

NHS Race and Health Observatory Learning and Action Network

Guidance for evaluating spread,
scale-up and sustainability

April 2026

Contents

List of acronyms	4
Acknowledgements	5
Foreword	6
1. Introduction	7
1.1: Background to the Maternal and Neonatal Learning and Action Network (LAN)	7
1.2: The evaluation partners	7
1.3: Purpose of this document	8
1.4: Definitions of key terminology	9
1.5: Rationale for evaluating spread, scale-up and sustainability	10
2. Planning an anti-racist evaluation	11
2.1: An anti-racist approach to evaluating QI	11
3. Co-production with racially minoritised lived experience partners	13
3.1: Guiding principles for co-production with racially minoritised lived experience partners	13
4. Developing a programme theory	16
4.1: Guidance for articulating the programme theory	16
4.2: Useful resources for developing a programme theory	17
5. Recommendations for evaluation design	18
5.1: Frameworks and possible evaluation questions	18
5.2: Possible data collection methods	22
5.3: Measurement and indicators	22
5.4: Analysis plan	23
5.5: Ethical considerations	24

5.6: Anticipated limitations	24
6. Reporting and dissemination	25
6.1: Audiences	25
6.2: Alignment with RHO communications strategy	25
6.3: Informing future scale-up	25
6.4: Recommendations	26
6.5: Transparency of reporting and impact tracking	26

List of acronyms

HIN	Health Innovation Network South London
IHI	Institute for Healthcare Improvement
LAN	Learning and Action Network
MUSIQ-AR	Model for Understanding Success in Quality - Anti-Racism
RE-AIM	Reach, Effectiveness, Adoption, Implementation, and Maintenance
RHO	(NHS) Race and Health Observatory
QI	Quality Improvement
UCLP	UCLPartners

Acknowledgements

This document was prepared by Dr. Lucy Goulding¹, with significant contributions from Fola Afolabi¹, Kate Cheema¹, Rob Lee², Helen Sheldon², Rohit Syed-Hasan¹ and Dr. Andrew Walker².

Affiliations

1. UCLPartners Evaluation and Insights Team
2. Health Innovation Network South London

The evaluation of the Maternal and Neonatal Learning and Action Network (LAN) was guided by an evaluation advisory group whose input has shaped the work presented here. Members included: Camille Aznar, Alima Ashfaq, Tracey Bignall, Manju Chandiramani, Nicola Crowley, Catherine Dale, Hebe Davies-Colley, Vicky Kerr, Abigail Mensah.

Sincere thanks to Jo Setters and Louise Thomas from UCLPartners who supported the delivery of the evaluation and to Katie Harrison (UCLPartners) and George London (HIN South London) for support with communications.

We are very grateful for significant input and guidance from members of staff at the NHS Race and Health Observatory (RHO) and Institute for Healthcare Improvement (IHI), including: Dr. Nandi Simpson, Laura Mallalieu, Sarindi Aryasinghe, Minara Chowdhury, Tricia Bolender, Corinna Parisi. The RHO Maternal and Neonatal Health Advisory Group also provided valuable feedback on the LAN evaluation.

We would like to thank Professor Julie Reed and Dr. Amar Shah for their expert review of an early version of the Model for Understanding Success in Quality - Anti-Racism (MUSIQ-AR), and the LAN quality improvement (QI) teams who tested MUSIQ-AR and gave their time to share their perspectives and support the evaluation.

Microsoft 365 Copilot was used to support the development of parts of this document under the instruction of the evaluation team and with close oversight. Specifically, Copilot was used to help draft initial wording of the co-production principles (section [3.1](#)) which were then significantly adapted and strengthened by the evaluation team; to refine draft suggestions in section [6](#) on reporting and dissemination; to synthesise information; and to refine some of the language for clarity and consistency.

Foreword

The NHS Race and Health Observatory is an independent organisation, established to explore ethnic inequalities in access to healthcare, experiences of healthcare, health outcomes, and inequalities experienced by black and minority ethnic members of the health and care workforce. The Observatory collates evidence, and commissions research to address evidence gaps and, importantly, works with stakeholders across the healthcare system, to mobilise change and develop tools and resources to address racial and ethnic inequities in health and healthcare.

The guidance presented in this document furthers our aim to provide resources to support the NHS to progress race equity; it draws on the expertise of evaluation partners from the Health Innovation Network South London (HIN) and UCL Partners (UCLP), and their experience of taking an explicitly anti-racist approach to evaluating the NHS Race and Health Observatory (RHO), maternal and neonatal Learning and Action Network (LAN). The RHO LAN aimed to develop an anti-racism focused QI approach, that supports practitioners to identify and address racism within maternity and neonatal services; and identify, scale and spread improvement approaches that embed anti-racism into services and improve maternal and neonatal health outcomes. The guidance presented here compliments the RHO LAN evaluation and outlines considerations for evaluating scale and spread.

As the NHS undergoes further large-scale transformation, we know that the risk of de-prioritisation due to resource constraints is likely to disproportionately affect race equity initiatives. Evaluation as a source of evidence to support understanding of effective anti-racism interventions therefore becomes acutely important to retain a focus on race equity and support organisations as they seek to move from high-level commitments, to redressing racial inequities in health access, experience and outcomes for people from Black, Asian and ethnic minority backgrounds. Drawing on the breadth of expertise of HIN and UCLP evaluation teams, and rooted in MUSIQ AR and RE-AIM frameworks, the guidance and considerations presented here are directly applicable to evaluation of initiatives aimed at spread, scale and sustainability of the RHO LAN programme; but also, to evaluation of spread and scale-up of anti-racism QI initiatives more broadly, including beyond maternal and neonatal health.

The persistence of racial inequities in health calls for determined, coordinated action, to enhance anti-racist improvement practice. Key to this is an understanding of how contextual factors influence effectiveness of interventions, and transferrable lessons for scale and spread. We commend UCLP and HIN colleagues on producing this guide in support of that aim, and we hope that you find it helpful.

Dr Nandi Simpson
Director of Implementation NHS Race and Health Observatory

1. Introduction

1.1: Background to the Maternal and Neonatal Learning and Action Network (LAN)

In 2024, the NHS Race and Health Observatory (RHO), in partnership with the Institute for Healthcare Improvement (IHI) and supported by the Health Foundation, launched an innovative maternal and neonatal Learning and Action Network (LAN), which aimed to address ethnic health inequalities via an anti-racism focussed Quality Improvement (QI) programme. The programme was scheduled to run for 15 months, with formal support from the programme team ending in June 2025.

Ten teams, from eight Integrated Care Systems, in four regions participated. The teams each focused on one of four conditions where evidence highlights significant ethnic inequalities in health outcomes: post-partum haemorrhage; preterm birth; maternal mental health; and gestational diabetes.

The QI programme supported these teams to make and sustain changes via improvement coaching, anti-racism webinars, clinical webinars, and online and in-person sessions during which teams shared and developed their learning around implementing anti-racism focussed QI.

1.2: The evaluation partners

The Health Innovation Network South London (HIN) together with UCLPartners (UCLP) were commissioned to evaluate the LAN. The HIN and UCLP are both Health Innovation Networks - regional partnerships that connect the NHS, communities, academia, and industry to accelerate the adoption and spread of healthcare innovations that improve patient outcomes and system efficiency.

The evaluation was delivered by a core team and supported by a wider advisory group, including members with relevant lived experience.

The evaluation report is published separately to this document.

1.3: Purpose of this document

At the outset, the evaluation of the LAN had two overarching aims:

-
- 1. To understand the feasibility and acceptability of an anti-racism focused improvement approach to address ethnic inequalities in maternal and neonatal health.*
 - 2. To work with RHO and IHI teams to develop an evaluation protocol that enables them to understand contextual and implementation factors that affect the effectiveness and scalability of interventions developed through the anti-racism focused LAN approach, to inform subsequent phases of the programme.*
-

This document directly relates to the second aim. However, it should be noted that the second aim is also addressed in the main evaluation report which has a key focus on understanding contextual factors that impact anti-racism focussed QI. The evaluation methodology and findings are described in detail in the evaluation report. We cross-reference to the evaluation report within this document.

As of April 2026, implementation plans for subsequent phases of the LAN programme have not been established. However, significant related national work is underway via the Perinatal Equity and Anti-Discrimination Programme, commissioned by NHS England, and the National Maternal Care Bundles for post-partum haemorrhage and maternal mental health. The findings of the LAN evaluation, and the guidance for evaluating spread and scale-up presented here, should be of direct relevance to these national initiatives.

Therefore, rather than presenting a protocol that provides step-by-step instructions for evaluating subsequent phases of the LAN programme, within this document we present considerations and guidance for evaluating the spread and / or scale-up of anti-racism focussed quality improvement programmes. This advice is informed by the findings of the evaluation and expertise that UCLP and HIN have on this topic.

1.4: Definitions of key terminology

We use the terms spread, scale-up and sustainability within this document. We use the following definitions of these terms from THIS Institute¹.

Spread

(sometimes referred to as adoption)

The term 'spread' refers to **taking improvement interventions that have been successful in one setting and implementing them somewhere else**. This can be done with or without modifying the original intervention. The original setting doesn't have to be within healthcare - you can take ideas and inspiration from other industries such as hospitality for improving team culture or the transport industry for scheduling.

Scale-up

Scale-up involves **developing the infrastructure needed to enable widespread implementation of improvement interventions**. Think about the broad range of infrastructure and process needed to successfully scale-up the intervention. It might include technology, systems, behaviour, people, culture, and leadership.

Sustainability

Sustainability refers to improvement interventions being **maintained over time**, so that new ways of working and improved outcomes become the norm.

Spread, scale-up, and sustainability do not happen in a neat, linear process. They tend to be overlapping and interdependent processes.

THIS Institute, 2025

Recommendations for evaluators and those commissioning an evaluation

- Provide clear definitions of key terms (e.g. spread, scale-up, sustainability) and use these terms consistently.
- During the early stages, establish if the evaluation is assessing spread, scale-up, sustainability, or a combination of these.

¹ THIS Institute (2025) [Explain THIS: Spread, Scale-Up, and Sustainability - THIS Institute - The Healthcare Improvement Studies Institute](#) Accessed 04/12/2025

1.5: Rationale for evaluating spread, scale-up and sustainability

Spread refers to the replication and adoption of proven interventions across different settings. In programmes focusing on anti-racism, spread should seek to achieve reduction in health inequity in communities that experience disproportionate health burdens due to racism. The ‘interventions’ to be spread may include both the specific change ideas that are implemented (e.g. developing effective educational materials and guidelines on post-partum haemorrhage for a specific racially minoritised community), and also the anti-racism focused QI methodology that is used to plan, implement and assess the impact of these change ideas (e.g. the adapted, anti-racism focused Model for Improvement). Evaluating spread can help drive wider adoption. By understanding the contextual factors that enable successful dissemination of anti-racism-focused QI methods and interventions (and how these elements may need to be adapted for different local settings) evaluation findings can inform strategies for broader implementation. This, in turn, strengthens efforts to reduce ethnic health inequalities across the system

Scale-up involves creating the infrastructure and capacity, such as workforce development, governance structures, technology, and cultural alignment, necessary for sustained implementation at system level. Evaluating scale-up would determine whether anti-racism focused change ideas and QI methodologies are not only being adopted but also supported by the organisational and policy frameworks required for long-term success. This process would highlight whether improvement programmes are implemented equitably, identify barriers to adoption, and ensure cultural relevance across diverse contexts.

Evaluating sustainability of anti-racism quality improvement programme interventions will determine whether anti-racism focused change ideas and QI methodologies endure beyond the life of the programme and whether adaptations are made to suit local contexts over a longer time horizon. Anti-racism work involves cultural and structural shifts that require time, resources, and leadership commitment to embed into everyday practice.

Recommendations for evaluators and those commissioning an evaluation

- Evaluating the spread, scale-up and sustainability of anti-racism-focused quality QI initiatives is essential to ensure that successful programmes extend beyond isolated pilots and deliver meaningful, population-level impact.
- Robust evaluation of both spread and scale-up provides evidence relevant to sustainability, cost-effectiveness, and real-world outcomes, which are critical for informing policy and resource allocation.
- The Health Foundation have produced a briefing ‘From innovation to widespread adoption: key principles’² which evaluators and implementers may find helpful.

² The Health Foundation (2025) From innovation to widespread adoption: key principles. Available from: <https://q.nhsconfed.org/evidence-and-insights/evidence-and-briefings/from-innovation-into-wide-spread-adoption-key-principles-and-tools> Accessed 18/12/2025

2. Planning an anti-racist evaluation

Any team working on the evaluation of spread and scale-up of an anti-racist focused QI programme should plan and ensure delivery of an actively anti-racist evaluation.

2.1: An anti-racist approach to evaluating QI

The team who worked on the evaluation of the LAN aspired to deliver an anti-racist evaluation. The team drew on guidance from WestEd, a non-profit organisation in the US.³ The ethnic backgrounds and experiences of the evaluation team were considered at the earliest stages in order to improve diversity. The team regularly reflected on cultural humility and their positionality and sought expert input in this area. An advisory group was convened to guide the work of the evaluation team; this group was diverse in terms of ethnicity, professional experience and lived experience. An anti-racism approach was interwoven into the evaluation methodology via integration of the Race and Health Observatory's seven anti-racism principles⁴ into the Model for Understanding Success in Quality - a published framework used to understand how contextual factors influence the success of quality improvement initiatives. The resulting evaluation framework, named the Model for Understanding Success in Quality: Anti-Racism (MUSIQ-AR) which was used to guide data collection and analysis, therefore had an explicit anti-racism lens. This work is further described in section [5.1.1](#), and within the evaluation report.

Recommendations for evaluators and those commissioning an evaluation

- Evaluation teams and evaluation commissioners may find the guidance from WestEd² helpful.
- Proactively compose ethnically diverse evaluation teams and advisory groups from the earliest stages of the evaluative process. Reflect on ethnic diversity, and diversity of professional and lived experience throughout.
- Use MUSIQ-AR to inform data collection and analysis of evaluation findings. The

³ WestEd. (2021). Anti-racist evaluation strategies: A guide for evaluation teams. Available from: https://wested2024.s3.us-west-1.amazonaws.com/wp-content/uploads/2024/07/11171326/anti-racist-evaluations-R10_FINAL-ADA.pdf Accessed 02/12/2025

⁴ NHS Race and Health Observatory (2024) Seven Anti-racism principles. Available from: [7 PRINCIPLES GRAPHIC.v6](#) Accessed 02/12/2025

NHS Race and Health Observatory's seven anti-racism principles³ are integrated within this model. Evaluators should reflect on how MUSIQ-AR is applied and adapted so that it is appropriate for their context, especially given the fluidity and differences across space and time when it comes to systems of race / racialisation. This may include consideration of acceptable terminology for the context e.g. use of 'racially minoritised communities' which is taken from the RHO principles.

3. Co-production with racially minoritised lived experience partners

Co-production of an evaluation means “a way of working, thinking, and approaching each step of your health services evaluation in partnership with end-users” (Young *et al.* 2025, p496)⁵. In this context, the ‘end users’ are racially minoritised individuals who have their own lived experience of utilising maternity and neonatal services, and also other stakeholders within the maternity and neonatal healthcare system, including NHS workers, representatives of Integrated Care Boards and Local Maternity and Neonatal Systems, representatives of Voluntary, Community and Social Enterprise organisations, and programme and policy stakeholders. This section will focus on co-production with racially minoritised people who have experience of services, termed ‘lived experience partners’.

Evaluators, or commissioners of an evaluation, should ensure meaningful, equitable, and sustained involvement of racially minoritised communities with lived experience in the evaluation of the spread and scale-up of anti-racism focused QI programmes. Here we present initial co-production principles that can be considered and further refined by those working on implementation and evaluation of spread / scale-up (including lived experience partners). We recognise that the resources available for evaluation may impact feasibility. Linked to resource is timescales. Meaningful co-production will require time to build trust between lived experience partners and evaluators and partner organisations, including NHS stakeholders. At the outset it would be important to ensure that all stakeholders have the skills needed to be able to build trust.

3.1: Guiding principles for co-production with racially minoritised lived experience partners

- **Co-production:** Lived experience partners should work together with evaluators on evaluation design, data interpretation, development of recommendations and dissemination of findings.
- **Anti-racism:** Actively challenge racism by sharing power, centring racially minoritised

⁵ Young, A.M., Christoffersen, A., Trevor, D., Pateman, K. (2025). *Co-production in Health Services Evaluation*. In: Kaehne, A., Feather, J. (eds) *Handbook of Health Services Evaluation*. Springer, Cham. https://doi.org/10.1007/978-3-031-87869-5_26

voices, and using findings to drive equity and systemic change.

- **Cultural humility:** Maintain a practice of continuous critical reflexivity, explicitly considering positionality and the potential for unconscious bias to influence the evaluation process. Hold regular review and challenge sessions to actively mitigate against racial bias across all stages of the evaluation and to ensure the interpretive lens prioritises lived experiences.
- **Equity and inclusion:** Engagement must go beyond tokenistic representation and embed voices of racially minoritised communities in decision-making
- **Governance and oversight:** Ensure good representation of individuals from racially minoritised communities within governance structures, for example in advisory or steering groups. Ensure these spaces are inclusive and that lived experience partners from racially minoritised communities hold power within the governance structures, moving beyond tokenistic representation. Clear roles and responsibilities should be co-defined with lived experience partners, these may be included within Terms of Reference, which lived experience partners should co-develop.
- **Cultural Safety:** All engagement and evaluation processes should create an environment where people feel respected, understood, and free from harm or discrimination. This means:

Respecting cultural norms: Acknowledge and embrace traditions, beliefs, and practices that are important to participants.

Meeting language needs: Understand and be responsive to the language needs of lived experience partners so that everyone can fully participate. For example, provide interpretation, translation, or plain language materials.

Recognising lived realities: Understand and respond to the social, economic, and historical contexts that shape people's experiences, ensuring that involvement does not reinforce inequalities or cause distress.

However, we recognise there are environments where harm and discrimination may arise. We recommend planning for mitigation strategies and what reparative practice looks like.

- **Incentives:** Understand how lived experience partners want to be recognised for their time and expertise. Often this may involve payment, but recognise that not all individuals want, or are able to receive payment, for example if this may affect benefits. Offer to reimburse expenses, e.g. travel and childcare costs, to mitigate barriers to participation.
- **Accessibility:** Consider accessibility from a range of angles, e.g. geographical, physical, information and digital accessibility. Provide information in a format that is accessible to each lived experience expert.
- **Training and support:** Identify what training and support is needed together with lived experience partners and be responsive to this. For example, this could involve

providing training on QI principles, anti-racism, evaluation methods, and data interpretation tailored for non-technical audiences. Offer regular opportunities for lived experience partners to debrief.

4. Developing a programme theory

Programme theories explain how and why a programme is expected to work. A programme theory is an overarching conceptual framework that encompasses tools like theories of change, logic models and driver diagrams.

Programme theories state:

- the issue or problem that the programme seeks to address and for whom;
- the resources to be provided;
- the actions that will be taken;
- the mechanisms of change (why the actions will lead to improvements);
- the anticipated outcomes and impacts;
- and any assumptions about the context that might influence whether or not the programme is successful.

The programme theory should be tightly linked to the evaluation measurement framework, so that the evaluation captures whether the programme worked as intended (or not) and why. For each element of the programme theory (from resources through to impacts), define what success looks like, and derive measurement domains. A plan for measuring both processes (the actions that are taken) and outcomes (what happens as a result - clinical, service and implementation outcomes) will thus be created. For the evaluation of anti-racism focussed quality improvement programmes, it is likely to be appropriate to include both quantitative and qualitative assessment (see section [5.2](#)).

It is recommended that implementation teams clearly articulate the programme theory for spread / scale up. This could be done with support from colleagues with expertise in evaluation.

4.1: Guidance for articulating the programme theory

- Once initial implementation plans for spread and / or scale-up have been developed, it may be helpful if the programme theory (demonstrating logical links between programme resources, activities, outcomes and anticipated impacts) is captured in diagrammatic form with accompanying narrative to explain the diagram. A Theory of Change, logic model or driver diagram may be used to display the visual part of the programme theory.
- This articulation of how the programme is expected to work should include reference to the evidence base. However, teams should be anti-racist in how they consider the evidence, acknowledging that evidence bases can reflect systemic racism / bias, and

accepting lived experience and decolonial knowledge as valid forms of evidence.

- The potential impact of contextual factors should be considered. The findings from the evaluation of the LAN will be useful.
- A range of stakeholders, including lived experience partners, should be involved in creating the programme theory. Ascertain and act on the access needs of stakeholders, for example provide training, have plain language summaries, consider the appropriateness of online versus in person meetings. The resulting output should describe who was involved in this process and how they were involved.
- The programme theory should be updated during the course of the programme, in line with any changes to programme implementation. Log updates and key decisions to help future learning.
- The articulation of programme aims should be time-bound, so that it can easily be understood what change is expected during the QI programme and what impacts might be observed longer term. This would help stakeholders to reflect on whether programme aims have been achieved and help project teams to plan projects with aims that clearly align with programme aims and aspirations.

4.2: Useful resources for developing a programme theory

- Davidoff F, Dixon-Woods M, Leviton L, Michie S. Demystifying theory and its use in improvement. *BMJ Qual Saf.* 2015 Mar;24(3):228-38. doi: [10.1136/bmjqs-2014-003627](https://doi.org/10.1136/bmjqs-2014-003627). Epub 2015 Jan 23. PMID: 25616279; PMCID: PMC4345989.
- WHO (2025) [How to develop an evidence-informed theory of change for health: WHO technical guidance](#) . Accessed 08/12/2025;
- Community Toolbox, [Chapter 2., Section 1. Developing a Logic Model or Theory of Change - Main Section | Community Tool Box](#). Accessed 08/12/2025

5. Recommendations for evaluation design

5.1: Frameworks and possible evaluation questions

We recommend using two theoretical frameworks that together will help to explore contextual factors that influence QI project success (MUSIQ-AR), and adaptations and dynamics that influence scale and spread (RE-AIM). RE-AIM will provide focussed enquiry on *what happened*, whereas MUSIQ-AR will explain *why it happened*, with an integrated anti-racism lens. These frameworks can be used to develop evaluation questions, inform data collection materials, and structure analysis and reporting.

Evaluation questions should ideally be developed in collaboration with stakeholders including lived experience partners. A long-list of potential evaluation questions for evaluating spread and scale-up, that are grounded in the proposed theoretical frameworks, are presented below. The questions are based on professional expertise and experience of delivering the evaluation of the LAN. It was not feasible to include a wider group of stakeholders including lived experience partners in developing this long-list. The evaluation questions should therefore be further refined with these stakeholders, depending on programme implementation plans, and in light of the resource available for evaluation.

5.1.1: MUSIQ-AR

Understanding context is critical when evaluating spread and scale-up, as what works in one context may not work elsewhere. The evaluation framework must therefore enable the capture of context.

The team working on the evaluation of the LAN created a framework – The Model for Improving Success in Quality–Anti-Racism (MUSIQ-AR) – to understand how contextual factors impact anti-racism focused QI work.

Existing frameworks within the improvement and implementation science literature for understanding contextual factors (for example CICI⁶, CFIR⁷) did not include a focus on anti-racism. The evaluation team therefore adapted the Model for Understanding

⁶ Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, Wahlster P, Polus S, Burns J, Brereton L, Rehfues E. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implementation Sci.* 2017 Feb 15;12(1):21. [doi: 10.1186/s13012-017-0552-5](https://doi.org/10.1186/s13012-017-0552-5). PMID: 28202031; PMCID: PMC5312531.

⁷ Damschroder, L.J., Reardon, C.M., Widerquist, M.A.O. et al. The updated Consolidated Framework for Implementation Research based on user feedback. *Implementation Sci* 17, 75 (2022). <https://doi.org/10.1186/s13012-022-01245-0>

Success in Quality^{8,9} to incorporate the RHOs 7 anti-racism principles³. This was achieved over a series of workshops and consultations, involving people from racially minoritised communities, people with relevant lived experience, and experts in improvement science.

We recommend using MUSIQ-AR to inform the development of data collection materials and to facilitate analysis. Further, the MUSIQ-AR calculator (an Excel-based tool for scoring contextual factors) could be used by implementation teams and QI project teams to assess readiness for conducting anti-racism focussed QI at the outset, helping teams to maximise contextual enablers and mitigate contextual barriers at an early stage. The MUSIQ-AR calculator was tested with LAN QI project teams at the final in person learning session in March 2025 with refinements made based on teams' feedback. The current MUSIQ-AR calculator is available upon request from the evaluation team.

The domains of MUSIQ-AR could be used to pose evaluation questions as follows:

External environment

- How have external policies, regulations, or incentives influenced adoption?
- Were effective structures in place to support cross-system working on anti-racism focused QI?
- Did externally provided support and expertise facilitate capability building and provide opportunities for sharing learning with teams working on similar challenges in other areas?

Organisation

- What is the organisational level of anti-racism maturity? Is there an organisational culture of naming racism, and identifying racial bias, for example within policies?
- What organisational resources (financial, human, technical) were allocated to enable spread?

Microsystem (e.g. the maternity or neonatal unit)

- How did the local team culture and readiness for change affect implementation at each site?
- At the microsystem level, was there a culture of naming racism and identifying racial

⁸ Kaplan HC, Provost LP, Froehle CM, Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Quality & Safety*. 2012;21(1):13-20. <https://qualitysafety.bmj.com/content/21/1/13>.

⁹ Reed JE, Kaplan HC, Ismail SA. A new typology for understanding context: qualitative exploration of the Model for Understanding Success in Quality (MUSIQ). *BMC Health Services Research*. 2018;18:584. DOI: <https://doi.org/10.1186/s12913-018-3348-7>

bias?

QI project team

- How did the composition and capability of QI teams influence success in scaling up? Were teams intentionally diverse in ethnicity, lived experience, and professional experience?

Infrastructure

- Did teams have access to good quality, ethnicity disaggregated service and clinical outcome data? What has been done at national and organisational levels to support data access and utilisation?
- Did the adapted Model for Improvement support spread / scale-up?

5.1.2: RE-AIM

We also suggest employing a framework that provides a structured way of assessing spread, scale-up, and sustainability. Our specific recommendation is to use RE-AIM¹⁰ (Reach, Effectiveness, Adoption, Implementation, and Maintenance).

The five domains of RE-AIM could be used to pose evaluation questions as follows:

Adoption

- To what extent has the programme been scaled across NHS regions? How widely has the programme been adopted?
- Are new organisations willing and able to join anti-racism focussed QI programmes in maternal and neonatal health? Which sites do not join and why?
- What factors influence adoption across different sites? What are the reasons for adoption?
- How has the anti-racism focussed Model for Improvement QI approach been implemented and adapted in new contexts?
- How have specific quality improvement interventions that were developed by teams who participated in the LAN programme been implemented and adapted in new contexts?
- Are the evaluation recommendations around composition of QI project teams successfully applied?
- How is fidelity to the original LAN model maintained or adapted?

¹⁰ [RE-AIM - Home - Reach Effectiveness Adoption Implementation Maintenance](#) Accessed 04/12/25

Reach

- What was the selection process for acceptance onto the anti-racism focussed QI programme?
- To what extent are underserved / racially minoritised communities being reached as the programme expands?
- Is the programme accessible and equitable across different settings?

Implementation

- What investment (time and costs) is needed to support implementation locally and at scale?
- What infrastructure was required to support scale-up?
- What exactly has been implemented locally? What adaptations are made to suit local contexts?

Effectiveness

- Is the programme effective in improving process and outcomes in new contexts?
- Are there unintended consequences when the programme is scaled?
- Have improvement programmes built capacity for anti-racism focussed QI?
- What impact has spread and scale-up had on reducing ethnic health inequalities?

Maintenance (sustainability) *N.B. the time horizon of the evaluation will influence feasibility of studying sustainability*

- Are systems in place to support long-term delivery?
- How long do the implemented change ideas stay in place, and are they adapted over time?
- Are changes in processes and outcomes sustained over time in new locations?

The [RE-AIM website](#) is also likely to be useful for those planning and supporting programme implementation, as evidence-based strategies are provided to help implementers to enhance reach, effectiveness, adoption implementation and maintenance.

5.2: Possible data collection methods

Data collection plans will be dependent upon the resource and timelines available for evaluation, and the number of sites that a subsequent anti-racism focussed QI programme is spread / scaled to.

A mixed methods approach, involving qualitative and quantitative data collection and analysis, is likely to be most appropriate for answering the range of possible evaluation questions posed in section [5.1](#).

The methods employed during the evaluation of the LAN were successful in meeting the evaluation objectives. These methods comprised: interviews with key stakeholders, qualitative analysis of programme level documentation, observation, use of the MUSIQ-AR Excel tool, case studies (comprising qualitative analysis of project level documentation and focus groups with QI project teams).

Surveys and online focus groups may provide a pragmatic means to reaching a larger pool of participants, in line with the enhanced reach of a scaled-up programme.

Scoping work could be undertaken to ascertain the feasibility of secondary analysis of process and outcome data across sites that are working on the same clinical challenge. An important first step would be to establish governance requirements around data sharing.

5.3: Measurement and indicators

Evaluators could draw on RE-AIM to develop indicators that demonstrate the measurable difference that a spread / scale-up programme is making towards it's intended outcomes and impacts. Suggestions for consideration are presented here:

- **Reach:** Proportion of eligible sites that have / haven't adopted the programme; diversity of these sites / representativeness of populations that are disproportionately affected by a specific health inequality; number or percentage of staff and women / birthing people exposed to the QI intervention(s); number of new sites added during the scale-up process.
- **Effectiveness:** changes in process or outcome data, changes in access, experience, and outcomes for racially minoritised groups, staff perceptions and experiences.
- **Adoption:** proportion of sites reporting adaptations made and their rationale; Variation in adoption rates across different settings (urban vs rural, large vs small hospitals).
- **Implementation:** percentage of sites delivering all core components of the QI programme as intended; average cost per site to implement the programme; savings associated with adverse events avoided.
- **Maintenance:** Percentage of sites maintaining QI interventions one year after initial

implementation, percentage of sites showing sustainable integration into routine practice and policy; budget allocation for continued support beyond any initial funding.

5.4: Analysis plan

The analysis plan will be dependent upon the evaluation methods selected by evaluators / evaluation commissioners, so high level suggestions rather than a detailed plan are presented here.

Recommendations for evaluators and those commissioning an evaluation

- Plan analysis early, for example when finalising the evaluation framework during a scoping phase.
- Involve stakeholders, especially those with lived experience.
- The analysis plan needs to address the evaluation objectives and be underpinned by a clear rationale. Plan to synthesise and triangulate findings from data collected via different methods. The evaluation of the LAN provides an example of how this can be undertaken.
- Develop an intersectional approach to analysis to enable understanding of if/how outcomes differ within racially minoritised groups. This would recognise being truly anti-racist requires examination of how structural racism intersects with other forms of oppression.
- For qualitative data analysis, we recommend drawing on MUSIQ-AR to understand contextual factors with an integrated anti-racism lens, and RE-AIM to understand what happened in relation to spread, scale-up and sustainability.
- A framework analysis approach could be helpful. This would provide a structured process for managing and interpreting large volumes of qualitative data while remaining flexible enough to incorporate both inductive (emerging from the data) and deductive (based on pre-set objectives) themes. Gale *et al.* (2013)¹¹ present step-by-step instructions for framework analysis and appraise strengths and limitations of the method.
- The RE-AIM website has tools and resources to help guide analysis: <https://re-aim.org/resources-and-tools/>

¹¹ Gale *et al.* (2013) Using the framework method for the analysis of qualitative data in multi-disciplinary health Research, *BMC Medical Research Methodology*, 13:117, <http://www.biomedcentral.com/1471-2288/13/117>

5.5: Ethical considerations

Across all sections of this document, we have made recommendations that aim to put equity and inclusivity at the heart of evaluations of anti-racism focused QI programmes.

Plans should be made to ensure the psychological safety of evaluation team members, including lived experience partners, and evaluation participants. Openly recognise that undertaking work to tackle racism can be emotive and that difficult feelings can be unexpected or re-traumatising. Provide regular opportunities to debrief with an appropriate person. Be prepared to signpost people to appropriate sources of support.

Evaluators must adhere to data protection legislation, obtain informed consent from evaluation participants, maintain confidentiality, and transparently report the evaluation process and findings.

Resources for evaluators and those commissioning an evaluation

- The National Institute for Health Research Applied Research Collaboration West have developed guidance on best ethical practice in evaluations: [Full-ethics-guidelines-revised-Nov-2020.pdf](#)
- The 15 Health Innovation Networks in England have developed 21 principles for doing high-quality real-world evaluations. This is based on National Institute for Health and Care Excellence (NICE) principles for real world evidence generation: [Health-Innovation-Network-Real-World-Evaluation-Considerations-2025.pdf](#)

5.6: Anticipated limitations

It is well documented that the availability and quality of data disaggregated by ethnicity can be poor. The NHS England [Ethnicity Recording Improvement Plan](#) sets out actions to strengthen the quality, consistency and completeness of ethnicity data recording, but it is expected that poor data will persist and present challenges for both QI teams and evaluators.

6. Reporting and dissemination

The approach to reporting and dissemination of the evaluation findings and learning should be planned as an integral part of the evaluation process to ensure that this learning is shared widely and used to drive meaningful change.

6.1: Audiences

The findings need to reach a range of audiences, including programme stakeholders such as project teams, communities, organisational leaders, policy makers, commissioners, and funders, as well as national and regional bodies like NHS England, Integrated Care Boards, and professional networks. Feedback should be tailored to different audiences and presented in accessible formats. Academics, community researchers and healthcare educators should be engaged to support knowledge mobilisation and future research.

6.2: Alignment with RHO communications strategy

Dissemination should align with the Race and Health Observatory (RHO) communications strategy to ensure consistent messaging and maximise reach. This means using RHO's established channels—such as its website, newsletters, and social media—to publish reports, blogs, and case studies, and timing these outputs to coincide with national campaigns on equity and anti-racism for greater impact. Integration with RHO's strategy will position the programme within a broader national conversation about tackling ethnic health inequalities. However, to maximise reach, the dissemination strategy might need to expand beyond RHO's strategy to align with the communication preferences of different communities.

6.3: Informing future scale-up

The evaluation should not only inform improvements in racialised disparities within maternity and neonatal services but also generate insights that can be applied to other areas of healthcare. Reporting should highlight transferable lessons, identify adaptations needed for different contexts, and provide a clear framework for assessing readiness and building capability in future programmes.

6.4: Recommendations

Recommendations should be a core output of the dissemination process. These should translate evaluation findings into actionable proposals for NHS leadership, voluntary, community and social enterprise organisations, community stakeholders and government, linking them to existing national priorities such as the NHS Equity and Equality Action Plans and MBRRACE reports. Recommendations should include practical guidance on embedding anti-racism and lived experience involvement into quality improvement policy and practice. Recommendations should be co-produced with stakeholders, especially those from the global majority and with lived experience.

6.5: Transparency of reporting and impact tracking

Feedback to communities and stakeholders is essential for transparency and trust. Findings should be shared in accessible formats—such as infographics, short videos, and translated summaries—and validated through community feedback sessions. These sessions provide an opportunity to check assumptions and interpretations, maintain accountability, and demonstrate how input from lived experience has shaped decisions. Knowledge mobilisation activities, such as webinars, podcasts, and learning collaboratives, can further extend the reach of the findings and support their application in practice. Involvement of those with lived experience in the preparation of lay summaries and in knowledge mobilisation activities can ensure the findings reflect their priorities.

Finally, dissemination should prioritise transparency by publishing full evaluation reports alongside concise executive summaries. Impact tracking should be built into the process to monitor how dissemination influences practice and policy over time. This impact tracking should have an anti-racism focus to enable monitoring of how practices and policies are changing racialised disparities. By combining clear communication, inclusive engagement, and strategic alignment, reporting and dissemination can ensure that the evaluation delivers lasting benefits to address ethnic health inequalities in maternity and neonatal services and beyond.