200.

Net Present Social Value = £3,982,200 − £1,800,000 = £2,182,200.

Benefit-Cost Ratio = £3,982,200 ÷ £1,800,000 = 2.21.

Sensitivity analysis would then be undertaken using alternative assumptions regarding the duration of benefits, healthcare utilisation impacts and trial attrition.



Worked Example 4: Cost-Effectiveness Analysis of a Community Singing Programme

A local authority commissions a 12-week community singing programme for older adults experiencing loneliness and mild mental health difficulties. The programme involves weekly group singing sessions delivered by a trained facilitator and is attended by 100 participants.

The total cost of delivering the programme, including staffing, venue hire, administration and materials, is £30,000.

Since the intervention is relatively small-scale, to assess effectiveness, participants completed surveys before and after the programme on self-reported GP visits and A&E attendances and the counterfactual adjustment methodology - as described in Section 4.2 - was applied. The participant attribution estimate was 18% deadweight which is lower than the benchmark deadweight figure of 22.5% and hence 22.5% was used as the deadweight estimate for the intervention.

The analysis finds that participants reported an average of 2.0 GP visits during the pre-intervention period and 1.5 GP visits during the post-intervention period. Across 100 participants, this represents a reduction of 50 GP visits. It is estimated that 38.75 of these avoided GP visits are attributable to the intervention (50 * 0.775).

Similarly, participants reported an average of 0.20 A&E attendances during the pre-intervention period and 0.12 A&E attendances during the post-intervention period. Across 100 participants, this represents a reduction of 8 A&E attendances. After applying additionality and deadweight adjustments, the intervention is estimated to have prevented 6.2 A&E attendances over 100 people (8 * 0.775).

The cost-effectiveness ratios are then calculated as follows:

Cost per GP visit avoided = £30,000 ÷ 38.75 = £774

Cost per A&E attendance avoided = £30,000 ÷ 6.2 = £4,839

These figures indicate the cost required to prevent one additional GP visit and one additional A&E attendance respectively.

Evaluators may also estimate the financial impact on NHS services. Using the NHS unit costs of £48 per GP visit and £220 per A&E attendance, the estimated NHS savings are:

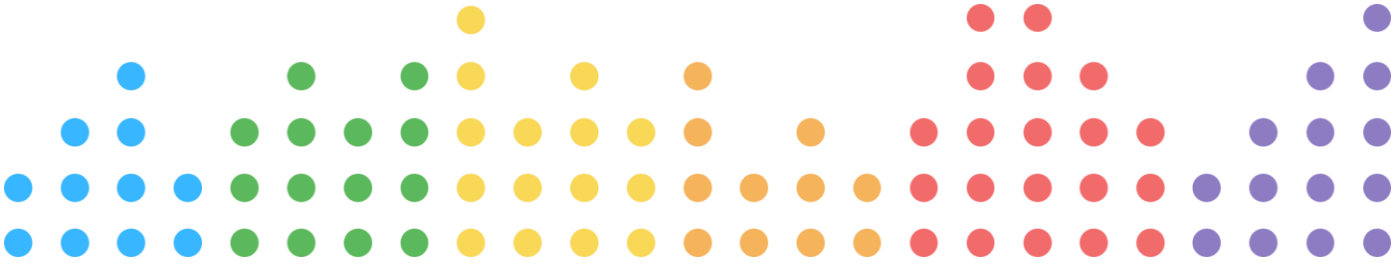
GP savings = 38.75 × £48 = £1,860

A&E savings = 6.2 × £220 = £1,364

Total NHS savings = £3,224

The net NHS cost of the programme = £30,000 − £3,224 = £26,776

Sensitivity analysis would then be undertaken using alternative assumptions regarding additionality, deadweight and healthcare utilisation impacts to assess the robustness of the findings.



7.2 Applied analyses

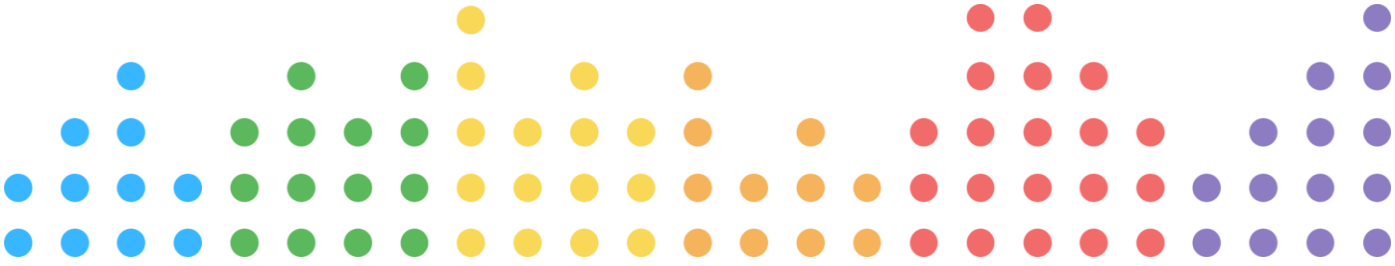
In contrast to the hypothetical worked examples in Section 7.1, this section illustrates how the economic evaluation methods set out in these Guidelines can be applied to real-world creative health programmes using the data currently available to delivery organisations. It draws on a separate quantitative study undertaken alongside the development of the Guidelines, which is presented in full in Annex C, using data from two existing creative health programmes. The purpose of the exercise was to explore how far existing programme data could support a Green Book-aligned economic evaluation as per these Guidelines and to demonstrate what can, and cannot, be estimated where the available evidence falls short of the data that would ideally be collected for a robust evaluation.

The analysis should not be interpreted as a model of best-practice, a benchmark for future evaluations, or a definitive assessment of the two programmes considered. It is deliberately a pragmatic example of what can be done with the best available information currently collected by creative health programmes. The exercise is therefore as informative about current data limitations as it is about the estimated economic results. In particular it shows the analytical steps that can be taken when standardised wellbeing data are available, while also illustrating why stronger evaluation design and more systematic data collection are required to support robust commissioning and investment decisions.

Selecting programmes, assembling the data and analytical approach

A standardised data template was developed and shared with creative health organisations and practitioners. The template requested information on programme characteristics, costs, participant demographics and pre- and post-intervention measures, including life satisfaction, EQ-5D, health service utilisation, and the Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS). The intention was to identify real programmes with data that could support an applied economic evaluation rather.

In practice, most programmes could not provide the combination of data required. Common issues included reliance on qualitative feedback, collection of outcomes only after participation, use of non-standardised measures, and the absence of pre- and post-intervention health or wellbeing data. Two programmes - ***Flip of the Coin*** at the Easter Ross REALITIES hub and 'Wellbeing Your Way' by ***Odd Arts*** - offered the most suitable datasets and were therefore selected for the applied analyses.



However, it is noted that even for these programmes, the datasets remained limited in sample size and completeness.

None of the core outcomes proposed in these Guidelines was assessed in the programmes evaluated and hence SWEMWBS was the principal outcome in the analysis as this was collected in both programmes. SWEMWBS is one of the Supplementary outcomes recommended in these Guidelines.

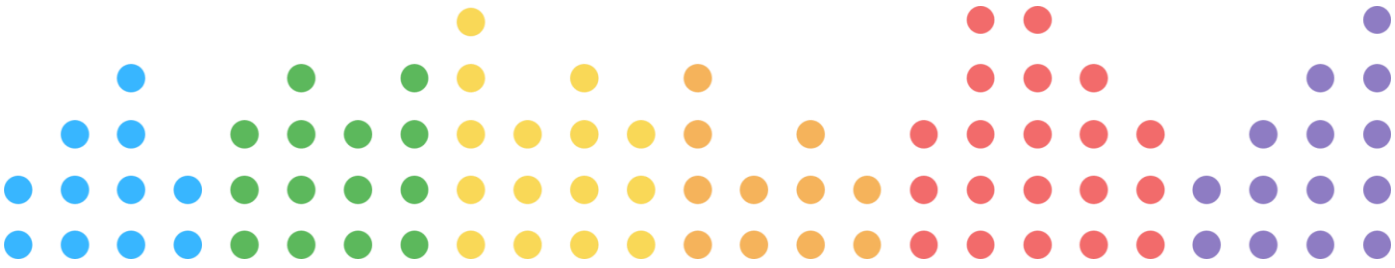
Because SWEMWBS is not a preference-based measure and does not have a direct monetary value for Cost-Benefit Analysis, we constructed a 'valuation bridge' using a method developed by Fujiwara et al. (2017). Using the Understanding Society dataset the statistical relationship between SWEMWBS and life satisfaction was estimated and translated into monetary values using the WELLBY framework. This allowed observed changes in mental wellbeing to be expressed as an indicative social value estimates within a Green Book-consistent framework.

Neither programme collected direct data on healthcare utilisation or workplace productivity. Secondary benefits were therefore estimated indirectly. Relationships between SWEMWBS and GP visits, outpatient use, hospitalisation and presenteeism were estimated using Understanding Society and then applied to the programme-level changes in SWEMWBS. These changes were monetised using relevant NHS unit costs and earnings data.

The impact of the programme on SWEMWBS was estimated using propensity score matching to construct an external comparison group from Understanding Society, followed by a Difference-in-Differences approach on the matched individuals to compare changes over time between programme participants and matched non-participants.

Key limitations

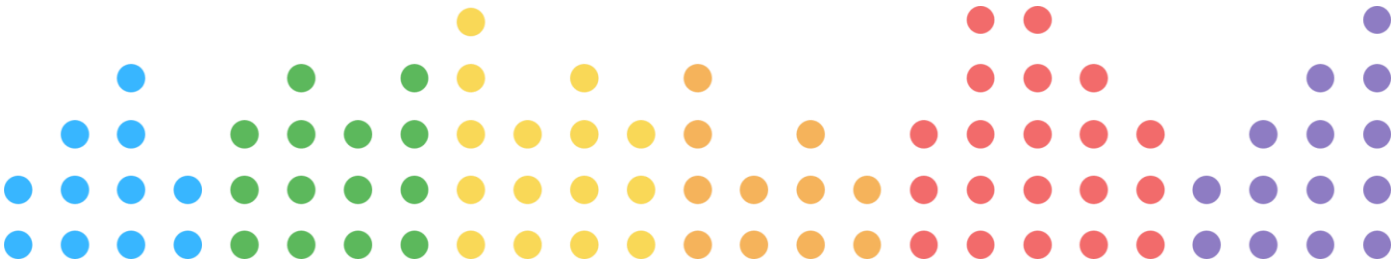
The principal constraint was the quality and quantity of the programme data. The analytical samples were small - 33 participants for Flip of the Coin and 60 participants in the final matched Odd Arts analysis - and important demographic and socioeconomic variables were missing. Neither programme had collected a prospective, contemporaneous control group. The comparison groups therefore had to be constructed from Understanding Society and the closest available waves of data that include SWEMWBS data did not align precisely with the intervention periods for the two programmes.



Programme-specific issues add further uncertainty. Flip of the Coin did not collect baseline wellbeing prospectively and was still ongoing at follow-up. For Odd Arts, many participants were in a custodial setting, a population not represented in Understanding Society. In both programmes, healthcare utilisation and productivity benefits were inferred rather than observed directly, and some costs or additionality factors could not be fully measured.

The results should therefore be treated as a pragmatic approach to economic evaluation constrained within the current availability of data. They are indicative of the possible order of magnitude of impact and value.

Future studies would be strengthened by following the recommendations in these Guidelines and in particular in relation to prospective outcome measurement, data collection on core outcomes including health utility, life satisfaction and direct healthcare-use data, larger samples and credible comparison groups.



Applied Example 1: Flip of the Coin (REALITIES, Easter Ross)

REALITIES is a creative health initiative working across Scotland to support healthier and more resilient communities in areas affected by health inequalities. The applied evaluation used data from Flip of the Coin at the Easter Ross hub, which works with people with lived experience of intergenerational trauma, poverty, poor mental health, substance use and contact with the criminal justice system. It was selected because it provided the most complete available pre- and post-intervention wellbeing data.

The dataset contained 33 participants. The hub collected the 14-item WEMWBS, which was converted to the seven-item SWEMWBS used in the valuation model. Average SWEMWBS increased from 17.99 at baseline to 25.94 at follow-up, an improvement of 7.94 points. To estimate the counterfactual, participants were matched to respondents in Scotland from Understanding Society using baseline wellbeing and available demographic and socioeconomic characteristics. A Difference-in-Differences model estimated an additional 6.74-point increase in SWEMWBS relative to the matched comparison group.

The estimated SWEMWBS change was monetised using the WELLBY valuation approach. Secondary benefits associated with reduced GP and outpatient use, lower probability of hospital admission and reduced presenteeism were estimated indirectly from the relationship between SWEMWBS and these outcomes in Understanding Society. A six-month duration adjustment was applied to reflect uncertainty about the persistence of the wellbeing effect.

The results are subject to substantial uncertainty. The sample is small; baseline wellbeing was not collected prospectively at programme entry; follow-up occurred while the programme was ongoing; and the external comparison group was observed over a different time period. Healthcare utilisation and productivity were not measured directly in the programme. The estimates should therefore be read as indicative rather than definitive.

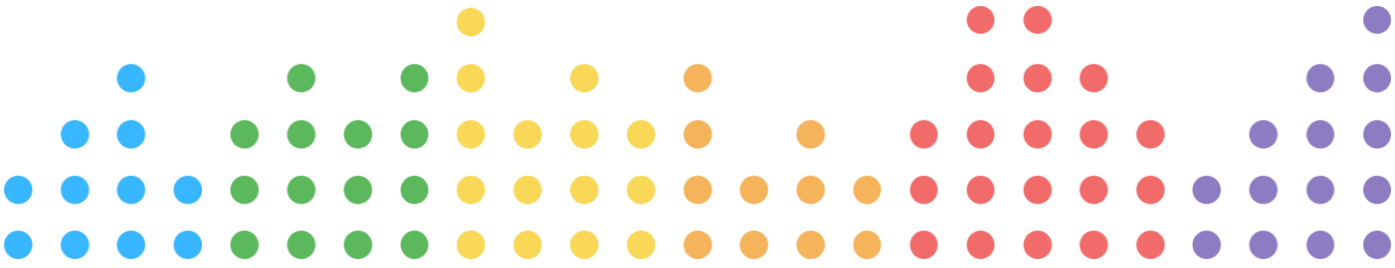
Indicative results:

Estimated attributable change in SWEMWBS: +6.74 points

Monetary value (change in SWEMWBS plus healthcare utilisation and productivity): approximately £333,969 in total; £10,120 per participant

Estimated cost per beneficiary: £756

Indicative benefit-cost ratio: 13.38



Applied Example 2: Odd Arts – Wellbeing Your Way

Odd Arts is a theatre and arts organisation working with people experiencing high levels of inequality and disadvantage. The applied evaluation considered its Wellbeing Your Way programme, delivered during 2024 in a custodial facility, a community organisation and schools. The programme was selected because it had pre- and post-intervention SWEMWBS data, together with some information on participant characteristics.

The dataset contained 61 participants with baseline and follow-up SWEMWBS scores collected around 12 weeks apart. Average SWEMWBS increased from 14.54 to 22.03, an improvement of 7.49 points. One participant was excluded from the matching analysis because gender information was missing, leaving 60 participants. A matched comparison group was constructed from Understanding Society respondents in North West England using baseline SWEMWBS and gender, followed by the same Difference-in-Differences approach used for Flip of the Coin. The estimated attributable improvement was 4.30 SWEMWBS points.

The wellbeing improvement was monetised through the WELLBY framework. Estimated secondary benefits included reductions in GP visits, outpatient visits and hospital admission. Presenteeism was not monetised because most participants were not in employment at baseline, particularly within the custodial setting. The same six-month duration adjustment was applied to the primary and secondary benefits.

This example presents particular counterfactual challenges. People living in custodial institutions are not represented in Understanding Society, so there is no directly comparable population in the national household survey. Some imbalance in baseline wellbeing remained after matching. Programme cost information may also omit some resource categories, meaning that the estimated cost per participant is likely to be a lower bound. As a result, the benefit-cost ratio may overstate value and should be interpreted only as an indicative order-of-magnitude estimate.

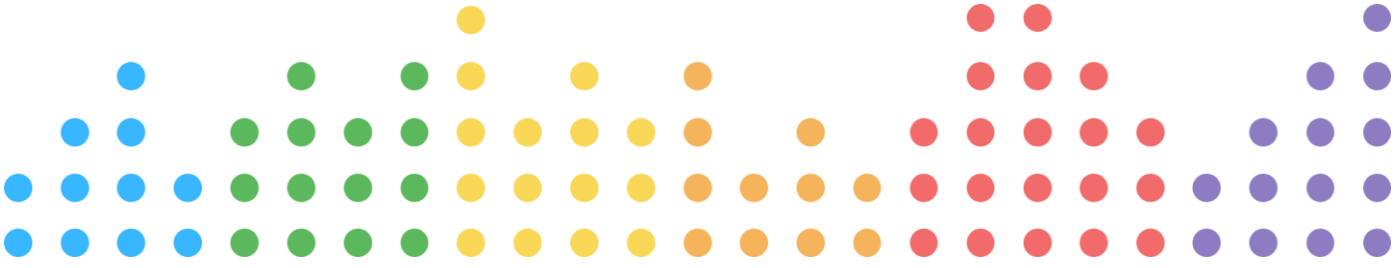
Indicative results:

Estimated attributable change in SWEMWBS: +4.30 points

Monetary value (change in SWEMWBS plus healthcare utilisation and productivity): approximately £550,815 in total; £9,030 per participant

Estimated cost per participant: £707

Indicative benefit-cost ratio: 12.77



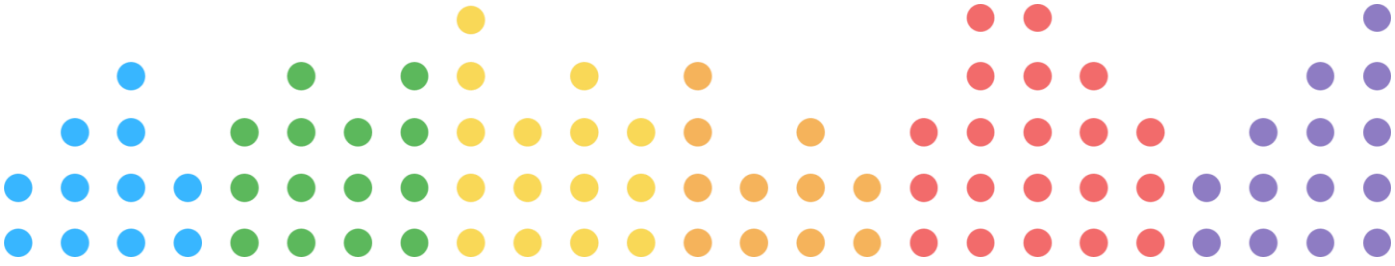
8. Economic evaluation template

This section provides a standardised template for undertaking and reporting economic evaluation in creative health in accordance with these Guidelines. It is intended to translate the methodological recommendations set out in the preceding sections into a concise and practical framework that can be applied across different types and scales of creative health interventions.

The template guides users through the main stages of an economic evaluation: defining the intervention and the decision that the evaluation is intended to inform; setting out the Theory of Change and relevant outcomes; selecting between Cost-Benefit Analysis (CBA) and Cost-Effectiveness Analysis (CEA); determining a proportionate approach to impact evaluation; measuring outcomes and costs; estimating attributable impacts; and reporting the economic results, assumptions and uncertainty. In doing so, it provides a consistent structure through which evaluations can be planned, undertaken and communicated.

The template should be regarded as a minimum framework rather than a comprehensive specification for every evaluation. Users are encouraged to undertake additional analysis, collect additional outcomes or apply more sophisticated evaluation methods where these are relevant, feasible and proportionate to the intervention and the decision being informed. In particular, larger or strategically important programmes may require substantially more detailed impact evaluation, economic modelling, sensitivity analysis and reporting than can be captured within this concise template.

Not every field in the template will be relevant to every intervention. The template should therefore be applied proportionately, with users referring to the relevant sections of these Guidelines for further methodological detail.



Creative Health Economic Evaluation Template

Purpose: This is a concise planning and reporting template to record the key decisions, evidence and results of an economic evaluation for creative health interventions. It is intended to assist the user in carrying out an economic evaluation aligned with the Guidelines.

Instructions: The template should be completed at the outset and updated as evidence becomes available. The level of analysis should be proportionate to the intervention and the decision being informed.

1. Intervention and decision context

Question	Complete
Intervention / programme	
Target population	
Scale and duration	
Problem or need being addressed	
Decision this evaluation will inform	
Primary decision-maker / audience	
Evaluation timing	☐ Ex-ante ☐ Ex-post

2. Theory of Change and outcomes

In a few sentences, describe how the intervention is expected to improve participants' health, wellbeing and lived experience and then complete the Theory of Change table below:

_____

_____

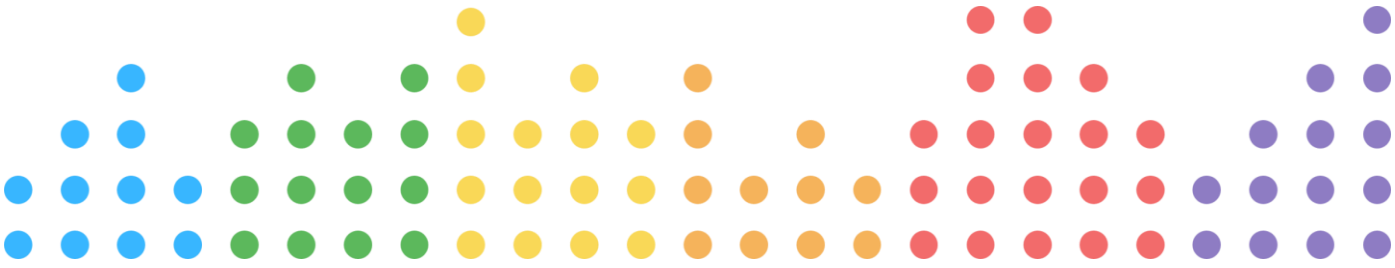
Theory of Change element	Complete
Main activities and outputs	
Primary outcome(s)	
Other important outcomes	
Who is expected to benefit?	

Core outcomes considered, where relevant and feasible:
☐ Health-related quality of life / QALYs
☐ Life satisfaction
☐ GP utilisation
☐ A&E utilisation
☐ Other(s): _____
_____

3. Choosing the type of economic evaluation

☐ Do you want to measure the intervention's full benefits and assess whether they justify its costs? → Cost-Benefit Analysis
☐ Do you want to understand the cost of achieving a specified outcome (e.g. reduced healthcare utilisation)? → Cost-Effectiveness Analysis

Selected approach: ☐ CBA ☐ CEA
Reason: _____



4. Impact evaluation

Here consider the scale, cost, participant numbers, strategic importance and the decision the evidence will inform in order to understand the level of analysis required:

☐ **Higher evidential requirement** — experimental or appropriate quasi-experimental design with a credible control/comparison group.
☐ **Proportionate evidential requirement** — pragmatic impact evaluation where a formal counterfactual is disproportionate.

Impact evaluation method:
☐ RCT
☐ Quasi-experimental
☐ Interrupted Time Series
☐ Pre-post + counterfactual adjustment
☐ Other: _____

5. Outcomes and impact

Outcome	Measure / data source	Observed change	Attributable impact

If using pre-post measurement with counterfactual adjustment:

Item	Value
Participant deadweight estimate	_____ %
Benchmark deadweight	22.5%
Deadweight used (higher figure of the two)	_____ %
Other material attribution / displacement adjustment	_____ %

6. Costs

Main resource / cost	£
Staff, practitioners & volunteers (volunteer time to be valued at minimum wage)	
Programme development costs & training costs	
Venues facility costs	
Materials & equipment	
Management & administration	
Other relevant costs	
TOTAL INTERVENTION COST	£

Cost per participant: £_____

7. Economic evaluation results

Complete the relevant section only:

A. Cost-Benefit Analysis (CBA)

Benefit	Attributable impact	Unit value	Total value
		£	£
		£	£
		£	£
TOTAL BENEFITS			£

Total costs: £_____ NPSV: £_____ BCR: _____ : 1

☐ Overlapping outcomes have been considered to avoid double counting.

B. Cost-Effectiveness Analysis (CEA)

Item	Result
Primary effectiveness outcome	
Attributable outcome achieved	
Total intervention cost	£
Cost-Effectiveness Ratio	£_____ per _____

8. Uncertainty and conclusion

Key assumptions / sources of uncertainty:

_____

Sensitivity analysis — how do results change under reasonable alternative assumptions?

_____

Final quality check	✓
Decision and CBA/CEA choice clearly specified	☐
Theory of Change and relevant outcomes specified	☐
Impact method is proportionate; counterfactual/deadweight addressed	☐
Costs and benefits/outcomes appropriately measured; double counting checked	☐
Assumptions, uncertainty and limitations reported	☐

Headline findings:

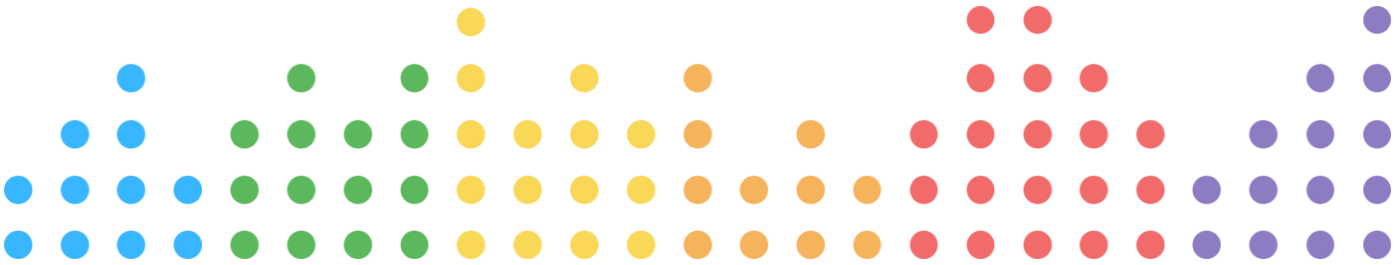
_____
_____
_____
_____

What does this mean for the commissioning / investment decision?

_____
_____
_____

Principal limitations of the evidence and caveats:

_____
_____
_____
_____



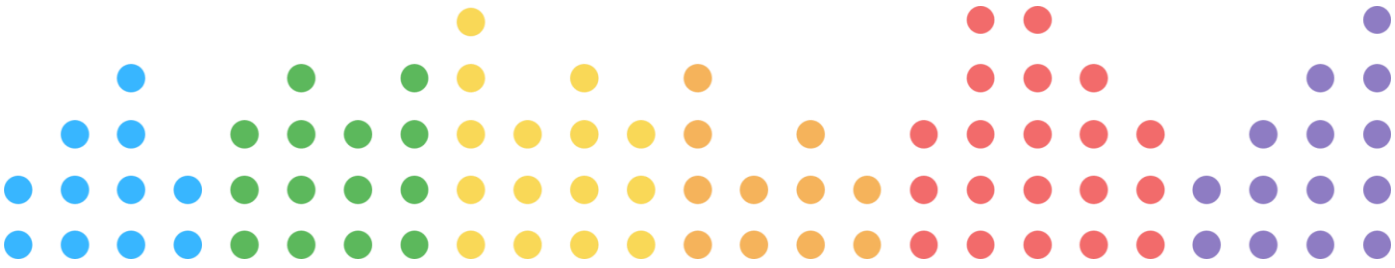
Annexes

Annex A. Examples of experimental and quasi-experimental methods that can be used for creative health evaluations

Randomised Controlled Trials (RCTs) are experimental methods that estimate the causal impact of an intervention by randomly assigning eligible participants to either an intervention group or a control group. The control group provides an estimate of the counterfactual. Because assignment is determined by chance rather than through participant or provider choice, the two groups are expected to be the same in both observed and unobserved characteristics before the intervention begins. This ensures that any differences in outcomes observed after the intervention between the two groups can be attributed to the impact of the intervention itself. RCTs are therefore widely regarded as the most robust approach for establishing causal impact and are often considered the benchmark against which other evaluation methods are compared. However, they may not always be feasible, practical, or ethically appropriate in creative health settings, particularly where programmes are already established, resources are limited, or withholding services from eligible participants is not possible.

Regression Discontinuity Design (RDD) is a quasi-experimental method that estimates the impact of an intervention by exploiting situations where eligibility is determined by a clearly defined threshold point. For example, a creative health intervention may only be offered to individuals with a depression score below a certain level, to people above a particular age, or to those living in areas meeting specific deprivation criteria. Individuals just above and just below the threshold are often very similar in most respects, except that one group receives the intervention while the other does not. The group that does not receive the intervention therefore serves as the control group. By comparing outcomes for individuals on either side of the eligibility threshold, the method can generate credible estimates of causal impact. When the eligibility rule is strictly applied and individuals cannot manipulate their position relative to the threshold, RDD is widely regarded as one of the strongest quasi-experimental approaches and can produce evidence that approaches the robustness of an RCT.

Difference-in-Differences (DiD) is a quasi-experimental method that estimates impact by comparing changes in outcomes over time between an intervention group and a control group. The control group serves as an estimate of the counterfactual by showing how outcomes would have changed over time in the absence of the intervention. Rather than comparing outcome levels at a single point in time, the method examines whether outcomes improve more for participants receiving the intervention than for a similar



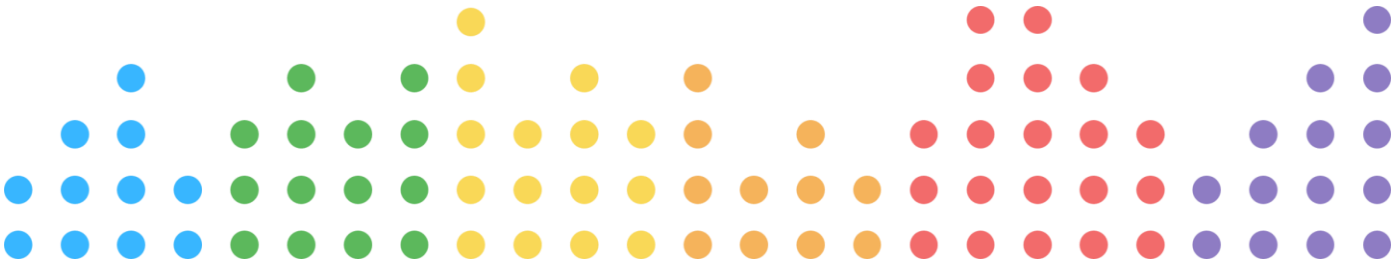
group that did not receive it. This approach helps control for fixed differences between groups and for broader changes that affect both groups equally, such as seasonal effects, policy changes, or wider economic and social trends. The credibility of the method depends on the assumption that, without the intervention, both groups would have followed similar trends over time, often tested by looking at whether historic trends in the outcomes are similar. When this assumption is plausible, DiD can provide strong evidence of causal impact and is widely used in policy and programme evaluation.

Propensity Score Matching (PSM) is a quasi-experimental method that seeks to estimate the counterfactual by creating a control group that closely resembles the intervention group. The method begins by estimating each individual's likelihood of participating in the intervention based on observable characteristics such as age, health status, prior service use, socioeconomic circumstances and so on. Participants are then matched with non-participants who have a similar probability of receiving the intervention. These matched non-participants serve as the control group and provide an estimate of what outcomes might have occurred in the absence of the intervention. By comparing outcomes between these similar groups, the method attempts to reduce selection bias and improve the credibility of causal conclusions. However, because matching can only account for observed characteristics, differences in unmeasured factors may still remain. As a result, PSM provides a useful approximation of the counterfactual but cannot eliminate all sources of bias.

Multivariate Regression is a quasi-experimental method that estimates the relationship between an intervention and an outcome while accounting for other factors that may influence that outcome. In causal analysis, the method seeks to approximate the counterfactual by statistically controlling for differences between participants who receive the intervention and those who do not. Regression uses information on observable characteristics - such as age, gender, health status, socioeconomic circumstances, or prior outcomes – to control for the effect of these factors to better single out the influence of the intervention.

The validity of this approach depends on whether all important factors influencing both participation and outcomes have been measured and included in the analysis. While regression can help reduce bias arising from observable differences, it cannot fully account for unmeasured factors such as motivation, personal preferences, or social support. Consequently, regression analysis can strengthen evidence of impact but generally provides weaker evidence of causality than experimental or stronger quasi-experimental approaches such as DiD.

Interrupted Time Series Analysis (ITS) is a quasi-experimental method that estimates the impact of an intervention by analysing repeated interventions over time for a group



of individuals. For each intervention episode, outcomes are assessed before and after the intervention is introduced. For example, a care-home singing programme is delivered on a cycle whereby it is provided for a period of three months and then paused for one month, before being reinstated again for three months and so on. Residents' health outcomes would be measured monthly throughout and if health consistently improves during periods when the programme is running and declines or stabilises when it is withdrawn, the repeated interruptions provide evidence that the observed changes are attributable to the intervention rather than to underlying trends or external factors as long as there is no other external factor that fluctuates at the same time that could also explain the trends.

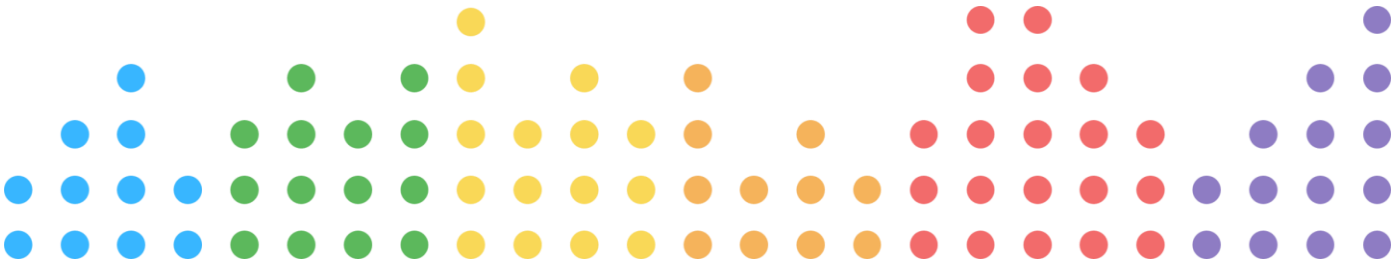
For example, if wellbeing improves during a youth creative programme delivered during the summer school holidays each year, some of that improvement might reflect the effect of being on holiday rather than the intervention itself. Care should be taken in these circumstances and where seasonality or other time-varying confounding cannot adequately be addressed, this should be reported explicitly and causal claims should be qualified accordingly.

ITS analysis does not necessarily require repeated observations of the same individuals as the method can be implemented using either longitudinal data from the same participants over time or repeated cross-sectional samples drawn from the same underlying population. The key requirement is that the delivery of the same intervention is repeated and that there are sufficient observations before and after the interventions.

Unlike other experimental and quasi-experimental methods, ITS does not require a separate control group. Instead, in ITS the counterfactual is constructed using the pattern and trend observed in the outcome prior to the interventions on a repeated basis and it is this repetition that can provide the evidence of impact. This is because a repeated correlation between the intervention and outcomes would tend to rule out other factors (e.g. other policies, programmes, or macroeconomic impacts) driving the outcomes because it would be unlikely that they are fluctuating at the exact same rate as the intervention.

Because no control group is used, ITS relies on a sufficiently large number of repeated interventions and observations before and after the interventions to establish a reliable trend and distinguish intervention effects from normal fluctuations over time and the impacts of other factors such as other programmes. This means that it requires long time series data. Under these circumstances ITS can provide evidence of causal impact in real-world settings where randomisation or control group analysis are not feasible.

Case-Control Studies are quasi-experimental methods that originated in epidemiology and are designed to investigate the causes of particular outcomes by working



backwards from the outcome to potential explanatory factors. Unlike experimental studies, which begin by assigning individuals to an intervention and then observing subsequent outcomes, case-control studies start with individuals who have experienced the outcome of interest (the cases) and compare them with similar individuals who have not experienced that outcome (the controls). Researchers then examine whether prior exposure to the intervention explains (part of) the differences in outcomes between the two groups. The case group and control group are matched as closely as possible using a range of key characteristics such as age; gender; location; prior health status; and socioeconomic status.

Case-control studies are particularly useful when the outcome of interest is relatively rare. For example, many early studies examining the relationship between smoking and cancer compared individuals with and without cancer and investigated differences in prior exposure to tobacco.

Case-control studies select controls that are similar to cases on relevant characteristics, but unlike methods such as DiD, PSM and Regression analysis, they typically rely on more limited matching procedures and therefore provide less protection against bias arising from differences between groups.

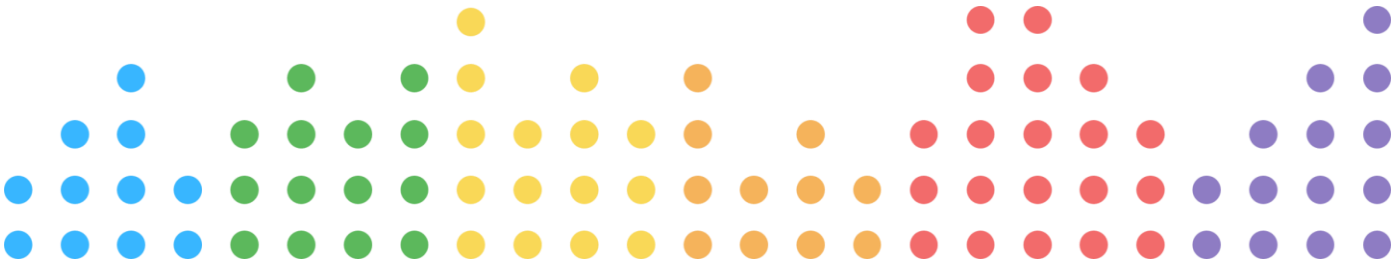
Annex B. Non-market valuation methods

1. Revealed Preference (RP) Valuation methods

RP methods estimate the value of non-market goods by observing individuals' behaviour in related markets where choices involve trade-offs. The core principle is that individuals reveal their preferences through actual decisions, allowing analysts to infer their (WTP) for non-market outcomes based on observed market behaviour.

Two commonly used approaches are the **travel cost method** and **hedonic pricing**. The travel cost method estimates the value of recreational or cultural goods by analysing how much individuals are willing to spend in terms of time and money to access them. For example, travel expenses and time costs incurred to attend a cultural venue or participate in a creative activity can be used as a proxy for the value placed on that experience. Hedonic pricing, by contrast, examines how non-market attributes, such as access to cultural amenities or environmental quality, are reflected in market prices, such as property values or wages.

Revealed preference methods have the advantage of being based on actual behaviour, which reduces concerns about hypothetical bias. However, their applicability is limited



to contexts where relevant market behaviour exists and can be clearly linked to the non-market good of interest. In creative health, this may restrict their use, as many outcomes, such as improvements in health or reductions in loneliness, are not directly reflected in observable market choices. Additionally, these methods require strong assumptions and econometric methods to isolate the value of specific attributes.

2. Stated Preference (SP) Valuation methods

SP methods estimate the value of non-market goods by asking individuals directly about their preferences through structured surveys or experiments. These methods are particularly useful when no suitable market behaviour exists from which values can be inferred.

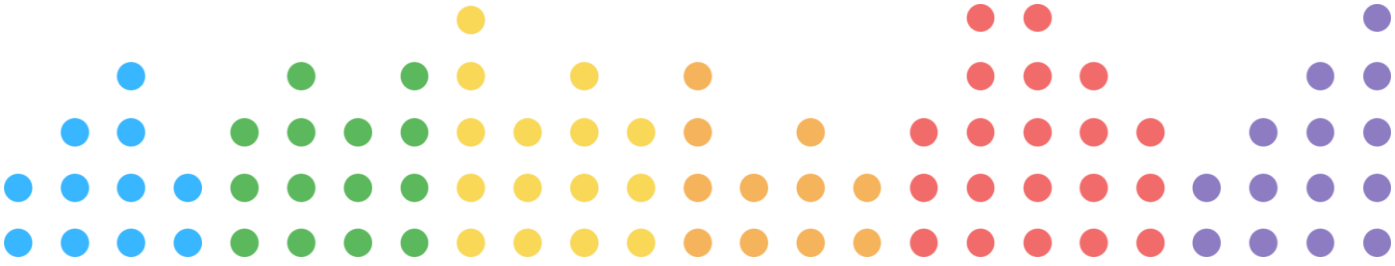
The two main approaches are **contingent valuation**, which asks respondents their maximum WTP for a specified change (or their minimum WTA if the change is negative), and **discrete choice experiments**, which ask respondents to choose between alternative scenarios with different attributes and costs.

By presenting carefully designed hypothetical scenarios, stated preference methods can capture both **use values** (e.g. participation in a creative health programme) and **non-use values** (e.g. valuing the existence of such programmes). This makes them especially relevant for creative health, where many benefits, such as improved wellbeing or social inclusion, are not directly linked to market behaviour.

However, these methods are sensitive to survey design and may be affected by biases such as strategic responses or hypothetical bias. The Green Book therefore emphasises the need for robust design and clear framing of scenarios. When well implemented, stated preference methods provide a flexible and widely applicable approach to valuing non-market outcomes.

3. Wellbeing Valuation (WV)

WV estimates the monetary value of non-market outcomes by analysing how changes in life circumstances affect individuals' life satisfaction. Using large datasets, statistical models are used to estimate the relationship between income and wellbeing, and between the outcome of interest (such as reduced loneliness or improved mental health) and wellbeing. By comparing these relationships, analysts can calculate the amount of money that would produce an equivalent change in wellbeing, thereby deriving a monetary value for the outcome.



This approach is applicable to creative health interventions as it can capture a wide range of outcomes, including mental health, social connection, confidence and quality of life.

Unlike stated preference methods, wellbeing valuation does not rely on hypothetical questions but instead uses observed associations between life circumstances and wellbeing. However, the method relies on strong assumptions about causality and the comparability of wellbeing effects across individuals. It also depends on the availability of robust data and appropriate model specifications. Despite these challenges, wellbeing valuation is increasingly used in UK policy appraisal and is consistent with the Green Book's emphasis on capturing broader measures of social value beyond traditional economic indicators.

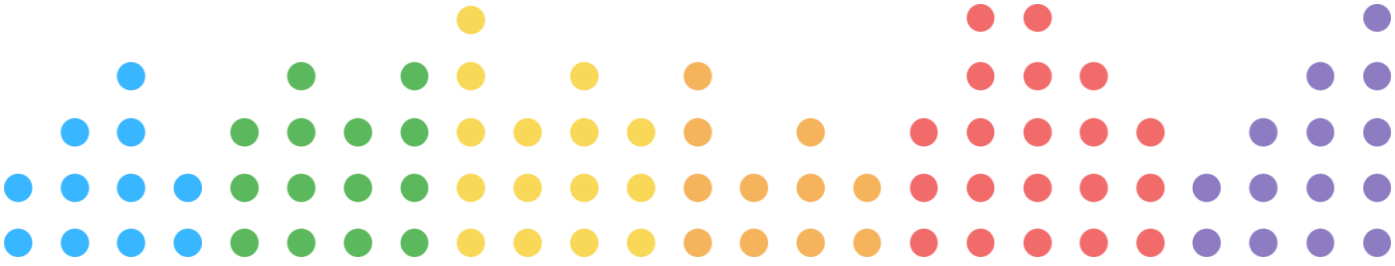
4. Health Valuation

In supplementary Green Book guidelines, the Department of Health and Social Care (DHSC) recommends the use of **Quality-Adjusted Life Years (QALYs)** as the primary metric for measuring health outcomes in economic evaluation. QALYs combine both the quantity and quality of life into a single measure, where one QALY represents one year of life in full health.

Health-related quality of life is typically measured through instruments such as the **EQ-5D**, which classify individuals' health across key dimensions covering mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each health state measured by the EQ-5D is assigned a value between 0 (death) and 1 (perfect health), based on **stated preference** surveys administered with the general population. QALYs are calculated by combining these values with the duration of time spent in each health state.

The Guidelines recommend aggregating QALYs across individuals and over time to estimate the total health impact of an intervention. These QALY gains or losses can then be used directly in CEA or converted into monetary values for use in CBA. The Green Book provides a standardised value per QALY.

Limitations associated with QALYs include the fact that mental health is not as extensively covered as physical health and the QALY is limited to health impacts. These limitations are particularly relevant to creative health because outcomes extend beyond health status to include broader social benefits.



Annex C. Examples of applied economic evaluation of Flip of the Coin and Odd Arts creative health programmes

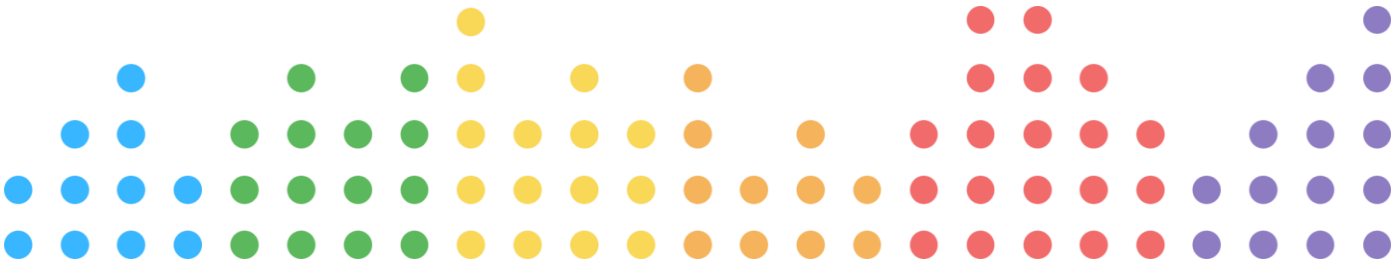
1. Introduction

Creative health refers to the use of arts, culture and creative activities -such as music, visual and scenic arts, dance and gardening- to support and improve health and wellbeing (National Centre for Creative Health, 2023). Delivery varies by setting: some creative health activity is delivered through formal United Kingdom National Health Service (NHS) pathways such as social prescribing, while much is delivered independently by cultural and voluntary-sector organisations without a clinical referral route. Across both routes, creative health activity engages with the social, cultural and environmental determinants of health, alongside clinical care.

This study develops and applies an economic evaluation framework broadly based, where possible, on the principles and methodological standards of the HM Treasury Green Book and the Economic Evaluation Guidelines for Creative Health, to estimate the social value generated by creative health programmes. A key aim of this study was to assess the extent to which current data collection and evaluation practices within the creative health sector are sufficient to support high-quality economic evaluation, while demonstrating what can currently be achieved using the best available programme data.

Economic evaluation is the systematic assessment of the costs and consequences of projects, policies, or interventions to determine which provides the best use of limited resources. It supports decision-making by assessing efficiency and value for money. **Cost-Benefit Analysis (CBA)** is main type of economic evaluation that underpins the Green Book. CBA expresses both costs and benefits in monetary terms and compares total expected benefits with total expected costs, often using measures such as net present value and benefit-cost ratios. CBA helps determine whether an intervention provide value to society (social value) and allows comparison across different programmes and sectors, while considering factors such as time, uncertainty, and opportunity costs.

The analysis distinguishes between **primary and secondary benefits**, specific to healthcare interventions. Primary benefits capture the direct impact of creative health interventions on participants quality of life and wellbeing, measured through indicators such as Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS). These



represent the intrinsic value of improvements experienced by participants themselves, such as improved confidence, creative expression, enjoyment, reduced pain and better mental health. On the other hand, secondary benefits capture wider spillover effects arising from these improvements. These may include potential reductions in NHS healthcare utilisation and improvements in workplace productivity. They represent wider economic or potentially cashable benefits to taxpayers and businesses.

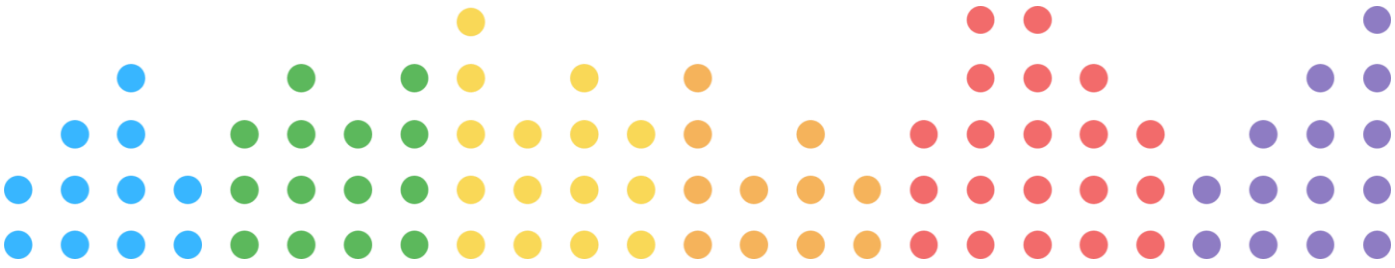
The outcomes examined in this study have been determined by the availability and quality of existing data. As part of the project, we engaged with creative health providers to identify existing programme and outcome data suitable for economic evaluation. A standardised data template was developed and shared with providers, setting out the information required, including programme characteristics, participant demographics and pre-and post- intervention measures of health and wellbeing. This process identified significant variation in the type, quality and consistency of data currently collected across the sector. In many cases, programmes relied on qualitative evidence, collected outcomes only after participation, or did not use standardised outcome measures suitable for economic evaluation. This meant that they could not be included in economic evaluations.

SWEMWBS is used to estimate primary wellbeing benefits, as it was the principal standardised wellbeing measure available in the programme data, with changes monetised using the WELLBY approach to allow inclusion within a Green Book-consistent framework.

Secondary benefits relating to healthcare utilisation and workplace productivity are indirectly estimated using the statistical relationship between SWEMWBS and relevant outcomes available in The UK Household Longitudinal Study 'Understanding Society', because health utilisation and productivity data itself was not collected in the programmes. The analysis considers **GP visits**, **outpatient use**, **hospitalisation** and **presenteeism**.

The **Flip of the Coin** and **Odd Arts programmes** provide the principal applied case studies because they offered the most suitable data for economic analysis. However, the available datasets have important limitations, including small sample sizes, limited quantitative outcomes measurement, incomplete demographic information, imperfect baseline measurement and the absence of data from control groups.

The analysis should, therefore, be regarded as an exploratory and pragmatic application of economic evaluation to the best available creative health data using



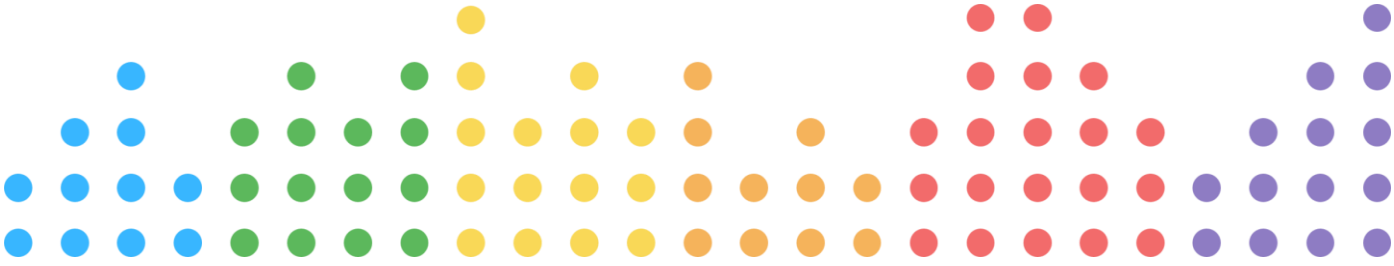
the principles of the Green Book wherever possible, rather than a definitive causal assessment of programme impact. Its purpose is both to provide indicative estimates of value and to demonstrate what can currently be achieved, while highlighting the improvements in data collection, outcome measurement, sample sizes and evaluation design required for more robust economic evaluation in the future.

The remainder of this report is structured as follows. Section 2 sets out the analytical framework underpinning the evaluation, including its alignment with the Green Book and the Theory of Change approach to creative health. Section 3 describes the methodology used to estimate and monetise primary wellbeing benefits and secondary benefits relating to healthcare utilisation and workplace productivity. Section 4 applies this methodology to the two case studies, Flip of the Coin and Odd Arts, presenting their respective costs, benefits and social value. The report concludes with a discussion of the findings and recommendations for future evaluation practice in the creative health sector in Section 5.

2. Analytical Framework

The benefits associated with creative health are supported by a growing evidence base, spanning improvements in mental and physical health (Fancourt & Finn, 2019; Wang, Mak & Fancourt, 2020), social outcomes such as reduced isolation (Lewandowska & Węziak-Białowolska, 2023), and broader effects on healthcare use and productivity (Dayson & Bashir, 2014). Assessing this value in a way that is consistent with the HM Treasury Green Book requires a clear account of how a creative health programme is expected to generate the benefits described above. This is set out through a **Theory of Change (ToC)**, which sets out how programme resources and activities are expected to generate outputs, outcomes and wider benefits. The ToC also helps identify who participates in the intervention and who experiences its benefits, providing a conceptual basis for assessing the social value generated by creative health programmes. Creative health interventions are theorised to generate these benefits primarily through participatory engagement, social connection and creative expression, mechanisms distinct from clinical treatment but which nonetheless support psychological and social wellbeing (Fancourt & Finn, 2019).

A creative health intervention's inputs, such as funding, creative practitioners, healthcare or community staff, venues, materials and partnerships with cultural or health organisations, are converted into outputs such as delivered sessions and participant engagement. These, in turn, generate outcomes, the changes experienced by participants, which may span improved health and wellbeing, reduced social



isolation, changes in healthcare utilisation and health-related quality of life. The framework therefore considers both financial and non-financial outcomes when assessing the value generated by creative health.

These effects can also extend beyond individual participants to wider health, care and community systems. Creative health interventions may form part of pathways, such as social prescribing, neighbourhood health models or wider programmes of support, and may contribute to broader objectives including reducing urgent care demand, supporting hospital discharge, addressing health inequalities and managing long-term conditions. The potential benefits of creative health should therefore be considered in relation to both the direct outcomes experienced by participants and the wider effects that may arise in the health and economic system.

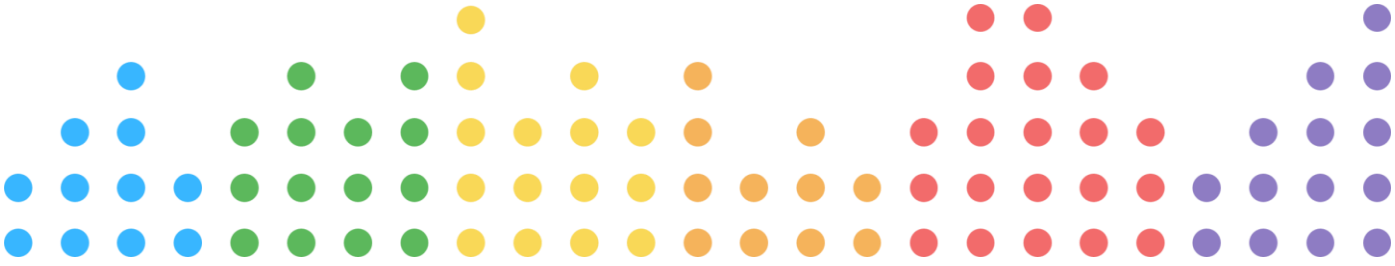
Alongside these benefits, economic evaluation requires consideration of the resources used to deliver creative health interventions. These may include programme delivery and staff time, payments to creative practitioners, venues and facilities, materials and equipment, training, volunteer contributions, healthcare resources and organisational overheads. In economic valuation these intervention costs are distinguished from any potential negative effects or outcomes, which should also be considered where relevant. The Green Book approach provides a basis for bringing these costs and benefits together within an economic evaluation, allowing the social value generated by creative health programmes to be assessed.

Overall, the framework provides a pathway from creative health inputs and activities, through outputs and outcomes, to the assessment of costs, benefits and social value. The Theory of Change provides the basis for identifying the outcomes and wider effects that are relevant to the economic evaluation, while the Green Book provides the overarching framework for assessing their costs and benefits. This provides the conceptual basis for the empirical analysis in this study.

3. Economic Valuation Methods

3.1. Outcome measurement: SWEMWBS

The **Short Warwick–Edinburgh Mental Wellbeing Scale (SWEMWBS)** is a seven-item instrument that measures subjective mental wellbeing across domains such as positive affect, social functioning, and sense of purpose, with scores ranging from 7 to 35. It is derived from the longer version, WEMWBS, which includes 14 items.



It is used as the main measure of primary benefit in our analysis since it has been widely used in evaluations of community arts and social prescribing programmes across the UK. However, the limitation of SWEMWBS is that it is not a preference-based metric and hence it does not generate utility weights and cannot be directly mapped onto the Quality-Adjusted Life Year (QALY) framework typically used in health economic evaluation. This means that is does not have a standardised monetary value that can be used in cost-benefit analysis (Jensen et al., 2024).

Its use in this study nevertheless reflects the data collected by the programmes - SWEMWBS was the only measure that allowed for a consistent assessment of impact across programmes. This limitation is acknowledged throughout the report.

To address this constraint, an empirical 'bridge' between SWEMWBS and monetary value is constructed using the WELLBY framework and longitudinal data from Understanding Society (USoc), following the methodology set out in Fujiwara et al. (2017). This approach estimates the statistical relationship between SWEMWBS and life satisfaction using a nationally representative panel dataset and derives shadow prices of changes in SWEMBS scores from that relationship.

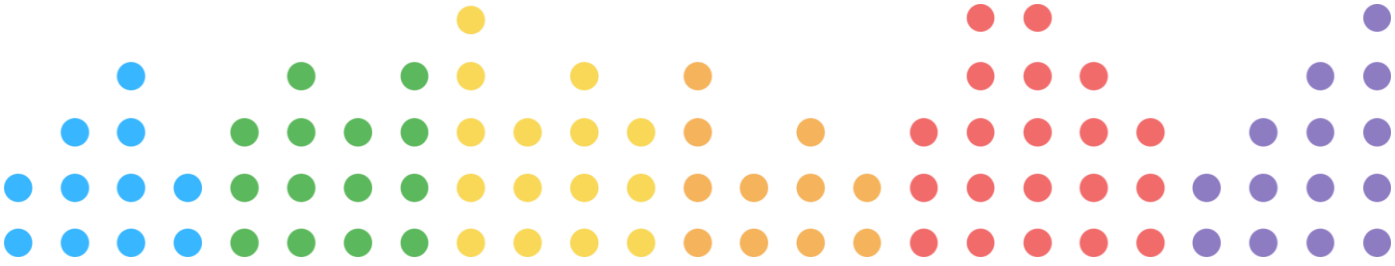
In addition to primary benefits estimation, as we discuss below, SWEMWBS outcomes are also used to estimate the potential impact on healthcare utilisation and productivity to provide an empirical basis for quantifying the secondary benefits. This is because the creative health programmes that we assessed do not directly collect health utilisation and employment impact data.

3.2. The monetary value of SWEMWBS

Improvements in mental wellbeing, as measured by SWEMWBS, are assumed to have a direct effect on life satisfaction. We model this through a standard life satisfaction econometric model which is used throughout academic research (Fujiwara & Campbell, 2011; Fujiwara, Oroyemi & McKinon, 2013) whereby life satisfaction (LS) is modelled as a function of SWEMWBS (S) and a vector (X) of individual characteristics - including income, employment status, and education- that are known determinants of subjective wellbeing:

$$LS_i = f(S_i, X_i) \quad (1)$$

Equation (1) is estimated using USoc, the UK's main longitudinal household panel survey.



USoc is a nationally representative longitudinal survey of individuals and households across the United Kingdom. Participants are followed annually, enabling analysis of changes in socioeconomic circumstances, employment, income, health, wellbeing, family relationships and housing over time. The study incorporates participants from the British Household Panel Survey and includes ethnic minority and immigrant boost samples. Its large, longitudinal sample provides a valuable resource for multidisciplinary research, population subgroup analysis and policy evaluation.

SWEMWBS is available in waves 1, 4, 7, 10, and 13 of USoc. Wave 13 has been excluded from the primary specification given that most interviews were conducted during a period marked by COVID-19 and rising inflation. These extraordinary circumstances distort the relationships between wellbeing, income and other covariates in ways that are unlikely to reflect normal conditions, making this wave unsuitable as a basis for estimation. For the same reason, wave 10 interviews conducted from March 2020 onwards have also been excluded.

The equation is estimated using individual and wave fixed effects. Individual fixed effects absorb time-invariant characteristics that may confound the relationship between SWEMWBS and LS. Wave fixed effects control for aggregate trends common to all individuals in a given period. A region-by-wave interaction is additionally included to account for time-varying regional shocks -such as differential economic conditions or local policy changes- that would otherwise go uncontrolled:

$$LS_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \theta_1 Individual_i + \theta_2 Wave_t + \theta_3 Wave_t \times Region_i + u_{it} \qquad (2)$$

Where:

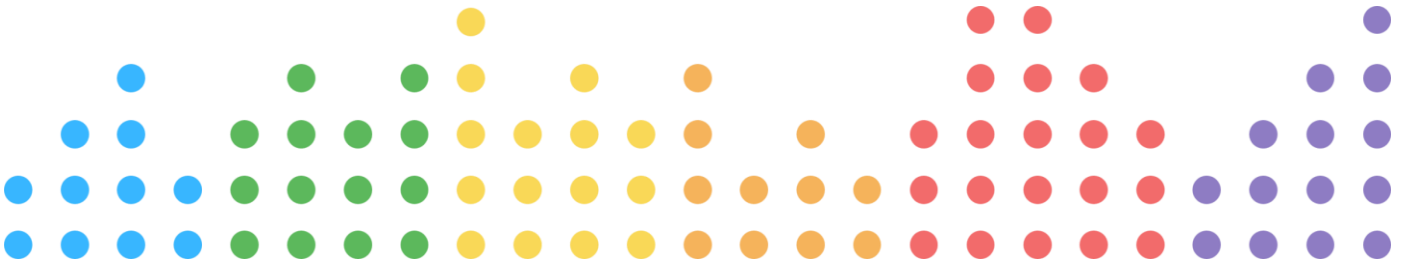
S_{it} = SWEMWBS score coded as the 12-point categories

X_{it} = Vector of controls (*Income, Employment status, Marital status, Education, Physical Health, Number of children*)

θ = Fixed effects

$Wave_t$ = Survey wave in USoc

This specification follows the approach set out in Fujiwara et al. (2017), with two extensions. First, the analysis uses a larger number of waves, increasing the total number of observations to 115,835 and second, a region-by-wave interaction is included as an additional fixed effect, alongside the individual and wave fixed effects



used in the original specification. This improves the model's ability to control for time-varying confounders at the regional level.

Several functional forms were tested to capture the relationship between SWEMWBS and LS. These included differing parametric specifications of the SWEMWBS variable: linear, logarithmic, and quadratic, as well as non-parametric alternatives in which SWEMWBS specification was categorised into 7,12, and 28 groups.

A 12-category specification for SWEMWBS was selected because it offers the best balance between statistical fit, measured by the highest R^2 and lowest RMSE, and parsimony, replicating the approach in Fujiwara et al. (2017). The 12-category score was developed by assigning a score of 1 for any SWEMWBS score in the lowest range from 7-14 and then for SWEMWBS scores from 15-35 the categories increased by every two-point interval making a 12-point scale.

The reference groups in the analysis was Category 1 and therefore all coefficients represent the additional life satisfaction gain associated with being in a given category relative to this baseline. The full model output is provided in section 6.

The estimated coefficients are used to derive the monetary value associated with each SWEMWBS category through the WELLBY framework by multiplying the coefficient by the value of one WELLBY in 2025 prices (£16,618)[9]. A scale adjustment of 10/6 is then applied to convert USoc's 1-7 LS scale to the 0-10 ONS scale underlying the WELLBY value. The resulting monetary values by SWEMWBS category are presented in Table 3.

We note that Our results are similar to the QALY values used in the Green Book (2022, p.87).[10] The Green Book recommends a QALY value of £70,000 in 2022 prices (approximately £81,000 in 2026 prices) based on the value of a statistical life (VSL). This represents the value of a year of life in perfect health. The value for category 12 of SWEMWBS (£71,608.74) in our analysis represents the value of a year of life with optimal mental wellbeing and functioning.

[9] Uprated from the HM Treasury Green Book reference value of £13,000 in 2019 prices, using the ONS implied GDP deflator, series MNF2

[10] Note that the 2026 version of the Green Book does not set out a QALY value.

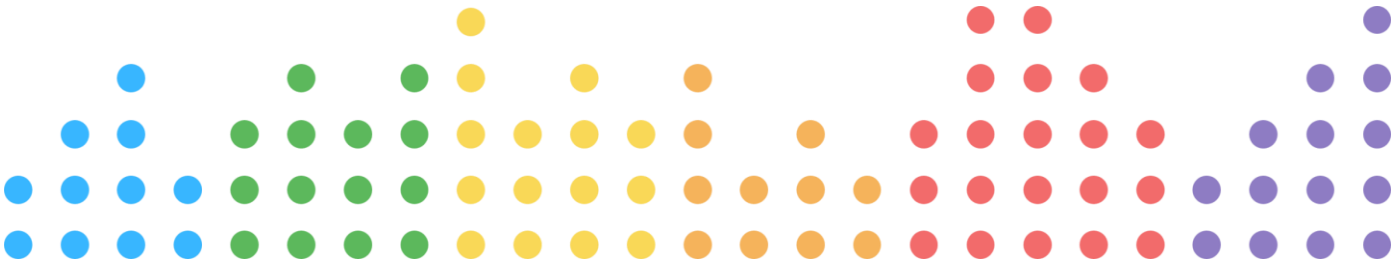


Table 3. Monetary value of mental wellbeing improvements by SWEMWBS category (2025 prices)

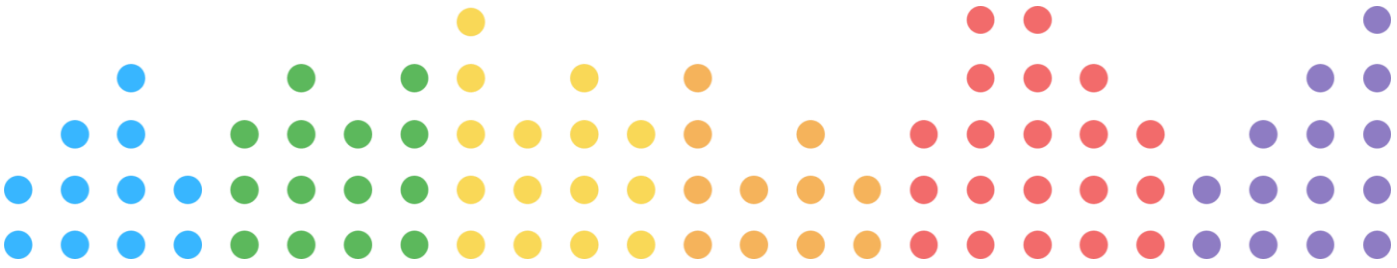
Category	SWEMWBS score	Coefficient (0-10 scale)	Monetary value
1	7-14	0.00 (reference)	0
2	15-16	0.60	£10,017.67
3	17-18	1.01	£16,823.55
4	19-20	1.70	£28,295.47
5	21-22	2.35	£39,111.70
6	23-24	2.85	£47,432.96
7	25-26	3.25	£53,963.47
8	27-28	3.56	£59,144.67
9	29-30	3.79	£62,989.85
10	31-32	3.96	£65,827.40
11	33-34	4.13	£68,634.04
12	35	4.31	£71,608.74

Source: Authors' own calculations

3.3. Secondary benefits of creative health interventions

We assess secondary benefits including healthcare utilisation, such as General Practitioner (GP) visits, outpatient visits and hospitalisations, and productivity-related outcomes, captured through presenteeism. Since this data was not directly collected or reported by creative health providers, the analysis uses an indirect method to infer changes in these variables from the change in the SWEMWBS score. This is based on the assumption that an improvement in mental wellbeing, measured through SWEMWBS, can lead to less health resource utilisation and improved labour market outcomes such as increased productivity.

We estimate the relation between SWEMWBS and each secondary outcome and then apply this relationship to the observed change in SWEMWBS among creative health participants. The secondary value of this impact is then measured using unit cost data sources. We tested both categorical and linear specifications of SWEMWBS across these models and the linear SWEMWBS measure is used in the main analysis because

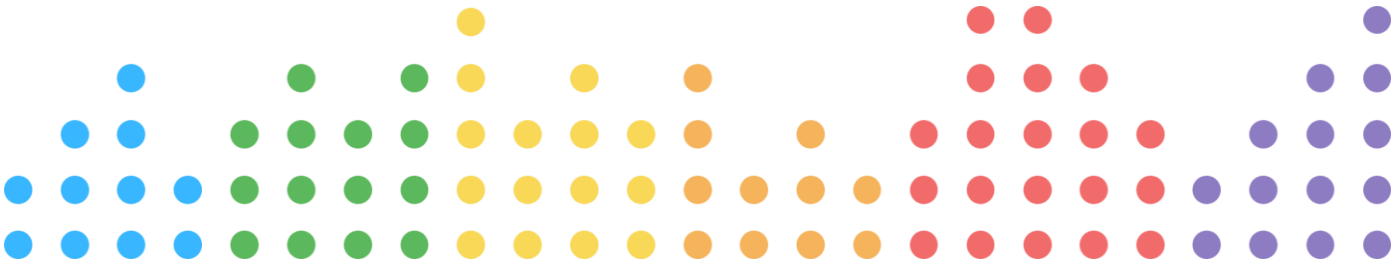


it provides a better model fit and produces more robust and stable estimates across the secondary benefit models.

These estimates should not be interpreted as a direct causal effect of creative health on healthcare utilisation and on labour market outcomes. Instead, they are indicative estimates based on the available data from creative health providers. Going forward these impacts should be estimated using direct health utilisation and labour market data.

The secondary benefit models are estimated using waves 7 and 10 of Understanding Society, due to data availability and concerns around COVID-related distortions in subsequent waves: GP visit information is only available from wave 7 onwards and we excluded wave 13 because it covers post-COVID recovery period and may therefore reflect exceptional changes in income, healthcare use and wellbeing.

To monetise the secondary benefits, estimated changes in healthcare use and productivity outcomes are combined with external unit cost and earnings data. Cost inputs for GP visits, outpatient visits, hospitalisation admission and presenteeism are drawn from NHS sources, the University of Kent/PSSRU, the Greater Manchester's Unit cost database, and the Office for National Statistics (ONS), allowing these effects to be translated into approximate monetary values.



Expected healthcare utilisation

We use Ordinary Least Squares (OLS) multivariate regression to estimate the association between changes in SWEMWBS -entered in a linear[11] form- and healthcare utilisation in the form of **GP visits and outpatient visits**. The model controls for individual fixed effects, survey wave effects, and region-by-wave effects, and provides coefficients that can be interpreted as changes in the number of GP or outpatient visits. The resulting specification is:

$$GP\ visits_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \theta_1 Individual_i + \theta_2 Wave_t + \theta_3 Wave_t \times Region_i + u_{it} \quad (3)$$

$$Outpatients_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \theta_1 Individual_i + \theta_2 Wave_t + \theta_3 Wave_t \times Region_i + u_{it} \quad (4)$$

Where:

S_{it} = SWEMWBS score entered as a continuous linear measure

X_{it} = Vector of controls (*Income, Employment status, Marital status, Education, Physical Health, Number of children*)

θ = Fixed effects

$Wave_t$ = Survey wave in USoc

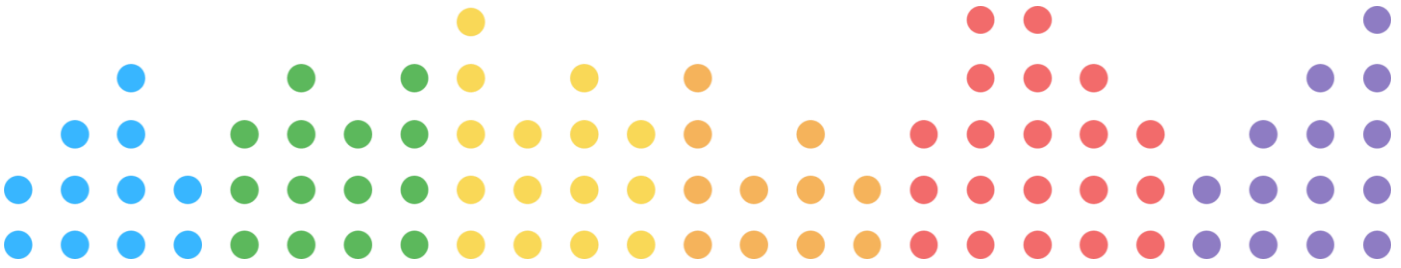
GP visits and outpatient visits are recorded on an ordinal scale from 0 to 4 (0 = none, 1 = one or two, 2 = three to five, 3 = six to ten, 4 = more than ten).

The coefficient β_1 , captures the association between SWEMWBS and healthcare utilisation.

The results show that higher SWEMWBS scores are associated with lower healthcare utilisation as expected. A one-point increase in SWEMWBS is linked to a 0.015-point reduction in GP visits and a 0.007-point reduction in outpatient visits, both statistically significant. Full model results are presented in section 6.

For monetisation, the estimated GP visits coefficient (-0.015) and the outpatient visits coefficient (-0.007) are multiplied by the participant's change in SWEMWBS to estimate the expected change in healthcare utilisation.

[11] The linear specification is used across all four secondary benefit outcomes. This choice is anchored in hospitalisation, where the binary, low incidence outcome rules out a robust categorical form. Applying linear consistently across all four benefits avoids mixing functional forms across the same set of estimates, and it carries no cost for GP, outpatient and presenteeism specifically, since fit is virtually identical between linear and categorical.



Hospitalisation is also considered as a secondary benefit, capturing more severe health events than GP or outpatient care. It is modelled as a binary outcome, indicating whether the individual was admitted to hospital during the recall period, rather than as the number of days admitted. This preserves the full estimation sample, while modelling days admitted would restrict the analysis to the subgroup of individuals who were hospitalised, which is too small for robust within-person estimation, as shown in section 6. The same fixed effects specification as in equations (3) and (4) is applied to this binary outcome. In this case, the model is estimated as a Linear Probability Model (LPM), so the coefficient is interpreted as a change in the probability of hospital admission.

$$HospitalAdmissions_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \theta_1 Individual_i + \theta_2 Wave_t + \theta_3 Wave_t \times Region_i + u_{it} \quad (5)$$

Where:

S_{it} = SWEMWBS score entered as a continuous linear measure

X_{it} = Vector of controls (*Income, Employment status, Marital status, Education, Physical Health, Number of children*)

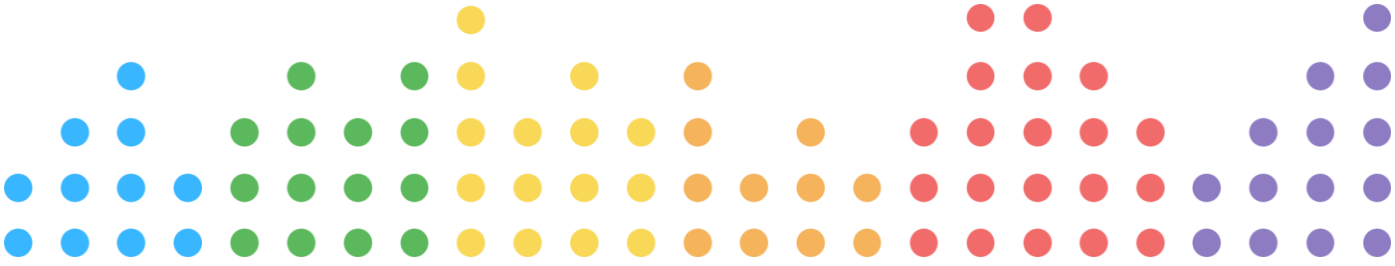
θ = Fixed effects

$Wave_t$ = Survey wave in USoc

The results show that an increase of one point in SWEMWBS is associated with a reduction of 0.001 points in the probability of hospital admission, statistically significant. Full model results are presented in the section 6.

For monetisation, the estimated coefficients are then multiplied by the participant's change in SWEMWBS to estimate the expected change in healthcare utilisation. As GP and outpatient visits are recorded on a banded scale rather than as visit counts, this estimated change is first converted into an equivalent number of visits using the midpoint value of each band. The hospitalisation coefficient, already expressed in probability units, requires no such conversion. The resulting change in utilisation is then combined with the corresponding NHS unit cost to produce a monetary value.

As a robustness check, we also estimated probabilistic (logit) specifications for each of the secondary benefit outcomes above, without individual fixed effects. These produced coefficients consistent in sign and statistically significant with the linear estimates reported here. They are not used for monetisation, however, as this



specification does not control for unobserved individual-level heterogeneity, unlike the fixed-effects approach applied consistently elsewhere in this analysis.

Presenteeism

A relevant secondary benefit is the potential reduction in work-related productivity losses. Ideally, this would include absenteeism, measured as days off work due to health problems. However, suitable absenteeism data are not available in USoc for the panel specification used here. As a result, presenteeism was used as the closest available proxy. Presenteeism captures reduced performance while still at work and therefore provides an alternative way to assess whether improvements in mental wellbeing are associated with productivity-related benefits.

Presenteeism is measured using the SF-12 Health Survey item asking how much of the time, in the past four weeks, the respondent accomplished less than they would have liked as a result of emotional problems. Responses are recorded on a five-category frequency scale (all of the time, most of the time, some of the time, a little of the time, none of the time), inverted so that higher values correspond to more frequent impairment.

The model is estimated only among employed individuals observed in waves 7 and 10, using individual, wave, and region-by-wave fixed effects, with standard errors clustered at the individual level. Employment status is not included as a control because the sample is already restricted to people at work. This yields the following OLS specification.

$$Presenteeism_{it} = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \theta_1 Individual_i + \theta_2 Wave_t + \theta_3 Wave_t \times Region_i + u_{it} \quad (6)$$

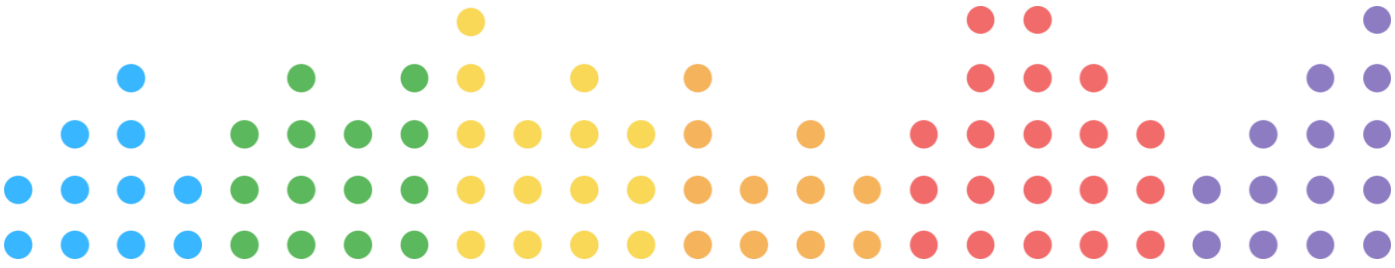
Where:

S_{it} = SWEMWBS score entered as a continuous linear measureX_i = Vector of controls (*Income, Employment status, Marital status, Education, Physical Health, Number of children*)

θ = Fixed effects

$Wave_t$ = Survey wave in USoc

We find that a one-point increase in SWEMWBS is associated with a 0.073-point reduction in the presenteeism score, statistically significant. These results indicate that better mental wellbeing is associated with lower presenteeism and improved work performance. Full model results are presented in the section 6.



The main challenge in monetising presenteeism is that the variable measures the frequency of reduced work performance, not its magnitude. Monetisation therefore requires an assumption translating each frequency category into an estimated percentage productivity loss.[12] This is informed by Work Productivity and Activity Impairment (WPAI)- and Work Limitations Questionnaire (WLQ)-based studies of mental health-related presenteeism, which report losses ranging from around 13-15% for occasional or subclinical impairment to around 45% for frequent or clinically significant impairment.

Each participant's change in SWEMWBS is multiplied by the coefficient (-0.073) to estimate the predicted change in presenteeism score. This change in score is then converted into a change in percentage productivity loss using the mapping described in footnote 4, and the resulting change in productivity loss is multiplied by earnings or labour cost data to estimate the monetary value of the productivity gain or loss.

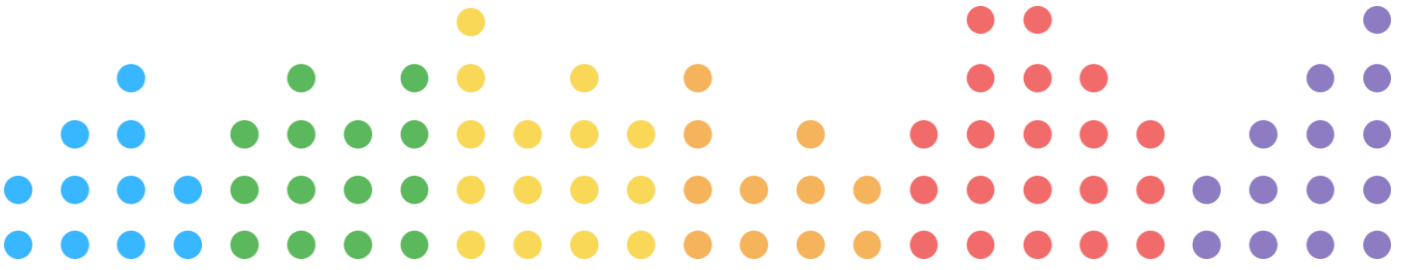
4. Programme evaluation

Section 3 developed the analytical framework for valuing changes in SWEMWBS and estimating the associated secondary impacts on healthcare utilisation and workplace productivity. This section applies these estimates to two creative health interventions to assess their social value impacts and demonstrate how the framework can be applied in practice.

A standardised data template was shared with creative health organisations and practitioners to identify suitable programme data for economic evaluation.[13] However, most organisations did not collect the required standardised pre- and post-intervention outcome data. The analysis therefore focuses on the two programmes for which suitable data were available: **Flip of the Coin (Easter Ross, REALITIES)** and **Odd Arts**,

[12] Each response category (none, a little, some, most, all of the time) is assigned an estimated percentage productivity loss, interpolated linearly between 0% (none of the time) and 45% (all of the time): 0%, 11.25%, 22.5%, 33.75% and 45% respectively. The 45% anchor reflects presenteeism levels reported for clinically significant depression/anxiety symptoms (Beck et al., 2011, 47%; Asami et al., 2015, 45.2%; Chokka et al., 2019–2020, 35.9-61.5%, all WPAI-measured). This gradient is corroborated at the lower end by presenteeism among workers with subclinical or low-severity symptoms (Asami et al., 2015, 13.4%).

[13] The template specified programme information, demographic characteristics, and pre- and post-intervention subjective wellbeing measures, including life satisfaction, EQ-5D, and SWEMWBS. Reasons for non-collection varied, including ethical concerns about administering certain questionnaires to beneficiaries, collecting only post-intervention measures, or relying on qualitative feedback rather than standardised instruments.



presented in Sections 4.2 and 4.3 respectively. All data used in the analysis were anonymised prior to analysis.

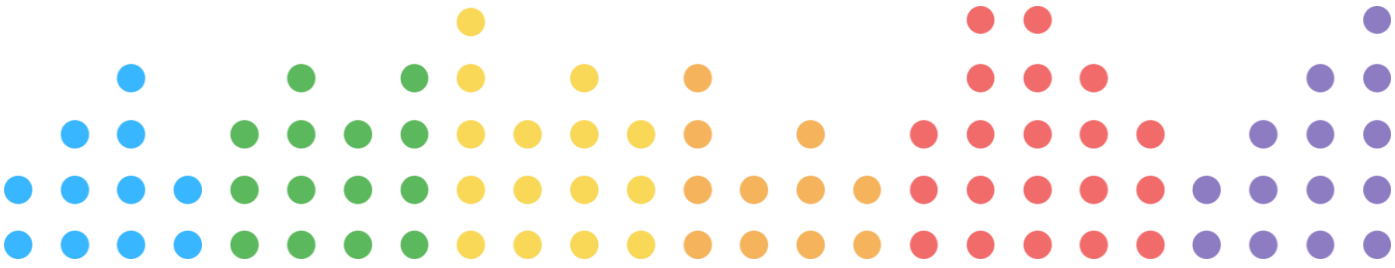
The available programme data have important limitations. Key demographic and socioeconomic variables were missing in both programmes alongside other characteristics that would normally support statistical analysis. The small size of both samples also limits how many characteristics can be controlled for simultaneously, as discussed further below. The analysis is therefore restricted to the available variables and cannot rule out remaining differences between participants and the comparison group.

There are also limitations in the Flip of the Coin outcome data. Baseline SWEMWBS scores were not collected prospectively at programme entry, and the timing of the baseline measure is uncertain. As the programme is ongoing, follow-up scores may also represent different stages of participation rather than a consistent post-intervention endpoint. These measurement issues create additional uncertainty around the estimated change in wellbeing.

The analysis should therefore be regarded as a pragmatic estimation based on the best available evidence rather than a fully Green Book-compliant technical economic evaluation. The approach applies relevant Green Book principles, including consideration of the counterfactual and the valuation of social and economic outcomes, but the limitations in sample sizes and data incompleteness constrain the robustness of the estimates. The results should consequently be interpreted as preliminary and indicative. More generally, the inability to include other programmes highlights the need for more consistent and systematic outcome data collection across the creative health sector to support higher-quality economic evaluation.

4.1. Flip of the Coin (REALITIES)

REALITIES is a creative health initiative bringing together communities across Scotland to support healthier, more resilient environments in areas affected by health inequalities, delivered through hubs that develop creative health interventions tailored to local needs. For this evaluation the Flip of the Coin project (Easter Ross hub) was selected as it provided the most complete pre- and post-intervention outcome data, including subjective wellbeing measures. The hub works with people with lived experience of intergenerational trauma, poverty, poor mental health, substance use, and contact with the criminal justice system.



The dataset contains information for 33 participants from two programme groups, all located in the Easter Ross area. The sample is mainly composed of women (25 of 33), and all participants are White British. Regarding employment status, the largest group consists of participants who are long-term sick or disabled, with smaller groups of retired, part-time employed and unemployed participants. The composition of the sample is therefore broadly consistent with the population targeted by the programme.

The Easter Ross hub collected the full 14-item WEMWBS, while the valuation model in Section 3 uses SWEMWBS, the 7-item version available in USoc. WEMWBS scores were therefore converted to SWEMWBS using the item correspondence in Stewart-Brown et al. (2009), summing the seven SWEMWBS items within the full scale.

Moreover, two data preparation steps were needed before this conversion. Non-integer answers in individual WEMWBS items (e.g. between 3 and 4) were resolved by averaging the two adjacent values, and where a single item was missing it was imputed using the participant's mean score across the remaining items, too avoid losing observations, given the small size of the sample.

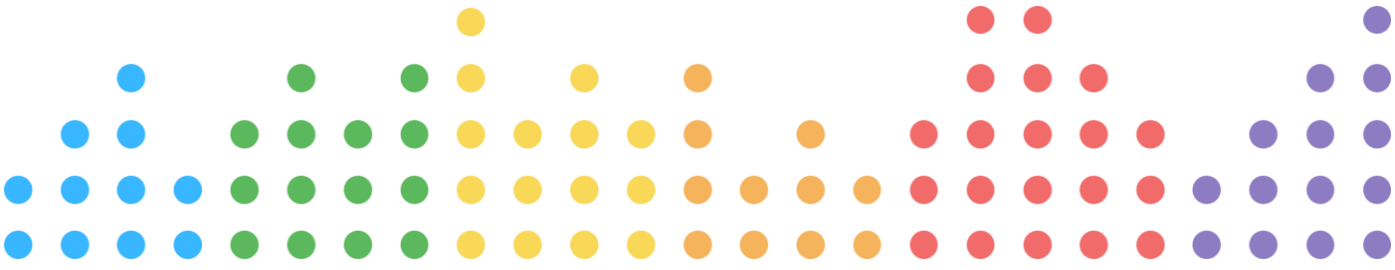
Baseline SWEMWBS scores averaged 17.99, indicating relatively low mental wellbeing at entry. Following the intervention, the average score increased to 25.94, representing a gross improvement of around 7.94 SWEMWBS points. As noted in Section 4, baseline scores were not collected prospectively at programme entry, an important limitation affecting the precision of the estimated change, addressed in the identification strategy below.

4.1.1. Methodology

To construct an external comparison group, USoc was used, as it provides longitudinal data and repeated measures of SWEMWBS, allowing changes in wellbeing to be compared over time and at the individual level.

A limitation is that SWEMWBS is only available in selected waves of USoc. Waves 10 (2018-2020) and 13[14] (2021-2023) were therefore used as the pre- and post-periods for the comparison group, as these were the closest available waves to the intervention period that provided the necessary data. This creates a temporal gap between the two

[14] Wave 13 was excluded from the WELLBY and secondary benefits models due to parameter instability in this period (see Section 3), likely reflecting atypical post-COVID economic and behavioural patterns. Its use here is different: it serves only to observe the comparison group's SWEMWBS at the post-intervention period as a counterfactual, not to estimate a structural relationship between variables, so the same concern does not apply.



periods: the control group is observed over approximately three years, whereas the Flip of the Coin intervention was delivered over a substantially shorter period. This asymmetry should therefore be considered when interpreting the estimated programme effect.

Propensity Score Matching (PSM) was used to construct the most comparable comparison group possible. **PSM** is a quasi-experimental technique used to estimate programme impacts where a randomised control group is not available; it matches participants to non-participants with similar observed characteristics, approximating the counterfactual of what would have happened in the absence of the intervention. PSM cannot account for unobserved differences between groups and therefore does not provide the same level of causal evidence as a randomised controlled trial, but it is widely used in policy evaluation and is generally considered one of the strongest statistical methods available when experimental designs are not feasible.

Participants were matched to respondents living in Scotland in wave 10 of USoc based on baseline SWEMWBS score, gender, employment status, marital status, age, income category and benefits receipt. A nearest-neighbour 1:3 matching procedure was applied, leaving 33 treated individuals and 99 matched controls, with no treated participants excluded. The resulting comparison group should be regarded as a pragmatic approximation of the best available counterfactual rather than a fully comparable control group.

Balance is assessed using the standardised mean difference (SMD), which indicates how similar the treated and comparison groups are for a given characteristic: values closer to zero indicate greater similarity. An absolute SMD below 0.25 is generally considered indicative of acceptable balance (Stuart, 2010). Balance improves for most variables, though not uniformly. Baseline SWEMWBS, gender, age, income and benefits receipt all fall within the 0.25 threshold. However, two employment categories remain above it after matching: sick/disabled (SMD = 0.32) and economically inactive (SMD = -0.30), reflecting the difficulty of finding comparable individuals in the general population for a sample in which long-term sickness and disability are highly prevalent.

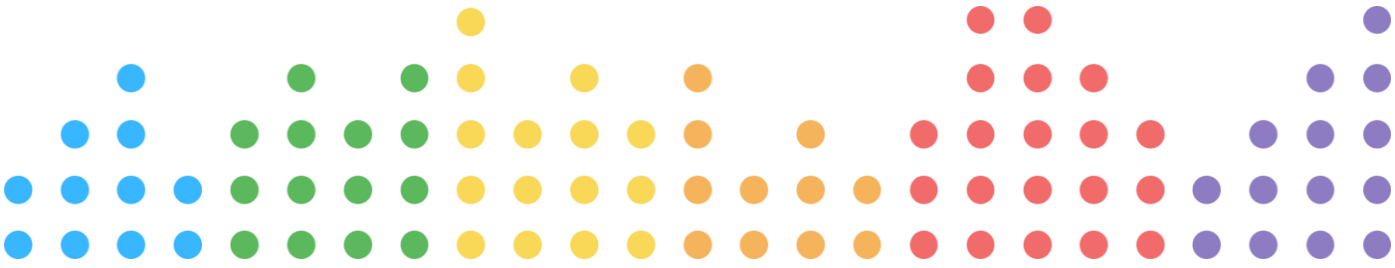


Table 4. Flip of the Coin: Balance diagnostics before and after matching

Variable	SMD before matching	SMD after matching
Baseline SWEMWBS	-0.99	-0.20
Gender (female)	0.46	0.19
Age	-0.09	-0.10
Employment: sick/disabled	1.33	0.32
Employment: employed	-1.67	-0.14
Marital: separated/divorced	0.33	0.02
Marital: married/cohabiting	-0.17	0.16
Income: £0–£14,999	0.91	0.10
Benefits receipt	1.23	0.09

Source: Authors' own calculations

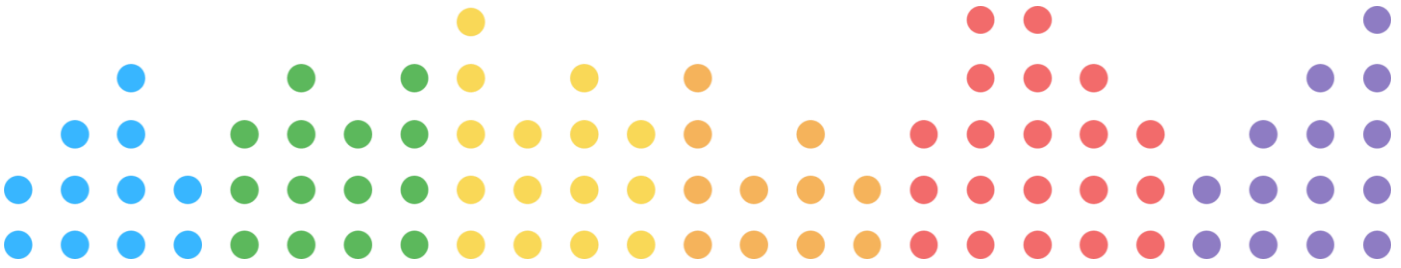
Programme impact was estimated using a **Difference-in-Differences (DiD)** model, which compares the change in SWEMWBS among Flip of the Coin participants with the corresponding change observed in the matched comparison group over time.[15] This approach estimates the improvement attributable to the programme after accounting for changes that would be expected in the absence of the intervention. Formally, the model is estimated as:

$$SWEMWBS_{it} = \beta_0 + \beta_1 Time_t + \beta_2 Treated_i + \beta_3 (Time * Treated)_{it} + u_{it} \quad (7)$$

Where $Time$ takes value 1 for the post-intervention period and 0 for the pre-intervention period. $Treated$ takes value 1 for Flip of the Coin participants and 0 for the matched controls. Therefore, β_3 captures the DiD estimator, the additional change in SWEMWBS attributable to programme participation above the trend observed in the control group. The matching procedure already controls for the variables that are available in both datasets, and for this reason no additional covariates are included in the regression.

A key limitation of this specification is the parallel trends assumption: the requirement that, in the absence of the intervention, the treated and control groups would have

[15] Other specifications were also tested. ANCOVA compares participants' wellbeing after the intervention while accounting for differences in their baseline wellbeing, rather than comparing changes over time, and therefore does not rely on the parallel trends assumption required by DiD. A deadweight-adjusted specification was also estimated, applying a 27% adjustment to the observed pre-post change, following Fujiwara et al. (2017) and the HCA guidance (Dancer, 2014).



followed the same trajectory. This is difficult to satisfy here for two reasons. First, the observation window for the control group spans approximately three years (wave 10 to wave 13), a substantially longer period than the Flip of the Coin intervention itself. Second, the programme is still ongoing, so the post-intervention measure does not necessarily capture its end point, but rather a point within a continuing process; the duration of participation therefore differs across participants, and the post score may reflect different stages of engagement. These limitations are taken into account when interpreting the results.

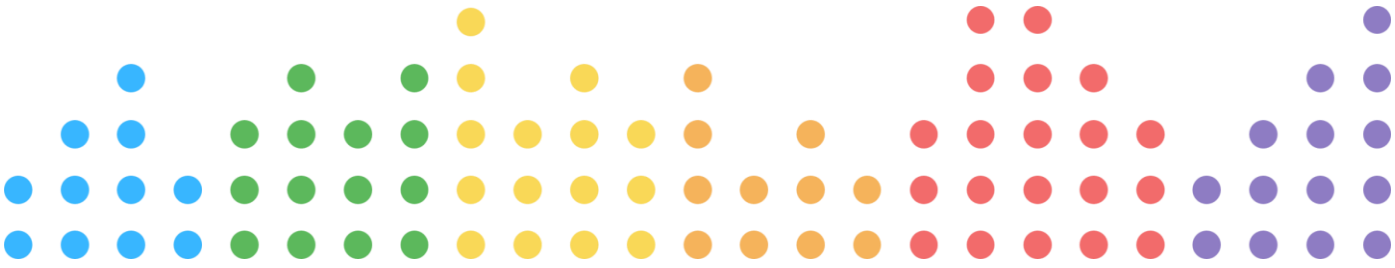
4.1.2. Results

Across the 33 Flip of the Coin participants, mean SWEMWBS scores increased from 17.99 at baseline to 25.94 at follow-up, representing a gross improvement of 7.94 points. This corresponds to a substantial improvement in mental wellbeing, with the average participant moving from category 3 (scores 17–18) to category 7 (scores 25–26) of the valuation scale. However, part of this improvement would be expected to occur even in the absence of the programme. The DiD model addresses this by comparing the change observed among programme participants with the change experienced by a matched comparison group, thereby estimating the component of improvement attributable to programme participation.

The DiD estimate indicates that participation in the Easter Ross hub was associated with an additional 6.74-point increase in SWEMWBS relative to the matched control group ($p<0.001$), a substantial improvement beyond the change expected in the absence of the intervention. The comparison group showed no significant change over the same period, supporting the reliability of the matched control group as a benchmark.

The DiD estimate of 6.74 SWEMWBS points was monetised using the 12-category WELLBY valuation described in Section 3. Each participant's baseline SWEMWBS category was compared to the category corresponding to their baseline score plus the estimated improvement, with the attributable benefit calculated as the difference in monetary value between these two categories, at a value of £16,618 per WELLBY (2025 prices). Because categories are not equally spaced in life satisfaction terms, the monetary value of the same point increase can differ depending on a participant's starting category. Applying this approach across all 33 programme participants produces an estimated total primary benefit of £662,883.11 for the programme.

Secondary benefits (GP visits, outpatient visits, hospitalisation and presenteeism) were monetised by applying fixed-effects coefficients estimated in Section 3 to each participant's DiD-estimated SWEMWBS improvement. For GP and outpatient visits, the



resulting change on the ordinal visits scale was converted into an estimated number of visits using category-specific midpoint gaps,[16] then valued at NHS unit costs.[17] Hospitalisation, already expressed in probability units, required no intermediate conversion and was valued at the average cost of a hospital episode. Presenteeism was converted from the ordinal scale into an estimated percentage productivity loss,[18] applied to each employed participant's estimated annual earnings.[19]

Applying this approach to Flip of the Coin's 33 participants, the estimated reduction in GP visits was 13 visits (£450), in outpatient visits 4 visits (£797), and in the probability of hospital admission 0.29 percentage points (£969). Presenteeism fell by an estimated 48 hours per year among the three employed (part-time) beneficiaries, valued at £946 per employed participant (£2,838 in total). Together, these secondary benefits amount to an estimated savings of £5,054.29 for the programme (£153.16 per beneficiary).

The primary and secondary benefits estimated above assume that the improvement in SWEMWBS is sustained for a full year. However, evidence on the persistence of wellbeing improvements from short creative health interventions suggests that effects are unlikely to be sustained at full strength for 12 months, and in line with HM Treasury Green Book guidance on scaling effects of shorter duration, a duration factor of 0.50 is applied, equivalent to six months of sustained effect.[20] Applying this factor, the estimated Flip of the Coin's primary benefit falls to £331,441.56 and secondary benefits to £2,527.15, giving a total adjusted social value of £333,968.70.

---

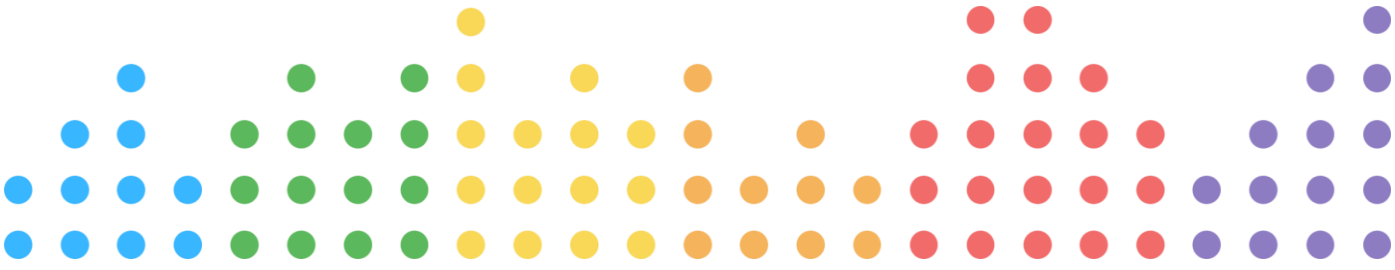
[16] Imputed from SWEMWBS category means in the matched USoc control group (wave 10, pre-intervention).

[17] £147 per hour of GP time, £195 per outpatient appointment and £3,256 per hospital episode (Greater Manchester Combined Authority, 2025, citing Jones et al., 2025); the GP figure is converted to a per-consultation cost of £35 assuming an average consultation length of 14.3 minutes, based on year-end 2025 national appointment duration data (NHS England Digital, 2025). All uprated to 2025 prices using the ONS GDP deflator, consistent with the approach applied to the WELLBY value.

[18] Linear mapping from 0% to 45%, informed by WPAI- and WLQ-based studies of mental health-related presenteeism (Beck et al., 2011; Asami et al., 2015; Chokka et al., 2021; Grasso et al., 2022), more on Section 3.3.

[19] Based on ONS median hourly earnings and average weekly hours by employment status (full-time/part-time).

[20] The WELLBY is defined as a one-point life satisfaction increase sustained over 12 months; this is a unit of measurement rather than a default duration assumption, and HM Treasury's supplementary wellbeing guidance requires shorter-lasting effects to be scaled proportionally. Evidence on the persistence of wellbeing improvements from creative health interventions remains limited, but Frontier Economics' review for DCMS's Culture and Heritage Capital Programme (2024) assumes a central duration of six months for short-term cultural engagement where longer-term follow-up evidence is unavailable, consistent with the six-month follow-up period used in comparable interventions such as Coulton et al.'s (2015) community singing trial. The same factor is applied to secondary benefits, which are derived from the same estimated SWEMWBS improvement and cannot be assumed to persist for longer than the primary effect.



Programme costs were compiled in line with Green Book principles, capturing the full economic cost of delivery rather than only the costs falling to the delivering organisation. Alongside staff time, practitioner fees, venue hire, materials and overheads, this includes volunteer time (valued at the ONS average hourly wage), in-kind contributions from partner organisations, and the time and travel costs incurred by participants in attending sessions free of charge. On this basis, total programme costs are estimated at £756.48 per beneficiary

Table 5. Flip of the Coin: estimated social value (2025 prices)

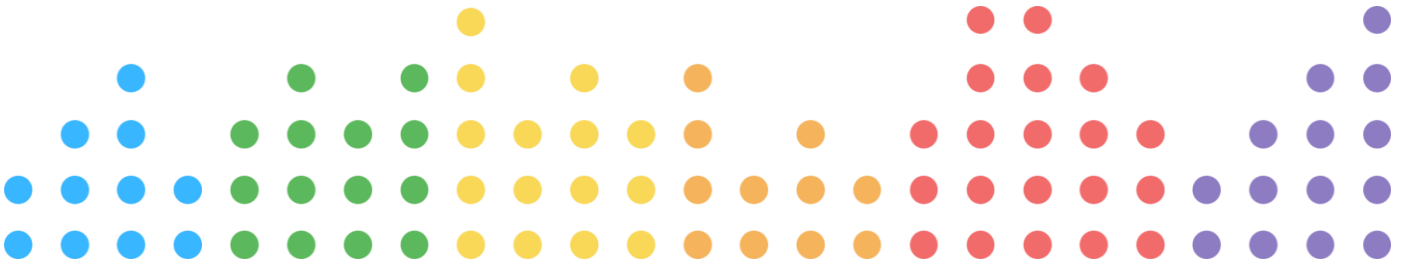
Component	Total	Per beneficiary
Primary benefit (before adjustment)	£662,883.11	£20,087.37
Secondary benefit (before adjustment)	£5,054.29	£153.16
Total benefit (before adjustment)	£667,937.40	£20,240.53
Total benefit (after 0.50 duration adjustment)	£333,968.70	£10,120.26
Programme cost	£200,467.86[21]	£756.48
Benefit-cost ratio		13.38

Source: Authors' own calculations. Primary benefits estimated via the WELLBY framework following Fujiwara et al. (2017). Secondary benefits estimated using USoc waves 7 and 10, with unit costs from Greater Manchester Combined Authority (2025), citing Jones et al. (2025). Data provided by Flip of the Coin SCIO (courtesy of Lucy Campbell, CEO).

Comparing the adjusted social value per beneficiary (£10,120.26) with the estimated cost per beneficiary (£756.48) gives a benefit-cost ratio of 13.38, meaning that **each pound invested in the programme is estimated to generate approximately £13.38 in social value**.

These results should be interpreted as indicative rather than definitive, reflecting the limitations of the underlying programme data and the unquantified additionality factors discussed in Section 5. The small sample size leaves the DiD estimate sensitive to individual extreme values, and since baseline SWEMWBS was not collected prospectively, retrospective reporting could bias the estimated change upwards. Secondary benefits carry further uncertainty, as GP and outpatient visit rates are imputed from the matched USoc control group rather than observed directly. The benefit-cost ratio above should therefore be read as indicative of the order of magnitude of Flip of the Coin's social value, not as a precise estimate.

[21] Total programme costs reflect all beneficiaries reported by Flip of the Coin over the evaluation period, whereas the benefit estimates are based on the 33 participants for whom SWEMWBS data were collected. The cost per beneficiary is applied to this subsample on the assumption that average costs do not vary systematically between participants included in the evaluation and the wider beneficiary population.



4.2. Odd Arts

Odd Arts is a theatre and arts organisation working with individuals who face high levels of inequality and disadvantage. Its work uses creativity as a way to generate positive and lasting changes.

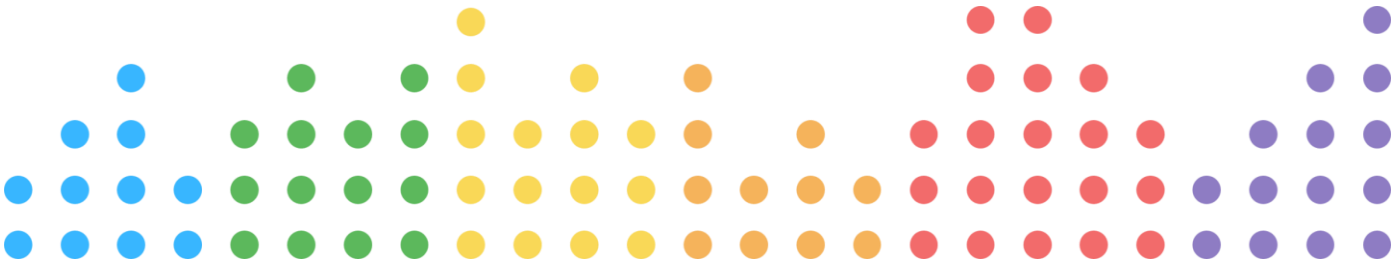
For this evaluation, Odd Arts provided data from the Wellbeing Your Way (WYW) programme, which was delivered during 2024 in three different settings: a custodial facility, a community organisation, and schools. The dataset includes 61 participants with pre- and post-intervention SWEMWBS scores, collected at the beginning of the programme and at completion, which was around 12 weeks later. Additional information was also available on age, gender, ethnicity, education, employment and training status, hours completed, and area of residence.

The sample is mainly composed of women, participants aged 25 or over, and individuals who were not in education, employment or training. Baseline SWEMWBS scores are low, with an average of 14.5. Most participants were located in the two lowest categories of the 12-category valuation scale used in this evaluation. This fact suggests that the programme was working with participants with very low levels of mental wellbeing at entry, and therefore with a considerable margin for improvement.

4.2.1. Methodology

As in Flip of the Coin, no control group was collected as part of the WYW programme evaluation and a comparison group was instead constructed using national data from USoc. This is particularly limiting for Odd Arts, as a substantial proportion of participants were in a custodial setting; since individuals living in custodial institutions are not included in household panel surveys, there is no directly comparable population available in USoc.

PSM was also used to construct a comparison group from USoc, restricting the sample to Northwest England. Matching was based on nearest-neighbour 1:3 matching using baseline SWEMWBS score and gender from wave 10. Other variables, including age, ethnicity and employment status, were explored but resulted in poorer balance between the treatment and comparison groups, reflecting the distinctive characteristics of this population, for whom comparable individuals are difficult to identify within a national household survey. One participant was excluded because gender information was missing, leaving a final analytical sample of 60 programme participants and 180 matched comparison individuals.



After matching, balance improves, although not for all variables to the same extent. Gender reaches an acceptable level of balance (SMD = 0.23), while baseline SWEMWBS remains more unbalanced (SMD = -0.41), reflecting the fact that a large share of Odd Arts participants starts from very low levels of mental wellbeing, which are less common among the general population sample.

The direction of this imbalance is important. The matched control group has a higher average baseline SWEMWBS score than the treated group, 16.19 compared with 14.52. Thus, the counterfactual group starts from a better wellbeing position than the Odd Arts participants. Thus, we are more likely to underestimate the programme effect. The results presented in Section 4.2.2 should therefore be interpreted taking this caveat into account.

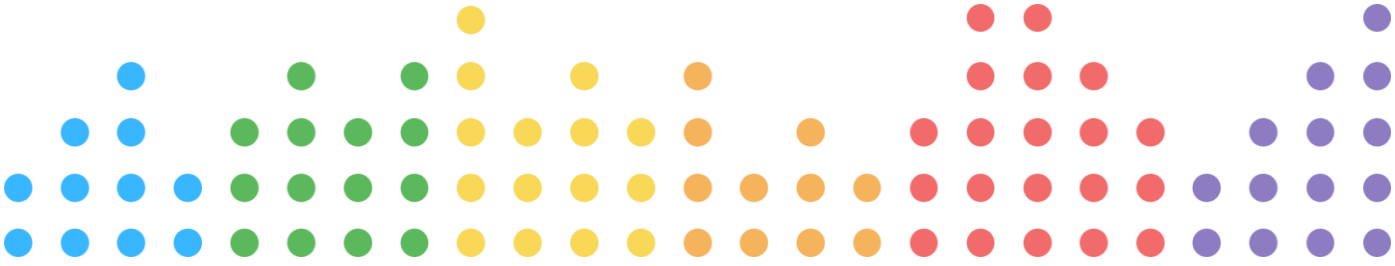
Table 6. Balance diagnostics before and after matching

Variable	SMD before matching	SMD after matching
Baseline SWEMWBS	-2.43	-0.41
Gender (female)	1.79	0.23

Source: Authors' own calculations

Programme impact was estimated using the same **DiD** model described for Flip of the Coin, which compares the change in SWEMWBS among Odd Arts participants with the corresponding change observed in the matched comparison group over time. The model takes the same form, with Treated equal to 1 for Odd Arts participants and 0 for matched controls. No additional covariates are included, since the prior matching step already controls for the variables available in both datasets. This is particularly relevant here, as a substantial proportion of Odd Arts participants had missing data on other characteristics such as age, ethnicity and Employment, Education and Training (EET) status, and their inclusion would have reduced the effective sample size considerably.

A key limitation of this specification is the parallel trends assumption -the requirement that, in the absence of the intervention, the treated and control groups would have followed the same trajectory. This assumption is difficult to satisfy here for two reasons. The observation window for the control group spans approximately three years (wave 10 to wave 13), while the Odd Arts intervention lasted approximately 12 weeks. Moreover, as noted, this population has no direct analogue in USoc. These limitations are reflected in the interpretation of results.



4.2.2. Results

Across the 61 participants, mean SWEMWBS scores increased from 14.54 at baseline to 22.03 at follow-up, a gross improvement of 7.49 points. This represents a substantial shift -the average participant moved from the lowest category of the valuation scale to category 5 (scores 21–22). However, a proportion of this change would be expected to occur in the absence of the intervention.

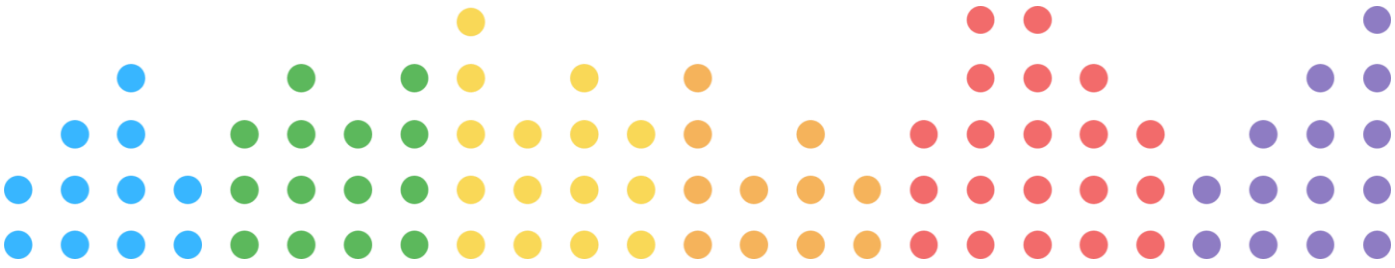
The same valuation approach set out for Flip of the Coin -WELLBY monetisation of primary benefits, fixed-effects estimation of secondary benefits, and the 0.50 duration adjustment- is applied here to Odd Arts's 61 participants.

The DiD estimate of 4.30 SWEMWBS points was monetised using the 12-category WELLBY valuation from Section 3, with each participant's attributable benefit calculated as the difference in monetary value between their baseline category and their baseline-plus-improvement category. This produces an estimated total primary benefit of £1,099,092.88, equivalent to £18,017.92 per participant.

Secondary benefits amount to an estimated reduction of 15 GP visits (£527), 4 outpatient visits (£867), and 0.35 percentage points in the probability of hospital admission (£1,142.72). Presenteeism was not estimated, as most participants were not in employment at baseline, predominantly due to the custodial setting, and this channel is therefore not relevant for the majority of the sample. Nevertheless, qualitative data point to a shift in EET status following the programme, mostly through roles within the custodial setting. Together, secondary benefits amount to an estimated £2,536.76, or £41.59 per participant.

Applying the duration adjustment reduces the combined primary and secondary benefit from £1,101,629.64 to £550,814.82, or £9,029.75 per participant.

The cost estimate for Odd Arts includes direct staffing and operational costs but may not fully capture all resource costs associated with delivery. Unlike Flip of the Coin, the financial data available for Odd Arts do not include costs such as materials and equipment or, for community-based delivery specifically, venue hire and participant travel; it is possible that some of these costs were incurred but not captured in the data provided, though for activity delivered in custodial settings such categories may genuinely not apply. Costs have been uprated to 2025 prices using the ONS GDP deflator. As a result, the cost per participant for Odd Arts (£707.30) should be treated as a lower-bound estimate. Comparing the adjusted social value per participant with this estimated cost per participant gives a benefit-cost ratio of 12.77, meaning that each



pound invested in Odd Arts is estimated to generate approximately £12.77 in social value, likely an overstatement to the extent that programme costs are incomplete.

Table 7. Odd Arts: estimated social value (2025 prices)

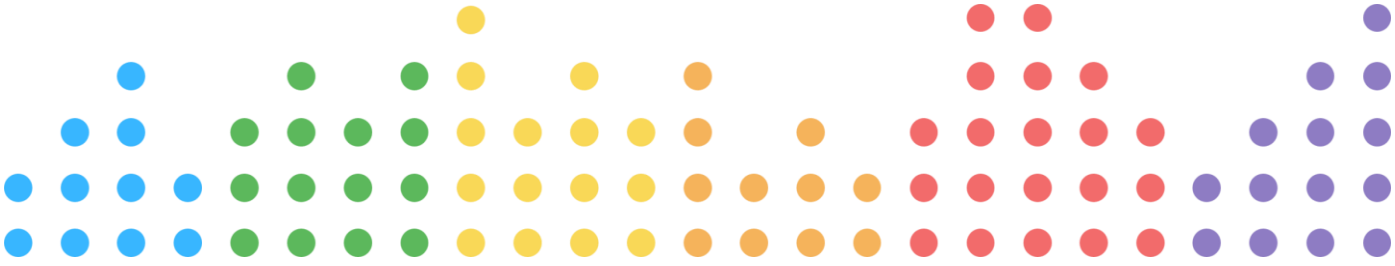
Component	Total	Per beneficiary
Primary benefit (before adjustment)	£1,099,092.88	£18,017.92
Secondary benefit (before adjustment)	£2,536.76	£41.59
Total benefit (before adjustment)	£1,101,629.64	£18,059.50
Total benefit (after 0.50 duration adjustment)	£550,814.82	£9,029.75
Programme cost	£43,145.44	£707.30
Benefit-cost ratio		12.77

Source: Authors' own calculations. Primary benefits estimated via the WELLBY framework following Fujiwara et al. (2017). Secondary benefits estimated using USoc waves 7 and 10, with unit costs from Greater Manchester Combined Authority (2025), citing Jones et al. (2025). Data provided by Odd Arts (courtesy of Chloe Atkar, Data & Monitoring Coordinator).

As with Flip of the Coin, these results should be interpreted as indicative rather than definitive, reflecting the limitations of the underlying programme data and the unquantified additionality factors discussed in Section 5. The parallel trends assumption is particularly difficult to test here, as the intervention took place in 2024 while the closest Understanding Society data with SWEMWBS are from an earlier period, and the specific profile of the Odd Arts sample makes constructing a fully comparable control group more difficult. As with Easter Ross, secondary benefit estimates rely on visit rates imputed from the matched control group rather than observed directly, and the benefit-cost ratio above should be read as indicative of the order of magnitude of Odd Arts' social value rather than as a precise estimate.

5. Conclusions

Creative health interventions are increasingly recognised as a way to improve health and wellbeing, while also contributing to wider social and economic outcomes. However, the evidence needed to demonstrate and monetise this value remains less developed than in other areas of health and public investment. This study addresses this gap by applying, to the greatest extent possible, a Green Book-aligned evaluation framework to two creative health programmes, using propensity score matching and



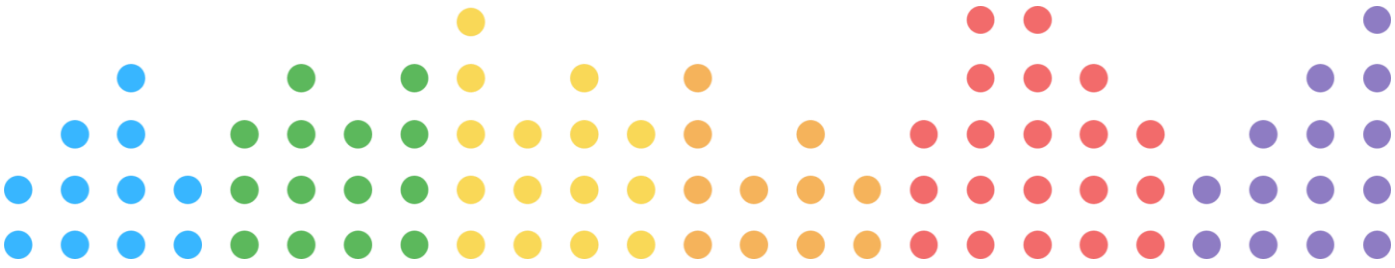
difference-in-differences to estimate the wellbeing change associated with participation.

The analysis highlights the limits of current data collection practice in the sector. Much of the data reviewed was not collected for economic evaluation. Even the two programmes selected as case studies, because they had the strongest available data, did not meet all the requirements of a fully Green Book-compliant evaluation. This is a practical issue as much as a methodological one. Creative health providers often work with people facing complex needs, and extensive data collection can place an additional burden on beneficiaries and delivery teams. Evaluation requirements therefore need to be realistic enough to work in practice, while still being strong enough to support credible decisions about investment.

A central limitation is that many of the wellbeing measures commonly used in creative health, including SWEMWBS, are not preference-based and do not have a direct monetary value for cost-benefit analysis. This study makes a methodological contribution by extending the wellbeing valuation approach set out in Fujiwara et al. (2017) to derive a monetary value for SWEMWBS. A similar approach is also used to estimate secondary benefits related to healthcare utilisation and productivity, where these outcomes were not directly observed. Compared with the original specification, the analysis draws on a larger number of Understanding Society waves and includes a region-by-wave fixed effect, improving control for time-varying regional factors.

This valuation bridge, whilst useful, should be understood as a second-best solution. It allows economic evaluation to be carried out with the type of data that many creative health organisations currently collect. It should not, however, replace better data collection. Where possible, future evaluations should collect outcome measures with a clearer direct link to economic value, as well as direct information on healthcare use, employment and productivity.

A further limitation is the absence of a genuine control group. A Green Book-compliant evaluation would ideally include a prospective and contemporaneous control group, collected alongside the intervention. Neither Odd Arts nor Flip of the Coin collected one. In both cases, a matched comparison group was constructed from Understanding Society. This provides a more credible counterfactual than a simple pre-post comparison, but it is still an approximation. As discussed in Section 4, the main limitations are the temporal gap between the intervention period and the available Understanding Society waves, and, for Odd Arts, the absence of a directly comparable population for participants in custodial settings.



Despite these constraints, the estimates provide a useful picture of the social value generated by both programmes. In both cases, the largest share of value comes from the primary wellbeing benefit, meaning the direct improvement in participants' quality of life. This is important. The value of creative health should not be reduced only to healthcare savings or productivity gains. These secondary benefits matter, but they are not the whole case for investment.

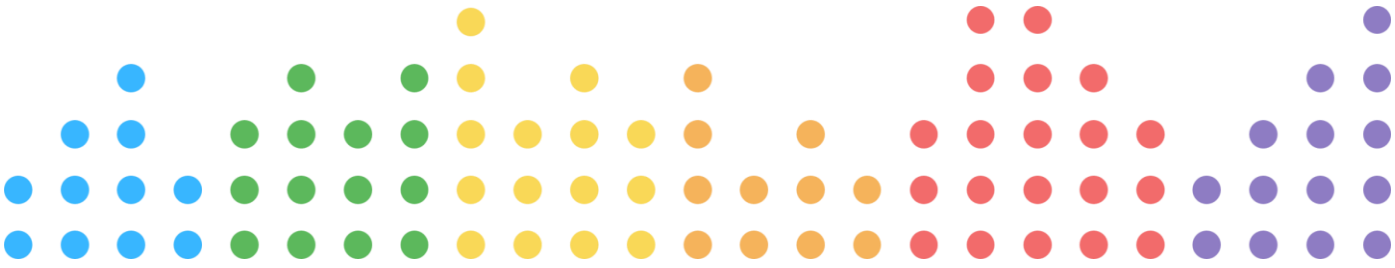
The secondary benefits are smaller and are likely to be understated. They are estimated indirectly from the relationship between SWEMWBS and healthcare and productivity outcomes in Understanding Society, rather than observed directly among programme participants. As a result, they capture only part of the wider spillover effects that may arise from participation. Even so, expressing both primary and secondary benefits in monetary terms makes it possible to combine them with programme costs and produce a benefit-cost ratio.

Applying this approach, Flip of the Coin generated an estimated social value of £333,968.54 across 33 participants, equivalent to £10,120.26 per beneficiary. Against an estimated programme cost of £756.48 per beneficiary, this gives a **benefit-cost ratio of 13.38**.

Odd Arts generated an estimated social value of £550,814.82 across 61 participants, or £9,029.75 per participant. Against an estimated cost of £707.30 per participant, this gives a **benefit-cost ratio of 12.77**.

In both cases, the estimated value is substantially above programme cost, even after applying a conservative duration adjustment to reflect the likely persistence of wellbeing effects. These figures should be read as indicative of the order of magnitude of social value generated, rather than as precise estimates.

The results should also be interpreted in light of the wider Green Book additionality framework. This evaluation accounts for deadweight and duration, but it was not possible to account for displacement or leakage, due to the absence of data on participants' engagement with other services. This is particularly relevant because both samples are drawn from populations with significant support needs. Easter Ross works with people with lived experience of intergenerational trauma, poverty, poor mental health, substance use and contact with the criminal justice system, while the Odd Arts sample includes a substantial share of participants from a custodial setting. Participants in both cases may therefore have been receiving support from other services at the same time. Some part of the observed wellbeing change may be linked



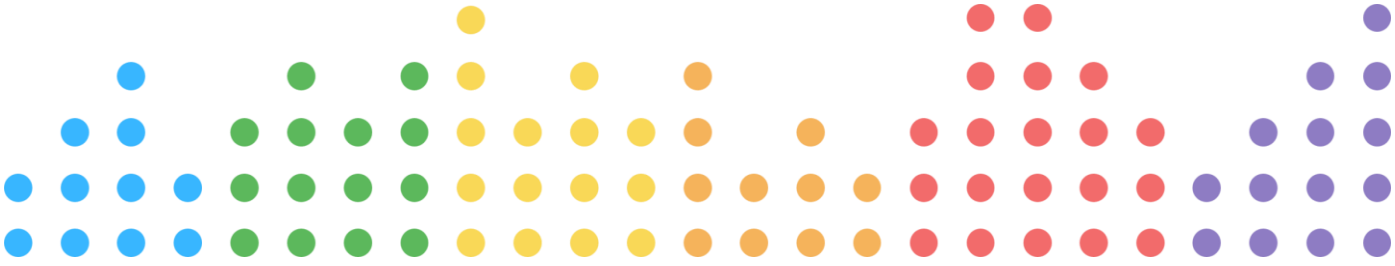
to this wider support rather than to the programmes alone. In the absence of data to quantify these effects, the estimates should be regarded as an upper bound on each programme's net additional social value.

Two further technical issues may also lead to overstatement. First, propensity score matching can only control for observed characteristics. It cannot remove unobserved differences between participants and the matched comparison group. Some imbalance also remains after matching: baseline SWEMWBS in Odd Arts, where the matched control group started from a more favourable position than participants, and employment status in Easter Ross, where two categories remained above the conventional balance threshold. Where these differences are related to changes in wellbeing, the estimated programme effect may partly reflect underlying differences between the groups.

Second, the treated samples are small: 33 participants in Easter Ross and 60 in Odd Arts. The DiD estimates may therefore be influenced by a small number of participants with large changes in SWEMWBS, particularly given the wide range of baseline scores observed in both samples. These concerns cannot be fully quantified with the data available, but they should be borne in mind when interpreting the magnitude of the results.

Taken together, this study should be read as a first exploratory approximation of what a Green Book-aligned economic evaluation of creative health can look like constrained by the availability of data. It is not a definitive assessment of either programme's value.

The next step is to strengthen the evidence infrastructure around creative health. This means more consistent outcome measurement, larger samples, direct data on healthcare use and productivity, and genuine control groups collected prospectively where possible. Together with the Economic Evaluation Guidelines for Creative Health, this report is intended to support that process: it offers a worked example that other organisations in the sector can draw on when designing their own data collection and evaluation practices in line with the Green Book. The estimates presented here suggest that creative health can generate meaningful social value beyond its direct delivery cost, but confirming this with confidence, and making the case for investment on equal terms with other forms of public spending, will require the stronger evidence base this study has sought to work towards.

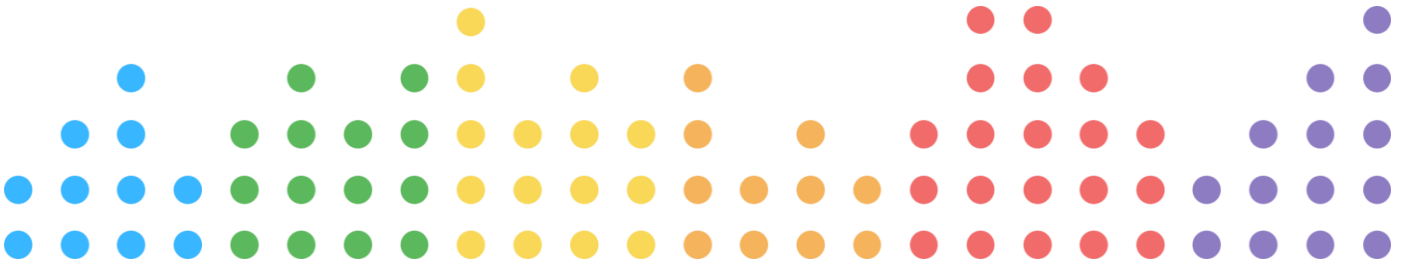


6. Full Statistical Output and Model Diagnostics

OLS fixed-effects model: Life satisfaction

Table 8. OLS fixed-effects model for life satisfaction

Variable	Coefficient	Std. Error	p-value
Swemwbs cat2	0.362***	0.046	<0.001
Swemwbs cat3	0.608***	0.043	<0.001
Swemwbs cat4	1.022***	0.041	<0.001
Swemwbs cat5	1.412***	0.040	<0.001
Swemwbs cat6	1.713***	0.041	<0.001
Swemwbs cat7	1.948***	0.041	<0.001
Swemwbs cat8	2.135***	0.041	<0.001
Swemwbs cat9	2.274***	0.042	<0.001
Swemwbs cat10	2.377***	0.045	<0.001
Swemwbs cat11	2.478***	0.049	<0.001
Swemwbs cat12	2.585***	0.054	<0.001
Log household income	0.032***	0.008	<0.001
Age	0.023	0.014	0.108
Employment-Unemployed	-0.164***	0.026	<0.001
Employment-Retired	0.114***	0.021	<0.001
Employment-Inactive	0.068**	0.024	0.005
Employment-Student	0.229***	0.031	<0.001
Employment-Sick/Disabled	-0.319***	0.042	<0.001
Employment-Other	-0.065	0.052	0.213
Marital status-Single	-0.026	0.024	0.267



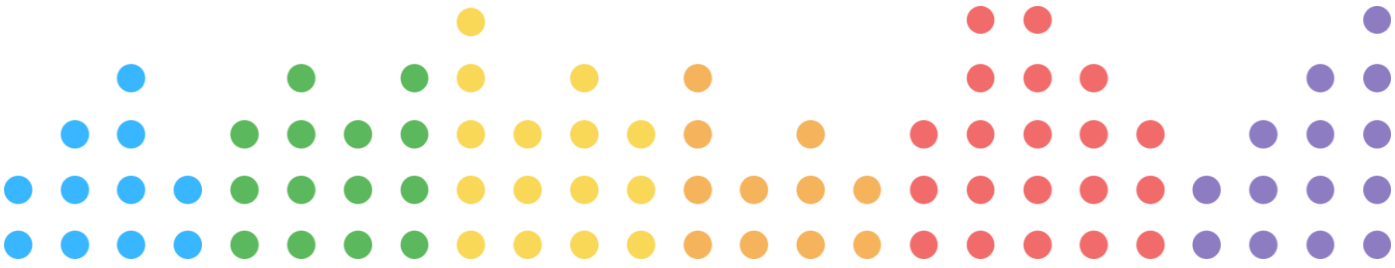
Variable	Coefficient	Std. Error	p-value
Marital status-Separated/Divorced	-0.103***	0.028	<0.001
Marital status-Widowed	-0.028	0.043	0.512
Education-Tertiary(Master;PhD)	-0.012	0.053	0.813
Education-A level	0.040	0.038	0.292
Education-GCSE	0.067	0.050	0.180
Educationother-Other	0.044	0.082	0.592
Education_qual-None	0.072	0.083	0.390
Physical health	0.007***	0.001	<0.001
Number of children	0.016.	0.009	0.076

Notes: Standard errors are clustered at the individual level. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.

OLS fixed-effects model: GP visits

Table 9. OLS fixed-effects model for GP visits

Variable	Coefficient	Std. Error	p-value
SWEMWBS	-0.015***	0.002	<0.001
Log household income	0.006	0.010	0.524
Age	0.014	0.024	0.556
Employment-Unemployed	0.185***	0.041	<0.001
Employment-Retired	0.099**	0.033	0.003
Employment-Inactive	0.200***	0.040	<0.001
Employment-Student	0.037	0.043	0.389
Employment-Sick/Disabled	0.416***	0.068	<0.001
Employment-Other	0.211**	0.074	0.005
Marital status-Single	-0.064.	0.036	0.079



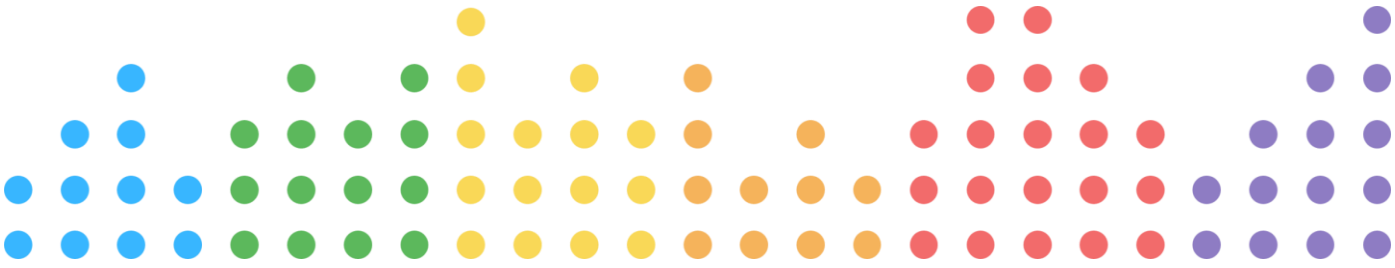
Variable	Coefficient	Std. Error	p-value
Marital status-Separated/Divorced	0.001	0.042	0.986
Marital_status-Widowed	-0.009	0.063	0.886
Education-Tertiary(Master, PhD)	-0.022	0.092	0.812
Education-A level	0.035	0.063	0.578
Education-GCSE	-0.019	0.085	0.828
Education-Other	-0.028	0.152	0.855
Education-None	-0.205	0.126	0.104
Physical health	-0.025***	0.001	<0.001
Number of children	-0.028.	0.017	0.096

Notes: Standard errors are clustered at the individual level. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.

OLS fixed-effects model: Outpatient visits

Table 10. OLS fixed-effects model for outpatient visits

Variable	Coefficient	Std. Error	p-value
SWEMWBS	-0.007***	0.002	<0.001
Log household income	0.005	0.010	0.593
Age	0.030.	0.018	0.097
Employment-Unemployed	0.142***	0.038	<0.001
Employment-Retired	0.149***	0.034	<0.001
Employment-Inactive	0.342***	0.044	<0.001
Employment-Student	0.024	0.037	0.513
Employment-Sick/Disabled	0.275***	0.066	<0.001
Employment-Other	0.067	0.075	0.371
Marital status-Single	-0.087*	0.035	0.013



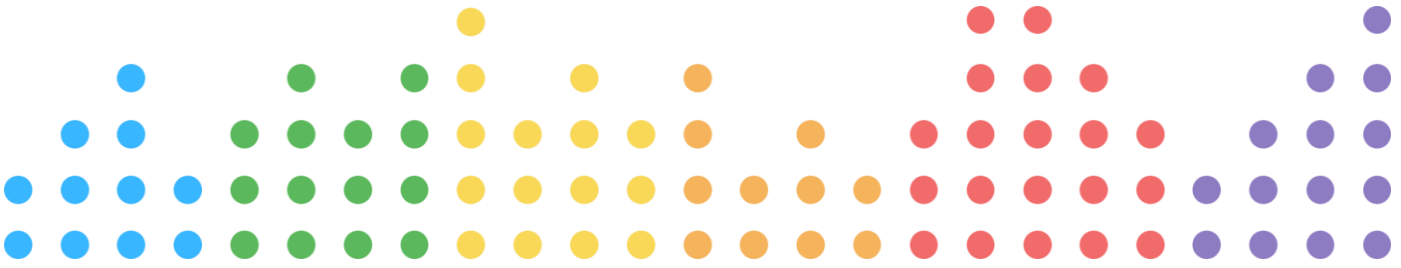
Variable	Coefficient	Std. Error	p-value
Marital status-Separated/Divorced	0.004	0.042	0.925
Marital status-Widowed	0.025	0.059	0.674
Education other-Tertiary (Master, PhD)	-0.123	0.079	0.122
Education-A level	0.073	0.057	0.196
Education-GCSE	0.107	0.073	0.144
Education-Other	0.206.	0.122	0.091
Education-None	0.189.	0.115	0.100
Physical health	-0.026***	0.001	<0.001
Number of children	-0.025	0.017	0.138

Notes: Standard errors are clustered at the individual level. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.

OLS fixed-effects model: Hospitalisation

Table 11. OLS fixed-effects model for hospital admission

Variable	Coefficient	Std. Error	p-value
SWEMWBS	-0.001*	0.001	0.012
Log Household Income	0.000	0.003	0.926
Age	0.008	0.005	0.147
Employment-Unemployed	0.060***	0.012	<0.001
Employment-Retired	0.046***	0.010	<0.001
Employment-Inactive	0.172***	0.016	<0.001
Employment-Student	-0.004	0.012	0.699
Employment-Sick/Disabled	0.091***	0.023	<0.001
Employment-Other	0.046.	0.024	0.053
Marital status-Single	0.006	0.012	0.635



Marital status-Separated/Divorced	-0.020	0.014	0.149
Marital status-Widowed	0.045*	0.021	0.033
Education other-Tertiary (Master, PhD)	-0.000	0.024	0.987
Education-A level	0.048**	0.016	0.003
Education-GCSE	0.050*	0.022	0.025
Education-Other	0.092*	0.037	0.014
Education-None	0.045	0.032	0.163
Physical health	-0.006***	0.000	<0.001
Number of children	0.026***	0.006	<0.001

Notes: Standard errors are clustered at the individual level. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.

Flip of the Coin: DID model for SWEMWBS

Table 12. Difference-in-differences model for SWEMWBS

Variable	Coefficient	Std. Error	t-value	p-value
Intercept	19.373***	0.566	34.218	<0.001
Post period	1.197	0.802	1.492	0.137
Treatment group	-1.383	1.132	-1.222	0.223
DID effect	6.743***	1.602	4.208	<0.001

Notes: The dependent variable is SWEMWBS. The DID effect corresponds to the interaction between treatment group and post period. R^2 = 0.137; Adjusted R^2 = 0.127; N = 263. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.

Odd Arts: DID model for SWEMWBS

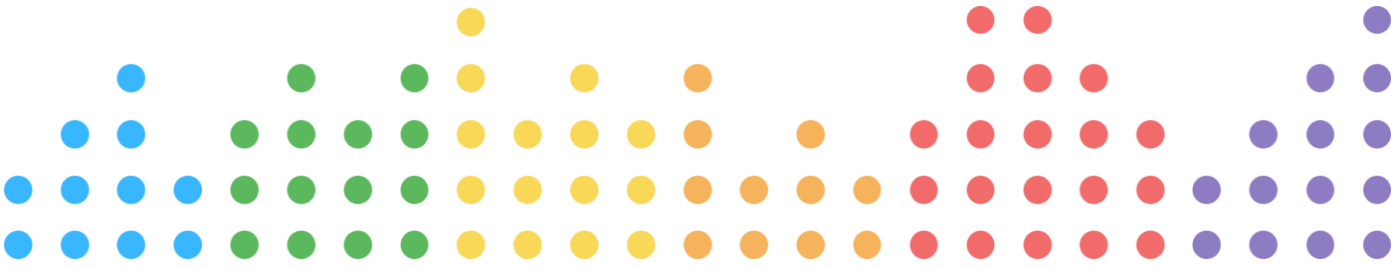
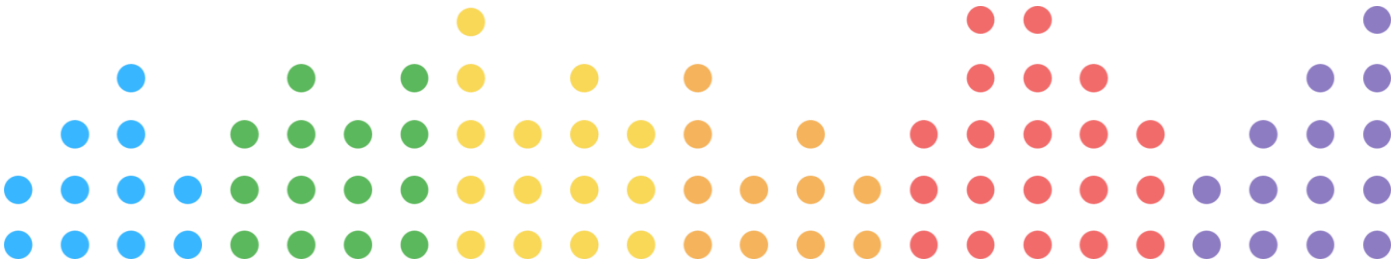


Table 13. Difference-in-differences model for SWEMWBS

Variable	Coefficient	Std. Error	t-value	p-value
Intercept	16.194***	0.307	52.694	<0.001
Post period	3.191***	0.438	7.279	<0.001
Treatment group	-1.653**	0.611	-2.707	0.007
DID effect	4.301***	0.866	4.968	<0.001

Notes: The dependent variable is SWEMWBS. The DID effect corresponds to the interaction between treatment group and post period. R^2 = 0.247; Adjusted R^2 = 0.243; N = 476. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, . p<0.10.



Bibliography

References cited in Guidelines:

All-Party Parliamentary Group on Arts, Health and Wellbeing (2017) Creative Health: The Arts for Health and Wellbeing Inquiry Report. London: APPG on Arts, Health and Wellbeing. https://ncch.org.uk/appg-ahw-inquiry-report (Accessed: 26/06/2026).

Chatterjee, H., & Noble, G. (2016). *Museums, Health and Wellbeing*. Routledge.

Davies, J., Flynn, G., Karkou, V. & Tudor Edwards, R. (2026) Assessing the Economic Impact of the Arts on Health and Healthcare Services in Wales: Arts Council of Wales. https://arts.wales/about-us/research/assessing-economic-impact-arts-on-health-and-healthcare-services-wales (Accessed: 26/06/2026).

Drinkwater, C., Wildman, J., & Moffatt, S. (2019). Social prescribing. *BMJ*.

de Andrade, M. and Angelova, N. (2018). Evaluating and Evidencing Asset-based Approaches and Co-production in Health Inequalities: Measuring the Unmeasurable? *Critical Public Health*, 30(2), pp.232–244. https://doi.org/10.1080/09581596.2018.1541229.

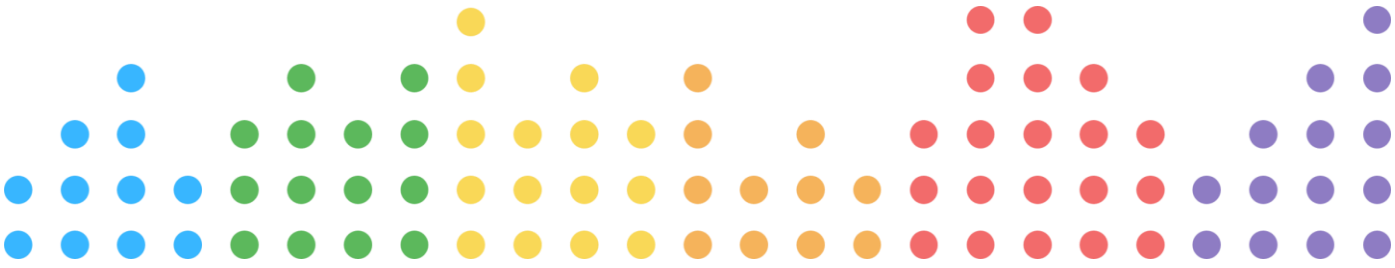
Estevao, C., Fancourt, D., Dazzan, P., Chaudhuri, K.R., Sevdalis, N., Woods, A., Crane, N., Bind, R., Sawyer, K. and Rebecchini, L. (2020) Scaling-up health-arts programmes: the largest study in the world bringing arts-based mental health interventions into a national health service. Psychological Medicine. Cambridge University Press. Published online 23 December 2020.

Fancourt, D., & Finn, S. (2019). *What is the Evidence on the Role of the Arts in Improving Health and Wellbeing?* WHO.

Hearst, J., Gorf, H., Penn, F., Opher, S., Mortimer, A., Davies, L., East, C., Garrett, L., Pearce, M., Rule, E., Waterson, H., & Chatterjee, H. J. (2025) Embedding Creative Health in Integrated Care Systems: Insights from Gloucestershire. London: University College London. https://ncch.org.uk/uploads/Embedding-CreativeHealth-in-Integrated-Care-Systems-Insights-fromGloucestershire.pdf (Accessed: 26/06/2026).

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M. and Geddes, I. (2010) Fair Society, Healthy Lives: The Marmot Review. London: Institute of Health Equity.

Marmot, M., Allen, J., Boyce, T., Goldblatt, P. and Morrison, J. (2020) Health Equity in England: The Marmot Review 10 Years On. London: Institute of Health Equity.



Mueller, C. E., & Gaus, H. (2015). Assessing the performance of the “counterfactual as self-estimated by program participants”: Results from a randomized controlled trial. *American Journal of Evaluation, 36*(1), 7–24. https://doi.org/10.1177/1098214014538487

Mueller, C. E., & Gaus, H. (2018). Treatment effect estimation using self-estimated counterfactuals under varying conditions: A meta-analytic exploration. *Journal of MultiDisciplinary Evaluation, 14*(30), 16–36. https://doi.org/10.56645/jmde.v14i30.484

Malley, J., Rand, S., Netten, A., Towers, A.-M., & Forder, J. (2019). Exploring the feasibility and validity of a pragmatic approach to estimating the impact of long-term care: The “expected” ASCOT method. *Journal of Long-Term Care, 2019*, 67–83. https://doi.org/10.31389/jltc.11

National Centre for Creative Health (NCCH) and APPG on Arts, Health and Wellbeing (2023) Creative Health Review: How Policy Can Embrace Creative Health. London: NCCH. https://ncch.org.uk/creative-health-review (Accessed: 26/06/2026).

NHS Strategic Commissioning Framework (2025) NHS England » Strategic commissioning framework (Accessed: 26/06/2026)

NHS England (2025) Neighbourhood health guidelines 2025/26. https://www.england.nhs.uk/long-read/neighbourhood-health-guidelines-2025-26/ (Accessed: 23/6/2026).

Thomson, L.J., Mughal, R., & Chatterjee, H.J. (2026). Tackling neighbourhood health inequalities in the UK through cross-sector collaborations and community engagement. Frontiers in Public Health. 14: 1832109. https://doi.org/10.3389/fpubh.2026.1832109

Thomson, L.J., Hume, V., Critten, A. & Chatterjee, H.J. (2025). Common features of environmentally and socially engaged community programmes addressing the intersecting challenges of planetary and human health. Frontiers in Public Health: Planetary Health, 13(1449317 https://doi.org/10.3389/fpubh.2025.1449317

Thomson, L.J. & Chatterjee, H.J. (2024) Barriers and enablers of integrated care in the UK: A rapid evidence review of review articles and grey literature 2018-2022. Frontiers in Public Health, 11(1286479). https://doi.org/10.3389/fpubh.2023.1286479

Thomson, L.J., Waterson, H. & Chatterjee, H.J. (2024) Successes and challenges of partnership working to tackle health inequalities using collaborative approaches to

community-based research. International Journal for Equity in Health, 23(135). https://doi.org/10.1186/s12939-024-02216-1

Thomson, L.J., Gordon-Nesbitt, R., Elsden, E. & Chatterjee, H.J. (2021). The role of cultural, community and natural assets in addressing societal and structural health inequalities in the UK: Future research priorities. BMC International Journal for Equity in Health, 20(249), 1–15. https://doi.org/10.1186/s12939-021-01590-4

WHO (2019). *Health Evidence Network Synthesis Report: What is the Evidence on the Role of the Arts in Improving Health and Wellbeing?*

References cited in Annex C:

Asami, Y., Goren, A., & Okumura, Y. (2015). Work productivity loss with depression, diagnosed and undiagnosed, among workers in an Internet-based survey conducted in Japan. Journal of Occupational and Environmental Medicine, 57(1), 105-110. doi: 10.1097/JOM.0000000000000310. PMID: 25563547

Beck, A., Crain, A. L., Solberg, L. I., Unützer, J., Glasgow, R. E., Maciosek, M. V., & Whitebird, R. (2011). Severity of depression and magnitude of productivity loss. Annals of Family Medicine, 9(4), 305-311. doi: 10.1370/afm.1260

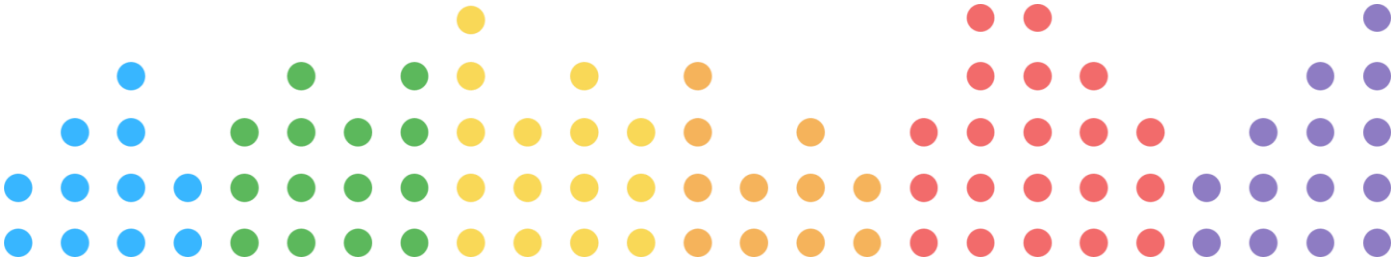
Coulton, S., Clift, S., Skingley, A., & Rodriguez, J. (2015). Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: randomised controlled trial. British Journal of Psychiatry, 207(3), 250-5. doi:10.1192/bjp.bp.113.129908

Chokka, P., Ge, H., Bougie, J., Ettrup, A., & Clerzius, G. (2021). Anxiety symptoms in working patients with major depressive disorder treated with vortioxetine: Associations with clinical and treatment outcomes in the AtWoRC study. Therapeutic Advances in Psychopharmacology, 11. doi:10.1177/20451253211013148

Dancer, S. (2014). Additionality guide (4th ed.). Homes and Communities Agency.

Dayson, C., & Bashir, N. (2014). The social and economic impact of the Rotherham Social Prescribing Pilot: Main evaluation report. Sheffield Hallam University.

Fancourt, D., & Finn, S. (2019). What is the evidence on the role of the arts in improving health and well-being? A scoping review. Health Evidence Network Synthesis Report 67. WHO Regional Office for Europe.



Flip of the Coin SCIO (2026). Dataset: Easter Ross SWEMWBS pre- and post-intervention data. Provided for analysis by Lucy Campbell, CEO, Flip of the Coin SCIO.

Frontier Economics. (2024). Culture and Heritage Capital: Monetising the impact of culture and heritage on health and wellbeing. Report for the Department for Culture, Media and Sport.

Fujiwara, D., & Campbell, R. (2011**).** Valuation Techniques for Social Cost-Benefit Analysis: Stated Preference, Revealed Preference and Subjective Well-Being Approaches. HM Treasury and Department for Work and Pensions.

Fujiwara, D., Oroyemi, P., & McKinnon, E. (2013). Wellbeing and civil society: Estimating the value of volunteering using subjective wellbeing data (DWP Working Paper No. 112). Department for Work and Pensions.

Fujiwara, D., Keohane, K., Clayton, V., & Hotopp, U. (2017). Mental health and life satisfaction: The relationship between the Warwick Edinburgh mental wellbeing scale and life satisfaction – a pilot study. HACT & Simetrica.

Grasso, J. R., Chen, S. Y., & Schneider, R. (2022). Examining changes in presenteeism and clinical symptoms in a workforce mental health benefits program. Journal of Workplace Behavioral Health, 37(4), 253–266. https://doi.org/10.1080/15555240.2022.2097087

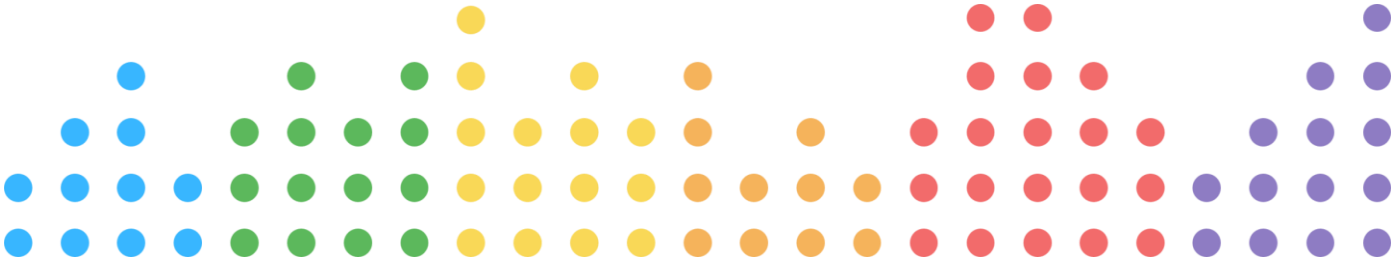
Greater Manchester Combined Authority. (2025). Unit cost database (Version 3.0) [Data set]. https://www.greatermanchester-ca.gov.uk/what-we-do/research/research-cost-benefit-analysis/

HM Treasury. (2026). Magenta Book: Central government guidance on evaluation. https://www.gov.uk/government/publications/the-magenta-book

HM Treasury. (2026). The Green Book: Central government guidance on appraisal. https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

Jensen, A., Holt, N., Honda, S., & Bungay, H. (2024). The impact of arts on prescription on individual health and wellbeing: A systematic review with meta-analysis. Frontiers in Public Health, 12, Article 1412306. https://doi.org/10.3389/fpubh.2024.1412306

Jones, K. C., Weatherly, H., Birch, S., Castelli, A., Chalkley, M., Dargan, A., Findlay, D., Gao, M., Hinde, S., Markham, S., Smith, D., & Teo, H. (2025). Unit costs of health and social care 2024 manual (Version 2) [Technical report]. Personal Social Services



Research Unit, University of Kent & Centre for Health Economics, University of York. doi: 10.22024/UniKent/01.02.109563

Lewandowska, K., & Węziak-Białowolska, D. (2023). The impact of theatre on social competencies: a meta-analytic evaluation. Arts & Health, 15(3), 306–337. doi:10.1080/17533015.2022.2130947

National Centre for Creative Health. (2023). What is creative health? https://ncch.org.uk/what

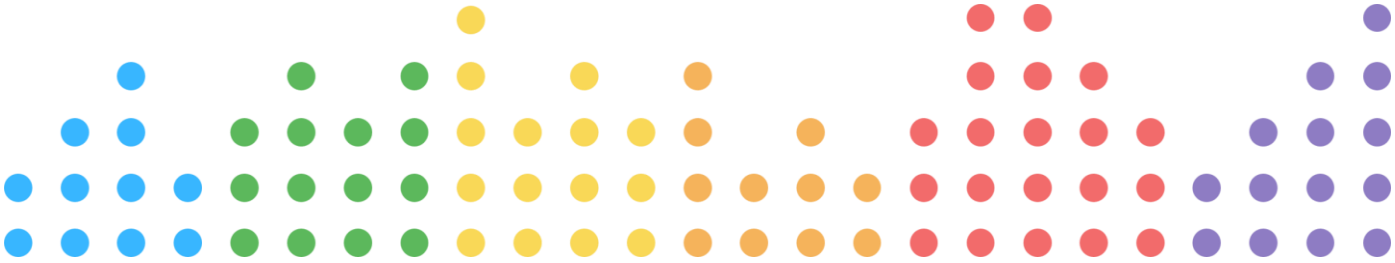
NHS England Digital. (2026). Appointments in General Practice, December 2025 [Data set]. https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/december-2025

Odd Arts. (2026). Wellbeing Your Way SWEMWBS & ONS data [Dataset]. Provided for analysis by the Wellbeing Your Way project.

Office for National Statistics. (2026). Gross domestic product at market prices: Implied deflator: SA [Data set]. Retrieved from https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/mnf2/pn2

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science, 25(1), 1–21. doi: 10.1214/09-STS313

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. Health and Quality of Life Outcomes, 7, Article 15. doi:10.1186/1477-7525-7-15

University of Essex, Institute for Social and Economic Research. (2026). Understanding Society: Waves 1-15, 2009-2024 and Harmonised BHPS: Waves 1-18, 1991-2009 [Data collection, 21st ed.]. UK Data Service.

Wang, S., Mak, H. W., & Fancourt, D. (2020). Arts, mental distress, mental health functioning & life satisfaction: fixed-effects analyses of a nationally-representative panel study. BMC Public Health, 20, 208. https://doi.org/10.1186/s12889-019-8109-y