



OPEN ACCESS

# Quality and reporting of large-scale improvement programmes: a review of maternity initiatives in the English NHS, 2010–2023

James McGowan ,<sup>1</sup> Bothaina Attal ,<sup>1</sup> Isla Kuhn ,<sup>1,2</sup> Lisa Hinton ,<sup>3</sup> Tim Draycott ,<sup>4</sup> Graham P Martin ,<sup>1</sup> Mary Dixon-Woods <sup>1</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2023-016606>).

<sup>1</sup>The Healthcare Improvement Studies Institute (THIS Institute), Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

<sup>2</sup>School of Clinical Medicine, University of Cambridge, Cambridge, UK

<sup>3</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

<sup>4</sup>Department of Women's Health, North Bristol NHS Trust, Westbury on Trym, Bristol, UK

## Correspondence to

Professor Mary Dixon-Woods, THIS Institute, University of Cambridge School of Clinical Medicine, Cambridge, UK; [md753@medschl.cam.ac.uk](mailto:md753@medschl.cam.ac.uk)

Received 3 August 2023

Accepted 13 November 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** McGowan J, Attal B, Kuhn I, *et al.* *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2023-016606

## ABSTRACT

**Background** Large-scale improvement programmes are a frequent response to quality and safety problems in health systems globally, but have mixed impact. The extent to which they meet criteria for programme quality, particularly in relation to transparency of reporting and evaluation, is unclear.

**Aim** To identify large-scale improvement programmes focused on intrapartum care implemented in English National Health Service maternity services in the period 2010–2023, and to conduct a structured quality assessment.

**Methods** We drew on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews guidance to inform the design and reporting of our study. We identified relevant programmes using multiple search strategies of grey literature, research databases and other sources. Programmes that met a prespecified definition of improvement programme, that focused on intrapartum care and that had a retrievable evaluation report were subject to structured assessment using selected features of programme quality.

**Results** We identified 1434 records via databases and other sources. 14 major initiatives in English maternity services could not be quality assessed due to lack of a retrievable evaluation report. Quality assessment of the 15 improvement programmes meeting our criteria for assessment found highly variable quality and reporting. Programme specification was variable and mostly low quality. Only eight reported the evidence base for their interventions. Description of implementation support was poor and none reported customisation for challenged services. None reported reduction of inequalities as an explicit goal. Only seven made use of explicit patient and public involvement practices, and only six explicitly used published theories/models/frameworks to guide implementation. Programmes varied in their reporting of the planning, scope and design of evaluation, with weak designs evident.

**Conclusions** Poor transparency of reporting and weak or absent evaluation undermine large-scale improvement programmes by limiting learning and accountability. This review indicates important targets for improving quality in large-scale programmes.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Large-scale improvement programmes are a key strategy for addressing unwarranted variations in quality and safety of care, but their impact is mixed and often limited.
- ⇒ Previous research suggests a number of features of improvement programmes that need to be optimised, but how well these quality criteria are routinely met remains unknown.

## WHAT THIS STUDY ADDS

- ⇒ Many large-scale maternity improvement initiatives in the English National Health Service—including some major national programmes of the last decade—lack an evaluation report.
- ⇒ Where an evaluation report was available, quality and design of programmes against prespecified criteria was highly variable, often demonstrating significant weaknesses.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Poor transparency of reporting and weak evaluation in large-scale improvement programmes undermine learning and accountability; explicit attention to features of quality is necessary to improve the design, conduct and impact of large-scale programmes.

## INTRODUCTION

Variations in quality and safety of health-care have remained troubling and persistent across health systems globally. Efforts

to address these challenges often take the form of large-scale improvement programmes,<sup>1–4</sup> including, for example, multiorganisational collaborative approaches, major initiatives commissioned by policy and professional bodies, implementation programmes and research projects. These programmes are variably effective, with often disappointing results.<sup>5–10</sup> Clarity is, however, now emerging on some of the key features of ‘what good looks like’ for such programmes.<sup>4 10–13</sup> In this article, we report a study that both identifies large-scale improvement initiatives in a clinical area experiencing major patient safety challenges, and offers a structured quality assessment of improvement programmes where an evaluation report was retrievable.

The available literature suggests that a number of features are especially important in large-scale improvement programmes. First, such programmes should be well specified and reported<sup>14–16</sup> to ensure shared understanding of what the programme comprises and its mechanisms.<sup>2 10 17</sup> A second feature of high-quality improvement programmes is that the interventions they use and their delivery should be supported by best available clinical evidence.<sup>11 18–21</sup> Third, high-quality programmes should recognise and meet the requirements for implementation support in participating organisations.<sup>2 22–25</sup> Such support needs to be sensitive to the highly heterogeneous nature of local capability, which has been implicated in variable responses to improvement programmes,<sup>2 10 16 22–24</sup> with lower performing organisations having distinctive support needs.<sup>11 13 25–31</sup> Fourth, consistent with published policy objectives,<sup>32–34</sup> programmes should explicitly address inequalities between socioeconomic and ethnic groups.<sup>35</sup> Fifth, patient and public involvement (PPI) has an important role in enhancing the impact of improvement efforts.<sup>36–42</sup> Sixth, improvement programmes benefit from use of formal published theories, models and frameworks from implementation science to guide their work.<sup>43</sup> Finally, an important feature of good improvement programmes is a commitment to sound evaluation,<sup>11 20 21</sup> including, where possible, assessment of effectiveness, process evaluation and economic evaluation.<sup>44–46</sup>

The extent to which large-scale improvement programmes routinely meet these seven criteria is unclear. Many programmes, including those commissioned or delivered by national-level organisations, are conducted in a context where expectations of some features (eg, programme specification) may be insufficiently explicit, and incentives for high-quality evaluation and reporting may be lacking.<sup>14 15</sup>

English National Health Service (NHS) maternity services are an important example of where quality problems are especially prominent in public discourse<sup>47–49</sup> arising from high-profile organisational failures,<sup>50–53</sup> evidence of persistent unwarranted variation in outcomes,<sup>54–56</sup> rising clinical negligence claims,<sup>49</sup>

culture and workforce challenges<sup>57 58</sup> and inequalities linked to socioeconomic status and ethnicity.<sup>33 35 59–63</sup> These challenges have not yet proved tractable, despite multiple large-scale improvement initiatives.<sup>64 65</sup> Maternity services are therefore an important setting in which to assess quality of large-scale improvement programmes, particularly in relation to their reporting and evaluation.

We aimed to identify large-scale improvement programmes that had been implemented in English NHS maternity services between 2010 and 2023 and, for those with an available evaluation report, to conduct a structured quality assessment based on the selected features identified above.

## METHODS

### Design

Our design was a review with two components: a search for large-scale improvement initiatives implemented in English NHS maternity services in the period 2010–2023 and, for those that met definitional criteria as improvement programmes and had a retrievable evaluation report, a structured quality assessment. We focused specifically on programmes that primarily addressed quality or safety of intrapartum care, which has been consistently implicated in variations in adverse clinical outcomes in maternity care.<sup>66–68</sup> Initiatives primarily focused on antenatal or postnatal care were therefore not in scope.

Our initial exploratory work found that the majority of maternity improvement initiatives were not research projects and had not been reported in the academic literature; relevant information was mostly available in diverse sources such as websites, policy documents and programme reports. To ensure that both our search and our assessment of programmes was nonetheless structured and systematic, our review was informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for Scoping Reviews guidance<sup>69</sup> (online supplemental material A). We also completed all five steps in the Arksey and O’Malley framework,<sup>70</sup> although not sequentially.

As our review was not intended as a full scoping review of research, we did not register the review in an online database; we did, however, produce a protocol that was used to guide the conduct of the review (online supplemental material B).

### Eligibility criteria

We developed prospective criteria to guide the identification of eligible improvement programmes and sources of evidence.

#### Identification of improvement programmes

Initial scoping identified that a large number of highly heterogeneous improvement efforts had taken place in English NHS maternity services in the period we were studying. The following types of initiatives,

**Box 1 Definition of 'improvement programme'**

For the purpose of our study, we defined a healthcare improvement programme as a set of planned activities:

- ⇒ Seeking to address a known quality or safety deficit; or seeking to implement evidence-based recommendations or standards of care or practice.
- ⇒ Implemented at scale, that is, in two or more healthcare organisations or clinical services.
- ⇒ With the characteristics of an organised programme, for example, with a structured set of goals, resources, a programme team and report.
- ⇒ Primarily concerned with improving clinical care quality or safety including structures, processes or outcomes.

strategies and interventions were excluded from our review: incident investigation and inspection programmes; national clinical audits and confidential enquiries (eg, the Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries - MBRACE programme)<sup>71</sup>; organisational restructuring, major system change and service transformation programmes; health technology assessments and trials of digital technologies without an existing evidence base; clinical guidelines or recommendations without an accompanying implementation programme; and single-site quality improvement projects. Additionally, programmes implemented in clinical specialties other than maternity, outside the English NHS, or implemented in full before 2010 were excluded.

Only initiatives that met our definitional criteria as improvement programmes, had an evaluation report available and focused on intrapartum care were eligible for quality assessment. For this purpose, we defined 'improvement programmes' as encompassing a set of planned activities applied at a scale larger than local quality improvement projects, and requiring participation of more than one organisation or clinical service (see [box 1](#) for our full definition).<sup>1-3 72</sup>

Only improvement programmes where an evaluation report could be retrieved were included in our quality assessment, since exploratory work indicated it would not be possible to make reliable judgements about programme quality and reporting without such a report. We defined an 'evaluation report' as a published assessment of programme design, implementation or outcomes, including formative and/or summative evaluation activities.<sup>44</sup> We classified reports as 'retrievable' where they were available for full-text review following their identification in search results, or were otherwise publicly available (eg, published on organisational websites). In determining eligibility for quality assessment, evaluation reports were included without regard to where they had been published (eg, in academic or grey literature) or to the design and quality of the evaluation.

To ensure comprehensive quality assessment of programmes where an evaluation report was available, we supplemented our analysis of evaluation reports with available information from policy reports, programme reports, website entries, peer-reviewed research articles, reviews and study protocols. We excluded editorials, viewpoints, commentaries and letters, non-English language articles and sources published before 2010. Consistent with our focus on intrapartum care, we also excluded sources that primarily addressed quality or safety of care in the antenatal or postnatal periods, or that had only limited focus on intrapartum care (eg, the Getting It Right First Time - GIRFT programme).

**Information sources and search strategy****Information sources**

We searched two research databases, determining that this number was both proportionate to our aims and consistent with published guidance regarding the conduct of scoping studies.<sup>73</sup> Database searches were performed in MEDLINE via Ovid and CINAHL via Ebsco from 1 January 2010 to 8 February 2023. These databases were chosen because of their high subject relevance to maternity care. Subject headings (eg, Medical Subject Headings) and free text search terms and synonyms were included. We did not apply restrictions to publication type. Filters were applied for 'England' or 'NHS', adapted from Ayiku *et al.*<sup>74</sup> The database searches were designed and performed by a health librarian (IK) in collaboration with JM.

A series of structured searches were performed to identify literature or other information relevant to maternity improvement programmes in Google, Google Scholar and websites of national organisations active in UK maternity quality and safety, supplemented by purposive hand searches. Online searches were based on the search strings 'maternity safety programme', 'maternity quality programme' and 'maternity improvement programme', and were performed independently by two researchers (JM and BA).

**Search strategy**

Our search strategy was based on a modified 'PICOS' framework:

- ▶ Patient population—women receiving NHS care during the intrapartum period.
- ▶ Intervention—quality and safety improvement programmes.
- ▶ Comparison—not applicable.
- ▶ Outcome—clinical and other outcomes related to quality and safety of maternity care.
- ▶ Setting—NHS maternity services in England.

Four studies identified during preliminary scoping searches were used as 'golden bullets'<sup>75</sup> to assess and improve the sensitivity and specificity of the search strings in identifying relevant literature.<sup>76-79</sup> We

supplemented structured searches with forward and backward citation tracking of a purposively selected sample of included studies to improve the sensitivity of the search.

The full list of information sources, search strategies for database and grey literature searches and search record templates are provided in online supplemental material C.

### Eligibility screening

Bibliographic database search results were deduplicated in EndNote, imported into Rayyan<sup>80</sup> and screened on the basis of title and abstract. Screening of search results from non-bibliographic sources was performed onscreen by JM and BA; for each search, the first 100 (Google Search and organisational websites) and 500 (Google Scholar) search results were screened and unique sources identified by consensus. Screening of all search results and full-text sources to identify improvement programmes eligible for assessment and relevant sources of evidence was performed independently by two researchers (JM and BA); disagreements regarding the eligibility of both programmes and sources were resolved by discussion.

### Data categories and charting

#### Data categories

For improvement programmes eligible for quality assessment, we charted data under the seven categories in [table 1](#). These categories correspond to selected

features of quality we had identified from the wider improvement literature, as well as identifying basic programme characteristics.

Our assessment of programmes was supported by published standards in two areas. First, we supported our assessment of programme specification (category 1 in the charting framework) and implementation support (category 3) by using a modified Template for Intervention Description and Replication (TIDieR) checklist,<sup>81</sup> which was adapted to apply to improvement programmes ([table 2](#)).

Second, our assessment of programme evaluation (category 7 in the charting framework) was supported by a multi-item checklist informed by UK government guidance on programme evaluation (the ‘Magenta Book’)<sup>44</sup> ([table 3](#)).

#### Charting process

The charting process was supported by a tool built in Microsoft Excel that mapped to the charting framework in [table 1](#) (online supplemental material D). Consistent with scoping review methodology,<sup>70</sup> we developed the data items for extraction into the charting framework iteratively, modifying them as new data were identified and analysis progressed. The charting tool was piloted using a small sample (n=2) of sources and amended prior to formal charting. Two researchers (JM and BA) independently charted all data for six of the seven data categories; data relating to use of theories, models and

**Table 1** Data charting framework

Data category	Description
1 Programme specification (including basic programme characteristics)	Our assessment of quality of programme specification was supported by a multi-item checklist based on the Template for Intervention Description and Replication (TIDieR) adapted for improvement programmes ( <a href="#">table 2</a> ). <sup>81</sup> A simple ordinal scale was used to grade completeness of description for each item: described in full; mostly described; some description; no description. Additionally, we charted data relating to the following basic programme characteristics: <ul style="list-style-type: none"> <li>▶ Time period and scale of implementation.</li> <li>▶ Clinical setting.</li> <li>▶ Target recipients.</li> <li>▶ Type of programme.</li> <li>▶ Use of named improvement tools or methodologies.</li> <li>▶ Key stakeholders involved in funding or commissioning the programme, programme development and service-level implementation support.</li> </ul>
2 Use of evidence-based interventions	We assessed whether programmes implemented interventions supported by published clinical evidence, including assessment of the source or level of evidence cited by report authors. <sup>18 19 21</sup>
3 Description of implementation support for services	We assessed descriptions of implementation support and activities as part of the modified TIDieR checklist ( <a href="#">table 2</a> ). <sup>81</sup> We assessed whether and how programmes customised implementation support for challenged services, which we defined as services in receipt of support from national regulators to improve quality and safety of care.
4 Commitment to reducing inequalities	We assessed whether programmes included the reduction of health or care inequalities as an explicit goal. <sup>32–34</sup>
5 Patient and public involvement (PPI)	We assessed the extent to which programmes involved patients and the public. <sup>41 42</sup> A simple ordinal scale was used to assess the extent of PPI: PPI throughout the programme; partial involvement; none.
6 Use of formal published theories, models or frameworks	Drawing on Nilsen’s taxonomy <sup>107</sup> and the work of Dadich <i>et al</i> , <sup>106</sup> we assessed the use of formal published theories, models and frameworks from implementation science <sup>43</sup> to guide programme design, implementation and evaluation.
7 Programme evaluation	Drawing on UK government guidance on evaluation, <sup>44</sup> we assessed the scope and design of programme evaluation supported by a multi-item checklist ( <a href="#">table 3</a> ).

**Table 2** Multi-item checklist for assessment of programme specification (modified TIDieR checklist)

Checklist item	Brief description of item
1 Programme goal(s)	Clarity of description of the programme goals
2 Mechanism of action/theory of change	Description of the mechanism(s) by which the programme will improve care
3 Materials	Description of physical and informational materials used in/required for programme delivery
4 Procedures and activities	Description of the key activities of the programme and its interventions
5 Modes of delivery	Description of how the programme was delivered, for example, in face-to-face meetings
6 Frequency, duration and time period of local implementation	Description of how often and for what periods of time the programme was delivered
7 Local tailoring and modifications	Description of how the programme or its implementation was modified by local clinical teams
8 Processes for assessing or maintaining fidelity	Description of how fidelity to the programme and its interventions was assessed or maintained, including any strategies used
9 Outcome of fidelity assessment	Description of the outcome of a fidelity assessment(s), that is, the extent to which the programme was delivered as planned
10 Implementation support and activities	Description of activities designed or intended to enable or support programme implementation locally

TIDieR, Template for Intervention Description and Replication.

frameworks were charted by JM. Disagreements in assessment gradings were resolved by discussion.

### Appraisal of evidence and reporting of findings

Though conducting a quality assessment of eligible programmes (based on the seven features in [table 1](#)) was a key objective of our analysis, we did not seek to review evidence of effectiveness of programmes, nor did we aim to conduct an appraisal of the methodological quality of individual sources of evidence, as these were not goals of our study. Consistent with the norm in scoping reviews,<sup>70</sup> we also did not formally aggregate or synthesise evidence, instead developing summaries of the data organised by the charting framework. Key findings and themes are reported under the categories of this framework in the Results section<sup>82</sup> and are summarised in a series of supplemental tables.

### Patient and public involvement

Patients and the public were not involved in the design or conduct of the review.

**Table 3** Multi-item checklist for assessment of programme evaluation

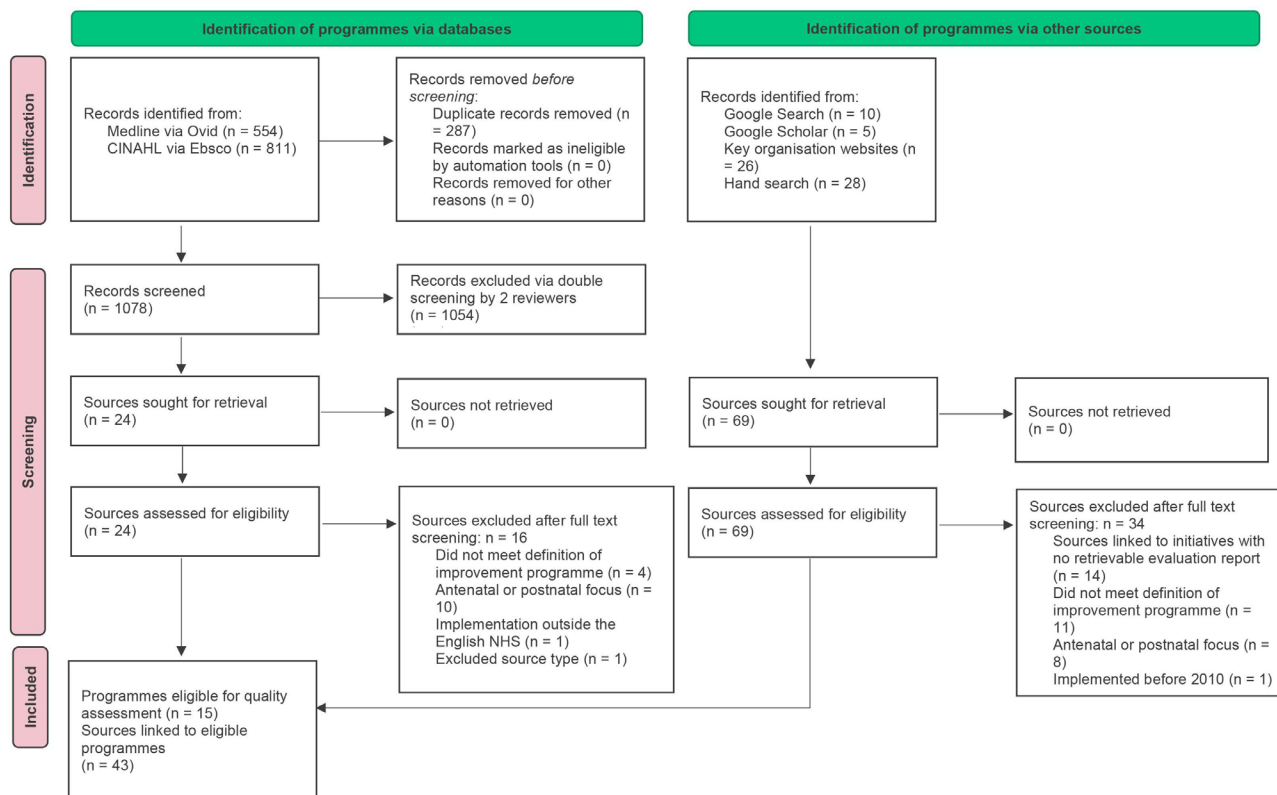
Category	Checklist items
Planning for evaluation	Is there evidence of a prospective evaluation plan? (eg, a published protocol) Was a pilot conducted?
Scope of the evaluation	Is there a clear description of what was assessed as part of the evaluation? Outcome(s) (including clinical and non-clinical outcomes) Processes Design or implementation Economic evaluation
Design	Did the authors name a study design? Named design or evaluation activities Was the evaluation randomised?

## RESULTS

We identified 1434 records via bibliographic databases and searches of the grey literature and other sources. From these, 93 full-text sources were retrieved and assessed for eligibility, including one evaluation report that was not initially available, but was subsequently retrieved through personal correspondence.<sup>83</sup> The process by which sources of evidence were identified and screened to determine their eligibility is summarised in our PRISMA diagram ([figure 1](#)).

Following full-text eligibility screening, we found 38 improvement initiatives (reported across 50 sources) that could not be included in the quality assessment. These initiatives are documented, together with exclusion reasons, in online supplemental material E. Of these, 14 initiatives—including most major initiatives in English maternity services that were implemented during the time period under study, such as the NHS England's Maternity and Neonatal Safety Improvement Programme and the Maternity Safety Support Programme—lacked a retrievable evaluation report. A further 13 initiatives did not meet our definitional criteria as improvement programmes (eg, because they did not report implementation in two or more clinical services). We also excluded nine initiatives that focused primarily on antenatal or postnatal care from the quality assessment, since our scope was limited to intrapartum care. Two other initiatives that had not been implemented in the relevant setting or time period were also excluded at this stage.

We identified 15 initiatives implemented in maternity services in England between 2010 and 2023 that had a principal focus on intrapartum care, met our definitional criteria as improvement programmes and had a retrievable evaluation report. These 15 initiatives were included in the structured quality assessment and were reported across 43 sources of evidence. Of these sources, grey literature constituted the majority (24). Peer-reviewed academic journal articles



**Figure 1** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram (adapted from Page *et al* [109]). NHS, National Health Service.

constituted 19 sources, with nine taking the form of original research articles reporting findings of evaluations.<sup>30 76–79 84–87</sup> Other article types included study protocols (n=5),<sup>88–92</sup> quality improvement reports (n=3)<sup>93–95</sup> and review articles (n=2).<sup>96 97</sup>

We report the findings of our quality assessment of the 15 programmes organised around the seven features of ‘what good looks like’ below, with additional data in online supplemental material F.

#### Feature 1: quality of programme specification

Reporting of the basic characteristics of the programmes included in our quality assessment was reasonable (online supplemental material F: table 1). For example, it was possible to identify the clinical setting, target recipients, programme type and source of funding for all 15 programmes. Information on the scale of the programme was available for 14 of the 15 programmes, and on time period of implementation for 13 programmes.

Beyond these basic characteristics, completeness of programme description varied markedly between programmes and was generally of low quality (online supplemental material F: table 2). No programme described all 10 items of the modified TIDieR checklist in full. Five of the included programmes provided no description for half of the checklist items. Two programmes (Saving Babies’ Lives Care Bundle and Prevention of cerebral palsy in pre-term Labour

- PReCePT) provided full description of six checklist items, while three programmes (Maternity Incentive Scheme, Maternity Safety Training Fund and the Safer Births Project) did not describe any checklist item in full.

Completeness of description also varied between checklist items. Goals (described in full for 11 programmes) were generally well described, as were mechanisms of action and theories of change (described in full for eight programmes). In contrast, frequency, duration and time period of local implementation, local tailoring and modifications, processes for assessing or maintaining fidelity and outcome of fidelity assessment were described poorly. For example, no programmes offered a full description of frequency, duration and time period of local implementation or of local tailoring and modifications, and only three programmes offered any level of description for the latter item. Items relating to fidelity were particularly poorly described; only two programmes (Saving Babies’ Lives Care Bundle and PReCePT) offered any level of description of a fidelity assessment outcome. Only the Saving Babies’ Lives Care Bundle provided a full description of both the processes and outcome of fidelity assessment.

#### Feature 2: use of evidence-based interventions

Programmes varied in the extent to which they explicitly based their improvement interventions on

evidence. Of the 15 programmes we assessed, only eight reported that their interventions were developed with explicit reference to published evidence (online supplemental material F: tables 3 and 4). Of these, most (n=6) based their interventions on recommendations from national clinical guidance or quality standards, for example, those published by Royal Colleges and the National Institute for Health and Care Excellence. Two programmes (PRCePT and Perinatal Excellence to reduce injury in preterm birth - PERIPrem) cited a range of studies in support of their interventions, including systematic reviews and meta-analyses, randomised controlled trials and cohort studies.<sup>79 95 98 99</sup> The remaining seven programmes did not cite the evidence base for interventions used, if any.

### Feature 3: description of implementation support for services

Completeness of description of the implementation support available to clinical services (final column, online supplemental material F: table 2) was poor. Only three programmes (Obstetric anal sphincter injury Care Bundle - OASI-CB, PRCePT and PERIPrem) were judged to have described implementation support and activities in full. Whether or not implementation support had been provided was unclear for seven programmes, as no level of description was offered.

We did not identify any programme report that described customisation of implementation support for challenged services. The PRCePT programme did offer enhanced implementation support to a subset of participating services, though the authors did not report whether this was targeted specifically at challenged services.<sup>100</sup> Recruitment to the Labour Ward Leadership programme was partially informed by Care Quality Commission (CQC) inspection reports, with 'Trusts with identified problems' given priority.<sup>83</sup> Two programmes<sup>91 100 101</sup> accounted for variation in local service contexts in their evaluation methods by using, among other criteria, CQC inspection ratings to inform sampling of study participants. The interim evaluation of the Maternity Incentive Scheme<sup>102</sup> stated that additional support was provided to Trusts that were not in compliance with all 10 incentivised actions to enable them to 'achieve full compliance', but we did not find evidence that this support was targeted to support challenged services. In contrast, NHS trusts in 'special measures' and those in receipt of support from national regulators were specifically excluded from participating in the Each Baby Counts Learn & Support programme.<sup>103</sup> An incidental finding was the potential for programmes to exacerbate inequalities of resourcing between services, as noted by a PERIPrem report (parentheses added):

In Trusts where there was a pre-existing QI (quality improvement) culture and a desire to embed new practice and change, implementation was easier. Those

with active hospital QI teams were able to access additional support and training for the project.<sup>104</sup>

### Feature 4: commitment to reducing inequalities

Although we identified several evaluation reports that adjusted for socioeconomic status, ethnicity or levels of deprivation in their analysis,<sup>78 79 84 95 100 101 105</sup> we did not find any examples among the 15 programmes we assessed that identified the reduction of health or care inequalities as an explicit goal (online supplemental material F: table 4).

### Feature 5: patient and public involvement

Of the 15 programmes we assessed, only seven made explicit use of PPI practices. These included five that demonstrated comprehensive attention to involvement at all stages of the programme (including design, development, implementation, evaluation and dissemination). The remaining eight programmes did not mention of the role of patients or the public (online supplemental material F: table 4).

### Feature 6: use of formal published theories, models or frameworks

A minority (n=6) of assessed programmes made explicit use of formal published theories, models or frameworks from implementation science to guide programme implementation or evaluation (online supplemental material F: tables 4 and 5).<sup>106</sup> We identified 12 documented examples of use of theories, models or frameworks from four of Nilsen's proposed five categories,<sup>107</sup> including implementation theories (n=6), classic theories (n=2), determinant frameworks (n=2) and evaluation frameworks (n=2).

### Feature 7: programme evaluation

Programmes varied in their reporting of the planning, scope, design and conduct of evaluation (online supplemental material F: table 6). Evidence of a prospective evaluation plan (eg, in the form of a published protocol) was identified for seven of the 15 programmes, and four programmes conducted a pilot that informed programme development or implementation.

The evaluations we assessed were dominated by weak designs, often relying on post hoc methods of data collection, such as self-report questionnaires and staff surveys. Of the 19 peer-reviewed research articles identified (which collectively reported on nine programmes), five<sup>76-79 84</sup> reported evaluations of effectiveness, but only two studies reported on cost-effectiveness.<sup>78 79</sup> Only three evaluation reports<sup>76 84 100</sup> (including one preprint<sup>100</sup>) employed a randomised design. Three evaluations employed quasiexperimental approaches.<sup>78 79 92</sup> Only four articles reported findings from process evaluations, implementation research or qualitative studies.<sup>30 85-87</sup>

## DISCUSSION

Large-scale improvement programmes have been a key strategy for addressing quality deficits in healthcare globally, but the programmes assessed in our review of one exemplar clinical area often fell short on key features of quality. Though a large number of improvement initiatives have been undertaken in a 13-year period in maternity services, a particularly challenged clinical specialty in the English NHS, many—including a large number of major national programmes of the last decade—did not meet a basic requirement of a retrievable evaluation report. This represents a major threat to learning and accountability. Among programmes with a focus on intrapartum care that offered some form of evaluation report and could be assessed, there was considerable variability and very often evident flaws in transparency and quality of programme specification, use of evidence-based interventions, implementation support, PPI, use of formal published theories, models and frameworks, and evaluation. Notably, no programme that we quality assessed had explicitly set reduction of inequality as a goal. Our findings are unlikely to be unique to maternity settings or to the English NHS. These findings, and the methods used to generate them, are likely to be of relevance to many other clinical areas targeted by large-scale improvement programmes in healthcare settings internationally.

A first step in improving the quality of improvement programmes is to ensure that they are sufficiently well specified to permit identification of their components and the mechanisms through which they work, not least so that they can be scaled with fidelity if shown to be effective,<sup>10 14 108</sup> and modified or abandoned if not effective. Programmes included in our quality assessment demonstrated highly variable completeness of programme description, with some demonstrating weaknesses across several items of a modified TIDieR checklist.<sup>81</sup> We also identified an important lack of transparency in reporting relating to the implementation support available to participating services, despite its recognised role in effective improvement,<sup>25 29</sup> including in maternity care.<sup>30 100</sup> Crucially, despite recurrent reported organisational degradations in NHS maternity services, we were unable to identify any examples of adaptation of implementation specifically to account for the context of challenged services, with one programme actively excluding challenged units.<sup>103</sup> The extent to which programmes explicitly grounded their interventions in evidence or drew on theories, models or frameworks from implementation science<sup>43</sup> to guide implementation and evaluation was also highly variable, with many examples of poor reporting practice.

Second, a commitment to evaluation and public reporting of all findings should be seen as fundamental to high-quality commissioning of large-scale improvement programmes in healthcare. Many (14)

of the initiatives we identified—including some of the high-profile national maternity programmes of the last decade—lacked a retrievable evaluation report. In these cases, it is not possible to determine whether the programmes worked (made a difference to outcomes) and should be scaled, to assess whether these programmes represent a good use of resources, or to identify how programme design or implementation might have been improved. Even where evaluation reports were available, they often demonstrated substantial weaknesses. Future programme design and delivery should be organised to facilitate the use of rigorous evaluation designs that allow reliable assessments to be made about effectiveness and cost-effectiveness, as well as what works, what doesn't and why across diverse clinical settings.<sup>44</sup>

Third, despite repeated policy commitments to improve equity,<sup>32 34</sup> our findings add to concerns about inequalities in NHS maternity care.<sup>33 35 59</sup> We found no examples of improvement programmes included in our quality assessment that identified the reduction of health and care inequalities as an explicit goal. Indeed, there was some evidence of programme design having potential to contribute to widening inequalities between high-performing and low-performing services, which is likely to impede efforts aimed at improving equity.<sup>33</sup> Finally, despite the emphasis placed by national policy on including those who use maternity services in the design and delivery of improvement programmes,<sup>41 42</sup> PPI appeared to be lacking in over half of the programmes we assessed. Improving the impact of large-scale improvement efforts on the quality and safety of care in future will likely require these gaps between these enshrined policy objectives and programme design to be closed.

## Strengths and limitations

Our adaptation of the principles of scoping review methodology was successful in addressing our aims, given that most programmes we identified were not research projects and therefore unsuitable for full scoping or systematic review designs. The design of our search strategy was improved by specialist librarian input, by pilot testing the search strings to test their capacity to identify relevant literature, by selecting databases with relevant scope and by supplementing our structured searches with extensive hand searches. We sought to lend rigour to the charting process by requiring that assessment decisions be agreed by two researchers, by pilot testing the data charting tool to improve its reliability and by basing our assessment criteria, where possible, on published standards relating to programme specification<sup>81</sup> and evaluation.<sup>44</sup>

Our selection of the TIDieR checklist<sup>81</sup>—developed to improve published descriptions of healthcare interventions, including complex interventions—on which to base our assessments of reporting was appropriate



given the sociotechnical nature of improvement programmes. Our modified checklist is likely to be valuable in future studies to enable better description of features specific to improvement programmes. Relatedly, our definition of ‘improvement programme’ offers clarity to those engaged in programme commissioning, design and evaluation regarding specific aspects of these interventions that distinguish them from other types of improvement intervention in healthcare.

We acknowledge several limitations. It is possible that relevant sources of evidence relevant to eligible programmes were missed by our search strategy. The list of quality features we identified was necessarily selective and may not be comprehensive. It remains possible that researcher-related factors may have biased our assessments in some non-transparent way. Practical considerations meant our search was limited to programmes implemented in England since 2010, meaning that potentially important learning from programmes implemented elsewhere in the UK and prior to 2010 was excluded. Restriction of scope to intrapartum care may have excluded some programmes with different characteristics of quality and reporting.

## CONCLUSIONS

Transparent reporting and high-quality evaluation are critical to learning and accountability in healthcare systems facing persistent quality of care challenges. This review of large-scale maternity improvement programmes in the English NHS since 2010 has identified widespread poor practice in programme design, transparency of reporting and evaluation. These findings are both cause for concern and unlikely to be unique to this clinical setting. Our study suggests important targets for improving the design, delivery, evaluation and reporting of large-scale programmes in healthcare to maximise their impact on quality and safety, ensure accountability, including for how resources are used, and better aggregate learning to improve care for patients.

**Twitter** James McGowan @jgmcgowan, Isla Kuhn @ilk21, Lisa Hinton @LisaHinton4, Graham P Martin @graham\_p\_martin and Mary Dixon-Woods @MaryDixonWoods

**Contributors** MD-W is the guarantor of the study. Conceptualisation: JM, MD-W. Data curation: JM, BA. Formal analysis: JM, BA. Funding acquisition: JM, MD-W. Investigation: JM, BA, IK. Methodology: JM, BA, IK, LH, TD, GPM, MD-W. Project administration: JM, IK. Resources: JM, IK, MD-W. Software: JM, IK. Supervision: GPM, MD-W. Validation: LH, TD, GPM, MD-W. Visualisation: JM. Writing—original draft: JM, MDW. Writing—review and editing: all authors.

**Funding** The Health Foundation's grant to THIS Institute.

**Competing interests** TJD: Research and Innovation lead for PROMPT Maternity Foundation and has a part-time appointment at NHS Resolution where he leads on a Safety Action within the Maternity Incentive Scheme.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Please contact the first author.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

James McGowan <http://orcid.org/0000-0001-6963-4106>  
Bothaina Attal <http://orcid.org/0000-0002-3650-7415>  
Isla Kuhn <http://orcid.org/0000-0002-2879-4020>  
Lisa Hinton <http://orcid.org/0000-0002-6082-3151>  
Tim Draycott <http://orcid.org/0000-0002-1825-4864>  
Graham P Martin <http://orcid.org/0000-0003-1979-7577>  
Mary Dixon-Woods <http://orcid.org/0000-0002-5915-0041>

## REFERENCES

- Ovretveit J, Gustafson D. Using research to inform quality programmes. *BMJ* 2003;326:759–61.
- Benn J, Burnett S, Parand A, *et al*. Studying large-scale programmes to improve patient safety in whole care systems: challenges for research. *Social Science & Medicine* 2009;69:1767–76.
- Ovretveit J. Evaluation of quality improvement programmes. *Qual Saf Health Care* 2002;11:270–5.
- Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in Healthcare: lessons from the health foundation's programme evaluations and relevant literature. *BMJ Qual Saf* 2012;21:876–84.
- Peden CJ, Stephens T, Martin G, *et al*. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *Lancet* 2019;393:2213–21.
- Power M, Tyrrell PJ, Rudd AG, *et al*. Did a quality improvement collaborative make stroke care better? A cluster randomized trial. *Implement Sci* 2014;9:40.
- Bion J, Richardson A, Hibbert P, *et al*. Matching Michigan: a 2-year stepped Interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. *BMJ Qual Saf* 2013;22:110–23.
- Benning A, Ghaleb M, Suokas A, *et al*. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *BMJ* 2011;342:d195.
- Benning A, Dixon-Woods M, Nwulu U, *et al*. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;342:d199.

- 10 Lenguerrand E, Winter C, Siassakos D, *et al.* Effect of hands-on Interprofessional simulation training for local emergencies in Scotland: the THISTLE stepped-wedge design randomised controlled trial. *BMJ Qual Saf* 2020;29:122–34.
- 11 Dixon-Woods M. How to improve healthcare improvement—an essay by Mary Dixon-woods. *BMJ* 2019;367:l5514.
- 12 Brandrud AS, Nyen B, Hjortdahl P, *et al.* Domains associated with successful quality improvement in healthcare – a nationwide case study. *BMC Health Serv Res* 2017;17:648.
- 13 Vindrola-Padros C, Ledger J, Hill M, *et al.* The Special Measures for Quality and Challenged Provider Regimes in the English NHS: a rapid evaluation of a national improvement initiative for failing healthcare organisations. *Int J Health Policy Manag* 2022;11:2917–26.
- 14 Jones EL, Lees N, Martin G, *et al.* How well is quality improvement described in the perioperative care literature? A systematic review. *Jt Comm J Qual Patient Saf* 2016;42:196–206.
- 15 Jones EL, Dixon-Woods M, Martin GP. Why is reporting quality improvement so hard? A qualitative study in perioperative care. *BMJ Open* 2019;9:e030269.
- 16 Skivington K, Matthews L, Simpson SA, *et al.* A new framework for developing and evaluating complex interventions: update of medical research Council guidance. *BMJ* 2021;374:n2061.
- 17 Back J, Ross AJ, Duncan MD, *et al.* Emergency Department escalation in theory and practice: A mixed-methods study using a model of organizational resilience. *Ann Emerg Med* 2017;70:659–71.
- 18 Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)* 2005;24:138–50.
- 19 Hempel S, Bolshakova M, Turner BJ, *et al.* Evidence-based quality improvement: a scoping review of the literature. *J Gen Intern Med* 2022;37:4257–67.
- 20 Marshall M, Pronovost P, Dixon-Woods M. Promotion of improvement as a science. *Lancet* 2013;381:419–21.
- 21 Dixon-Woods M, Martin GP. Does quality improvement improve quality *Future Hosp J* 2016;3:191–4.
- 22 Carter P, Ozieranski P, McNicol S, *et al.* How collaborative are quality improvement collaboratives: a qualitative study in stroke care. *Implement Sci* 2014;9:32.
- 23 Dixon-Woods M, Leslie M, Tarrant C, *et al.* Explaining Matching Michigan: an ethnographic study of a patient safety program. *Implement Sci* 2013;8:70.
- 24 Stephens TJ, Peden CJ, Pearse RM, *et al.* Improving care at scale: process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). *Implementation Sci* 2018;13:142.
- 25 Greenhalgh T, Robert G, Macfarlane F, *et al.* Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;82:581–629.
- 26 Kaplan HC, Brady PW, Dritz MC, *et al.* The influence of context on quality improvement success in health care: a systematic review of the literature. *Milbank Q* 2010;88:500–59. 10.1111/j.1468-0009.2010.00611.x Available: <http://doi.wiley.com/10.1111/milq.2010.88.issue-4>
- 27 Dixon-Woods M. The problem of context in quality improvement London The Health Foundation; 2014.
- 28 Howarth E, Devers K, Moore G, *et al.* *Contextual issues and qualitative research, in Challenges, solutions and future directions in the evaluation of service innovations in health care and public health.* NIHR Journals Library, 2016.
- 29 Greenhalgh T, Papoutsi C. Spreading and scaling up innovation and improvement. *BMJ* 2019;365:l2068.
- 30 Redwood S, Pithara-McKeown C, Stone T, *et al.* Scaling up an intervention to protect Preterm infants from neurodevelopmental disabilities — findings from a qualitative process evaluation comparing standard with enhanced quality improvement support packages for maternity units in England. *Implement Sci* 2023;18:19.
- 31 Vindrola-Padros C, Ledger J, Barbosa EC, *et al.* “The implementation of improvement interventions for “low performing” and “high performing” organisations in health, education and local government: a phased literature review”. *Int J Health Policy Manag* 2022;11:874–82.
- 32 NHS England. *Equity and equality. Guidance for local maternity systems.* London: NHS England and NHS Improvement, 2021.
- 33 Care Quality Commission. Safety, equity and engagement in maternity services. 2021.
- 34 NHS England. NHS pledges to improve equity for mothers and babies and race equality for staff. 2021.
- 35 Jardine J, Walker K, Gurol-Urganci I, *et al.* Adverse pregnancy outcomes attributable to socioeconomic and ethnic inequalities in England: a national cohort study. *Lancet* 2021;398:1905–12.
- 36 Greenhalgh T, Jackson C, Shaw S, *et al.* Achieving research impact through co-creation in community-based health services. *Milbank Q* 2016;94:392–429.
- 37 Ostrom E, Parks RB, Whitaker GP, *et al.* The public service production process: A framework for analyzing police services. *Policy Studies Journal* 1978;7:381. 10.1111/j.1541-0072.1978.tb01782.x Available: <https://onlinelibrary.wiley.com/toc/15410072/7/s1>
- 38 Langley J, Wolstenholme D, Cooke J. Collective making’ as knowledge mobilisation: the contribution of participatory design in the Co-creation of knowledge in Healthcare. *BMC Health Serv Res* 2018;18:585.
- 39 Robert G, Locock L, Williams O, *et al.* *Co-producing and Co-designing.* Cambridge: Cambridge University Press, 2022.
- 40 Robert G, Donetto S, Williams O. Co-designing Healthcare services with patients. In: Loeffler E, Bovaird T, eds. *The Palgrave Handbook of Co-production of Public Services and Outcomes.* Cham, Switzerland: Palgrave Macmillan Cham, 2021: 313–333.
- 41 Greenhalgh T, Hinton L, Finlay T, *et al.* Frameworks for supporting patient and public involvement in research: systematic review and Co-design pilot. *Health Expect* 2019;22:785–801.
- 42 Mockford C, Staniszewska S, Griffiths F, *et al.* The impact of patient and public involvement on UK NHS health care: a systematic review. *Int J Qual Health Care* 2012;24:28–38.
- 43 Eccles MP, Mittman BS. Welcome to implementation science. *Implementation Sci* 2006;1.
- 44 HM Treasury. *Magenta Book: Central Government guidance on evaluation 2020.* London,
- 45 Portela MC, Pronovost PJ, Woodcock T, *et al.* How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* 2015;24:325–36.
- 46 Parry G, Coly A, Goldmann D, *et al.* Practical recommendations for the evaluation of improvement initiatives. *Int J Qual Health Care* 2018;30:29–36.

- 47 Parliamentary and Health Service Ombudsman. Spotlight on maternity care: your stories, your rights. Manchester; 2023.
- 48 House of Commons Health and Social Care Committee. *The safety of maternity services in England, Fourth Report of Session 2021–22*. 2021.
- 49 Yau CWH, Leigh B, Liberati E, *et al*. Clinical negligence costs: taking action to safeguard NHS Sustainability. *BMJ* 2020;368:m552.
- 50 Kirkup B. The report of the Morecambe Bay investigation. 2015.
- 51 Kirkup B. Reading the signals. maternity and neonatal services in East Kent – the report of the independent investigation. 2022.
- 52 Ockenden D. Findings, conclusions and essential actions from the independent review of maternity services at the Shrewbury and Telford hospital NHS trust. *Our Final Report* 2022.
- 53 NHS England. Independent Review of Maternity Services at Nottingham University Hospitals. . 2022 Available: <https://www.england.nhs.uk/midlands/publications/independent-review-of-maternity-services-at-nottingham-university-hospitals/>
- 54 Care Quality Commission. The state of health care and adult social care in England 2021/22. 2022.
- 55 Proudlove NC, Goff M, Walshe K, *et al*. The signal in the noise: robust detection of performance “outliers” in health services. *Journal of the Operational Research Society* 2019;70:1102–14.
- 56 Geary R, Knight H, Carroll F, *et al*. A step-wise approach to developing indicators to compare the performance of maternity units using hospital administrative data. *BJOG* 2018;125:857–65. 10.1111/1471-0528.15013 Available: <https://obgyn.onlinelibrary.wiley.com/toc/14710528/125/7>
- 57 Care Quality Commission. Maternity survey. 2023.
- 58 Dixon-Woods M. Learning from maternity service failures at East Kent hospitals. *BMJ* 2022;379:o2755.
- 59 Knight M, Bunch K, Patel R, *et al*. *Saving Lives, Improving Mothers’ Care Core Report - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2018-20*. National Perinatal Epidemiology Unit, University of Oxford Oxford, 2022.
- 60 Birthrights. Systemic racism, not Brokenbodies. an inquiry into racial injustice and human rights in UK maternity care; 2022.
- 61 Gohir S. INVISIBLE, maternity experiences of Muslim women from Racialised minority communities, A summary report. 2022.
- 62 Peter M, Wheeler R. The Black Maternity Experiences Survey. a nationwide study of black women’s experiences of maternity services in the United Kingdom. 2022.
- 63 House of Commons Women and Equalities Committee. *Black maternal health, Third Report of Session 2022–23*. 2023.
- 64 Department of Health and Social Care. *Safer Maternity Care, Progress Report 2021*. London, 2021.
- 65 Department of Health, Safer Maternity Care. *The National Maternity Safety Strategy - Progress and Next Steps*. London, 2017.
- 66 NHS Litigation Authority. *Ten Years of Maternity Claims: An Analysis of NHS Litigation Authority Data*. London: NHS Litigation Authority, 2018.
- 67 NHS Resolution. The second report: the evolution of the Early Notification Scheme. 2022.
- 68 NHS Resolution. *Five years of Cerebral Palsy claims, A thematic review of NHS Resolution data*. London, 2017.
- 69 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for Scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
- 70 Arksey H, O’Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005;8:19–32.
- 71 Knight M, Felker BK, Patel A, *et al*. *Saving Lives, Improving Mothers’ Care Core Report - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2019-21*. Oxford: National Perinatal Epidemiology Unit, University of Oxford, 2023.
- 72 Mohr JJ, Batalden PB. Improving safety on the front lines: the role of clinical microsystems. *Qual Saf Health Care* 2002;11:45–50.
- 73 Peters M, Godfrey C, McInerney P, *et al*. Chapter 11: Scoping reviews section 11.2.5 search Strategy(2020 version). In: Aromataris E, Munn Z, eds. *JBIManual for Evidence Synthesis*. JBI, 2020.
- 74 Ayiku L, Levay P, Hudson T, *et al*. The MEDLINE UK filter: development and validation of a geographic search filter to retrieve research about the UK from OVID MEDLINE. *Health Info Libr J* 2017;34:200–16.
- 75 Zwakman M, Verberne LM, Kars MC, *et al*. Introducing PALETTE: an Iterative method for conducting a literature search for a review in palliative care. *BMC Palliat Care* 2018;17:82.
- 76 Ismail KMK, Kettle C, Macdonald SE, *et al*. Perineal assessment and repair longitudinal study (PEARLS): a matched-pair cluster randomized trial. *BMC Med* 2013;11:209.
- 77 Rahman N, Vinayakarao L, Pathak S, *et al*. Evaluation of training programme uptake in an attempt to reduce obstetric anal sphincter injuries: the SUPPORT programme. *Int Urogynecol J* 2017;28:403–7.
- 78 Widdows K, Roberts SA, Camacho EM, *et al*. Stillbirth rates, service outcomes and costs of implementing NHS England’s Saving Babies’ Lives care bundle in maternity units in England: a cohort study. *PLoS ONE* 2021;16:e0250150.
- 79 Edwards HB, Redaniel MT, Sillero-Rejon C, *et al*. National PRCePT Programme: a before-and-after evaluation of the implementation of a national quality improvement programme to increase the uptake of magnesium sulfate in preterm deliveries. *Arch Dis Child Fetal Neonatal Ed* 2023;108:342–7.
- 80 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan—a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
- 81 Hoffmann TC, Glasziou PP, Boutron I, *et al*. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:bmj.g1687.
- 82 Pope C, Mays N. *Qualitative Research in Healthcare, 4th edn*. 2020: Wiley Blackwell,
- 83 NHS England (Avoiding Term Admissions into Neonatal units programme), Royal College of Midwives, Royal College of Obstetricians and Gynaecologists. Labour ward leaders working together for safe care. pilot programmes: evaluation report; 2017.
- 84 Gurol-Urganci I, Bidwell P, Sevdalis N, *et al*. Impact of a quality improvement project to reduce the rate of

- obstetric anal sphincter injury: a multicentre study with a stepped-wedge design. *BJOG: Int J Obstet Gynaecol* 2020;128:584–92.
- 85 Lau YZ, Widdows K, Roberts SA, *et al.* Assessment of the quality, content and perceived utility of local maternity guidelines in hospitals in England implementing the Saving Babies' Lives care bundle to reduce stillbirth. *BMJ Open Qual* 2020;9:e000756.
- 86 Bidwell P, Sevdalis N, Silvertown L, *et al.* Women's experiences of the OASI care bundle; a package of care to reduce severe perineal trauma. *Int Urogynecol J* 2021;32:1807–16.
- 87 Bidwell P, Thakar R, Gurol-Urganci I, *et al.* "Exploring clinicians' perspectives on the 'Obstetric Anal Sphincter Injury Care Bundle' national quality improvement programme: a qualitative study". *BMJ Open* 2020;10:e035674.
- 88 Bick DE, Kettle C, Macdonald S, *et al.* Perineal assessment and repair longitudinal study (PEARLS): protocol for a matched pair cluster trial. *BMC Pregnancy Childbirth* 2010;10:10.
- 89 Bidwell P, Thakar R, Sevdalis N, *et al.* A multi-centre quality improvement project to reduce the incidence of obstetric anal sphincter injury (OASI): study protocol. *BMC Pregnancy Childbirth* 2018;18:331.
- 90 Widdows K, Reid HE, Roberts SA, *et al.* Saving Babies' Lives project impact and results evaluation (SPiRE): a mixed methodology study. *BMC Pregnancy Childbirth* 2018;18:43.
- 91 Edwards H, Redaniel MT, Opmeer B, *et al.* Evaluating an enhanced quality improvement intervention in maternity units: PRECePT trial protocol. *BMJ Open Qual* 2021;10:e001204.
- 92 O'Brien S, Lenguerrand E, Jordan S, *et al.* Simulation training for operative vaginal birth evaluation: study protocol for an observational stepped-wedge interrupted time-series study (STROBE). *BMC Pregnancy Childbirth* 2019;19:109.
- 93 Burhouse A, Lea C, Ray S, *et al.* Preventing cerebral palsy in preterm labour: a multiorganisational quality improvement approach to the adoption and spread of magnesium sulphate for neuroprotection. *BMJ Open Qual* 2017;6:e000189.
- 94 Edwards K, Impey L. Extreme preterm birth in the right place: a quality improvement project. *Arch Dis Child Fetal Neonatal Ed* 2020;105:445–8.
- 95 Glover Williams A, Tuvey S, McBain H, *et al.* Perinatal excellence to reduce injury in preterm birth (PERIPrem) through quality improvement. *BMJ Open Qual* 2022;11:e001904:11. .
- 96 Robertson L, Knight H, Prosser-Snelling E, *et al.* Each Baby Counts in 2018 – lessons learned and future directions. *Obstetrics, Gynaecology & Reproductive Medicine* 2018;28:253–7.
- 97 Jurczuk M, Bidwell P, Gurol-Urganci I, *et al.* The OASI care bundle quality improvement project: lessons learned and future direction. *Int Urogynecol J* 2021;32:1989–95.
- 98 West of England Academic Health Science Network (AHSN). PRECePT – preventing cerebral palsy in preterm babies, Available: <https://www.weahsn.net/our-work/transforming-services-and-systems/precept>
- 99 West of England Academic Health Science Network (AHSN). PERIPrem. n.d. Available: <https://www.weahsn.net/our-work/transforming-services-and-systems/periprem>
- 100 Edwards HB, Redaniel MT, Sillero-Rejon C. Evaluation of standard and enhanced quality improvement methods to increase the uptake of magnesium sulfate in preterm deliveries for the prevention of neurodisability (PRECePT study): a cluster randomized controlled trial. *Obstet Gynecol* [Preprint].
- 101 Grimwood T, Snell L. *The Maternity Safety Training Fund: An Evaluation*. University of Cumbria, 2018.
- 102 NHS Resolution. Maternity Incentive Scheme – an interim evaluation. 2020.
- 103 Royal College of Obstetricians and Gynaecologists. *Each Baby Counts + Learn & Support. Final Evaluation Report*. London, 2022.
- 104 McBain H, Tuvey S, Williams AG, *et al.* Barriers and Enablers to the implementation of a perinatal care bundle: the perinatal excellence to reduce injury in premature birth (PERIPrem) project. [Preprint] 2022.
- 105 Widdows K, Roberts S, Camacho E, *et al.* *Evaluation of the implementation of the Saving Babies' Lives Care Bundle in early adopter NHS Trusts in England*. Manchester, UK: Maternal and Fetal Health Research Centre, University of Manchester, 2018.
- 106 Dadich A, Piper A, Coates D. Implementation science in maternity care: a Scoping review. *Implement Sci* 2021;16:16.
- 107 Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci* 2015;10:53.
- 108 Renwick S, Hooke S, Draycott T, *et al.* PROMPT Wales project: national scaling of an evidence-based intervention to improve safety and training in maternity. *BMJ Open Qual* 2021;10:e001280.
- 109 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71:71:.