UCL DEPARTMENT OF APPLIED HEALTH
RESEARCH

Provision of Data Analysis and Economic Evaluation of the Incentivisation Scheme in support of Sign up to Safety on behalf of the NHS Litigation Authority

Final Report

Submitted to

NHS Litigation Authority (NHS LA)

Submitted by

UCL Consultants Limited (UCLC)

On behalf of:

Dr Elena Pizzo, Senior Health Economist

Prof Stephen Morris, Professor of Health Economics

Prof Tim Draycott, Consultant Obstetrician

Rachael Hunter, Senior Health Economist

Dr Laura Vallejo Torres, Senior Health Economist

Dr Caroline Clarke, Health Economist

Estela Capelas Barbosa, Health Economist

Dr Christopher Wai Hung Yau, Clinical Research Fellow

Corresponding author:
Dr Elena Pizzo, Senior Health economist
Email: e.pizzo@ucl.ac.uk, Tel: 020 3108 3250

17 January 2017

Contents

Executive Summary

This document constitutes the final report for the project "Provision of Data Analysis and Economic Evaluation of the Incentivisation Scheme (IS) in support of Sign up to Safety (SU2S) on behalf of the NHS Litigation Authority". The NHS Litigation Authority (NHS LA) is supporting Sign up to Safety, a campaign launched in June 2014 aiming to improve and strengthen the safety of patients in the NHS. As part of the scheme over £18 million has been awarded to Trusts across the country as financial incentives to implement interventions aimed at reducing harms leading to clinical claims.
In January 2015, the NHS LA received 249 bids from 114 member Trusts to the Sign up to Safety incentive funding scheme across a wide geographical reach. Sixty-seven bids were approved from a range of safety improvement areas, of which around £8.6 million was allocated to maternity units and £0.8 million to A&E units. The payment of the awarded funds was made between May and September 2015. Despite there being no official end date for the Sign up to Safety scheme, the initiative aimed to see impacts within 3 years.

The NHS LA subsequently commissioned a study to evaluate the cost-effectiveness of the incentivisation scheme and to explore the feasibility of developing a dataset to improve safety and avoid claims.

An economic evaluation was conducted to assess the impact of the incentivisation scheme and the funded interventions on incidents, health and claims. Impacts on health and on claims may not materialise until sometime after the incidents occur, perhaps several years afterwards. Therefore it was difficult to measure the difference in benefits and costs directly. For this reason we adopted a modelling approach, whereby longer-term costs and benefits were modelled from short-term impacts on incidents.

We also conducted a qualitative analysis to understand what Trusts thought of the scheme. Overall Trusts had a positive perception of the scheme. The main benefit they reported was the provision of financial resources for reducing harms, which given the current economic climate were viewed as extremely important; without the resources from the scheme the interventions would not have been implemented. One Trust reported: *"[…] being able to demonstrate quick wins and timely improvements/evaluations when presenting business cases can be challenging when there are so many demands on resources from all disciplines and whilst it is anticipated the improvements will achieve the desired effect there is no immediate reassurance that can demonstrate its success".*
Another perceived benefit was that the scheme was supported by the Department of Health, which helped Trust staff understand the investment and commitment to improvements and "*the fact that the staff can see, touch, benefit from the investment in the additional resources provides a direct link from them to the Department of Health. […] it is really valuable for the staff delivering the care to have that recognition and understanding"*.

Trusts also reported that the scheme raised awareness of errors, reporting and claims, improved the morale of teams, and had initiated a culture change. For example, one Trust reported: "*For organisations and NHS staff to truly learn from patient harm and medical errors it is important to create a safe environment which is less performance based and more values based, to allow staff and patients to work more collectively towards creating an open learning service and provide better care and treatment as a result".*

The main limitation of the scheme was seen by Trusts to be the short-term and non-recurrent nature of the funding and that without future funding impacts were unlikely to be seen in the long term. Trusts also recognised that their IT systems could be improved, to assist with routine collection of data and evaluate interventions in a more appropriate and timely way.

The main aim of the evaluation was to undertake 3 pieces of work:

(1) *Collecting robust 'cause of harm' data*
This aim was to explore the feasibility of developing a dataset for a range of clinical areas, including maternity and A&E, that could be used by a range of stakeholders to improve safety and help avoid claims.

(2) *Reducing missed fractures in A&E*
Funds were given to a number of Trusts to fund "out of hours" senior radiographers and other interventions. The aim was to evaluate the impact of the incentivisation scheme on reducing the number of missed fractures and subsequent claims for missed fractures.

(3) *Reducing intrapartum harm*
Funds were provided for 28 maternity units to implement a range of self-selected interventions including: additional cardiotocography (CTG) machines, additional supervisory staff and staff training. The aim was to evaluate the impact of the incentivisation scheme on reducing the number of adverse outcomes, including stillbirth, hypoxia, and brain injury at birth and downstream, the number of litigation claims for maternity services.

Here we summarise what we did and the results of the 3 pieces of work.

(1) *Collecting robust 'cause of harm' data*
In the first part of the study we explored the feasibility of developing a potential learning dataset. The main aim of the dataset would be to assemble data at Trust level on incidents and claims for a range of clinical areas that could be used by a range of stakeholders to learn from and help avoid claims. There are several datasets already available and we reviewed the following: the National Reporting and Learning System (NRLS), the Perinatal and maternity patient management data from BadgerNet, the National Neonatal Research Database (NNRD), the National Neonatal Audit Programme (NNAP), the risk management software, the NHS maternity statistics and the Hospital Episodes Statistics Admitted Patient Dataset, the Friends and Family Test Data for maternity services, MBRRACE-UK, the Maternity dashboard, the Each Baby Counts national improvement programme and data from the "Getting it Right First Time" project on orthopaedic errors.

We explored the availability of a broad range of datasets at national and local levels but most of these sources were not sufficient to provide complete data. To be useful, a database should collect data on the number and type of incidents and their association with health and claims, but should also provide extra information useful in reducing incidents and claims such as adverse events and near miss reporting, risk assessment, safety alerts, patient experience and feedback, complaints and concerns (e.g., as in risk management software tools and local incident reporting systems, MBRRACE-UK or the maternity dashboard).

Given the number and type of datasets currently available, the benefits of collecting new data should be clearly identified before embarking on new data collection activities. Any new data that are collected should be related to claims, and should be collected in a way that can minimise or avoid under-reporting. Finally, costs should be taken into account, not only to set up and maintain a database, but also in terms of resources and time spent to collect data at Trust level, extract them, clean them and analyse them. There appears to be a considerable amount of data available, but little effort to co-ordinate databases or improve their reliability. Therefore we recommend that more investment should be made to improve existing datasets (e.g., in terms of data linkage between them and improving reporting) rather than creating new ones.

(2) *Reducing missed fractures in A&E*

To evaluate the impact of the incentivisation scheme on reducing harm and claims in A&E units, we collected data at Trust level (from both Trusts that received funding from the scheme and those that did not)[1] on number of X-rays, number of missed fractures in A&E and claims due to missed fractures. However, the data were not consistent or complete mainly due to the difficulty in measuring missed fractures.

As it was not possible to quantify missed fractures from participating Trusts, we used evidence from published studies to assess the effectiveness and cost-effectiveness of the interventions implemented by the Trusts as part of the SU2S incentivisation scheme.

A decision model was built to assess the cost-effectiveness of 2 main interventions adopted in successful Trusts: continuous quality improvement (CQI) strategies; and, hot reporting of radiology imaging. Costs were measured in terms of the funding received by Trusts and the impact of the interventions on treatment costs and litigation costs. Outcomes were measured in terms of quality adjusted life years (QALYs) using published evidence. The results show that both interventions reduce overall costs and improve the health outcomes of patients. Annual cost savings across the 4 successful Trusts were estimated to be £250,000 and £678,000 with CQI and hot reporting interventions, respectively. These values are equivalent to a mean cost saving per patient of £2.50 and £6.70.

(3) *Reducing intrapartum harm*

To evaluate the impact of the Sign up to Safety (SU2S) financial incentive scheme in reducing intrapartum harm we collected data at Trust level (from both Trusts that received funding from the scheme and those that did not) on a broad range of clinical and process measures (stillbirths, low 5 minute Apgar score <7, therapeutic hypothermia (cooling), unexpected neonatal intensive care unit (NICU) admissions, 3rd and 4th degree tears, rates of instrumental delivery and Caesarean sections) that could be used to assess the impact of the interventions in reducing harm and future claims. We used these to undertake a cost-consequences analysis of the SU2S financial incentivisation scheme in maternity units.

A difference-in-differences (DiD) analysis was used to see how these measures changed before and after the introduction of the scheme in Trusts who received funding compared with the changes over the same time period seen in Trusts that were not part of the scheme.

From the quantitative analysis we found no evidence for any improvements in clinical outcomes in either the Trusts who received funding or those who did not. In Trusts who received funding, there was no statistically significant difference in any of the outcomes measured, except for a significant increase in the number of reported tears, which is likely to be related to increased identification with the scheme. These findings were confirmed using a series of sensitivity analyses.

Given these results, it was not possible to calculate cost savings associated with the interventions, reduced claims, or improvements in health-related quality of life and mortality.

Further analysis at a later date may be beneficial, after a longer period has elapsed since Trusts implemented the interventions. It is difficult to specify what this period might be. We speculate that interventions might take a year to implement fully from the start date and then at least a year to

[1] In the report we use the term “successful” and “unsuccessful” Trust to refer only to whether or not the Trust was successful in being awarded funding as part of the scheme (e.g., we do not refer to the ability of the Trust in improving outcome measures).

have an effect on incidents. For maternity care, in our data, the average time gap between the date of incident and the date a claim is made was 3.5 years, and so if future research was to measure the impact on claims directly in maternity services a substantial period from the date at which interventions were implemented needs to have elapsed. This time period might vary by the type of intervention and the clinical specialty. Further research on the appropriate time period for measurement would be useful.

In analysing data from the NHS LA Claims Database we calculated the minimum number of claims that would need to be avoided for the SU2S financial incentivisation scheme to be considered good value for money. Using the average cost for a successful cerebral palsy (CP) claim to the NHS LA of £4,745,295, and that in total just over £8 million pounds were invested in maternity interventions for the scheme, then it would be cost saving if there were two fewer CP claims across the 28 Trusts who received funding over the duration of the project. Based on published data, in order to achieve this it would be necessary to avoid: 18 cases of babies cooled, or alternatively 70 babies born with an Apgar score <7[5] minutes or 24 babies born with an Apgar score <4[5] minutes across these Trusts. This is likely to be possible as the mean Apgar score <7[5] minutes was 1.3% and there are reported rates in the UK of <0.5% that have been sustained for more than a decade[2].

An indirect benefit of the scheme reported by NHS LA was that the NHS LA partnered with NHS Supply Chain to assist the maternity units in collectively procuring their equipment, and a saving of £36k was achieved from sales of £227k.

We identified several reasons why the scheme might not have had an impact on reducing intrapartum harm:

- **Choice of interventions.** Taking the results of the statistical analysis at face value, the interventions implemented by Trusts within the evaluation timeframe, may not have been effective in reducing harms. The maternity units invested their funding in a broad range of interventions, and evidence of effectiveness for the interventions they implemented is limited. Also, the Trusts all implemented different interventions, so the overall result could reflect a mixture of effective (although not demonstrably so due to the short timeframe) and ineffective interventions.
- **Implementation problems.** At the time of the analysis (conducted 12-14 months after the Trusts had received funding from the scheme) not all Trusts had implemented all of the interventions they had originally proposed.
- **Short duration of follow up.** There is likely to be a delay before the benefits of the interventions are seen, so even if interventions are fully implemented then it may take longer than the time horizon of our evaluation (one year) to see a tangible benefit. In at least one unit level intervention with positive outcomes, a year was required to train all of the unit, and outcomes changed in the following year.
- **Data quality.** The results of our analysis may be limited by the quality of data provided. Maternity datasets are recognised to be some of the most accurate in the NHS, however many Trusts found it difficult to provide data from their IT department for this evaluation, and some used paper records to source data for us, which might have negatively impacted on the accuracy, reliability and completeness of the data. In addition, the small sample size (we only had aggregate data for 44 Trusts, and this was not always complete) and the lack of patient level data may also have affected the results.

[2] Draycott T, Sibanda T, Owen L, Akande V, Winter C, Reding S et al. (2006) Does training in obstetric emergencies improve neonatal outcome? BJOG: An International Journal of Obstetrics &Gynaecology. 113(2):177-82.

Summary and recommendations

The evaluation has provided two different results, both of which are limited by the availability of data. On one hand providing funding to reduce missed fractures in A&E has shown a reduction of the overall costs (including litigation costs) and an improvement in patients' health and health care, but it was not based on actual data from participating Trusts. On the other hand providing funding in maternity units did use data from participating Trusts but did not demonstrate any statistically significant impacts of the interventions on clinical outcomes.

On the basis of the evidence generated for this report it appears that the interventions regarding missed fractures in A&E may be cost-effective, but those to reduce intrapartum harms in maternity units may not be or it may be too early to tell; however, it is difficult to draw conclusions given the available data in both areas.

Recommendations

1. Given the timelines involved, further evaluation of the scheme in the future may be beneficial. This is likely to vary by the type of intervention and the clinical speciality, but is likely to be several years.

2. A balance needs to be made between making top-down recommendations about the interventions Trusts ought to implement, and giving Trusts autonomy to which interventions to implement in response to local needs. However, interventions in this and similar schemes should be based on good evidence of effectiveness and cost-effectiveness. We concur with the 2016 NHS England National Maternity review recommendation: *'Most importantly, any training undertaken must have been proven to be effective in improving outcomes or other aspects of quality, and its impact monitored locally'*.

3. Future schemes could include establishing evidence on the main causes of errors, and signposting effective interventions with support for local measurement and regional benchmarking of clinical outcomes (A&E, maternal and neonatal outcomes), process measures (interventions), and implementation (e.g., proportion of staff trained).

4. Recognition needs to be given to the timescale between allocation of funding and implementation time (e.g., implementation of intervention can be delayed by procurement time). Efforts should be made to ensure that Trusts awarded funding from this and similar schemes can act on that funding in a timely manner, so the interventions can be of maximum benefit to patients as soon as possible.

5. At present there is a considerable amount of data available on errors and claims but more effort is required to coordinate these data, improve their reliability and reduce under-reporting.

6. To assist with reporting, investment in providing integrated, flexible, efficient and user-friendly IT systems is needed to bring all Trusts up to a minimum standard. This will allow data to be collected in a timely, accurate, complete and reliable way.

7. The NHS LA could usefully partner with recognised academic, Improvement Science and clinical groups to improve the selection, implementation and evaluation of improvement initiatives in the future.

Acknowledgements

This is report was commissioned and funded by the NHS Litigation Authority (NHS LA). The executive summary has been reviewed and revised by the NHS LA.

During the study we have benefited from the advice and suggestions provided by the following people (in alphabetical order):

Catherine Anderson, NHS LA
Michelle Anstiss, NHS LA
Jeanette Beer, NHS LA
Denise Chaffer, NHS LA
Richard Colquhoun, Neonatal Data Analysis Unit, Imperial College London
Richard Evans, The Society and College of Radiographers
Alan Fenton, Newcastle Neonatal Service, Royal Victoria Infirmary
Chris Gale, Neonatal Data Analysis Unit, Imperial College London
Pelle Gustafson, Landstingens Ömsesidiga Försäkringsbolag, Sweden
David Gurusinghe, NHS LA
John Machin, NHS England
Michael Magro, NHS LA
Sally Masson, NHS LA
Alison McLellan, NHS LA
Catherine O'Sullivan, NHS LA
Donna Poole, NHS LA
Linda Prescott-Clements, National Clinical Assessment Service, NCAS
Eugene Statnikov, Neonatal Data Analysis Unit, Imperial College London
Michele Upton, NHS England
Suzette Woodward, Sign up to Safety

We would like to thank all the Trusts who were funded by the SU2S scheme for providing us with all the necessary data to run the analysis and those not funded that kindly agreed to be part of the study as comparator sites (in alphabetical order):

- Alder Hey Children's NHS Foundation Trust
- Barts Health NHS Trust
- Birmingham Women's NHS Foundation Trust
- Blackpool teaching hospital NHS Foundation Trust
- Burton Hospitals NHS Foundation Trust
- Calderdale and Huddersfield NHS Foundation Trust
- Central Manchester University Hospitals NHS Foundation Trust
- City Hospital Sunderland NHS Foundation Trust
- Colchester Hospital University NHS Foundation Trust
- Countess of Chester Hospital NHS FT Trust
- Doncaster & Bassetlaw Hospitals NHS Foundation Trust
- East Kent University Hospitals NHS FT Trust
- East Lancashire Hospitals NHS Trust
- East Sussex Healthcare NHS Trust
- Epsom and St Helier University Hospitals NHS Trusts
- Frimley Health NHS Foundation Trust
- Gateshead Health NHS Foundation Trust
- Gloucestershire Hospitals NHS Foundation Trust
- Guys and St Thomas NHS Foundation Trust

- Hampshire Hospitals NHS Foundation Trust
- Harrogate & District NHS Foundation Trust
- Ipswich Hospital NHS Trust
- Isle of Wight NHS Trust
- Kingston Hospital NHS Foundation Trust
- Leeds Teaching Hospitals NHS Trust
- Lewisham & Greenwich NHS Trust
- Liverpool Women's Hospital NHS Foundation Trust
- Milton Keynes University Hospital NHS Trust
- Newcastle Upon Tyne Hospitals NHS Foundation Trust
- Norfolk & Norwich University Hospitals NHS Foundation Trust
- North Middlesex University Hospital NHS Trust
- North Tees & Hartlepool NHS FT Trust
- Northern Lincolnshire & Goole Hospitals NHS FT Trust
- Pennine Acute Hospitals NHS Trust
- Queen Elizabeth Hospital King's Lynn NHS Foundation Trust
- Royal Berkshire NHS Foundation Trust
- Royal Cornwall Hospitals NHS Trust
- Royal Surrey County Hospital NHS Foundation Trust
- Royal Wolverhampton NHS Trust
- Salisbury NHS Foundation Trust
- Shrewsbury and Telford Hospital NHS Trust
- South Tees Hospitals NHS Foundation Trust
- Stockport NHS Foundation Trust
- Taunton and Somerset NHS Foundation Trust
- University Hospitals Southampton NHS Foundation Trust
- University Hospitals Morecambe Bay NHS Foundation Trust
- University Hospitals of Leicester NHS Trust
- West Hertfordshire Hospitals NHS Trust
- West Suffolk NHS Foundation Trust
- York Teaching Hospitals NHS Foundation Trust

We are very grateful for the considerable amount of work put in by the staff in all the Trusts to provide us with the information for this report.

We would also like to thank the Neonatal Data Analysis Unit (NDAU), Imperial College London, for providing us with the data used in the study.

The views expressed in this report are not necessarily those of the above individuals or the participating Trusts.

1. Background and aims

The NHS Litigation Authority (NHS LA) has supported Sign up to Safety (SU2S)[3], a campaign launched in June 2014 by the Secretary of State for Health aiming to improve and strengthen the safety of patient care in the NHS. The initiative aimed to save 6,000 lives, halve avoidable harm and halve the costs of harm in the NHS over the following three years. All health organisations have been asked to develop safety improvement plans to reduce harm and save lives by working to reduce the causes of harm and take a preventative approach.

The NHS LA provided support to the campaign offering financial incentives to those organisations that provided evidence that their Safety Improvement Plans would reduce harm and claims. A description of the SU2S incentivisation process is available in Appendix 1. In January 2015, the NHS LA received 249 bids from 114 member Trusts to the Sign up to Safety incentive funding scheme across a wide geographical reach. Sixty-seven bids were approved from a range of safety improvement areas; the top five focus areas were:

1. Maternity - purchase of cardiotocography (CTG) electronic monitors, STAN (ST analysis - fetal ECG monitors), recruitment for "second pair of eyes", central CTG monitoring and alert systems, remote access to tracings, training.
2. Safety Culture - a range of human factors and teamwork interventions.
3. Surgical - includes training and equipment, human factors training in a number of surgical specialties particularly orthopaedics and neurological surgery.
4. A&E - improving missed and delayed diagnosis, especially fractures, diagnostics, "hot" radiography reporting in 24 hours, performance feedback on missed diagnoses.
5. Deteriorating patient - early and improved recognition, electronic flags, improved management pathways.

The 67 bids received total funding of £18.7 million, of which around £8.6 million was allocated to maternity and £0.8 million to A&E.

The payment of the awarded funds was made in 2015, but at different points in time: 32 Trusts were given funding in May, 8 Trusts in June, 2 in July, 4 in August and 2 in September. Only Foundation Trusts were allowed to carry forward funding allocated, whereas others were required to implement plans of their bid within 2015. A detailed description of all the interventions is provided in the relevant chapters below. Despite there being no official end date for the Sign up to Safety scheme, the initiative aimed to see impacts within 3 years. Trusts were required to provide reports at six months and one year to monitor their progress.

The NHS LA subsequently commissioned a study to evaluate the cost-effectiveness of the incentivisation scheme and to explore the feasibility of developing a dataset to improve safety and avoid claims.

The main aim of the evaluation was to undertake 3 pieces of work:

(1) *Collecting robust 'cause of harm' data*
This aim was to explore the feasibility of developing a dataset for a range of clinical areas, including maternity and A&E, that could be used by a range of stakeholders to improve safety and help avoid claims.

[3] http://www.england.nhs.uk/signuptosafety/

(2)*Reducing missed fractures in A&E*
Funds were given to a number of Trusts to fund 'out of hours' senior radiographers and other interventions. The aim was to evaluate the impact of the incentivisation scheme on reducing the number of missed fractures and subsequent claims for missed fractures.

(3) *Reducing intrapartum harm*
Funds were provided for 28 maternity units to implement a range of self-selected interventions including: additional cardiotocography (CTG) machines, additional supervisory staff and staff training. The aim was to evaluate the impact of the incentivisation scheme on reducing the number of adverse outcomes, including stillbirth, hypoxia, and brain injury at birth and downstream, the number of litigation claims for maternity services.

We also conducted a qualitative analysis to understand what Trusts thought of the scheme.

The report is structured in seven sections: we first report the results of the qualitative analysis performed to capture the views of the Trusts about the scheme. In section 3 we provide an overview of the dataset already available and the feasibility of developing a new learning dataset to avoid harm. In this section we will highlight some of the problems encountered in collecting the data for the main analysis. After describing the overall methodology of the economic evaluation in section 4, we then present the results of the evaluation in A&E and maternity units in sections 5 and 6 respectively. We conclude the report in section 7 by discussing the results of the study and providing some recommendations.

2. What do Trusts think of the Sign up to Safety Bid Incentivisation Scheme?

Before exploring the cost-effectiveness of the financial incentive scheme, it is useful to understand the impressions of the Trusts regarding the scheme.

To this end, on 1st December 2016 we contacted by email both Trusts that received funding and those who applied for funding but were unsuccessful and asked them to complete and return a questionnaire aimed at capturing their impressions of the scheme (Appendix 2). The questionnaires were directed to the contact person referred by NHS Litigation Authority, most of whom occupied a senior management position, for example Director of Quality, Head of Corporate Risk or Governance Lead.

We asked the Trusts to tell us what they thought were the main positive aspects of the scheme, what were the main limitations and what could be done differently. A reminder was sent in the following weeks. Overall 15 of the 55 Trusts contacted responded (27%) by 20th December and two of these were unsuccessful Trusts.

Here we summarise what, according to these Trusts, are the positive aspects of the scheme, the limits and challenges, what could be changed or improved upon and some key messages to the Department of Health. A full list of comments is reported in Appendix 3.

Positive aspects of the Sign up to Safety Incentivisation Scheme
Overall the Trusts provided very constructive comments about the SU2S financial incentives scheme. When asked, Trusts identified five main positive aspects of the scheme:

- Availability of resources and funding: the scheme financially supported the organisations in taking forward some safety projects that would have otherwise not been possible without the funding. In the current economic climate Trusts find it extremely difficult to obtain

additional resources, therefore the scheme was essential to implement intervention that otherwise would have been difficult to achieve.

- Improved safety and outcomes: the scheme has allowed Trusts to get organised around safety and to focus on reducing avoidable harm. As a large majority of incidents and claims are due to human factors, being given the opportunity to understand how and why medical errors occur, and tackle these issues has been valued as extremely important.
- Support from NHS LA and communication: the Trusts appreciated the support received from NHS LA and the communication provided via webinars and weekly emails. Being part of the SU2S scheme has been extremely important for Trusts to increase their awareness of safety. Being connected with other organisations has allowed networking and exchange of ideas.
- Learning and sharing with other organisations: the scheme has allowed organisations to learn from each other, share good practice and improvements nationally.
- Impact on staff: the scheme has increased awareness among staff, increasing their confidence through training, but also promoting multidisciplinary working, bringing together staff at all levels across the NHS all facing similar challenges. During the NHS LA workstream events, staff have been given the chance to speak up and be listened to. The SU2S financial incentive scheme might have made a bigger impact than the one that can be quantified, in terms of team work, enthusiasm and change in culture. The scheme has sent a strong message to staff that the Department of Health recognises and cares about the work they do and that, despite the current cuts in the NHS, the SU2S financial incentives scheme has provided a tangible help in terms of investment.

Limits and challenges of the Sign up to Safety Incentivisation Scheme

Trusts raised the following limitations of the scheme:

- Funding process and monetary aspects: funding was limited and some Trusts could not get the necessary resources to implement new interventions at all or at least in some areas. Moreover the funding was only made available for a short time, limiting the potential to observe substantial long-term benefits.
- Sustainability of the scheme: related to the funding is the sustainability of the scheme and embedding it into everyday care. For example, Trusts that have used the funds to recruit new staff are now facing the risk of not being able to retain those posts without additional resources.
- Organisational priorities: organisations have to ensure that the scheme fits in with their strategic objectives, otherwise it will not get the Trust Board support and fail. Also, in many cases the need to bid for funding before the work began meant that the original project scope did not necessarily match what was actually needed, because needs can change overtime.
- Data collection: several Trusts found it difficult to collect the data necessary to evaluate whether or not the interventions implemented were effective at reducing harms.
- Many Trusts noted the time required to see improvement may be long: many interventions will show their impact in the long term, therefore the scheme length may not be sufficient to show significant change.

Proposed changes

Trusts provided the following suggestions to improve different aspects of the scheme if it were to be run again in the future:

- Bidding process: the bidding process was perceived as very onerous, therefore more time and support should be provided with the bid and project planning.
- Support and collaboration: almost all Trusts have requested additional support from the NHS LA but also from academics to plan the interventions and to evaluate the impact made. This could be achieved by the NHS LA collaborating with established groups of academics

who could work with both the NHS LA and the locally funded teams. Additional support is also required from an IT perspective, to make sure that the data required for the evaluation of the interventions are collected in an appropriate, timely and complete way.

- Funding and sustainability: the Trusts suggest that funds should be given for long term investment of at least 3 years, to enable embedding and sustainability of change. This might mean funding fewer Trusts, each for a long period of time.
- Selection of the interventions: according to some Trusts it would have been useful to focus on key aspects of patient safety that need improving according to evidence and select interventions which have shown to be effective.
- Evaluation of the scheme's impact: while some Trusts suggest that they should be free to provide their individual measures of success, others think that an evaluation strategy with clear pre-identified measures of outcome should have been available at the bidding stage.

3. Development of a learning dataset

In this section we explore the feasibility of developing a potential learning dataset. The main aim of the dataset would be to collect data at Trust level on incidents and claims for a range of clinical areas, including maternity and A&E that could be used by a range of stakeholders to learn from and help avoid claims.

To be useful a dataset should have the following components:

1. Data on the number and type of incidents and their association with health and claims.
A detailed dataset containing these data would be able to provide background epidemiological data on the causes of incidents and how they relate to claims. To be useful, the type of incidents being collected should have a proven relationship with health and/or claims, and should be collected in a routine way to minimise under-reporting.

2. Data on interventions to reduce incidents and claims.
Another useful component would be a database of interventions that have been shown to be beneficial in reducing incidents or claims. This could also include interventions that have been proven to be not effective so these could be avoided.

3. Data to routinely evaluate the impact of interventions when they are implemented.
Such a dataset would facilitate evaluation of similar schemes in the future and more precisely analyses of the reduction in incidents and extra costs of interventions, plus supplementary data that could be used for modelling improvements in health, reduction in claims, a reduction in the cost of treating incidents and reduction in spending on claims.

3.1. Methods

We first describe what data are available in the NHS LA claim dataset, then we explored the content of the dataset available at national and local levels with specific reference to maternity and A&E.

3.1.1. NHS LA claims dataset

The NHS LA dataset contains data on claims at a national level from 1995 on a broad range of specialties. The data include the ID code of the claim, the name of the Trust or setting, the specialty (e.g., obstetrics or A&E), the main type of injury (e.g., cerebral palsy or missed fracture), the second and third injury if present, any concatenated injury (e.g., deafness, brain damage, unnecessary pain), the main cause of error (e.g., fail to monitor 1st stage labour, fail to diagnose fracture), the description of incident, including the date of incident, date of case creation (claim to NHS LA), the payout/outcome of case, including indication of the total cost paid, that includes costs for damage, defence and claimant costs. Recently a new field has been included in the dataset reporting whether the claim was preceded by a formal complaint.

The analysis of the NHS LA claim dataset could provide useful information on the main type of injuries and most importantly on the main causes of incident.

A detailed analysis of the maternity and A&E claim data is in sections 5 and 6.

3.1.2. Other available databases at national and local level

We checked the feasibility of linking the data available from NHS LA and from the Trusts with other databases at national and local level. In particular we looked at databases that could provide information on safety, harm and incidents, or that could be linked to errors with specific reference to maternity and A&E. These include:

- National Reporting and Learning System (NRLS);
- Perinatal and maternity patient management data from BadgerNet;
- Neonatal data as derived from the National Neonatal Research Database, maintained by the Neonatal Data Analysis Unit (NDAU);
- National Neonatal Audit Programme (NNAP);
- Risk management software tool;
- NHS maternity statistics;
- Hospital Episodes Statistics Admitted Patient Dataset;
- Friends and Family Test Data for maternity services;
- MBRRACE-UK: Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK;
- Maternity dashboard;
- The Each Baby Counts national improvement programme;
- Data from the "Getting it Right First Time" for orthopaedics errors.

A detailed description of each database is provided below.

The National Reporting and Learning System (NRLS)

The National Reporting and Learning System (NRLS)[4] is a central database of patient safety incident reports. Since the NRLS was set up in 2003, over four million incident reports have been submitted.
All information submitted is analysed to identify hazards, risks and opportunities to continuously improve the safety of patient care.
The database includes data on incidents by:

- type of incident: this includes accident, implementation of care and ongoing monitoring or review, medication, treatment and procedure, access, admission, transfer or discharge (or missing patients), documentation (including electronic and paper records, identification and drug charts), infrastructure, clinical assessment (diagnosis, scans, tests, assessments), self-harming behaviour, consent, communication, confidentiality, disruptive, aggressive behaviour (including patient to patient), medical device or equipment, infection control incident or other incidents;
- setting: acute and general hospital, ambulance, mental health service, community nursing, medical and therapy service, learning disability service, ambulance service, community pharmacy, general practice, community and general dental service and community optometry service;
- degree of harm: no harm, low, moderate, severe, death; by care setting and type of incident.

Despite being very informative, the quality of the data still needs improvement in terms of both completeness and accuracy (there are still many gaps in data, spelling errors and inconsistencies)[5] and the current dataset is not necessarily reflective of other measures of hospital quality and safety[6].

[4] http://www.nrls.npsa.nhs.uk/; https://report.nrls.nhs.uk/nrlsreporting/

[5] NIHR Imperial Patient safety translational research centre (2016). NRLS Research and Development. www.imperial.ac.uk/media/...of.../IMPJ4219-**NRLS**-report_010316-INTS-WEB.pdf

[6] Howell AM, Burns EM, Bouras G, Donaldson LJ, Athanasiou T, Darzi A. (2015). Can Patient Safety Incident Reports Be Used to Compare Hospital Safety? Results from a Quantitative Analysis of the English National Reporting and Learning System Data. PLoS One;10(12):e0144107.

Perinatal and maternity patient management data from BadgerNet

BadgerNet Maternity[7] is a full end-to-end, paperless maternity system with an easy to use interface that allows real-time recording of all events wherever they occur: in the hospital, community or home. This includes both high risk and low risk pregnancy pathways. The BadgerNet system is managed by Clevermed Ltd, an authorised NHS hosting company.

The system includes:

- BadgerNet Maternity, with data on maternity events and pregnancy pathways occurring in hospital, community or home;
- BadgerNet Neonatal, with records of care for all babies within neonatal services and it is in use in over 250 hospitals in the UK;
- BadgerNet Paediatric Intensive Care (PICU), that records all events within a paediatric intensive care or high dependency unit.

Despite the system having been improved in the last few years, BadgerNet needs to be complemented with data from the Trusts as some are not complete. This is done by the National Neonatal Research Database.

The National Neonatal Research Database

The National Neonatal Research Database (NNRD)[8] has been created through the collaborative efforts of neonatal services across the country to be a national resource. The NNRD is maintained and managed at the Neonatal Data Analysis Unit (NDAU) at Imperial College London and Chelsea and Westminster NHS Foundation Trust. The NNRD contains a defined set of data items (the Neonatal Dataset) that have been extracted from the BadgerNet neonatal electronic health record of all admissions to NHS neonatal units. Contributing neonatal units are known as the UK Neonatal Collaborative. The NNRD is updated each quarter and approximately 80,000 new patient records are incorporated each year. Data are currently available from 2010 when 96% of neonatal units in England contributed; from 2012 there has been 100% contribution from English neonatal units. The main advantage of using NNRD is that it contains data from BadgerNet and data from the Trusts that are cleaned and linked, so that the proportion of missing data is reduced.

National Neonatal Audit Programme

The National Neonatal Audit Programme (NNAP)[9] was established in 2006 to support professionals, families and commissioners in improving the provision of care provided by neonatal services which specialise in looking after babies who are born too early, with a low birth weight or who have a medical condition requiring specialist treatment. The NNAP Annual Report highlights the key findings and recommendations from the analysis of the data provided by neonatal units on the admissions of babies for neonatal care in England, Scotland and Wales each year. Data in the NNAP include data on: body temperature, antenatal steroids, retinopathy of prematurity, feeding, consultation with parents, disability follow-up, bronchopulmonary dysplasia, cultures and infections.

Risk management software tool (vendor system)

To electronically submit patient safety incident reports to NRLS, all Trusts use a risk management software tool, such as Datix[10], Prism, Sentinel, Ulysses[11] or other locally designed systems. These are web-based patient safety software for healthcare risk management applications. Using a variety of integrated software modules these software deliver a broad range of elements related to safety, risk and governance (e.g., incident, adverse event and near miss reporting, patient relations,

[7] http://www.clevermed.com/badgernet/badgernet-maternity/

[8] https://www1.imperial.ac.uk/neonataldataanalysis/data/;

[9] http://www.rcpch.ac.uk/improving-child-health/quality-improvement-and-clinical-audit/national-neonatal-audit-programme-nn-3

[10] http://www.datix.co.uk/

[11] http://www.ulysses.co.uk/systems/

malpractice claim management, risk assessment, safety alerts, patient experience and feedback, accreditation self-assessment, complaints, compliments, comments and concerns).

NHS maternity statistics

NHS maternity statistics[12] contains a wide range of maternity information with details of all births taking place in NHS hospitals (in England) excluding home births and those taking place in independent sector hospitals. This includes a wide range of information such as details of how the baby was born (method of delivery, onset), mother age, level of deprivation, ethnicity, delivery and birth complications, birth weight and gestation.

Hospital Episodes Statistics Admitted Patient Dataset

Hospital Episodes Statistics (HES)[13] is a data warehouse containing records of all patients admitted to NHS hospitals in England. It contains details of every hospital stay in English NHS hospitals and English NHS commissioned activity in the independent sector. Each record in HES includes a wide range of information including details of the patient (age, gender, geographic details), when they were treated and what they were treated for. This is the most complete database on admissions data in England.

Friends and Family Test Data for maternity services

The Staff FFT[14] was introduced in April 2014 to allow staff to give feedback on NHS services based on recent experience. Staff are asked to respond to 2 questions: a) the care question asks how likely staff are to recommend the NHS services they work in to friends and relatives needing care; b) the work question asks how likely staff would be to recommend the NHS service to friends and family as a place to work. The survey is conducted on a quarterly basis. The data are not classified as Official Statistics.

MBRRACE-UK: Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK;

MBRRACE-UK[15] is a national collaborative programme of work involving the surveillance and investigation of maternal deaths, stillbirths and infant death, including the Confidential Enquiry into Maternal Deaths (CEMD). MMBRRACE-UK collects information about all mothers in the UK who die during pregnancy or within 12 months after giving birth and some mothers who experience a serious illness in pregnancy or soon after giving birth, all mothers of babies who are stillborn or whose baby dies in the first few weeks after being born or have serious illnesses, and all births and stillbirths across UK (data cover England, Wales and Scotland but arrangements are still ongoing for Northern Ireland).

Maternity dashboard

The maternity dashboard[16] is a tool used to monitor the care given by maternity service against what is considered to be good practice in relation to safety and quality. It was first developed in 2008/09 for use across Kent, Surrey and Sussex and it was reviewed in 2014 by the South East Maternity, Children and Young People's Strategic Clinical Network (SE MCYP SCN). The dashboard includes a range of metrics that look at process and outcomes from several sources: HES, the national Friends and Family Test, and locally supplied data. Metrics include data on care, type of delivery (vaginal birth, caesarean sections, instrumental births, epidural, and episiotomy), serious

[12] http://content.digital.nhs.uk/catalogue/PUB16725
[13] http://content.digital.nhs.uk/hes
[14] https://www.england.nhs.uk/statistics/statistical-work-areas/friends-and-family-test/friends-and-family-test-data/
[15] https://www.npeu.ox.ac.uk/mbrrace-uk
[16] https://www.rcog.org.uk/en/guidelines-research-services/guidelines/good-practice-7/

incidents, complaints, 3rd and 4th degree tears, emergency readmissions within 30 days of delivery and NICU admissions. It provides information about the quality and safety of maternity services across the South East and is designed as a quality improvement tool for hospital providers of maternity care and commissioners of maternity services. Dr Matthew Jolly, National Clinical Director for the National Maternity Review and Women's Health, is currently leading the innovative development of a national interactive maternity dashboard. This could in the future have potential links with NHS LA claims data.

The Each Baby Counts national quality improvement programme

Each Baby Counts[17] is the RCOG's national quality improvement programme to reduce the number of babies who die or are left severely disabled as a result of incidents occurring during term labour. The aim of the programme is to reduce unnecessary loss of life by 50% by 2020.

Since 2015 data on still birth, death of a new-born baby and the birth of a baby with brain injuries have been collected and analysed from all UK units to identify lessons learned to improve future care. This will allow making recommendations to be made on how to improve practice at a national level.

Data from the "Getting it Right First Time" for orthopaedics errors (as part of the Royal National Orthopaedic Hospital NHS Trust)

The 'Getting it right first time' (GIRFT)[18] report published by Professor Briggs in late 2012, considered the current state of England's orthopaedic surgery provision and suggested that changes can be made to improve pathways of care, patient experience, and outcomes with significant cost savings. The report takes the view that this approach has the potential to deliver a timely and cost effective improvement in the standard of orthopaedic care across England. Part of the project looked at claims in orthopaedics, but data were gathered from the NHS LA claim database. A new pilot will undertake a national review of baseline data and collect new data on clinical outcomes, processes, patients experience and pathways, network arrangements, financial impacts and waiting times.

3.2. Experience of data correspondence from this study

As part of the study we have tried to correlate data on incidents provided to us by the Trusts with data on claims from the NHS LA. However this proved to be extremely difficult as the Trusts are not currently provided with a computerised system to record this information.

Positive aspects of the data gathering to evaluate the scheme

Having to collect data to evaluate the scheme meant that some Trusts discovered anomalies across different systems they used, or glitches where data moves from one system to another. Based on the sanity checks that we ran on their data, Trusts had to look into discrepancies and are now able to look into fixing this and were grateful for the queries.

Main challenges

The main difficulties in obtaining sufficient data to analyse the potential harm and to be used to avoid incidents have been identified as the following:

- Lack of dedicated staff to register the data;
- Lack of time for current staff (in same cases people had to work during days off to collect the data);

[17] https://www.rcog.org.uk/eachbabycounts
[18] http://www.gettingitrightfirsttime.com/

- Lack of expertise in accessing the data already available and consequent requirement to involvement of more people (especially if data held with more than one department);
- Information is stored across different systems and in different departments (e.g., data on perinatal harm is stored in maternity units and neonatal units in different systems);
- Limitations of access to IT systems regarding what has been recorded;
- Lack of access to IT system altogether (this is rare, but it was reported in one Trust), meaning that people had to search through paper records for some information;
- Changing IT systems, meaning that figures are recorded differently between old and new systems and separate searches have to be done for different time periods.

Potential changes and improvements
According to Trusts some things could perhaps help change the current system and make data collection easier:

- Knowing ahead of time that they would be required to submit this information as part of the evaluation of the SU2S scheme would have meant that they could have tried to record it, at least for the last 2-3 years, in the right format, to make access simpler and faster;
- Providing funding for proper IT systems in some Trusts. It seems there is little point in trying to improve things if there is no way of measuring whether or not it has worked.
- Providing funding for extra or dedicated staff to collect and record data.

3.3. Discussion

A detailed dataset containing data on the type and number of incidents and their association with health and claim would provide background epidemiological data on the causes of incidents and help in designing strategies for reducing claims. Data to routinely evaluate the impact of interventions when they are implemented would facilitate the analyses of the reduction in incidents and extra costs of interventions.

Various software systems are being used in maternity units to collect and store electronic data, which can be easily retrieved and linked to other databases. However, each unit should have a designated person responsible to ensure accurate recording and maintenance of maternity data. The CNST clinical governance standards require maternity units to have a risk management midwife or a manager in place.

It is important to crosscheck the data to ensure accuracy; for example, the operation book in the operating theatre could be checked to verify the number of caesarean performed each month or week. Information about patient complaints could be obtained from the complaints manager of the Trust.

As mentioned, to be useful a dataset should have the following components:

1. Data on the number and type of incidents and their association with health and claims.
A detailed dataset containing these data would be able to provide background epidemiological data on the causes of incidents and how they relate to claims, which would be useful for designing strategies for reducing claims, plus evaluating them (see number 3, below). To be useful, the type of incidents being collected should have a proven relationship with health and/or claims, and should be collected in such a way so as to ensure reliable reporting (i.e., minimising under-reporting by making data collection routine). The costs of collecting, analysing and disseminating the data should not be prohibitive.

2. Data on interventions to reduce incidents and claims.
While having epidemiological data of the kind described above is useful it is unlikely to directly inform strategies to reduce claims. Therefore, another useful component would be a database of interventions that have could be beneficial in reducing incidents or claims. This could also include interventions that have been proven to be not effective so these could be avoided. Data would also ideally be recorded on the incremental cost-effectiveness of interventions, where these data exist, as well as their impact on incidents and claims, along with the available evidence. This is likely to consist of systematic reviews that are routinely updated and disseminated.

3. Data to routinely evaluate the impact of interventions when they are implemented.
Such a dataset would facilitate evaluation of similar schemes in the future and more precisely analyses of the reduction in incidents and extra costs of interventions, plus supplementary data that could be used for modelling improvements in health, reduction in claims, a reduction in the cost of treating incidents and reduction in spending on claims. As noted, this dataset should be able to accommodate a study design that can evaluate the impact of interventions that is not affected by confounding factors, and therefore data need to be collected across Trusts and over time.

Several datasets are currently available on incidents, health outcomes and claims. Given the number and type of datasets that are currently available, careful thought needs to be given to the costs and benefits of collecting new data and before doing so confidence is needed that it will provide additional new information that could be used to learn from and help avoid claims. Important questions to consider before embarking on a new data collection exercise might include:

1. What are the benefits of collecting new data?
    1.1. Is there evidence the data on incidents being collected are related to claims?
    1.2. Can the data be collected in such a way so as to minimise or avoid under-reporting?
    1.3. How, precisely, will the data being collected be useful to learn from and help avoid claims?
2. What are the costs of collecting the data? These are likely to include costs incurred to:
    2.1. Collect and record the data in Trusts.
    2.2. Extract the data from Trusts and assemble it into a central database for analysis, including dealing with data security and information governance issues.
    2.3. Analyse the data.
    2.4. Disseminate the data and analyses.

An alternative approach that should also be considered is to build on existing datasets, focusing on better linkage between them and improving data quality by encouraging accurate and timely reporting. We suggest that the way forward may be to invest more in coordinating and improving the reliability of what already exist, instead of starting to collect new data.

Information Governance issues need to be resolved so data can be easily interrogated and shared with researchers and Trusts, with appropriate data handling and security restrictions. These data could also be routinely disseminated, though impact on participation should be explored if data are not anonymised (see the Sentinel Stroke National Audit Programme[19] and the National Chronic Obstructive Pulmonary Disease (COPD) Audit Programme[20] for examples of widely disseminated routinely collected data).

[19] https://www.strokeaudit.org/
[20] https://www.rcplondon.ac.uk/projects/national-copd-audit-programme-starting-2013

4. Economic evaluation of the Incentivisation Sign up to Safety Scheme

In this section we describe the overall methodology we planned to adopt for the evaluation of the Sign up to Safety financial incentives scheme in maternity and A&E units.
More specifically we describe the:

- overall approach adopted
- identification of interventions and measures
- identification and assessment of the interventions' costs
- identification of comparators
- identification and assessment of the impact of the interventions in terms of errors
- Difference-in-Difference approach
- identification and estimation of the reduction in costs due to avoided errors/incidents and future claims
- estimation of the improvement in health due to a reduction of errors and incidents
- assessment of the difference in benefits and costs
- measurement of cost-effectiveness
- sensitivity analysis
- data collection
- final analysis

A more detailed description of the methods is provided in sections 5 and 6 where the A&E and the maternity interventions are specifically analysed.

4.1. Economic Evaluation Approach

The approach

Our proposed conceptual framework to evaluate the cost-effectiveness of the incentive funding scheme is summarised in Figure 1. The aim is to calculate the *incremental cost-effectiveness* of the scheme. The analysis is 'incremental' in the sense that cost-effectiveness is evaluated with respect to a comparator, the counterfactual (i.e. not participating in the scheme). The incremental cost-effectiveness of the scheme depends on the *difference in benefits* or outcomes associated with participating in the scheme compared with not participating in it, and the *differences in costs*.

Interventions and measures

The first phase of the evaluation consisted of the identification of the single interventions /actions proposed by each Trust to reduce harm and incidents in the maternity and A&E units.

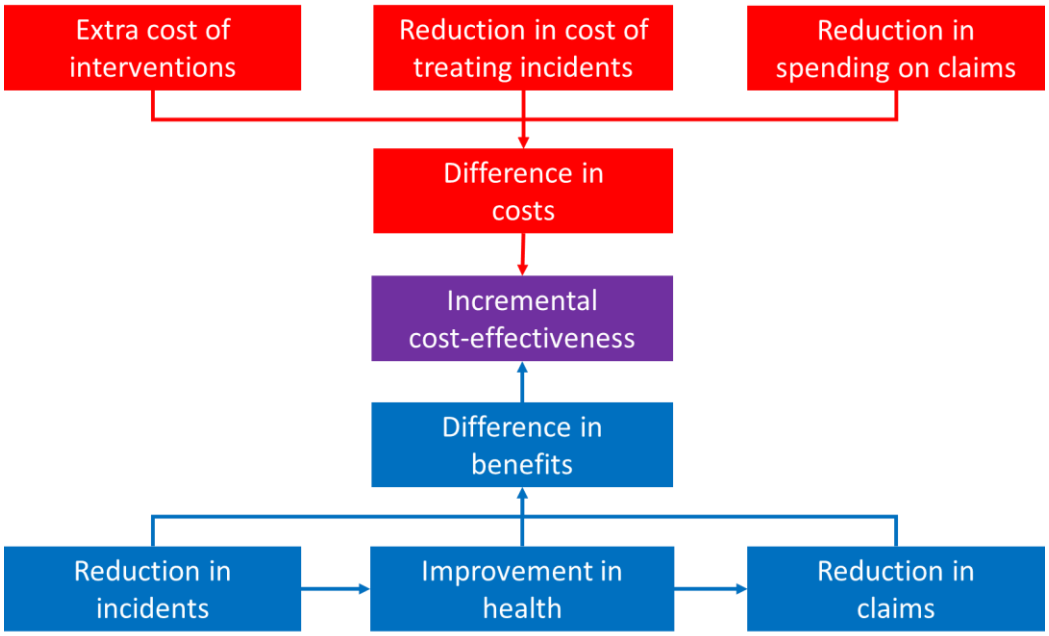
We contacted the successful[21] Trusts to provide us with a report summarising the specific areas they were addressing as part of the scheme, which interventions they were proposing, how the funds were allocated and spent across different interventions, what outcome measures they were planning to use and the type of information available on the impact of the scheme (number and types of incident and claims). The form used to collect these information is in Appendix 4.

Due to the heterogeneity of the interventions implemented in each Trust we grouped the interventions in 9 groups according to aim and similarity (e.g., computer hardware/software, fetal monitoring equipment, patient monitoring equipment, Q&S improvement administrator or staff/nurse/midwife, specialised obstetric equipment, training and development and neonatal

[21] Note that the terms "successful" and "unsuccessful" in the context of this report refer only to whether or not the Trust was successful in being awarded funding for improvements via the Sign up to Safety Scheme. We do not refer to the ability of the Trust in reducing errors or claims.

transport equipment). A more detailed description of the interventions will be provided in the specific sections 5 and 6 and in Appendices 4-6.

Figure 1 Evaluation framework



Assessment of the cost of activities/interventions

We assessed the extra costs of activities/interventions introduced by Trusts as part of the incentive funding scheme by looking at the data available in the applications they submitted to get the funds. However, we also contacted individual Trusts to determine if the actual expenditure on interventions reflected the costs included in the original application. This allowed us to get further detail on exactly what the intervention entailed (e.g., staff training, monitoring) for those activities and interventions whose cost is not that straightforward to value.

Identification of comparators

To assess whether the Sign up to Safety scheme has had an impact in reducing harms and errors we needed to identify a comparator. We used both the "time period before the introduction of the scheme" and the "situation in absence of scheme" as comparators. This allowed us to compare the costs and benefits of participating in the scheme with those of not participating in it. Specifically we looked at the extra costs of interventions and the reduction in incidents, as all the other benefits and costs can be estimated indirectly from the latter. The extra cost of interventions and the reduction in incidents were collected directly by the Trusts or derived using the measures they identified, and equivalent data were available for the comparator. We used Trusts that were not awarded funding in the Sign up to Safety scheme and collected data before and after the introduction of the scheme. This allowed us to apply a difference-in-differences (DiD) approach and take into account confounding factors that might have occurred during the implementation of the scheme. A more detailed description of the approach is provided in the methodology section 6.

Assessment of the impact of the interventions in terms of errors and incidents

Because of the wide range of interventions implemented, their effects on incidents are not the same and do not have the same intensity. Therefore, we tried to disaggregate incidents as much as possible by the degree of harm they cause (e.g., no harm, low harm, moderate harm, severe harm, death) and/or by type of incident.

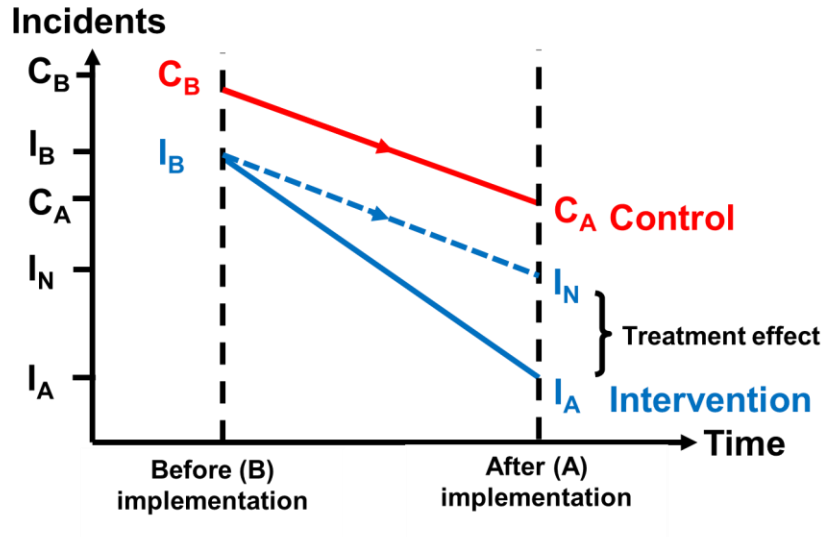
In order to adopt a DiD approach, similar data had to be collected before the implementation of the scheme and in other Trusts not participating (the control group), therefore we selected specific outcome measures that could be collected from all Trusts' routine data.

The Difference-in-Differences approach

We used a Difference-in-Differences (DiD) approach for the analysis of the difference in incidents before and after the introduction of the Sign up to Safety scheme in maternity units.

This is a statistical technique that attempts to mimic an experimental research design using observational study data. It calculates the effect of a treatment (e.g., participating in the incentive funding scheme) on an outcome (e.g., the number of unexpected admissions of term babies to neonatal intensive care unit or the number of term babies born with a low Apgar score) by comparing the average change over time in the outcome for the treatment group (e.g., successful (funded) Trusts) to the average change over time for the control group (unsuccessful (non-funded) Trusts). This method may eliminate possible confounding effects as it measures the difference-in-differences between the treatment and control group over time. DiD requires data measured at two (or more) different time periods for the two options being compared. In Fig. 2, the intervention group is represented by the solid blue line (I_B I_A) and the comparator group is represented by the solid red line (C_B C_A). Both groups are measured on the outcome (number of incidents) over time, before the intervention group has implemented the intervention (shown by I_B and C_B in Fig. 2). The intervention group then participates in the intervention and both groups are again measured after this point (I_A and C_A). Not all the change in outcomes over time for the intervention group (I_A minus I_B) is an effect of the intervention, because some benefits were achieved by the comparator group over time due to other factors (e.g., other interventions that were introduced at around the same time), and these would also have been achieved by the intervention group. DiD therefore calculates the change in the outcome variable over time for the intervention group, subtracting from this the change over time that was achieved by the comparator group due to other factors, shown by the dashed blue line (note that the slope of the line connecting I_B and I_N is the same as the slope connecting C_B to C_A). Hence the treatment effect of the incentive funding scheme is ((I_A minus I_B) minus (I_N minus I_B)), which is equal to (I_A minus I_N). This method is more demanding in terms of data because it requires data on incidents for Trusts participating in the scheme before and after implementing the interventions; it also requires equivalent data over the same time period for Trusts not participating in the scheme. However, given the probability of generating more accurate estimates of the impact of participating in the scheme we decided to adopt this approach.

Figure 2. Difference-in-differences approach



Identification and estimation of the reduction in costs due to avoided errors/incidents

Once we estimated the reduction in number of incidents and errors (by type and degree) we measured the monetary value of treating those incidents using payment by results data and, where not available, evidence from the literature and cost-of-illness studies. This focused on the short-run costs of dealing with incidents, rather than the long-term financial impacts, which are captured below.

Assessment of the reduction in claims

The relationship between incidents and claims was based on data disaggregated by clinical area (e.g., maternity, A&E) and by type of incident, as the number of claims as a percentage of incidents is unlikely to be constant across types of incident.

We decided to use specific data at Trust level instead of using the NRLS as the latter contains data on reported safety incidents, which may underestimate the actual number of incidents for several reasons:[22]

- Persistence of the "blame culture"[23];
- Fear of litigation or prosecution[23];
- Lack of response to previous reports[24];
- Lack of appropriate reporting systems[25];
- Lack of contractual incentives[22];
- Poor understanding of what to report as an incident[25];
- Lack of knowledge about how to report an incident [24]; and,
- Lengthy and complicated reporting processes[26].

For example, in a survey undertaken by the National Audit Office it was found that "*on average around 22 per cent of incidents went unreported and 39 per cent of near misses*". We acknowledge there may also be some degree of under-reporting in the data we collected directly from Trusts, but we preferred to use local data.

Given the specified time frame of the evaluation we could not measure any direct effect of the scheme on improving health outcomes and, in particular, in reducing claims. This is because there is a time lag from the time of perceived harm or incident, to settlement of a claim. We evaluated this time frame as part of the analysis of claims contained in specific sections 5 and 6.

Estimation of the reduction in costs due to avoided claims

The reduction in spending on claims has been measured focusing on the value of claims rather than the number, i.e., applying a mean value per claim. The mean value is specific to the type of incident (e.g., cerebral palsy, tears, missed fracture etc.) and reported degree of harm, since the value of the claim will vary by these factors. In addition, as the mean value of claims is not constant over time, it has been valued using the most recent data.

Estimation of the improvement in health due to a reduction of errors and incidents

Reduction in errors and incidents should translate into improvements in health. Generally the most common outcome measure in economic analyses is the Quality Adjusted Life Year, a measure that combines the length of life and the quality of that life. The use of this measure is recommended as it allows to comparison of interventions in different clinical areas. As Trusts are not capturing such

[22] http://www.publications.parliament.uk/pa/cm200809/cmselect/cmhealth/151/15108.htm
[23] Committee of Public Accounts, *A safer place for patients*, p 5
[24] Centre for Patient Safety and Service Quality, Imperial College, commissioned research
[25] National Patient Safety Agency, *Seven Steps to Patient Safety-Step 4: Promote reporting*, August 2004, p 97
[26] National Audit Office, commissioned research

information, improvements in health have been modelled, being predicted from reductions in incidents. To illustrate, suppose we have a patient population of 100 people. If patients experience an adverse incident then their health outcomes are 1 quality adjusted life year (QALY) each; if they do not experience an incident their outcomes are 10 QALYs each. In the absence of treatment via the Sign up to Safety incentive funding scheme 10 patients experienced an adverse incident and 90 do not. Therefore, the total QALYs available in the absence of the SU2S scheme across all patient are 10 patients*1 QALY + 90 patients*10 QALYs = 910 QALYs and the mean QALYs per patient are 910 QALYs/100 patients = 9.1.

Suppose now that the SU2S scheme pays for an intervention to improve safety, and this reduces the number of patients experiencing incidents from 10 to 5, the total QALYs across all patients are now 5*1 + 95*10 = 955 and the mean QALYs per patient are now 9.55. Hence, the intervention produces a mean improvement in health per patient of 9.55-9.1 = 0.45 QALYs.

Calculations of this kind required an estimate of the impact of the intervention on the reduction in the number of incidents, as above. They also required mapping the number of incidents to some measure of health, e.g., QALYs. To do this in the A&E setting, we used data from published sources on the health consequences of adverse incidents.

Assessment of the difference in benefits and costs to calculate the incremental cost effectiveness

As mentioned above, the proposed conceptual framework to evaluate the cost-effectiveness of the incentive funding scheme is summarised in Fig. 1. The aim is to calculate the *incremental cost-effectiveness* of the scheme. The analysis is 'incremental' in the sense that cost-effectiveness is evaluated with respect to a comparator, the counterfactual (e.g., Trusts not participating in the scheme). The incremental cost-effectiveness of the scheme depends on the *difference in benefits* or outcomes associated with participating in the scheme compared with not participating in it, and the *differences in costs*.

Measuring cost-effectiveness

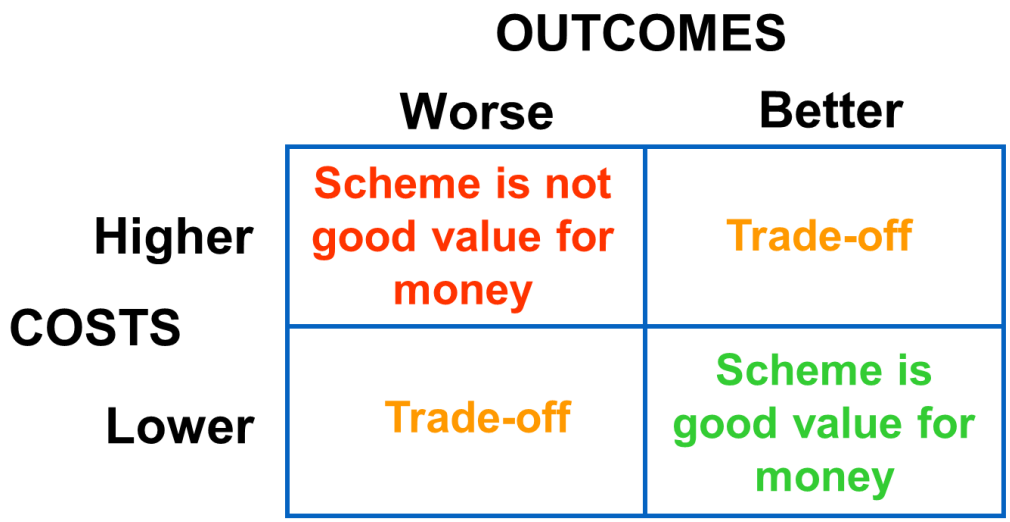
The framework described in the previous section is designed to measure the incremental cost-effectiveness of the Sign up to Safety incentive funding scheme. For two or more interventions or strategies (e.g., participating in the scheme versus not participating in it), an incremental cost-effectiveness analysis compares the resources used by each alternative (the costs) with the number of incidents, health, claims or other outcomes achieved (the effectiveness). Costs may be higher or lower with the scheme, and outcomes may be better or worse; the combinations of these differences are shown in Fig. 3. If the scheme has lower costs and better outcomes than the alternative (falling into the bottom right hand quadrant of Fig. 3), then it represents good value for money and will look attractive to decisions makers. Or, if the scheme incurs higher costs and worse outcomes than the alternative (top left hand quadrant of Fig. 3) then it does not represent good value for money. If the scheme is more effective than the alternative but only at a higher cost (top right hand quadrant of Fig. 3), then to decide whether or not the scheme represents good value for money requires a judgement as to whether the improved outcomes are worth the extra costs. If it is judged that the extra benefits are worth the extra costs then the scheme represents good value for money. Note that an intervention does not necessarily need to save money to be cost-effective. However, it does need to generate outcomes that are 'worth' paying for.

One way to weigh up the relative differences in outcomes and costs and make the trade-off is to measure outcomes using a generic health outcome measure, such as QALYs as discussed above. Since most interventions designed to improve health affect quality of life or length of life or both then this measure can be used across all clinical areas, allowing broad resource allocation decisions to be made. The National Institute for Health and Care Excellence (NICE) recommends a cost-

effectiveness threshold value for QALYs: an intervention that costs more than £30,000 per QALY is unlikely to be considered to be 'worth it' because it is believed that the money could be better spent elsewhere in the NHS.

As noted in the previous sections, improvements in health associated with the scheme can theoretically be measured in terms of QALYs; in this case it is possible to judge whether the improvements in health associated with the scheme are good value for money – are cost-effective – by applying the NICE threshold. Alternatively, it is possible to use the cost-effectiveness threshold to convert QALY gains associated with the Sign up to Safety incentive funding scheme into monetary terms, meaning that it is possible to compare the different in costs and the improvement in health associated with the scheme directly. If improvements in health are quantified using some measure other that QALYs then a cost-effectiveness threshold value is unlikely to exist for that measure, meaning that it will be difficult to judge if the scheme is cost-effective if it falls into the top right hand quadrant of Fig. 3. Empirically, measuring the impact of the scheme on health outcomes (QALYs or otherwise) is quite difficult. Nonetheless, we tried to measure possible improvements in health associated with the scheme in terms of QALYs using published sources.

Figure 3. Measuring incremental cost-effectiveness



Sensitivity analysis

Due to the uncertainty in each component of the modelling process we used univariate and probabilistic sensitivity analysis to explore and quantify uncertainty. Univariate sensitivity analysis has been performed varying all model inputs one at a time within plausible ranges to investigate the impact on incremental cost-effectiveness. Probabilistic sensitivity analysis (PSA) applies appropriate distributions to reflect the uncertainty with each parameter value in the model. The PSA can be used to present confidence intervals around the point estimates of cost-effectiveness, and be used to create a cost-effectiveness acceptability curve.

Data collection

We asked the Trusts to produce a report containing relevant information about the interventions. The report included information on the type of intervention, the date funds were awarded and the amount spent to implement each intervention (Appendix 4).

We also asked Trusts to provide some outcome measures that could help evaluating the effect of the interventions in reducing incidents and therefore the number of future claims (Appendix 4-7).
We asked Trusts to provide these data for the period before the implementation of the interventions and afterwards (up to the most recent date). Different questionnaires were administered for the maternity and A&E analyses (Appendix 8-9).

Data on claims before and after the implementation of interventions have been collected at Trust level by accessing the NHS LA claims database.

Applying the methodology
As noted, impacts on health and on claims may not materialise until sometime after the impact on incidents, perhaps several years afterwards. Therefore it is difficult to measure the difference in benefits and costs directly.

We therefore measured the reduction in incidents and extra costs of interventions directly, and use the reduction in incidents plus data from other sources to model improvements in health, reduction in claims, cost of treating incidents and reduction in spending on claims. We are aware that this approach is much more speculative and uncertain and does not measure impacts on health and claims directly.

Following the results of the DiD approach we planned to undertake an incremental cost-effectiveness analysis to compare the resources used by each intervention (the costs) with the number of incidents, health, claims or other outcomes achieved.
A sensitivity analysis has been performed to take into account of uncertainty in the input data and test the robustness of the results.

In the following sections we provide a more detailed analysis of the evaluation of the scheme in maternity units and A&E.

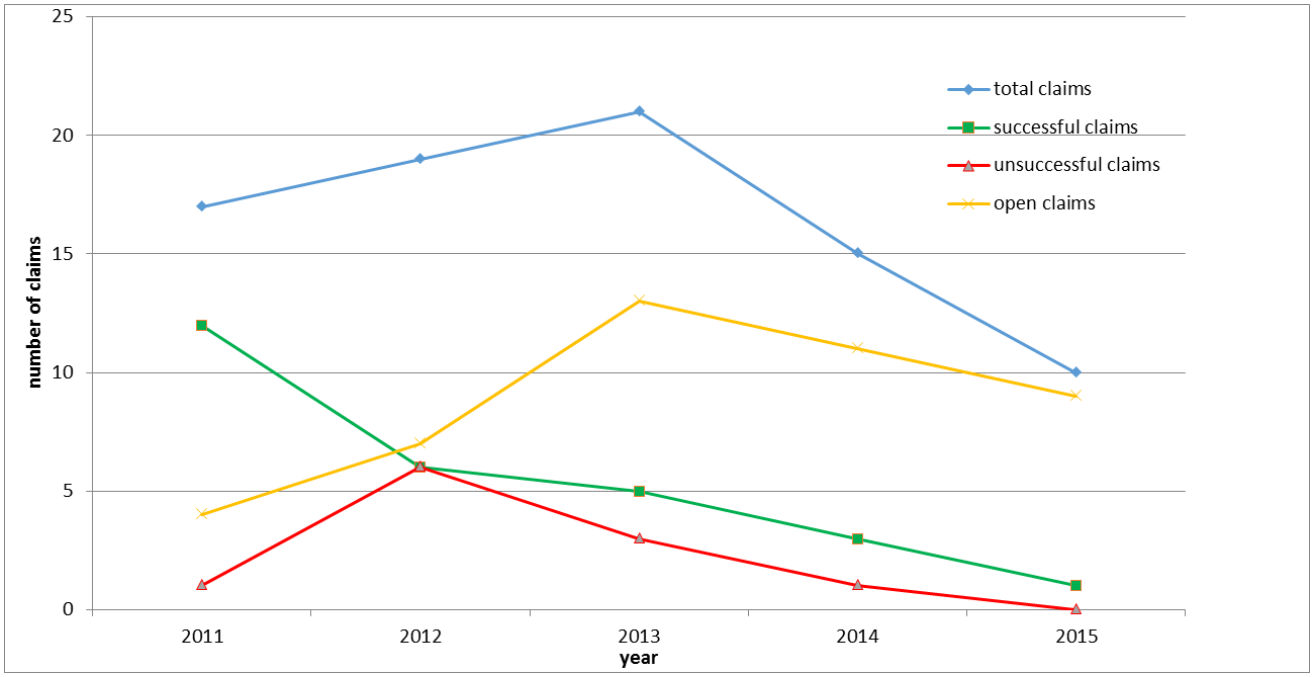
5. Evaluation of the Incentivisation Scheme in A&E: missed fractures

5.1. Overview of funding and claims

In total, 13 Trusts from across a wide geographical locations applied to the Sign up to Safety incentive funding scheme for interventions to reduce missed fractures in A&E. Out of these, 5 Trusts were successful in obtaining the funds, while for 8 Trusts funds were not awarded. Funds were given to Trusts that proposed to improve 'out of hours' services, to aid rapid diagnosis and to reduce the number of missed fractures and subsequent claims for missed fractures.

Between 2011 and 2016, the NHS LA received 146 claims related to missed fractures in A&E for the 13 Trusts that applied to the Incentive Scheme (68 claims were successful, 26 were unsuccessful and 51 are still open, most of which refer to claims made from 2014 onwards). Figure 4 shows the trends in total, successful, unsuccessful and open claims referring to incidents that occurred between 2011 and 2015. Over time we notice an increased trend until 2014, when there is a sudden decrease. Despite the funding that occurred in 2015 we cannot attribute a reduction to the financial incentive scheme mainly because the decline shown is probably due to the fact that many incidents from 2014/15 may not have been registered as claims yet. As the average time gap between the incident and claim is approximately 2 years, we can infer that the number of claims referred to incidents occurred in 2014 is still not complete and therefore we are unable to predict the trend in the following years.

Figure 4. Trend of claims for missed fractures due to incidents occurred in 2011-2015

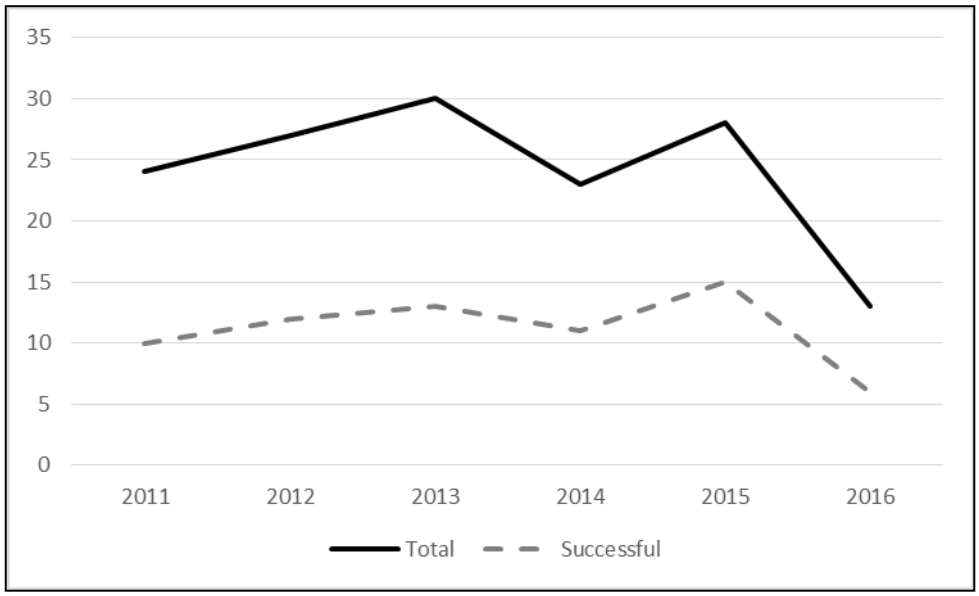


Source: our analysis based on NHS LA claim data

The number of claims due to missed fractures or related incidents in the Trusts that received the funds is 67, which corresponds to 45.9%. Figure 5 shows the total number of claims and claims in successful Trusts by year. The trend presented should be considered cautiously, particularly for 2016, for two reasons: (1) considering the average time for a claim to be made of two years, it is possible that not all claims from 2015 and 2016 have been opened yet and (2) given that the data provided only refers to claims made up until August 2016. Nonetheless, from the Figure, we observe that after 2014, the successful Trusts are increasing their relative proportion of claims, narrowing

the gap between the solid and dotted lines. The funds for the Sign Up to Safety financial incentives Scheme were granted in 2015, when a sudden drop in trends is observed. However, as mentioned previously this may be partly attributable to the fact that some claims for the last two years are yet to be opened, so this drop cannot be attributed to the scheme.

Figure 5. Number of claims by year: Total and Successful Trusts



Source: our analysis based on NHS LA claim data

The main causes of incident are: failure or delay in diagnosis (52% of all cases), failure to interpret X-Ray (15% of all cases) or failure to X-Ray (13% of all cases).

On average the total cost of a claim for missed fracture is £51,456, including the unsuccessful cases. The average cost of a successful case is £41,969, but it can vary from a minimum of £1,200 up to £270,870. The cost of the total damage, on average £17,677, is very low compared to a maternity claim and can reach maximum £150,000. The defence costs are generally very low, £3,000, whereas the claimant costs are higher – on average £21,000 but could reach £127,000.

5.2. Descriptive analysis of interventions

We asked the 5 successful Trusts to complete a standardised report (Appendix 4) to understand what was the main problem they wished to address (e.g., missed fractures) and how they were planning to address it through the scheme. In the same report we also asked Trusts to list some outcome measures that could help evaluate the effect of the interventions.

All five have completed the report, but a preliminary analysis showed that one successful Trust had been granted funds to improve alternative outcomes, including Sepsis and Pressure Ulcers. Therefore, the information from this particular Trust was not used to define the interventions put in place to reduce the number of missed fractures on A&E, and the consequent modelling performed in the next sections.

The four remaining Trusts had each chosen a very different approach to reducing missed fractures. One Trust invested in radiology imaging equipment for rapid interpretation and analysis of x-ray images, also known as hot reporting in out-of-hours A&E services; two invested in employing a nurse to focus on quality improvement, analysis of incidents and data recording, and one invested in

support training for staff in the interpretation of x-rays. Finally, one of the Trusts that chose to employ a nurse was also awarded funds to employ a radiographer allowing for hot reporting 7 days a week. Thus, in the 4 Trusts, five different types of interventions were introduced.

In total, for improvements on A&E focusing on missed fractures £388,208 were granted by NHS LA[27]. The most expensive intervention was the purchase of radiology imaging equipment, which received £257,194 in funding, whilst the cheapest intervention was the employment of one WTE radiographer for one year, which received £18,304. The relatively large sum invested in the radiology imaging equipment can be understood due to its capital nature and considering that the expected life use of the equipment is between 5 and 10 years.

Overall 19 of the 25 planned interventions have been implemented from 2014 and the majority of them were fully implemented within 2015 (74%), or by July 2016 (26%). It is worth noticing that some interventions (12%) were partially implemented or not implemented (20%) at the time the evaluation started.

5.3. Data collection

In order to measure the impact of the adopted interventions in reducing the number of missed fractures using the DiD approach described in the previous section we requested data from successful and unsuccessful Trusts before and after the introduction of the scheme. To do so we sought expert advice from The Society and College of Radiographers, to define the necessary information to be collected from the Trusts in order to perform this evaluation.

All five successful Trusts and eight non-successful Trusts were contacted by phone and email, and a questionnaire on the number and severity of missed fractures and claims as well as on the number of x-rays taken at their A&E departments was sent to all Trusts (Appendix 8). The questionnaires were directed to the contact person referred by NHS LA, and most occupied a senior management position, including Director of Quality, Head of Corporate Risk and Governance Lead. The data gathering process took several months, and the research team had to engage with Trusts due to difficulties in data gathering. The leads in each Trust involved other members of staff, including clinical leads in an attempt to fill in the data to the best of their capacity. Having consulted many of the responsible staff in each Trust, it became clear that most (both participating and non-participating) do not consistently collect data on missed fractures. Some consider this data collection not feasible, given that if a fracture was missed it is unlikely that the patient will return to the same A&E where an error has occurred, and thus missed fractures will not be identified in their Trust unless a claim was presented.

In the end, we received three incomplete datasets from participating Trusts, with one Trust stating that none of the requested data could be collected. Similarly, only one Trust that was not awarded funding provided (incomplete) data.

5.4. Methods

We undertook an analysis of the effectiveness and cost-effectiveness of the interventions put in place to reduce the number of missed fractures in the participating Trusts funded by the Sign up to Safety incentive funding scheme. The effectiveness of the interventions could not be measured from the data provided by the Trusts, as most Trusts did not respond, and those which did provided incomplete reports, as explained above. Therefore, we were unable to adopt the methodological approach described in section 7. Instead, the effectiveness of the interventions aimed at reducing

[27] In total £0.8 milion was awarded to A&E Trusts, but only £388,208 was for interventions to reduce missed fractures and the remaining £400,000 was for interventions to avoid sepsis, ulcers, infections etc.

the number of missed fractures in A&E departments was based on the effect measured in previous studies identified by a means of a systematic literature review undertaken by the research team. This review is presented in the next section.

Based on this information, a cost-effectiveness analysis was conducted to estimate the costs and health consequences of implementing the interventions to reduce the number of missed fractures at A&E compared with standard practice. In this analysis we simulated and compared the costs and health outcomes of a cohort of patients attending A&E due to a suspected fracture at each of the 4 funded Trusts with and without the interventions in place. The analysis took the perspective of the National Health Service (NHS). Effectiveness was measured using quality-adjusted life years (QALYs). The analysis followed the standard assumptions of the National Institute for Health and Care Excellence (NICE) reference case including applying a discount rate to future costs and benefits of 3.5%. A lifetime horizon was used, but no costs were modelled beyond the first year and therefore discounting was only applicable to health effects. As described in section 4, cost-effectiveness was summarized by the incremental cost-effectiveness ratio (ICER), defined as the incremental cost divided by the incremental effectiveness of two competing alternatives. The ICER represents the additional cost required to achieve one additional unit of effectiveness. The ICER is then compared with the decision makers' willingness to pay threshold in order to draw conclusions about the cost-effectiveness of the intervention. NICE recommends a cost-effectiveness threshold value for QALYs: an intervention that costs less than £20,000-£30,000 per QALY gained is considered to be cost-effective.

In the following sections we describe the literature review undertaken to measure the effectiveness of the interventions, and the cost-effectiveness model developed to estimate the cost and health outcomes of such interventions compared to standard care. This allows us to provide a comprehensive evaluation of the potential impact of Sign up to Safety Scheme in A&E missed fractures.

5.5. Literature review

5.5.1. Search aim

A systematic review of international literature was undertaken with the aim of identifying the effect of the interventions adopted in participating Trusts aimed at reducing missed fractures in A&E departments.

5.5.2. Search methods

Study identification

We used the following search terms: "missed fractures" AND ("A&E" OR "emergency department" OR "accident and emergency") in the full text. The literature search was carried out using the electronic databases Scopus, the Cochrane Library and the NHS Economic Evaluation Database held by the Centre for Reviews and Dissemination (http://www.crd.york.ac.uk/CRDWeb/). We also searched the database of NICE Guidelines. The search covered the literature back to the beginning of each database forward until October 2016. Titles and abstracts of all the articles identified were reviewed and relevant studies were obtained.

After removing duplicates, the search yielded 180 results. The reference lists of articles were searched to identify additional relevant citations.

Study selection

Papers that contained data on the impact of the interventions adopted by participating Trusts (i.e., related with hot reporting of radiology imagining and/or with continuous quality improvement strategies (CQI)) to reduce missed fractures in patients attending A&E departments were considered

for this review. CQI is a term used in the literature to describe interventions that seeks to improve the provision of services by establishing a system to identify improvement needs, to set new processes to reduce failures and means to measure them, as well as educational and results measurement protocols. We included systematic reviews, randomised and non-randomised controlled trials, before-and-after studies and economic evaluations. Papers written in languages other rather than English or Spanish were excluded. These inclusion/exclusion criteria led to the selection of 3 papers (Preston et al, 1998 [28]; Espinosa and Nolan, 2000 [29]; Hardy et al, 2013 [30]) – 17 were selected based on title/abstract, of which 15 were rejected after reviewing full papers, and 1 additional article was included based on cited references of identified studies.

5.5.3. Search results

We present the characteristics and outcomes of interest reported in the selected papers in Table 1. We identified 2 studies that provided information on the effect of CQI strategies at A&E using a before-and-after design, and one paper that evaluated the effect of hot reporting based on a randomised control trial (RCT).

The interventions evaluated in the identified studies are similar to the interventions adopted by participating Trusts, however, there are might be differences as the specific characteristics of the CQI interventions and the measures to achieve hot reporting of radiology imaging may vary. Therefore, we acknowledge that the effect of the interventions identified in the literature might not necessarily correspond to the effect of the interventions put in place in the participating Trusts. In order to alleviate this limitation we have excluded from the literature review studies focusing on interventions not related to those applied in the participating Trusts, according to the description provided in the standardised reports (Appendix 4 and 5) of the interventions and actions that were implemented with the funds granted by the scheme. Also, we do not include the impact of one intervention evaluated in one of the identified studies (see intervention 2 in Espinosa, 2000 in Table 1) as we did not find an equivalent approach in the participating Trusts. It is also worth noting that some of the studies evaluated the impact among all interpretive errors of radiographs, not only missed fractures. While we are able to identify the effect corresponding to missed fractures only in some of the studies, others did not provide sufficient evidence to calculate the effect on missed fractures alone. Nevertheless, these studies stated that the most common errors were related to missed fractures.

The evidence from the identified studies indicates that both types of interventions (i.e., CQI and hot reporting) have a significant impact in reducing the number of missed fractures (i.e. reducing false negative cases). The effect is larger for interventions related with hot reporting. In Hardy et al, 2013 the number of false positive cases (i.e., patients diagnosed with a fracture when they did not have one) is also reported and evaluated, finding also a significant reduction in this outcome as well. We also point out that most of these findings were obtained under research conditions rather than real-world conditions, and that in only one study were the data from the UK. Hence, the generalisability of the findings to the present context should be treated with caution. We investigate this uncertainty more in sensitivity analyses. The effects estimated in these papers are used to populate a cost-effectiveness analysis of the interventions aimed at reducing the number of missed fractures in the participating Trusts, as we present next.

[28] Preston CA, Marr JJ, Amaraneni KK, Suthar BS. (1998). Reduction of “callbacks” to the ED due to discrepancies in plain radiograph interpretation. Am J Emerg Med. 16(2):160-162.

[29] Espinosa JA, Nolan TW. (2000). Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. Bmj.;320(7237):737-740.

[30] Hardy M, Hutton J, Snaith B. (2013). Is a radiographer led immediate reporting service for emergency department referrals a cost effective initiative? Radiography. 2013;19(1):23-27.

Table 1. Characteristics of selected publications and main results

Author, year	Country	Intervention	Outcome	Results	Notes
Preston, 1998	USA	**CQI intervention** (1) Review and discussion of clinically significant film discrepancies by ED physicians and radiologists within 24 hours, (2) periodic joint retrospective review by both departments of all interpretive discrepancies regarding frequency and type, (3) encouragement of ED physicians to have a lower threshold for consulting the radiologist for any questionable findings while the patient is in the ED, and (4) agreement by the radiologist to be more available to review films even after hours.	Patients undergoing radiographic studies in the ED who required further follow-up as a result of discordant radiograph interpretation	False negatives Before CQI: 30/13,200 = 0.22% After CQI: 37/31,680 = 0.12% RR = 0.514 95% CI = 0.318 to 0.832	This study included other final diagnoses, such as chest mass and pneumonia, but only the values related to fractures were used to calculate the effect of the intervention
Espinosa, 2000	USA	**Intervention 1 - CQI** (1) Creation of a file of clinically significant errors. (2) Mandatory study of the entire file becomes part of the orientation of all new staff. (3) Overall departmental patterns of error are identified from this file, and a focused review of these patterns take place at staff meetings **Intervention 2 - Rearrangement of X-ray interpretation by ED physicians & radiographers** All standard radiographs were to be brought directly to the ED for immediate interpretation. A radiologist would provide an interpretation within 12 hours as a quality control measure. When a clinically significant misinterpretation was found by the radiologist, staff from the ED would contact patients and ask them to return	The rate of clinically significant misinterpretation A false negative interpretation that would have resulted in a change in the patient's care	**Intervention 1** False negatives Before CQI: 3% (2.8%-3.2%) After CQI: 1.2% (1.03% - 1.37%) RR = 0.400 95% CI = 0.354 to 0.469 **Intervention 2** After rearrangement of X-ray interpretation: 0.3% (0.26%-0.34%)	Errors might include fractures, foreign bodies, and other misinterpretations of radiographs. The most common (but not specific data provided) are fractures. The effect of intervention 2 was not considered as this intervention was not implemented in any of the participating Trusts.
Hardy, 2013	UK	**Intervention - Hot reporting** *Immediate reporting arm:* Patients wait in the radiology department following radiographic examination while the report is generated. The X-ray report returned to the ED at the same time as the patient for the ED clinician to review alongside the images. *Delayed reporting arm:* Patients return to the ED following radiographic examination to await review of the images by the referring ED clinician. X-ray report are issued by the radiology department at a later time and returned to the ED as was standard practice at each site.	Concordance in the interpretation of radiographs between radiology and ED. False negative cases based on ED interpretive errors. False positive cases based on ED interpretive errors.	False negative Treatment arm: 1/752 = 0.13% Control arm: 12/750 = 1.6% RR = 0.083 95% CI = 0.011 to 0.647 False positive Treatment arm: 14/752 = 1.86% Control arm: 36/750 = 4.6% RR= 0.388 95% CI = 0.217 to 0.734	Include all interpretive errors, not only fractures

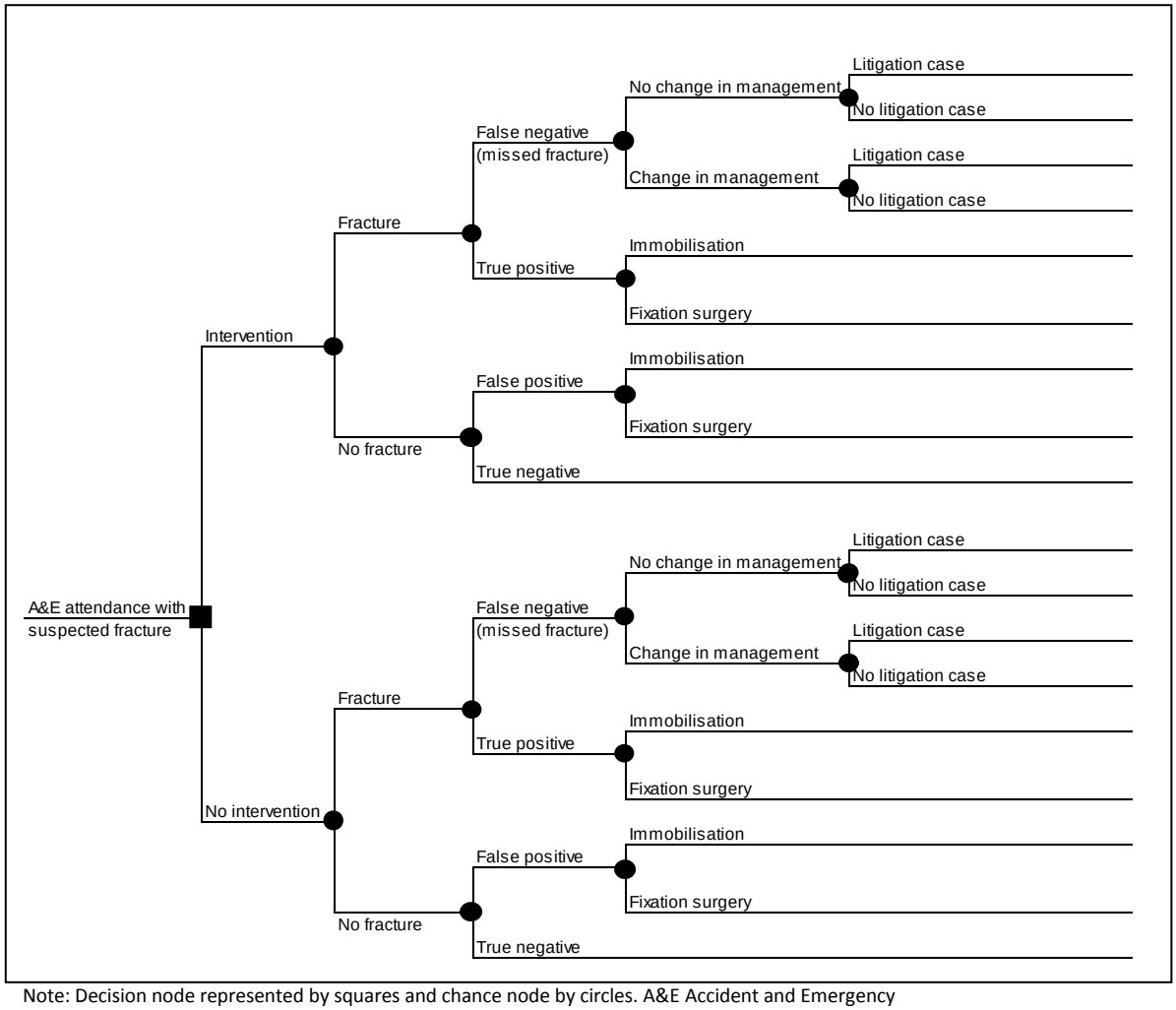
Note: ED = Emergency department; CQI = Continuous Quality Improvement; RR = Relative Risk;

5.6. Cost-effectiveness analysis

5.6.1. Model structure

The cost-effectiveness analysis was based on a decision analytical model. The model took the form of a decision tree shown in Figure 5. The structure of the tree is similar to that used in the cost-effectiveness analysis of imaging alternatives for suspected scaphoid fractures, undertaken in the recent NICE Guidelines on "Fractures (non-complex) assessment and management"[31] (NICE, 2016).

Figure 5. Structure of the decision tree



Note: Decision node represented by squares and chance node by circles. A&E Accident and Emergency

Under both the intervention and non-intervention arms, patients presenting at the A&E department with a suspected fracture might follow one of the following pathways: patients having a true fracture might be incorrectly diagnosed leading to a false negative case (missed fracture). Among these patients, some will require a change in management, and some cases will result in a litigation case. Patients with a fracture who are correctly identified and diagnosed will be treated accordingly to the severity of the fracture; with main treatment options consisting in immobilisation and fixation surgery. Among patients who do not truly have a fracture, some might be incorrectly diagnosed (i.e. false positive) and unnecessarily treated for a fracture. Patients that are correctly identified as not having a fracture (i.e., a true negative) will have no further consequences.

[31] National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management Fractures: diagnosis, management and follow-up of fractures. NICE Guideline NG38. February 2016

The possible pathways and the probabilities of each of the outcomes in the tree would be the same for patients under the intervention and non-intervention arm, with the exception of the probability of a false negative and a false positive case in the intervention arm, to which we applied the relative risks estimated by the studies identified in the literature review (see section 5.5).

5.6.2. Data

In order to populate the model the following set of parameters are required: 1) probabilities and relative risks; 2) health care resource use and unit costs associated to the diagnosis and treatment of suspected fractures; and 3) utilities associated with fracture-related outcomes and life expectancy values.

Probabilities and relative risks

Data on the probabilities and relative risks used in the model are summarised in Table 2. The probability of having a fracture among patients presenting at A&E departments with a suspected fracture was estimated using information on the number of x-rays plain films taken at A&E attendances during 2014-15, according to the data collected by Hospital Episode Statistics (HES)[32]; and the number of diagnosis of fractures recorded at Accident and Emergency Statistics[33] in the same year. We also considered the number of missed fractures based on the estimated rate of missed fractures at A&E departments.

Table 2. Probabilities and relative risks

Probabilities	**Value**	**SE**	**Source**
Prevalence of fractures among suspected cases	0.19693	0.00312	HES/Baker, 2016
False negative	0.01028	0.00079	Thomas, 1992
False positive	0.00443	0.00052	Thomas, 1992
Sensitivity	0.94780	0.00393	Thomas, 1992
Specificity	0.99448	0.00065	Thomas, 1992
Missed fractures requiring change in management	0.53293	0.03849	Thomas, 1992
Missed fractures leading to litigation case	0.02159	0.00534	Guly, 2001
Fractures treated with fixation surgery	0.20000	0.02821	NICE, 2016
Relative risks (RR)	**Value**	**Var ln(RR)**	**Source**
RR of false negative with CQI	0.40807	0.00480	Preston, 1998 Espinosa, 2000
RR of false negative with Hot Reporting	0.08311	1.08067	Hardy, 2013
RR of false positive with Hot Reporting	0.38785	0.09654	Hardy, 2013

Note: SE= Standard error; RR= Relative risks; HES= Hospital Episode Statistics

Therefore, the formula used to compute the prevalence of fractures among suspected cases is as follows.

$$Prevalence\ among\ suspected\ cases = \frac{Diagnosed + missed\ fractures\ at\ A\&E}{Number\ of\ xray\ plain\ films\ at\ A\&E}$$

[32] The Health and Social Care Information Centre, Hospital Episode Statistics for England, Accident and Emergency (A&E) statistics, 2014-15

[33] Baker C. Accident and Emergency Statistics: Demand, Performance and Pressure. Briefing paper umber 6964. November 2016. House of Commons Library.

This formula assumes that x-rays taken at A&E departments are primarily for the diagnosis of potential fractures; however, some patients might have an x-ray for a different reason, e.g., a potential chest infection. On the other hand, some patients with a fracture or a potential fracture might never have an x-ray undertaken during the A&E attendance. Nevertheless, this formula provides an approximation of the number of fractures among suspected cases, and a similar approach have been used in previous studies to estimate number of missed fractures at A&E departments (Lee and Bleetman, 2004)[34].

The literature review conducted to identify studies on the impact of the evaluated interventions also allowed us to identify studies that provided information of UK estimates on the underlying rates of false positive and false negative fractures at English A&E departments, as well as missed fractures requiring a change in management and the proportion of missed fractures that lead to a litigation case (Thomas et al., 1992 [35]; Guly, 2001 [36]). With respect to the proportion of fractures treated by a means of fixation surgery versus immobilisation, we applied the same assumption used in the aforementioned cost-effectiveness analysis conducted in the NICE guidelines on fractures[37]. These values are reported in Table 2.

The relative risks estimated from the literature review were used to evaluate interventions related with CQI and interventions related with hot reporting separately. Therefore, we provide two separate sets of results; 1) using the effect of the interventions among the two studies that evaluated CQI interventions which were combined by a means of a meta-analysis, and 2) using the effect estimated in the study that evaluated hot reporting. Note that while the former studies only estimated the effect on false negative cases, the latter also estimated the impact on false positive fractures.

Resource use and unit costs

The funds allocated by each of the participating Trusts to the interventions aimed at reducing missed fractures are presented in Table 3. In order to compute the cost on a per patient basis we considered the number of A&E attendances in each of these participating Trusts that are due to suspected fractures. This was computed considering the total number of attendances and the total number of x-rays taken at A&E departments, according to HES data[38]. Based on this information we compute that the cohort of patients attending A&E departments at the four participating Trusts with a suspected fracture in a given year is 100,957 patients. Considering this cohort and the total cost of the implemented interventions, the cost per patient is estimated in under £4 when we considered all the Trusts combined; the cost per patient varied from £1.5 to £9.6 among the four funded Trusts.

[34] Lee C, Bleetman A. (2004). Commonly missed injuries in the accident and emergency department. Trauma;6(1):41-51.

[35] Thomas HG, Mason AC, Smith RM, Fergusson CM. (1992). Value of radiograph audit in an accident service department. Injury;23(1):47-50.

[36] Guly HR. (2001). Diagnostic errors in an accident and emergency department. Emerg Med J.;18(4):263-269.

[37] National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management. Fractures: diagnosis, management and follow-up of fractures. NICE Guideline NG38. February 2016

[38] The Health and Social Care Information Centre, Hospital Episode Statistics for England, Accident and Emergency (A&E) statistics, 2014-15

Table 3. Cost per patient of interventions implemented by Trusts

Cost of interventions implemented by Trusts	**Value**	**Source**
Trust 9	£45,704	Trust report
Trust 4	£257,194	Trust report
Trust 10	£20,310	Trust report
Trust 28	£65,000	Trust report
Total cost (all Trusts combined)	***£388,208***	***Trust reports***
Patients attending A&E in each Trust a year		
Trust 9	124,007	HES
Trust 4	131,209	HES
Trust 10	68,129	HES
Trust 28	171,455	HES
Proportion of A&E attendances that are due to suspected fractures	0.204	HES
Patients attending A&E due to suspected fracture (all Trusts combined)	***100,957***	***HES***
Cost per patient attending A&E due to suspected fracture		
Trust 9	£1.81	Own calculation
Trust 4	£9.61	Own calculation
Trust 10	£1.46	Own calculation
Trust 28	£1.86	Own calculation
Cost per patient (all Trusts combined)	***£3.85***	

Note: HES = Hospital Episode Statistics. Trusts have been anonymised and numbered using the coding for the analysis

The unit costs of the health care items required for the management of patients are included in Table 4. We considered that every patient receives an x-ray and therefore, this cost is not included in the model. We assumed that missed fracture cases will have an additional ED visit, and if the patient requires a change in management this will include a fracture clinic visit and salvage surgery in the cases that lead to a litigation case, and immobilisation in the cases that do not lead to a litigation case. We applied the same unit costs as the ones used in the NICE guidelines cost-effectiveness model (NICE, 2016), most of which are based on NHS Reference Cost data. The cost of a litigation case was estimated as the mean total cost of claims notified between 2011 and 2016 for Sign up to Safety successful and unsuccessful Trusts where one of the injuries was 'Fracture' or 'Poor Outcome - Fractures etc' or where the incident details mention 'missed fracture' or 'misdiagnosis of fracture'.

Table 4. Unit costs of health care services and litigation costs

Unit costs	**Value**	**Source**
Cost of immobilisation	£10	NICE, 2016 – assumption
Cost of fixation surgery	£1,373	NICE, 2016 HRG: HA54Z (Day case), NHS Reference Costs
Cost of salvage surgery	£1,549	NICE, 2016 HRG: HA52Z (Day case), NHS Reference Costs
Cost of fracture clinic visit	£128	NICE, 2016 HRG: WF01B (Trauma and Orthopaedics)
Cost of ED visit	£120	NICE, 2016 HRG: WF01B (Accident and Emergency)

Effectiveness measure

Effectiveness was measured using QALYs. We used the same Quality of Life (QoL) values and similar assumptions regarding the duration of the reduced QoL effect on patient with identified and missed fractures than those applied in the cost-effectiveness analysis conducted by NICE (NICE, 2016). These values are presented in Table 5.

Table 5. Quality of life and life expectancy values

QoL weights	**Value (SE)**	**Source**
EQ-5D at 1-year post fracture	0.819 (0.020)	NICE, 2016 Mapped from PRWE scores from MacDermid,1998[39]
EQ-5D general population for 30 years old	0.930 (0.009)	NICE, 2016 – Kind, 1999[40]
Duration of effects	**Value**	**Source**
Duration of fracture-related QoL for identified fractures and missed fractures that do not lead to	1 year	NICE, 2016 - assumption
Duration of fracture-related QoL for missed fractures that lead to a litigation case	Lifetime	NICE, 2016 - assumption
Mean age at time of injury	30 years	NICE, 2016 - assumption
Mean age at death	80 years	NICE, 2016 - Interim life tables

Note: QoL= Quality of Life; SE= Standard error; PRWE= Patient Rated Wrist Evaluation

5.6.3. Sensitivity analysis

We conducted a one-way sensitivity analysis where we varied the following parameters: intervention cost per patient (we used the lowest and the highest costs estimated in each Trusts); the probability of a missed fractures (we used information provided by one participating Trust that allowed us to compute this parameter instead of the value identified in the literature); probability that a missed fracture leads to a litigation case (similarly, using data from the same participating Trusts); and the time horizon (assuming shorter time horizons of 5, 10 and 15 years, respectively). The remaining probabilities (i.e. prevalence of fracture, missed fractures requiring change in management; and proportion of fractures treated with fixation surgery) as well as the value of the relative risks in reducing false positive and false negative cases were also varied one at the time by applying a value equivalent to half and to double that used on the base case. We used the same approach to vary each of the treatment unit costs and the QoL values.

We also undertook a probabilistic sensitivity analysis using 1000 simulations in a Monte Carlo analysis to compute cost-effectiveness acceptability curves (CEACs). CEACs indicate the probability that an intervention is cost-effective for different values of the willingness to pay for an extra unit of outcome taking into account the overall uncertainty in the model parameters. For this we need to apply probability distributions to each of the parameters that depend on the nature of the parameter. Probabilities were characterized by a beta distribution. Resource use data inputs were characterized using a gamma distribution, while uniform distributions were applied to unit costs

[39] MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. (1998). Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma;12(8):577-86

[40] Kind P, Hardman G, and Macran S. UK population norms for EQ-5D. York. Centre for Health Economics, 1999. Available from: http://www.york.ac.uk/che/pdf/DP172.pdf

parameters; in both cases, we used upper and lower limits of 20% around the mean values. We used beta distributions to characterise the uncertainty around the utility values.

5.7. Results

Base case

The results of the analysis are presented in Table 6, showing the costs and QALYs for a cohort of 100,957 patients attending the A&E with a suspected fracture at the four participating Trusts in a given year. We present two separate sets of results; 1) using the effect of the interventions among the two studies that evaluated CQI interventions, and 2) using the effect estimated in the study that evaluated hot reporting.

Both interventions are found to reduce overall costs and to improve the health outcomes of patients; yielding in both cases to the conclusion that the intervention strategies dominate standard practice, i.e. the interventions are less costly and more effective compared to standard care. Cost savings in this annual cohort are estimated at £250,000 and £678,000 with CQI and hot reporting interventions, respectively. These values are equivalent to a mean cost saving per patient of £2.5 and £6.7. The number of total QALY gained are estimated in 18.2 and 34.2 for each type of intervention; equivalent to a QALY gain per patient of 0.0002 and 0.0003, respectively.

Table 6. Base case estimates of costs and QALYs for a cohort of 100,957 patients attending A&E with suspected fractures at the four participating Trusts in a year

	Intervention	**No intervention**	**Incremental (Intervention vs. No Intervention)**
CQI Intervention			
Costs	£6,573,333	£6,823,815	-£ 250,482
QALYs	2,276,154.5	2,276,136.4	18.2
Cost per QALY (Base case for CQI intervention)			Intervention dominates
Hot reporting intervention			
Costs	£6,145,311	£6,823,815	-£ 678,504
QALYs	2,276,170.6	2,276,136.4	34.2
Cost per QALY (Base case for hot reporting intervention)			Intervention dominates

Note: CQI= Continuous Quality Improvement; QALY= Quality-Adjusted Life Year.

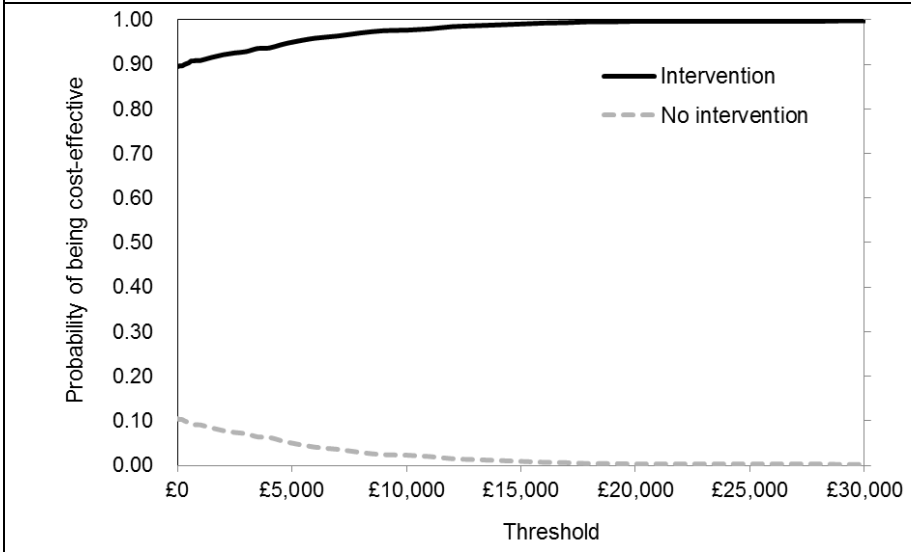
Sensitivity analysis

Our extensive one-way sensitivity analysis found very robust results in the case of the hot reporting intervention; in each of the analyses conducted the result indicated that the intervention was both more effective and less costly than standard care (results not shown). This was also the case in most of the analyses of the CQI interventions. However, in this case we found three instances where the intervention, while still leading to a better health outcome, was more costly than the standard case: 1) when the mean cost of a litigation claim was assumed to be half of that used in the base case (ICER estimated to be £4,997 per QALY gained); 2) when the relative risk for reducing missed fractures was half of that estimated in the identified papers (£33,638 per QALY gained); and when the cost of the intervention was assumed to be equal to the cost of the Trust found to have the largest cost (£18,232 per QALY gained).

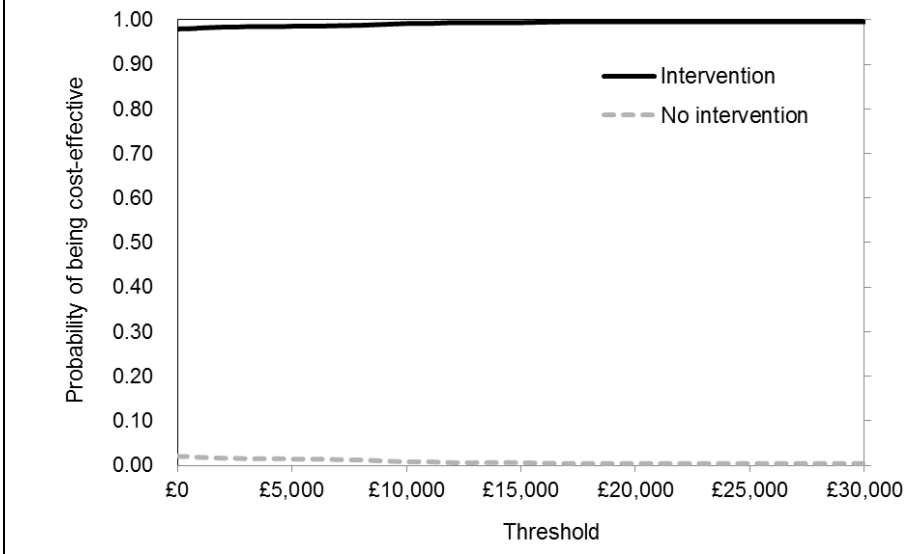
Figure 6 shows the CEACs computed based on the probabilistic analysis. At a threshold value of £20,000-£30,000 per QALY, the probability that the interventions are cost-effective compared with no intervention reaches over 99%.

Figure 6. Cost-Effectiveness Acceptability Curves

CQI intervention



Hot reporting intervention



5.7.1. Concluding remarks

In this evaluation of the Sign up to Safety financial incentives Scheme in A&E missed fractures we first identified the implemented interventions and developed a data collection questionnaire to gather information to estimate the impact of such interventions in reducing missed fractures in A&E. However, most Trusts were not capable of providing the required information, arguing in some cases that this data is not feasible to collect, as by definition the fractures of concern are missed. Therefore, Trust are required to enhance efforts to identify and record missed fractures at A&E departments in order to allow a proper evaluation of the impact of the interventions that are funded with the aim of reducing such incidents.

Given the difficulties in using the data from the Trusts to estimate the effect of the interventions funded by the Sign up to Safety scheme, we decided to review the scientific literature in order to estimate the effectiveness of these interventions based on the impact found in previous studies. We identified three relevant studies that evaluated interventions similar to the ones applied in the successful Trusts. However, there are might be differences between the interventions evaluated in these studies and those applied in the Trusts. Therefore, we acknowledge that the effect of the interventions identified in the literature might not necessarily correspond to the effect of the interventions put in place and funded by the scheme. This evaluation provides thus a tentative analysis of the potential effectiveness and cost-effectiveness of the interventions implemented in the Trusts, and the results should be treated with caution.

We found evidence of the effectiveness of these interventions in reducing the number of missed fractures, especially for intervention related to hot reporting, rather than those concerned with CQI. We then developed a decision analytical model to synthesise all available information into a cost-effectiveness analysis. We estimated that the intervention are not only more effective, but also less costly than standard care, yielding the conclusion that funding these types of interventions are a potential good value for money.

6. Evaluation of the Incentivisation scheme in maternity units

6.1. Overview of funding and claims

In total 56 Trusts from across a wide geographical area applied to the Sign up to Safety Incentive scheme for funding to reduce harms in maternity services. Out of these, 28 Trusts were successful in obtaining funds, while 28 were not. In brief, funds were given preferably to Trusts that proposed to improve intrapartum monitoring, staff improvement, training and to buy equipment (see below for further details).

In total, for improvements in maternity and intrapartum harm around £8 million was awarded by the NHS LA. The most expensive intervention was the purchase of K2 central monitoring equipment, which received £735,000 in funding, whilst the cheapest intervention was £340 for teaching materials. The large figure invested in the monitoring equipment can be understood due to its capital nature and considering that the expected life use of the equipment is between 5 and 10 years.

An indirect benefit of the scheme reported by the NHS LA was that NHS LA partnered with NHS Supply Chain to assist the maternity units in collectively procuring their equipment, and a saving of £36k was achieved from sales of £227k.

In Figure 7 we show the trend of claims for cerebral palsy registered between 1996-2016 related to incidents occurring between 1995-2015. The graph shows that in 15 years the number of claims has decreased over time, halving from 236 claims in 1996 to 104 in 2009, with a higher proportion of unsuccessful cases compared to the successful ones over time.

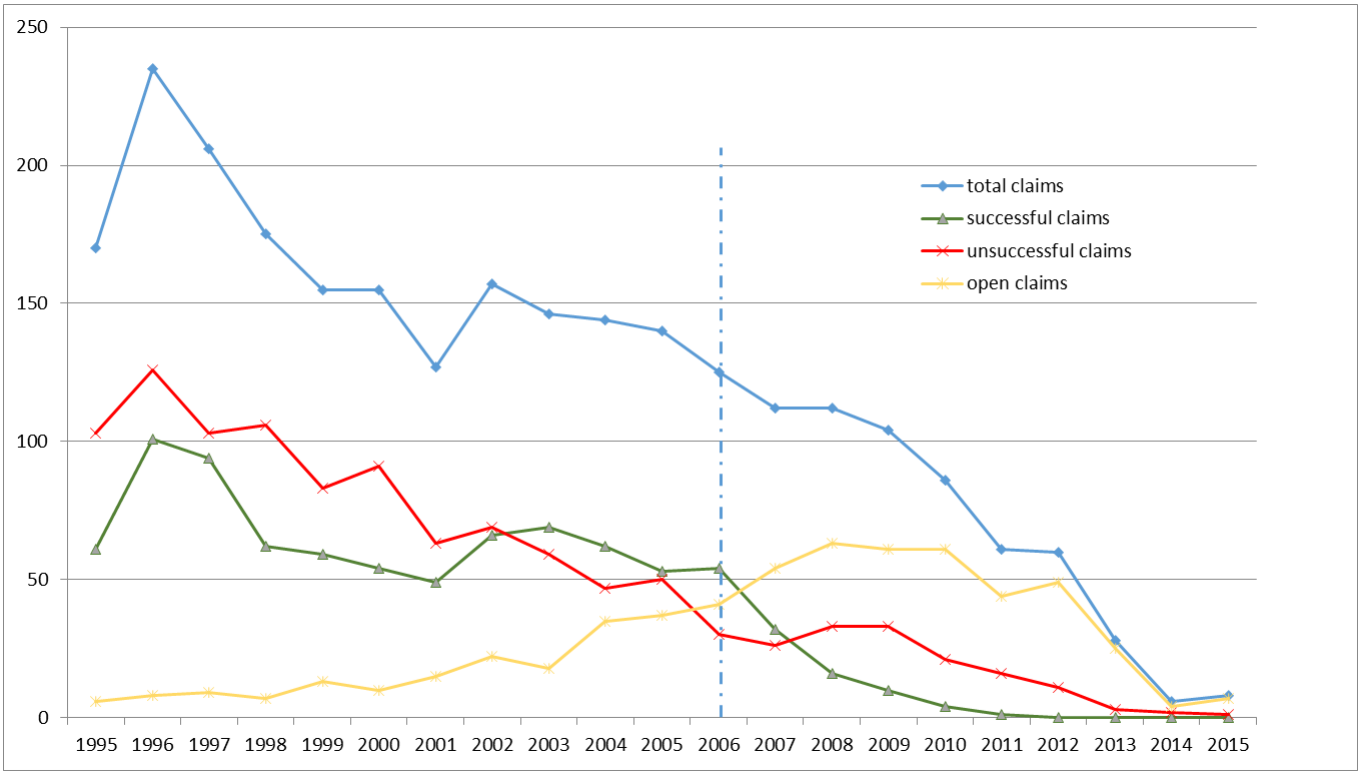
The time gap between the date of incident and the date of claim is on average 3.5 years and it has remained constant over time (at least until 2011). This means that the data in the last 4 years could still be incomplete, as there is still time for a claim to be made on incidents occurred in the past.

Looking more precisely at the trend of successful claims between 1995 and 2011 (when the time gap would suggest that no further claims should be registered for incidents occurring in that years), it seems they are decreasing from 2006. This could have been a disincentive factor for future claims and explain why the trend in claims is overall decreasing.

In addition to cerebral palsy other registered injuries are brain damage (49), blindness (4 cases) amputation (1 case), Erb's palsy (1 case) and uterine rupture (1 case).

Among the main reported causes of incident, 30% were due to failure to respond to an abnormal fetal heart rate (FHR) (767 cases out of 2512), 28% were due to failure to monitor 2nd stage labour (416 cases) and 1st stage labour (297 cases), 8% were due to a delay or failure in performing an operation (116 and 103 respectively), 7% were due to failure/delay in treatment (198 cases), and 6% were due to failure to recognise a complication (145). Other less common causes of claims include birth defects (65), failure of antenatal screening (51), failure to monitor dose of syntocinon (39), inappropriate use of forceps/failure to correctly apply forceps (32), failure to act on abnormal tests (27), failure to diagnose pre-eclampsia (22), failure/delay admitting to hospital (18) or failure to warn/obtain informed consent (12).

Figure 7. Trend of claims for cerebral palsy due to incidents occurred in 1995-2016



Source: analysis of NHS LA claims data

The average total cost of claims since 1995 is £2,716,684, including the unsuccessful claims (for which the costs have been paid by the claimants). A successful claim costs on average £4 million, but the cost can vary between £22,655 and £18million. The main part of the cost is represented by the cost of damage, estimated to be on average £3.6 million (but it can vary between £1,000 and £15.5 million), whereas the claimant cost is on average £258,000 (but can vary between £3,200 and £1.8 million) and the defence cost is on average £103,000 (with a maximum value of £778,750).

We also looked at the trend in claims for perineal tears but the numbers are very small: overall 41 claims have been registered between 1995 and 2014 with an average total cost of £89,229 for successful cases.

6.2. Descriptive analysis of interventions

First we asked the successful and unsuccessful Trusts to produce a report (Appendix 4) to understand what was the main problem they wished to address and how they were planning to address it through the scheme. In the same report we asked the Trusts to include some outcome measures that could help evaluate the effect of the interventions.

Eighty-three interventions were identified which can be grouped into the following categories:

- Cardiotocography (CTG)/intrapartum monitoring (fetal monitoring equipment)
- Staff
- Ultrasound equipment/software
- IT infrastructure
- Equipment (specialised obstetric equipment (episiotomy scissors), neonatal transport equipment)
- Training and development (mostly related to CTG interpretation)

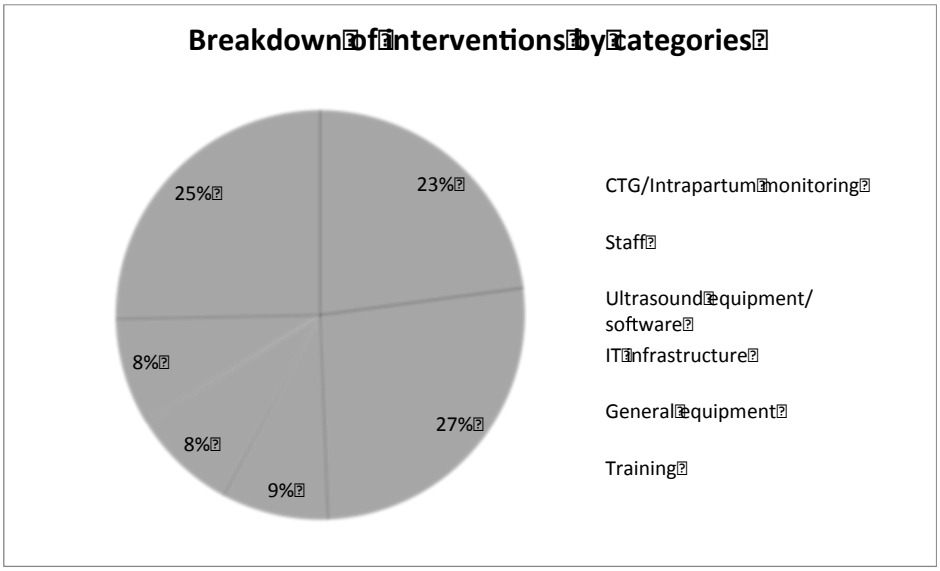
As maternity units often chose interventions in more than one of the above categories, and sometimes multiple interventions within categories, we have used the total number of interventions as the denominator, rather than total number of maternity units in the following descriptive analysis of the range of interventions proposed.

Between March 2015 and September 2016, 54% of the planned interventions were completely implemented and 25% only partially. However, 21% of the interventions were not implemented at all at the time we started the analysis.

General overview

Seventy-five percent of the interventions were focused on training, intrapartum monitoring and staff (Figure 8).

Figure 8. Breakdown of interventions by categories



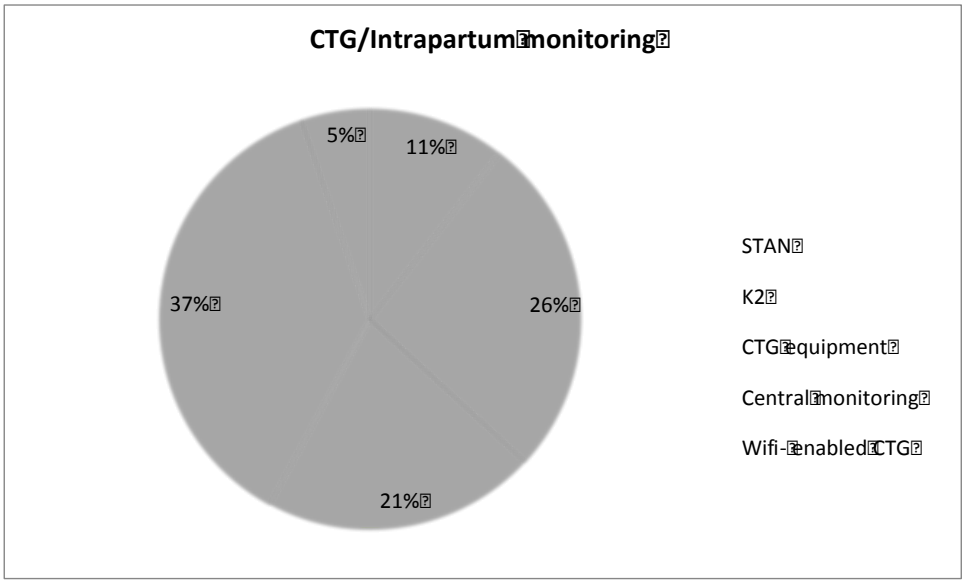
CTG/intrapartum monitoring

CTG and intrapartum monitoring represent the 27% of all interventions. In particular, central monitoring (37% over all CTG monitoring systems) was the most popular intervention within the CTG/intrapartum monitoring category (Figure 9). This figure is likely to be even higher, as the K2 Portal system can also be used as a form of central monitoring.

New staff

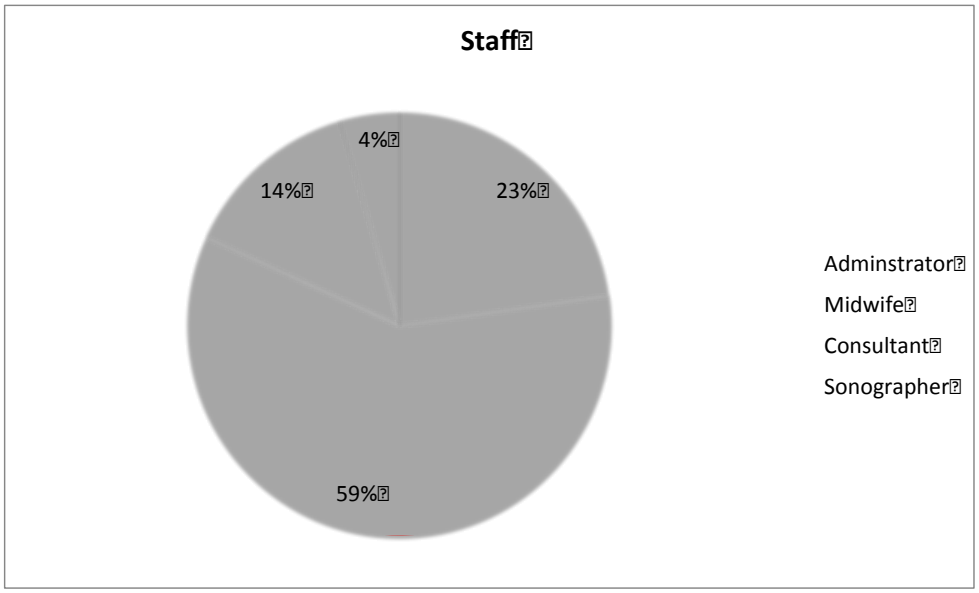
The 27% of all interventions in maternity units were related to recruitment of new staff. Midwives made up almost 60% of the staff that were funded. There was a variation in the job titles submitted by the units that decided to use the money for staffing. Only 27% of the roles reported had detailed job specifications and remits.

Figure 9. CTG Intrapartum monitoring interventions



Source: our data analysis

Figure 10. New staff



Source: our data analysis

Table 6. Job titles as described in the Trusts report

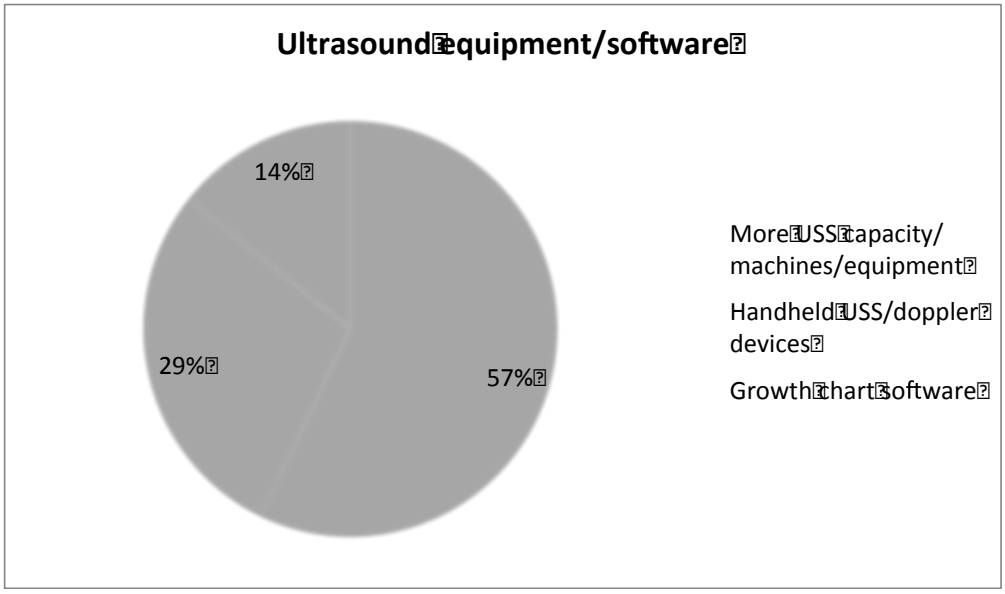
Different job titles	
Clinical Improvement Facilitator	Consultant
Intrapartum clinical practice educator	Fetal well-being midwife
Midwife	Project manager
Midwife lead	Admin support
Project lead	Sign up to safety campaign lead
Clinical champion	Human factors midwife
Management consultant	Specialist midwife for safer and active birth

Source: our data analysis

Ultrasound equipment/software

Of the interventions aimed at purchasing ultrasound equipment or software, more than half (57%) were on ultrasound scan (USS) capacity, machine and equipment, 29% on hand held USS and Doppler and 14% on growth chart software.

Figure 11. Ultrasound equipment

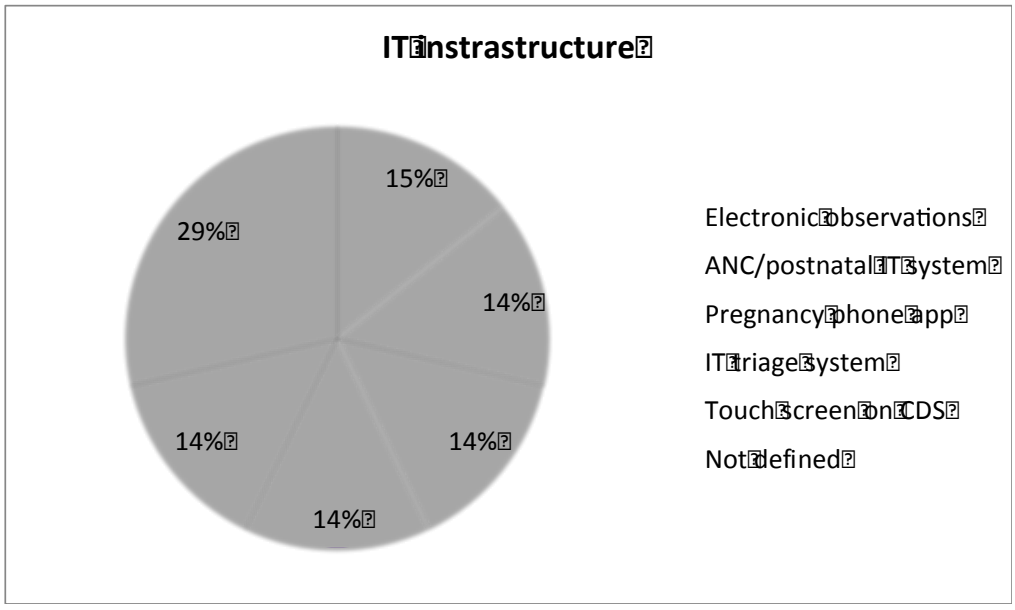


Source: our data analysis

IT infrastructure

Interventions on IT structure included electronic observations, ANC/postnatal IT system, electronic triage or pregnancy phone app or touchscreen on CDS.

Figure 12. IT infrastructure

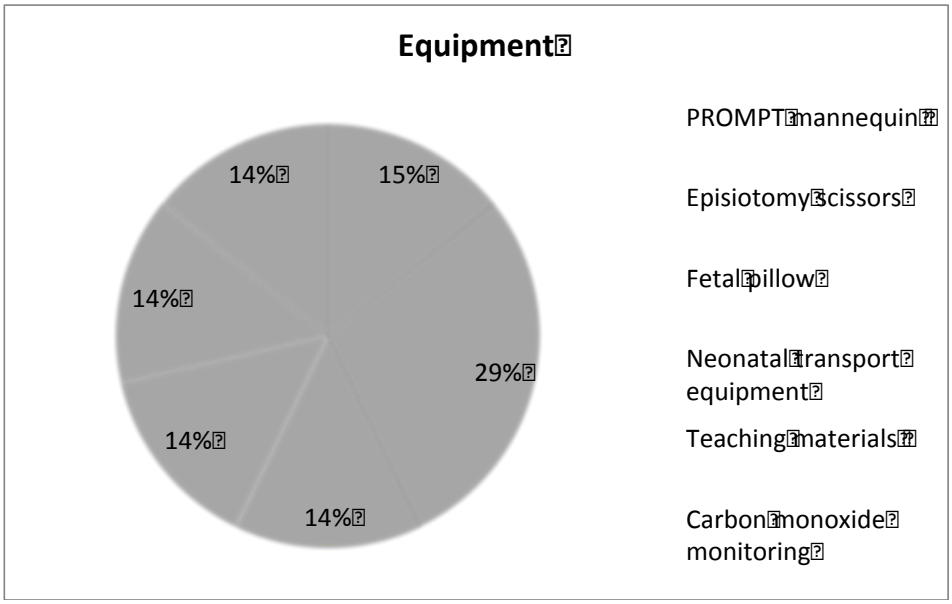


Source: our data analysis

Equipment

8% of all interventions were aimed at purchasing new equipment. The most popular equipment were the episiotomy scissors and the PROMPT mannequin. Other types of equipment include fetal pillows, neonatal transport equipment, teaching materials and carbon monoxide monitoring.

Figure 13. Equipment

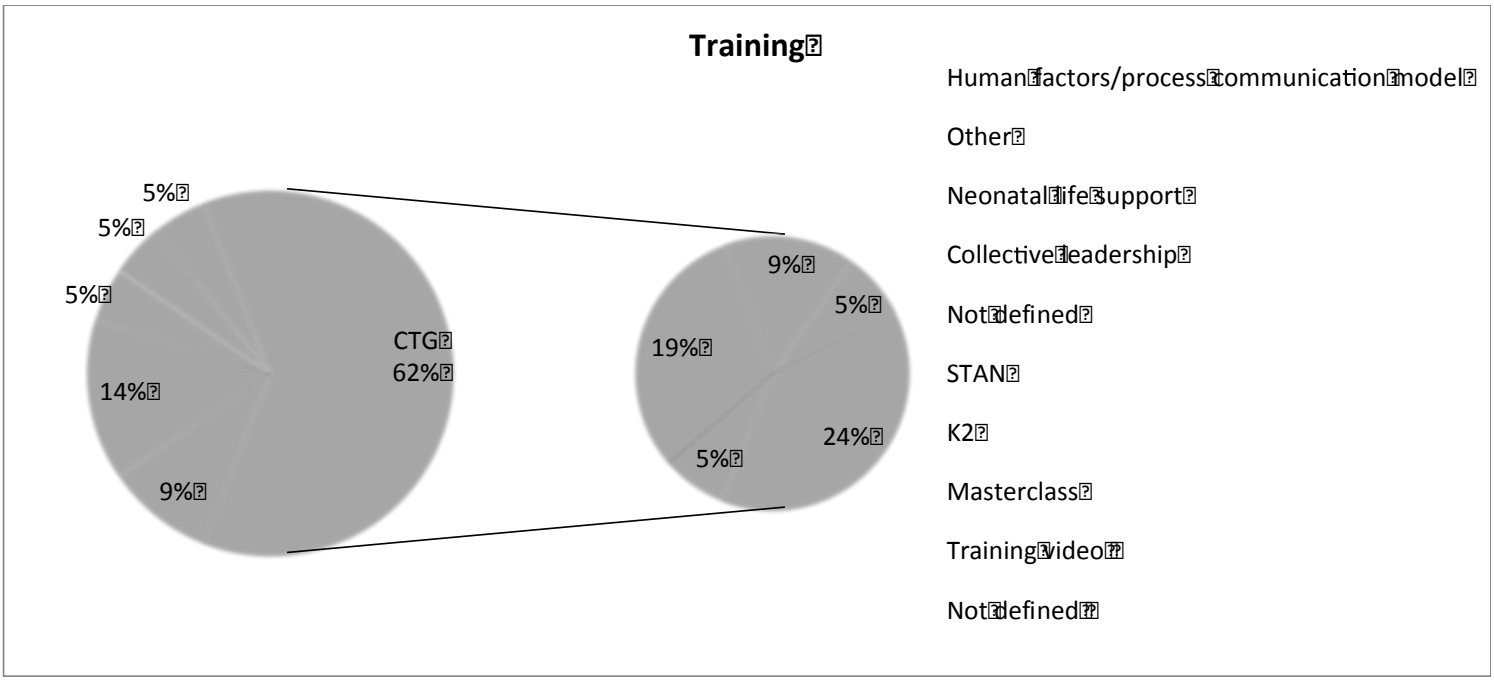


Source: our data analysis

Training

CTG interpretation was the most commonly chosen form of training, representing 62% of this category. Whilst established CTG training packages such as STAN, K2, and the CTG master class, make up around 50% of the CTG training, 38% was not clearly defined.

Figure 14. Training



Commentary on the interventions selected

The aim of the Sign up to Safety incentive funding scheme was to reduce harm and therefore litigation. Issues with CTG interpretation represent 15% of the total value of a decade of maternity claims in England (Anderson, 2013)[41] and there is a national ambition to reduce stillbirth and intrapartum asphyxia (National Maternity Service Review, 2016)[42]. It therefore makes sense that almost 25% of the interventions were focused on intrapartum monitoring, and 62% of the training on CTG interpretation.

However, the successful maternity units have invested their money in a broad range of interventions. Such diversity could suggest that there are no clear-cut solutions or obvious choices. Alternatively, this could be due to maternity units identifying different areas for improvement.

Four different forms of CTG training have been defined but 38% of CTG training was not clearly defined. Training is not always effective and can sometimes be harmful (Draycott et al, 2015).[43] Evidence of clinical benefit therefore has an important role in guiding and signposting towards the most effective interventions. There is currently a lack of evidence base for the CTG training programmes that have been adopted and this could explain the 'scattergun' approach.

With regards to funding for new staff, both the range of job titles described and the lack of clarity in their responsibilities illustrate the lack of detail about roles and responsibilities.

Another important finding was that 46% of the units (13/28) had not yet fully implemented or started introducing their interventions at the time of the final reports. This may mean that the impact of the scheme may be underestimated, however, even if all the maternity units implemented all of their interventions by the time of the reports, it is unlikely that there will be much change after only one year. One training programme allowed one year for training and then observed improvements over two years after the training was introduced (Draycott et al., 2006)[44]. The Victorian Managed Insurance Authority (VMIA) mandated the same training programme in Victoria, Australia and provided financial support to maternity units for implementation. In contrast to before, an initial evaluation demonstrated a significant reduction in one outcome (Apgar <7 at 1 minute) during the year of intervention (Shoushtarian et al., 2014)[45]. Finally, there were improvements in outcomes reported after shoulder dystocia in 4 years (Draycott et al., 2008)[46], but outcomes continued to improve for up to a decade after the introduction of training (Crofts et al., 2015)[47].

---

[41] Anderson A. (2013). Ten years of maternity claims: an analysis of the NHS Litigation Authority data – key findings. Clinical Risk;19(1):24-31.

[42] National Maternity Services Review. Better Births. Improving outcomes of maternity services in England. London; 2016.

[43] Draycott TJ , Collins K J, Crofts JF, Siassakos D, Winter C, Weiner CP, et al. (2015). Myths and realities of training in obstetric emergencies. Best Pract Research Clin Obstet Gynaecol;29(8):1067-76

[44] Draycott T, Sibanda T, Owen L, Akande V, Winter C, Reading S et al. (2006). Does training in obstetric emergencies improve neonatal outcome? BJOG: An International Journal of Obstetrics &Gynaecology. 113(2):177-82.

[45] Shoushtarian M, Barnett M, McMahon F, Ferris J. (2014). Impact of introducing Practical Obstetric Multi-Professional Training (PROMPT) into maternity units in Victoria, Australia. BJOG: An International Journal of Obstetrics & Gynaecology121(3):1710-8.

[46] Draycott T, Crofts J, Ash JP, Wilson LV, Yard E, Sibanda T, et al. (2008). Improving neonatal outcome through practical shoulder dystocia training. Obstet Gynecol.;112(1):14–20.

[47] Crofts J, Lenguerrand E, Bentham GL, Tawfik S, Claireaux HA, Odd D, et al. (2015).Prevention of brachial plexus injury-12 years of shoulder dystocia training: an interrupted time-series study. BJOG: An International Journal of Obstetrics & Gynaecology;123(1):111–8.

These 46% could also be considered as 'slower starters' and may have benefited from additional support.

There have recently been a number of reports from well designed, robust evaluations of national safety interventions abroad, none of which have been associated with improved clinical outcomes.

TOSTI Study (Netherlands)
A multicentre, cluster randomised controlled trial investigated whether simulation-based obstetric team training in a simulation centre improves patient outcomes (Fransen et al, 2016) [48]. The study was based on MOET (Managing Obstetric Emergencies and Trauma) course (Howell et al, 2007) [49]. 24 units were randomised to interventions and control combining a total of 28,657 women. In total 471 medical professionals received the training. The results show no improvements in clinical outcomes and the training did not reduce obstetric complications.

National Perinatal Safety programme (Sweden)
The programme involved all 46 obstetric units in Sweden. The programme, based on peer review process and local implementation of guidelines was initiated in 2008 and included a web-based fetal monitoring programme. A study conducted to evaluate the impact and effects of the national programme to improve safety for the new-borns show no significant improvement in outcomes, no change in Apgar score lower than 7 at 5 minutes and a doubled risk of incautious management of oxytocin (Luthander et al., 2016) [50].

CTG education programme (Denmark)
A national study conducted in Denmark with historical controls over 331,282 births tested new national CTG programme. Overall 53 courses were developed and 97% of maternity carers trained. The analysis of the impact of the training programmes found no significant effect, with no change in Apgar score lower than 7 at 5 minutes and reduced operative vaginal birth rates by 14% (Tellesen, 2016).[51]

The Sign up to Safety (SU2S) incentivisation scheme devolved the identification and choice of interventions down to unit level to local Trust clinical leaders, instead of making top-down recommendations, e.g., at the national level.

Review of the evidence base for the selected interventions
The recent NHS England National Maternity Services review[52] observed: "*…….any training undertaken must have been proven to be effective in improving outcomes or other aspects of quality, and its impact monitored locally".* In particular, we should endeavour to put women and their families at the centre of these programmes to ensure that staff are trained in a way that improves outcomes, and uses finite resources in a useful way.

[48] Fransen AF, van de Ven J, Schuit E, van Tetering A, Mol BW, Oei SG. (2016). Simulation-based team training for multi-professional obstetric care teams to improve patient outcome: a multicentre, cluster randomised controlled trial. BJOG: An International Journal of Obstetrics & Gynaecology; 123(11):1753-60.
[49] Howell C, Grady K, Cox C. (2007). Managing Obstetric Emergencies and Trauma. The NMOET Course Manual. 2nd edition. Cambridge University Press
[50] Luthander CM. (2016). The national perinatal patient safety programme: the challenges of implementation and evaluation. BMC Health Serv Res.;12(1):274.
[51] Tellesen L, 2016. PhD thesis. Submitted to University of Copenhagen
[52] National Maternity Services Review. Better Births. Improving outcomes of maternity services in England. London; 2016.

Cardiotocography (CTG)/ Intrapartum monitoring

Cardiotocography (CTG), or electronic recording of the fetal heart rate, is a way of assessing fetal wellbeing during labour and despite the limited evidence of clinical benefit, CTG monitoring is widely used as a screening tool for fetal distress/hypoxia in labour. Failures related to fetal monitoring are reported in all countries who routinely use it and there have been many attempts to reduce the error rate to improve outcomes.

CTG monitoring is known to have a low sensitivity for predicting intrapartum fetal hypoxia, meaning that an abnormal CTG trace does not always indicate that fetal hypoxia is present (NICE, 2007)[53].

ST analysis (STAN) can be used in combination with CTG monitoring to help improve the sensitivity and detect fetal heart ischemia. STAN software can analyse changes to the ST or T waves of the fetal electrocardiogram (ECG) that may suggest fetal heart hypoxia/ischemia. A recent Cochrane review concluded that when compared to electronic fetal monitoring alone, the use of adjunctive STAN demonstrated no improvements in numbers of babies with severe metabolic acidosis at birth or babies with neonatal encephalopathy (Neilson, 2006)[54]. There was also no improvement in the numbers of babies with low Apgar scores at 5 minutes (Olofsson et al., 2014)[55].

Staff

A number of Trusts opted to recruit more midwives or consultant obstetricians. There are some data that have demonstrated that higher numbers of midwives per births and a higher ratio of consultant obstetricians to midwives were associated with a lower probability of postnatal readmissions to hospital (Gerova et al., 2010)[56]. We did not investigate this outcome but it might be a useful measure in future evaluations. However, it is unlikely that funding would have significantly improved the staff/birth ratio and moreover increasing consultant presence alone does not appear to improve perinatal outcomes (Knight et al., 2016)[57].

Many Trusts employed personnel in a quality improvement capacity. Both the range of job titles described and the lack of clarity in their responsibilities illustrate the problems of 'work-as-imagined versus work-as-done'. These non-specific roles represent a slightly aspirational 'work-as-imagined', or work that should happen according to those completing the applications compared to what actually needs to happen (or the 'work-as-done') on the front line to promote improvement (Braithwaite et al., 2015)[58].

Finally, there is also a risk that those assigned to quality improvement roles may not have the appropriate skills, influence and resources to initiate the necessary changes for improvement. As a result, they may miss the problems that need addressing and instead focus on temporary solutions (Dixon-Woods and Martin, 2016)[59].

---

[53] Intrapartum Care. NICE Clinical Guideline. 2007;55(1):1–65.

[54] Neilson JP. Fetal electrocardiogram (ECG) for fetal monitoring during labour. The Cochrane Library. 2006.

[55] Olofsson P, Ayres-de-Campos D, Kessler J, Tendal B, Yli BM, Devoe L. (2014). A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part II: the meta-analyses. Acta Obstet Gynecol Scand.;93(6):571–86.

[56] Gerova V, Griffiths P, Jones S, Bick D. (2010). The association between midwifery staffing and outcomes in maternity services in England: observational study using routinely collected data. Preliminary Report and Feasibility.

[57] Knight HE, van der Meulen JH, Gurol-Urganci I, Smith GC, Kiran A, Thornton S, et al. (2016). Birth "Out-of-Hours": An Evaluation of Obstetric Practice and Outcome According to the Presence of Senior Obstetricians on the Labour Ward. Myers JE, editor. PLoS Med.;13(4):e1002000–15.

[58] Braithwaite J, Wears RL, Hollnagel E. From Safety-I to Safety-II: A White Paper. University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia; 2015 Aug pp. 1–43.

[59] Dixon-Woods M, Martin GP. (2016). Does quality improvement improve quality? Future Hosp J. Royal College

Episiotomy scissors

Two Trusts purchased specialised episiotomy scissors that are designed to help achieve an episiotomy at 60 degrees from the perineal midline, thereby reducing the incidence of severe perineal tears. However, there are very few data supporting their use. The current RCOG guideline[60] recognises that these scissors may improve the angle of episiotomy, but that there are conflicting data about the protective effect of episiotomy and there are no data supporting the use of these scissors in current practice to reduce severe perineal tears.

Training

Four different forms of CTG training were chosen but 38% (5/13) of CTG training was not clearly defined. Training is not always effective and can even sometimes be harmful (Draycott et al, 2015)[61].

There is currently almost no evidence for the CTG training programmes that have been adopted. One study evaluating the K2 interactive computer-based training package demonstrated that the programme improved participant knowledge, but there was no assessment of outcomes (Beckley et al., 2000)[62]. The recent national Danish study (Tellesen, 2016) [63] did not demonstrate any improvement in outcomes after CTG education after training 97% of maternity staff.

There is a plethora of other data, including some large randomised trials, related to different elements of CTG use published recently: standardisation of CTG assessment (FIGO guidelines), the Infant study, the use of the Sis-Porto system and also the Swedish national perinatal safety programme. None of these studies demonstrated any clinically significant improvements in perinatal outcome despite the different approaches taken: standardisation of assessment (FIGO), computerised decision support (Infant & Sis-Porto) and a unit level 'human' intervention in Sweden.

Improving outcomes is likely to be more complex than CTG interpretation alone, and certainly more complex than some form of knowledge transfer. Other programmes with positive outcomes have employed cognitive aids (stickers), learning in communities of practice and normalisation process theory, all of which is likely to be required for improvement (National Maternity Service Review, 2016)[64].

This is reinforced in a recent editorial on the negative results of a "skills and drills" intervention in India that proposed some possible explanations, particularly the need to recognise behavioural or organizational barriers related to hierarchical structures, roles, and team formation (Ricca, 2016)[65].
The current lack of signposting to the evidence base for CTG training could explain the Trusts' 'scattergun' approach and their valorising overly simplistic 'solutions'.

---

of Physicians;3(3):191–4.

[60] Royal College of Obstetricians & Gynaecologists. Green-top Guideline No.29. The Management of Third- and Fourth- Degree Perineal Tears. London: RCOG; 2015.

[61] Draycott T, Collins KJ, Crofts J, Siassakos D, Winter C, Weiner CP, et al.(2015). Myths and realities of training in obstetric emergencies. Best Practice & Research Clinical Obstetrics and Gynaecology.

[62] Beckley S, Stenhouse E, Greene K. (2000). The development and evaluation of a computer-assisted teaching programme for intrapartum fetal monitoring. BJOG.;107(9):1138–44.

[63] Tellesen L, 2016 PhD thesis. Submitted to University of Copenhagen

[64] National Maternity Services Review. Better Births. Improving outcomes of maternity services in England. London; 2016.

[65] Ricca J. (2016). Limits of "Skills And Drills" Interventions to Improving Obstetric and Newborn Emergency Response: What More Do We Need to Learn? Global Health: Science and Practice; 4(4):518-521.

Training is a complex intervention, for an even more complex system, but at its simplest: accoucheurs should ensure they use evidence-based training programmes to help them provide the best possible care and outcomes for mothers and babies.

Summary

Most of the programmes were based on training, or training was a significant part of the interventions chosen. More and better training has been an almost ubiquitous recommendation for almost two decades. Robust evaluation using scientifically rigorous study designs is essential because training for obstetric emergencies, however well intentioned, is not cheap (Yau et al., 2016)[66] and nor is it always associated with improvements in clinical outcomes (Draycott et al, 2015)[67] as we have demonstrated.

In particular, isolated staff training for fetal monitoring was not successful in any of these national programmes.

In 2016 a review of all the 23 obstetric emergencies training programmes that investigated clinical outcomes were more positive still in their conclusion that: '*…..training…. can improve quality of life and save lives*' (Bergh, 2015)[68]. However, the authors recognised that not all training was associated with improvements in outcome and training should be locally based in the maternity unit.

Most recently, a commentary on the Netherlands trial observed: "*Currently, the evidence supports local, multi-professional training, with integrated clinical and teamwork/human factors elements, for all staff annually"* (Draycott, 2016)[69].

This model of training has been very successful in Australia where a project supported by the Victorian Managed Insurance Authority (VMIA) has been associated with improvements in clinical outcome (Shoushtarian et al., 2014)[70] as well as a parallel reduction in claims, sufficient for the VMIA to return funding to the State health service.

[66] Yau CWH, Pizzo E, Morris S, Odd DE, Winter C, Draycott TJ. (2016).The cost of local, multi-professional obstetric emergencies training. Acta Obstet Gynecol Scand.; 95(10):111-9.
[67] Draycott T, Collins KJ, Crofts J, Siassakos D, Winter C, Weiner CP, et al. (2015). Myths and realities of training in obstetric emergencies. Best Practice & Research Clinical Obstetrics and Gynaecology; 29(8):1067-76.
[68] Bergh A-M, Baloyi S, Pattinson RC. (2015). What is the impact of multi-professional emergency obstetric and neonatal care training? Best Practice & Research Clinical Obstetrics and Gynaecology; 29(8):1028-43.
[69] Draycott T. (2016). Not all training for obstetric emergencies is equal, or effective. BJOG: An International Journal of Obstetrics & G. (in press).
[70] Shoushtarian M, Barnett M, McMahon F, Ferris J. (2014). Impact of introducing Practical Obstetric Multi-Professional Training (PROMPT) into maternity units in Victoria, Australia. BJOG: An International Journal of Obstetrics & Gynaecology;121(13):1710-8.

6.3. Methods

In this section we describe the methods adopted for the evaluation of the Sign up to safety scheme in maternity units. More specifically we will clarify which approach has been adopted, how costs and outcomes have been assessed, and we will explain why we ended up running a cost-consequence evaluation instead of a cost-effectiveness analysis as described in section 4.

Overview

In order to carry out an economic analysis of the scheme, we calculated the incremental costs and effects in successful Trusts who were awarded funding via the scheme, compared to the control group of Trusts who were not awarded funding via the scheme, using the DiD approach mentioned in section 4 .

Further details of the evaluation and the literature review are provided in Appendix 10.

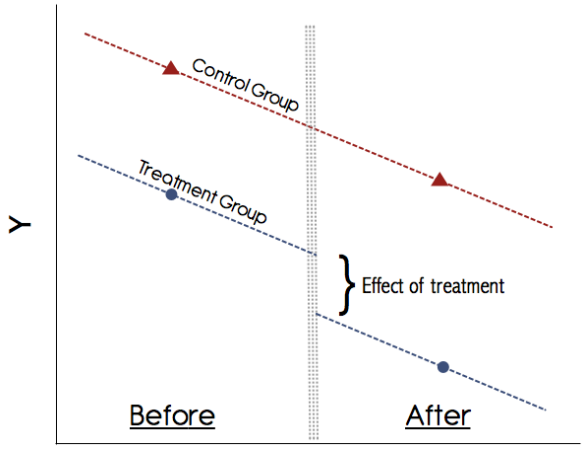
6.4. Model

Difference-in- differences analysis

The difference-in-differences (DiD) analysis requires data from the intervention group (i.e. successful (funded) Trusts, abbreviated to "Y" in this section) and the control group (unsuccessful (unfunded) Trusts, abbreviated to "N" in this section) in order to test whether the intervention (i.e. implementing any specific interventions, or receiving funds for the SU2S incentive funding scheme) has had an effect (Figure 15). We also need data from each group from time periods before and after the "boundary date" when the intervention is deemed to have come into effect in order to control for any underlying differences between the two groups, or for any effects of external influences that might take place at certain dates. This allows us to clarify whether any effect seen is in fact due to underlying changes with time that would have taken place regardless, or whether it is really due to receiving money from the NHS LA as part of the SU2S scheme.

Figure 15. Schematic of the Difference in Difference analysis design.

Point estimates of the numbers of incidents in the before and after periods are compared, with an adjustment for the difference between control and intervention/treatment groups.



A boundary date must therefore be chosen which separates the "before" and "after" time periods. We have two possible dates: the date on which the first of the specific interventions was implemented in each Trust (information taken from the reports sent to us by the Trusts; "intvn date"), and the date on which monies were received by the Trust from the NHS LA (information given to us by NHS LA, i.e. 3 working days after the money left the NHS LA account; "BACS date").

These two dates are different for each of the successful Trusts, meaning that the plots in our results are not as neat as the example above, as the before and after periods for each Trust do not necessarily correspond to the same calendar periods. We examined the results obtained using each of the two dates to check if the overall results are robust to the choice of boundary date.

We received Trust-level aggregated data, by calendar month, from 44 Trusts. This imposes some limitations on the conclusions that we can draw as it is not patient-level data, meaning that for example it is not possible to know what proportion of the mothers suffering perineal tears did in fact have instrumental assistance during birth and what proportion did not. Also, we do not know the proportion of babies admitted unexpectedly to NICU that were also cooled, or who also had a low Apgar score at 5 minutes. We only have the overall proportions for each Trust by month, and no further associations between variables. In order to account for the natural variation between Trusts, we have considered this dataset to be a panel dataset, meaning that each Trust is seen as an individual, which produces outcome values per month. We note also that a sample size of 44 is fairly small, so this is likely to limit the power of the analysis.

The variables which are thought might vary as a result of the interventions implemented are given in Table 7 below, along with the expression by which they are calculated.

Table 7. Outcome variables that might be affected by the specific maternity interventions implemented (and/or by signing up to the scheme itself), and expressions used to calculate them.

Variable	Expression
Proportion of term singleton stillbirths	No. term singleton stillbirths ÷ No. term singleton births
Proportion of term singleton newborns with Apgar score<7 at 5 minutes	No. term singleton newborns with Apgar<7 at 5 mins ÷ (No. term singleton births - No. term singleton stillbirths – No. term singleton babies whose Apgar not known/not recorded at 5 mins)
Proportion of term singleton babies therapeutically cooled after birth	No. term singleton babies therapeutically cooled ÷ (No. term singleton births - No. term singleton stillbirths)
Proportion of term singleton newborns admitted unexpectedly to NICU	No. term singleton babies unexpectedly admitted to NICU ÷ (No. term singleton births - No. term singleton stillbirths)
Proportion of mothers with 3rd degree perineal tears (excluding Trusts that only reported combined figures)	No. 3rd degree tears ÷ (No. mothers delivered – No. mothers delivered via Caesarean section)
Proportion of mothers with 4th degree perineal tears (excluding Trusts that only reported combined figures)	No. 4th degree tears ÷ (No. mothers delivered – No. mothers delivered via Caesarean section)
Proportion of mothers with 3rd or 4th degree perineal tears (including all Trusts)	No. 3rd or 4th degree tears ÷ (No. mothers delivered – No. mothers delivered via Caesarean section)

Therefore, the variables for which we have collected data, by Trust and by month, are (note that "term" indicates gestation ≥37 weeks):

- Total number of all births (i.e. all babies)
- Total number of mothers delivered of any birth
- Number of instrumental vaginal births
- Number of Caesarean sections
- Number of mothers with 3rd degree perineal tears
- Number of mothers with 4th degree perineal tears
    - Note that for three Trusts this information was given as a combined number of 3rd and 4th degree tears due to reporting restrictions, so we have performed sensitivity analysis where all 3rd and 4th degree tears were combined into a single variable for all Trusts, allowing inclusion of those three Trusts' data
- Number of singleton births (live and stillbirth)
- Number of term singleton births (live and stillbirth)
- Number of singleton stillbirths
- Number of term singleton stillbirths
- Number of term singleton newborns with Apgar score<7 at 5 minutes
- Number of term singleton newborns with Apgar score not known at 5 minutes
    - Note that this variable was requested as if babies do not have their Apgar score recorded we cannot assume that the unknown Apgar score is ≥7, so the number of unknown scores was subtracted from the denominator when calculating the proportion of babies with a low Apgar score (<7) at 5 minutes
- Number of term singleton babies therapeutically cooled after birth
- Number of term singleton newborns admitted unexpectedly to NICU (not for e.g., congenital malformations or social reasons)

Data collected for certain other variables around the outcomes required for analysis (e.g., total all births) were used to perform sanity checks to test the data and ensure that each dataset was coherent and contained the information that we had requested. Different Trusts' reporting systems filtered the data in different ways with different assumptions, so it was important to test this. Not all Trusts could provide data for all the variables, or for all months, particularly before around 2013.

The rate of Caesarean sections is not expected to change as a result of the interventions, but these figures are required to calculate the correct denominator for the proportion of mothers with tears, who have had vaginal births. We assume that mothers who have had Caesarean deliveries are not at risk of tears. Similarly, the rate of instrumental births is not expected to change with the interventions, but we investigated controlling for this variable when considering rates of tears.

We combined all the Trusts' data into a single panel data set, where each Trust was considered to be an individual, which reported various outcomes every month. We performed a multi-level analysis looking at the difference in differences between successful and unsuccessful Trusts in the period after the boundary date compared to before. The numbers calculated as part of this analysis were the following:

- Raw proportions: Average differences in proportions of each outcome separately before and after the intervention, in each of the successful and unsuccessful groups (i.e. four groups).
    - We calculated the point estimate and the 95% confidence intervals[71] for these values.
- Difference (Y): The difference in average differences in proportions of each outcome in the successful (Y) group, before and after the boundary date (after minus before)
    - We calculated the point estimate, and estimated the 95% confidence intervals for these values using the 95%CIs of the four separate results above
- Difference (N): The difference in average differences in proportions of each outcome in the unsuccessful (N) group, before and after the boundary date (after minus before)
    - We calculated the point estimate, and estimated the 95% confidence intervals for these values using the 95%CIs of the four separate results above
- Difference in differences: The difference in these two differences. This is the main DiD result.
    - We calculated the point estimate, and estimated the 95% confidence intervals for this value using the various 95%CIs of the results above

In order to provide meaningful numbers, these proportions were then converted into rates per 100,000 patients (i.e. per 100,000 mothers or per 100,000 babies), and we report the point estimate and 95% confidence intervals for these figures.

Using these rates or outcomes per 100,000 patients, we were able to calculate the consequences of the money given to Trusts as part of the Sign up to Safety scheme, in terms of the costs of extra outcomes arising, i.e. extra costs for procedures done or health care resource used, and differences in consequences in terms of health-related quality of life at a point in time around the birth, or mortality outcomes (for stillbirths). The results are in Appendix 10.

What is the intervention and what are the boundary dates?

"The intervention" can be thought of in either one of two ways: as granular specific interventions, or as the overall SU2S scheme. Challenges with the first approach are that all the Trusts implemented different interventions, at different times, including some implemented at different times within the same Trust. All of this means that for each specific intervention, there is insufficient data to perform an analysis looking at its effect in this context. For this approach, the cost of the intervention would correspond to how much has been spent on purchasing the specific interventions that the Trusts planned to buy using the NHS LA funding, and the date at which it took place is taken as the date of the first specific intervention implemented in each Trust. The abbreviation used in the Results section to indicate this scenario is "Intvn".

In the second case, this presumes an argument that the act of giving money, control and responsibility to a Trust to implement quality-related measures, with the aim of improving outcomes and reducing claims made, might be an effective intervention in itself, at a higher level. The cost of the intervention in this case would be the funds provided by the NHS LA to Trusts,

[71] The results throughout this section of the report are reported as point estimate with its 95% confidence interval. This means that the point estimate is an estimate of the "true" answer, and there is a 95% chance that the true answer lies somewhere within that 95% confidence interval. We can never state the point estimate without also stating its confidence interval, as that would be misleading.
When the confidence interval also contains zero (i.e. the lower bound is negative and the upper bound is positive), then the result that we have obtained from our data is not statistically significant. This means that it is possible that the "true" answer, or the true difference in outcomes before and after the scheme, is actually zero. One way to reduce the size of the confidence interval and obtain a more precise answer (reduce uncertainty) is to use a larger sample size (more Trusts), and so the small sample size is an important limitation of this analysis. On the other hand, if the true effect on using the scheme is in fact very large, then the difference in outcomes can be significant even with a small sample size.

regardless of whether or not they had been spent, and the date of the intervention would be the date on which the monies transferred from the NHS LA arrived in the Trust's bank account. The abbreviation used in the Results section to indicate this scenario is "BACS".

We have therefore carried out two parallel analyses, looking at the impact of each of these types of intervention, i.e. the specific interventions purchased ("Intvn"), and the overall scheme as an intervention ("BACS"). The unsuccessful (control) Trusts did not receive funds from the NHS LA so do not have their own intervention dates, but a date is needed for the DiD analysis. Therefore, the control group, i.e. the unsuccessful Trusts, were assigned an "intervention date" corresponding to the median intervention date reported by the successful Trusts. The second case considers the date on which funds were received by each Trust via BACS from NHS LA, which is assumed to be three working days after the money was sent[72]. Again, the control Trusts were assigned the median date of the successful Trusts for this payment date.

We have assumed that in the case where the intervention is taking part in the SU2S scheme in general, rather than the implementation of specific interventions listed by the Trusts, the same short-term proxy outcomes for the number of claims made as discussed in the overview above might be influenced. For further details on the reasons behind the choice of these outcomes as proxy outcomes for litigation claims, please see Appendix 10.

Outcome data collection

We prepared and sent a data entry sheet (see Appendix 9) to 28 successful and 23 unsuccessful Trusts requesting information on these outcomes from June 2011 to July 2016, inclusive. Not all Trusts sent complete data, and there was large variation in the source data that Trusts used to collate responses to our questions. Most used an in-house or externally provided electronic database or dashboard, and some Trusts sourced data from more than one system across different hospital departments. One Trust retrieved the data from paper copies of the monthly report. This has previously been discussed in the previous section, and it should be noted that a future funding priority should aim to provide integrated, flexible, efficient and user-friendly IT systems to bring all Trusts up to a minimum standard as a high priority, in order to both improve data quality and save time, which might then be spent with patients. We performed basic checks on all datasets received to test their face validity, corresponding with Trusts to ensure that the data received were as accurate as possible within the constraints of Trusts' IT systems, and time and staffing limitations. We then put the data received into a single dataset with Trusts identified only by code numbers so that their individual responses are not identifiable, and only aggregated data are reported.

Intervention dates

We have carried out two parallel analyses, firstly looking at the impact of the specific interventions purchased, using the implementation dates provided by each Trust regarding when the first of their interventions was implemented. The control group, i.e. the unsuccessful Trusts, were assigned an "intervention date" corresponding to the median intervention date of the successful Trusts (8 May 2015). The second analysis considers the date on which funds were received by each Trust via BACS from NHS LA. Again, the unsuccessful Trusts were assigned the median date of the successful Trusts for this payment date (29 Sep 2015). The Sign up to Safety dates of implementation of specific interventions in the successful Trusts fall between March 2015 and September 2016 (the latest date was an anticipated date stated in a report that we received in the spring of 2016). The BACS payment dates for the successful Trusts fall between May and September 2015.

[72] Information provided by NHS LA

6.5. Results

Of the 28 successful Trusts (Y), we received information from all 28 Trusts and were able to include data from 26 Trusts in our final dataset. The remaining 2 contained inconsistencies that had not been resolved before the data analysis was finalised.

Of the 23 unsuccessful Trusts (N), we received data from 21, and were able to include data from 18 in our final dataset. The remaining 3 contained inconsistences that had not been resolved before the data analysis was finalised.

The results are reported separately for babies and mothers. We firstly show histograms to illustrate how common each outcome is in each group (successful and unsuccessful).

The results are reported in tables, where we give the proportion of events of each outcome type, averaged over the Trusts. Then we calculate the before and after difference in each group, Difference (Y) and Difference (N), and then we use these values to calculate the difference in differences, which is the right-most column in Tables 8, 9, 12 and 13)

These proportions are then multiplied up to give the number of events per 100,000 term singleton babies for the outcomes relating to babies, and per 100,000 mothers for the outcomes relating to mothers (Tables 10, 11, 14 and 15). The uncertainty in all the results is very large, partly due to the small sample size, and also due to the low incidence of some of the outcomes, especially cooling, stillbirth, and 4th degree tears.

The raw proportions, split into the four groups (Before N; After N; Before Y; After Y), are shown in plots of the proportions against time, along with their linear regression lines, to illustrate more clearly the large amount of uncertainty in this dataset and the wide variation of results over time.

Finally, using the unit costs and the unit reductions in HRQOL from the NHS Reference Costs and the literature, we present the total differences in costs and HRQOL per outcome per 100,000 babies or mothers, and relate this to the amount spent by the NHS LA.

We requested data from June 2011 to July 2016, and we received the full dataset in some cases but not all. In particular, some Trusts could not extract data for the early part of that period for all outcomes, meaning that the level of missing data was higher in the earlier years. To mitigate any selection or other bias arising from this missing information, we considered only the data reported in the 12 months before and after the boundary dates, and the base case analysis is presented using only this one-year data. An added advantage to this is that we do not need to control for annual variations if we are only including data from one year before and after, and this simplifies the analysis.

Babies' outcomes

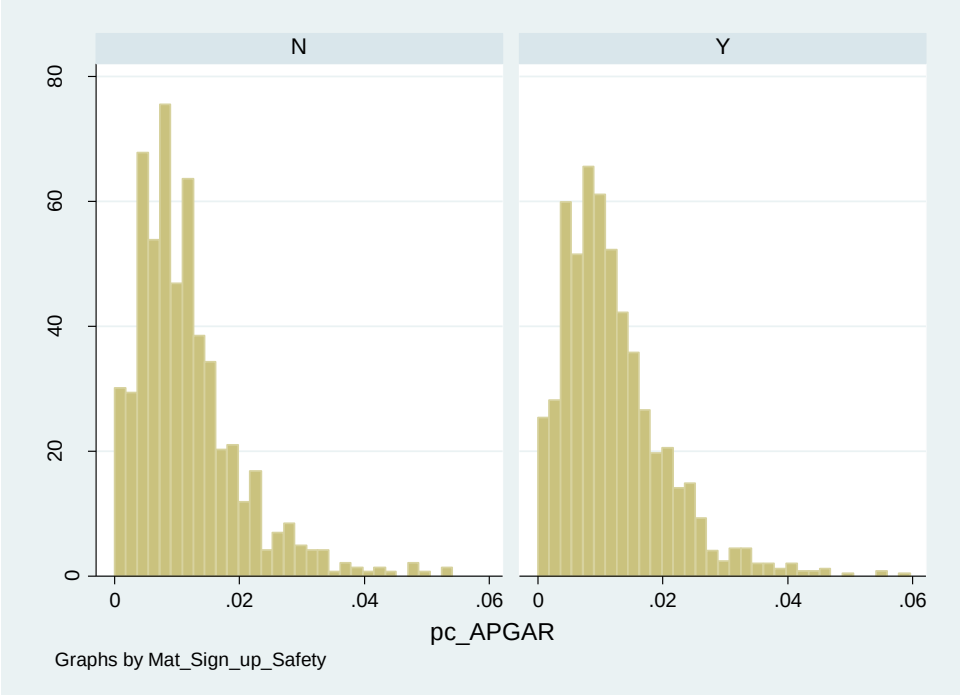
Histograms (N=unsuccessful Trusts, Y=successful Trusts)

These graphs show

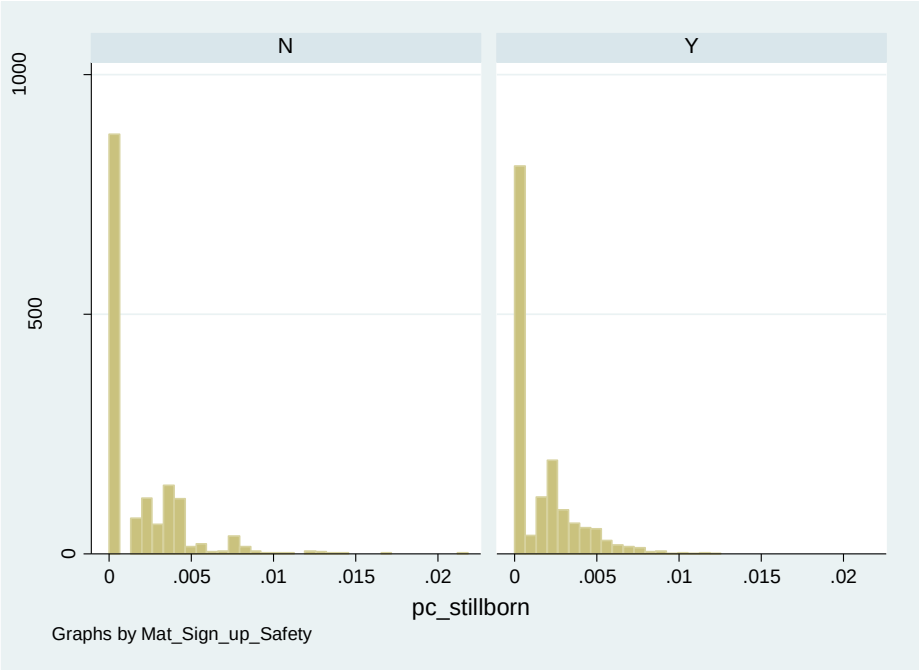
- "pc_APGAR": the average proportions of low Apgar scores at 5 minutes (with reported numbers of unknown Apgar scores subtracted from the denominators as detailed in Table 7 above), split by successful/unsuccessful Trusts (Graph 1);
- "pc_stillborn": the average proportions of stillbirths per month, in successful and unsuccessful Trusts (Graph 2);
- "pc_cooled": the average proportions of babies therapeutically cooled per month, in successful and unsuccessful Trusts (Graph 3);
- "pc_NICU": the average proportions of babies unexpectedly admitted to NICU per month, in successful and unsuccessful Trusts (Graph 4).

The x-axis in each plot represents the proportion of the specified outcome (i.e. 0.01 = 1%), and the y-axis, called "Density", represents the frequency of each value on the x-axis; for example, in both the N Trusts and the Y Trusts, the most frequent (highest column) monthly rate of low Apgar score recorded is about 1%, and the most frequent rate of stillbirth or cooling is close to zero.

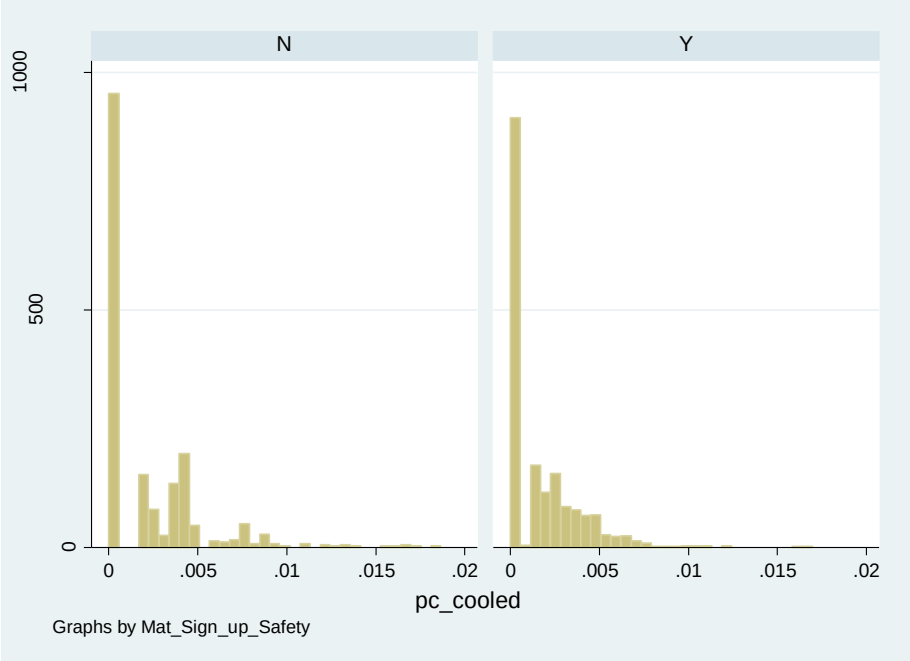
Graph 1. Average proportion of low Apgar scores (<7) at 5 minutes in successful and unsuccessful Trusts



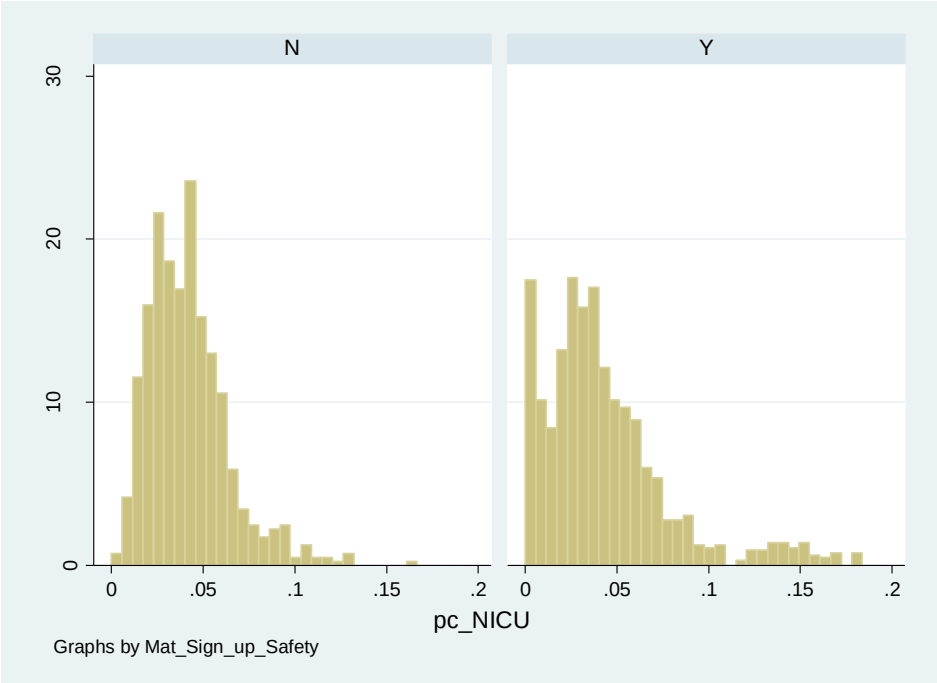
Graph 2. Average proportion of stillbirth per month in successful and unsuccessful Trusts



Graph 3. Average proportion of therapeutically cooled babies per month, in successful and unsuccessful Trusts



Graph 4. Average proportion of babies unexpectedly admitted to NICU per month, in successful and unsuccessful Trusts



Differences in proportions of babies' outcomes in our sample

These results are shown in Table 8 (using the date of implementation of the first specific intervention, or the median of that date) and Table 9 (using the date on which the BACS payment arrived at the Trust, or the median of that date). The data are described using the next set of tables (Table 10 and Table 11) as it is easier to discuss when considering numbers of events per 100,000 babies than when considering the proportions which are given here. The proportions are reported however as they relate to Graphs 1-4.

Table 8**.** Proportions of outcomes, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the specific intervention implementation date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Proportions	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	0.00133	0.00118	-0.00016	0.00141	0.00160	0.00019	-0.00035
lower 95%CI	0.00102	0.00083	-0.00050	0.00103	0.00122	-0.00019	-0.00073
upper 95%CI	0.00164	0.00152	0.00019	0.00179	0.00199	0.00058	0.00004
Low Apgar score	0.01240	0.01389	0.00149	0.01234	0.01301	0.00068	0.00081
lower 95%CI	0.00988	0.01130	-0.00110	0.00928	0.00995	-0.00239	-0.00225
upper 95%CI	0.01492	0.01648	0.00408	0.01539	0.01608	0.00374	0.00387
Cooled	0.00154	0.00135	-0.00018	0.00190	0.00221	0.00031	-0.00049
lower 95%CI	0.00113	0.00090	-0.00063	0.00139	0.00169	-0.00021	-0.00101
upper 95%CI	0.00194	0.00181	0.00027	0.00241	0.00273	0.00083	0.00003
Unexpected NICU	0.04177	0.04357	0.00180	0.04336	0.04286	-0.00050	0.00230
lower 95%CI	0.03276	0.03449	-0.00728	0.03249	0.03198	-0.01138	-0.00859
upper 95%CI	0.05077	0.05265	0.01088	0.05423	0.05375	0.01039	0.01318

Table 9. Proportions of outcomes, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the BACS payment arrival date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Proportions	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	0.00142	0.00130	-0.00013	0.00157	0.00153	-0.00004	-0.00009
lower 95%CI	0.00112	0.00101	-0.00043	0.00118	0.00116	-0.00043	-0.00048
upper 95%CI	0.00172	0.00158	0.00017	0.00196	0.00190	0.00035	0.00031
Low Apgar score	0.01192	0.01366	0.00175	0.01134	0.01278	0.00144	0.00031
lower 95%CI	0.00960	0.01137	-0.00057	0.00835	0.00982	-0.00155	-0.00268
upper 95%CI	0.01423	0.01596	0.00406	0.01433	0.01574	0.00443	0.00330
Cooled	0.00174	0.00156	-0.00018	0.00206	0.00213	0.00007	-0.00025
lower 95%CI	0.00129	0.00112	-0.00064	0.00145	0.00154	-0.00054	-0.00086
upper 95%CI	0.00219	0.00199	0.00027	0.00266	0.00271	0.00068	0.00035
Unexpected NICU	0.04195	0.04488	0.00294	0.04372	0.04263	-0.00108	0.00402
lower 95%CI	0.02951	0.03247	-0.00950	0.02755	0.02648	-0.01725	-0.01215
upper 95%CI	0.05439	0.05730	0.01538	0.05989	0.05878	0.01509	0.02019

Differences in numbers of events per 100,000 babies

These results are shown in Table 10 (using the date of implementation of the first specific intervention, or the median of that date) and Table 11 (using the date on which the BACS payment arrived at the Trust, or the median of that date).

When using the Invtn date, the stillbirth and cooling rates seem to both reduce slightly in the Y group (Diff (Y) is negative), and increase slightly in the N group (Diff (N) is positive). These changes are not however statistically significant. The rate of low Apgar score increases slightly (and not significantly) in both groups, leading to an overall insignificant increase. The rate of unexpected NICU admission seems to increase (insignificantly) in the Y group, decrease (insignificantly) in the N group, with an overall insignificant increase. As these changes are not significant, we cannot conclude that there is any change at all.

When using the BACS date, the results are virtually identical, and there is no alteration in the conclusion drawn, which is that we can detect no change. This does not mean that there definitively is no change, instead it means that with the data we have, and the small sample size, there is no detectable change, so either it is small, or it is non-existent.

Table 10. Numbers of events per 100,000 babies, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the specific intervention implementation date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Events per 100,000 babies	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	133	118	-16	141	160	19	-35
lower 95%CI	102	83	-50	103	122	-19	-73
upper 95%CI	164	152	19	179	199	58	4
Low Apgar score	1240	1389	149	1234	1301	68	81
lower 95%CI	988	1130	-110	928	995	-239	-225
upper 95%CI	1492	1648	408	1539	1608	374	387
Cooled	154	135	-18	190	221	31	-49
lower 95%CI	113	90	-63	139	169	-21	-101
upper 95%CI	194	181	27	241	273	83	3
Unexpected NICU	4177	4357	180	4336	4286	-50	230
lower 95%CI	3276	3449	-728	3249	3198	-1138	-859
upper 95%CI	5077	5265	1088	5423	5375	1039	1318

Table 11**.** Numbers of events per 100,000 babies, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the BACS payment arrival date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Events per 100,000 babies	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	142	130	-13	157	153	-4	-9
lower 95%CI	112	101	-43	118	116	-43	-48
upper 95%CI	172	158	17	196	190	35	31
Low Apgar score	1192	1366	175	1134	1278	144	31
lower 95%CI	960	1137	-57	835	982	-155	-268
upper 95%CI	1423	1596	406	1433	1574	443	330
Cooled	174	156	-18	206	213	7	-25
lower 95%CI	129	112	-64	145	154	-54	-86
upper 95%CI	219	199	27	266	271	68	35
Unexpected NICU	4195	4488	294	4372	4263	-108	402
lower 95%CI	2951	3247	-950	2755	2648	-1725	-1215
upper 95%CI	5439	5730	1538	5989	5878	1509	2019

Plots of mean differences in proportions against calendar month, separated by before vs. after and successful vs. unsuccessful Trusts

Note that the DiD results in Table 10-11 above have calculated the mean proportion in the after period, i.e. with the average taken over the 12 months after the boundary date (or for as many months as there is data available), and the mean proportion for the before period (average over 12 months before the boundary date), and subtracted one from the other. The plots shown in Figures 16, 17, 18, 19 below show trends with time, and thus they might be able to convey some more meaning or direction in the data. However, the trend lines show that there are no clear trends in the outcomes measured for this sample.

The average proportion in each of the four groups (Before N; After N; Before Y; After Y) corresponds to the values in the first pair of results tables (Tables 10 and 11). Each point in the scatter plots that are shown here (Figures 16-19) is the average proportion of all Trusts in that group over that calendar month. For example, the green triangle point in the left-hand plot of Figure 16 here below, at month 50 and lying just below 0.006 along the y-axis, tells us that, of the Y Trusts, in the before period (BACS boundary date), the average proportion of low Apgars in July 2015 was just under 0.6%.

Low Apgar scores

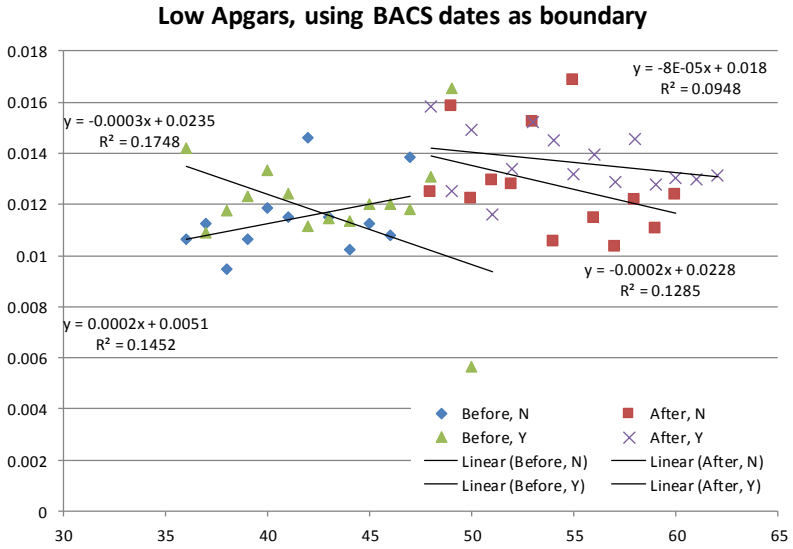
These graphs show the progression of the rates of low Apgar scores (<7 at 5 minutes) as a function of time (months) for each of the four groups: before and unsuccessful ("Before, N"); after and unsuccessful ("After, N"); before and successful ("Before, Y"); after and successful ("After, Y"). No adjustment can be made to shift the time points such that all Trusts implement at a false 'zero', as this would mean that information regarding external influences on the low Apgar rates, e.g. seasonal variations, or changes in policy or funding that happened in a specific month, would be lost. The linear trend lines are the regression lines that fit each of the four groups. A negative gradient implies that the outcome became less frequent over time, and a positive gradient the reverse, although it is important to note that the R^2 value, denoting the goodness of fit, is close to zero for all outcomes, indicating that there is no real trend and in fact the values are scattered

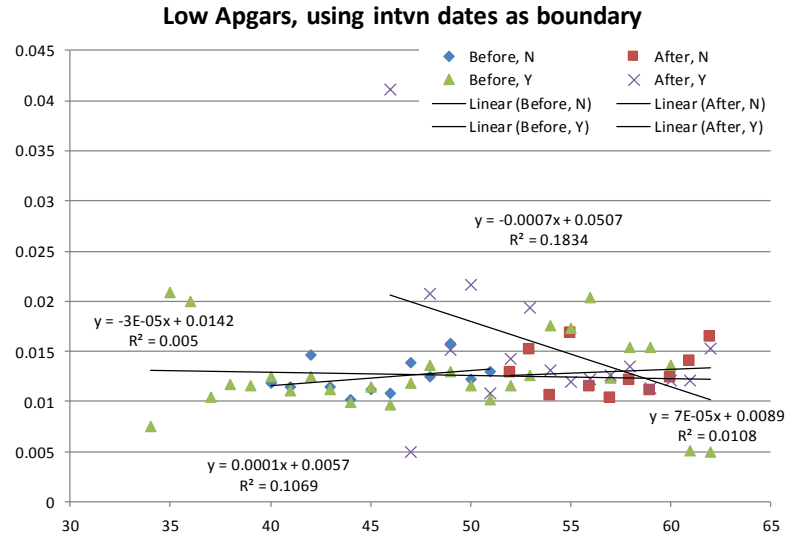
almost randomly. This uncertainty is also reflected in the results tables in this section (see Table 10 and Table 11 above).

Note: The y-axis in each plot in Figures 16-19 indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Figure 16. Plots for the proportions of babies with low Apgar scores as a function of time, with the numbers of unknown Apgar scores removed from the denominator.

The four sets of points in each plot correspond to (i) unsuccessful Trusts' data (N) before the boundary month, (ii) successful Trusts' data (Y) before the boundary month, (iii) unsuccessful Trusts' data (N) after the boundary month, (ii) successful Trusts' data (Y) after the boundary month. The four linear regression lines correspond to each set of points. The left-hand plot used the BACS dates as the boundary dates, and the right-hand plot used the first specific intervention dates as the boundary dates.

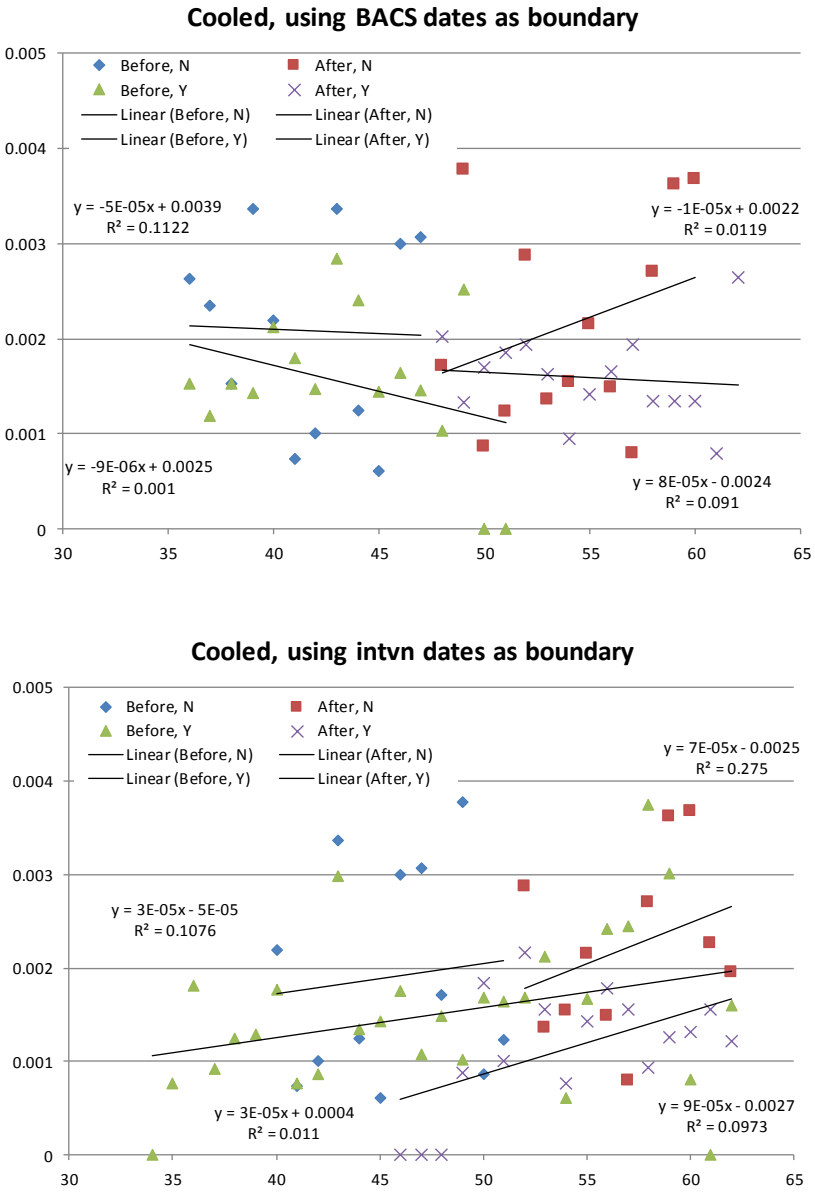




Cooling

These plots, along with the values in Table 10-11, also show no significant changes. The lack of significance is given in the wide 95%CI, straddling zero (see Table 10-11 above), and in the low values for goodness of fit shown by the linear regression lines in Figure 17 below.

Figure 17. Plots for the proportions of babies that are therapeutically cooled. The top plot used the BACS dates as the boundary dates, and the bottom plot used the first specific intervention dates as the boundary dates.

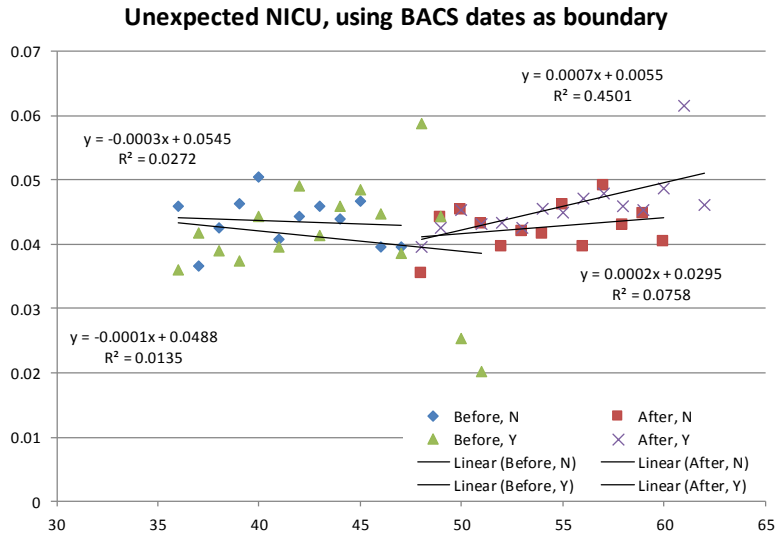


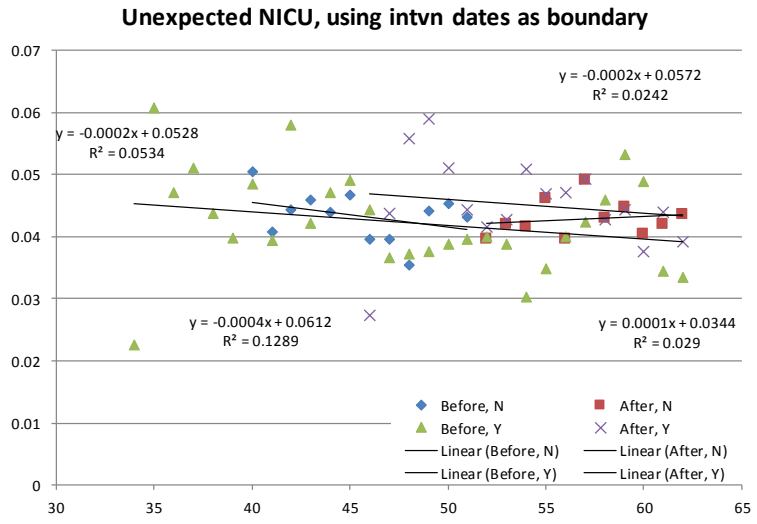
Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Unexpected admissions in NICU

These plots, along with the values in Table 10-11, also show no significant changes. The lack of significance is given in the wide 95%CI, straddling zero (see Table 10-11 above), and in the low values for goodness of fit shown by the linear regression lines in Figure 18 below.

Figure 18. Plots for the proportions of term babies that are unexpectedly admitted to NICU. The top plot used the BACS dates as the boundary dates, and the bottom plot used the first specific intervention dates as the boundary dates.



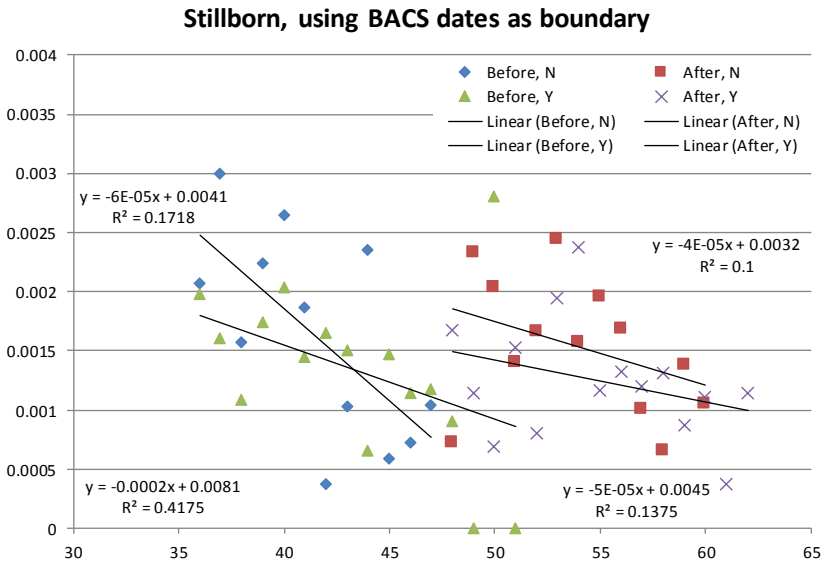


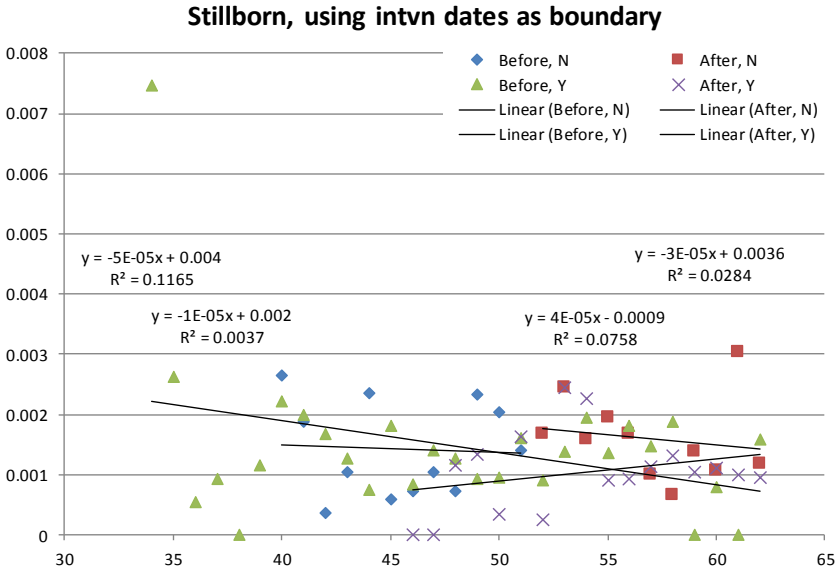
Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Stillbirth

These plots, along with the figures in Table 10-11, show no significant change with time. The lack of significance is given in the wide 95%CI, straddling zero (see Table 10-11 above), and in the low values for goodness of fit shown by the linear regression lines in Figure 19 below. Possibly a slight increase from numbers in Table 10-11, but the trends in the graph suggest a decrease through each time period. This disagreement simply adds to the lack of certainty over any meaningful trend or conclusion.

Figure 19. Plots for the proportions of babies that are stillborn. The top plot used the BACS dates as the boundary dates, and the bottom plot used the first specific intervention dates as the boundary dates.





Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Mothers' outcomes

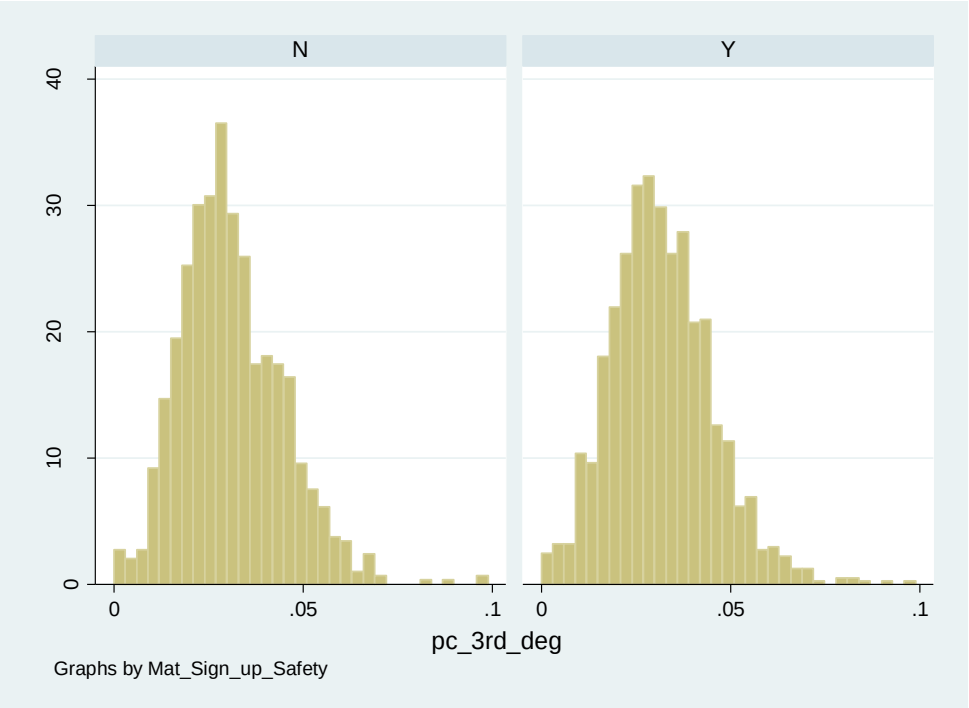
Histograms (N=unsuccessful Trusts, Y=successful Trusts)

These graphs show the average proportions of:

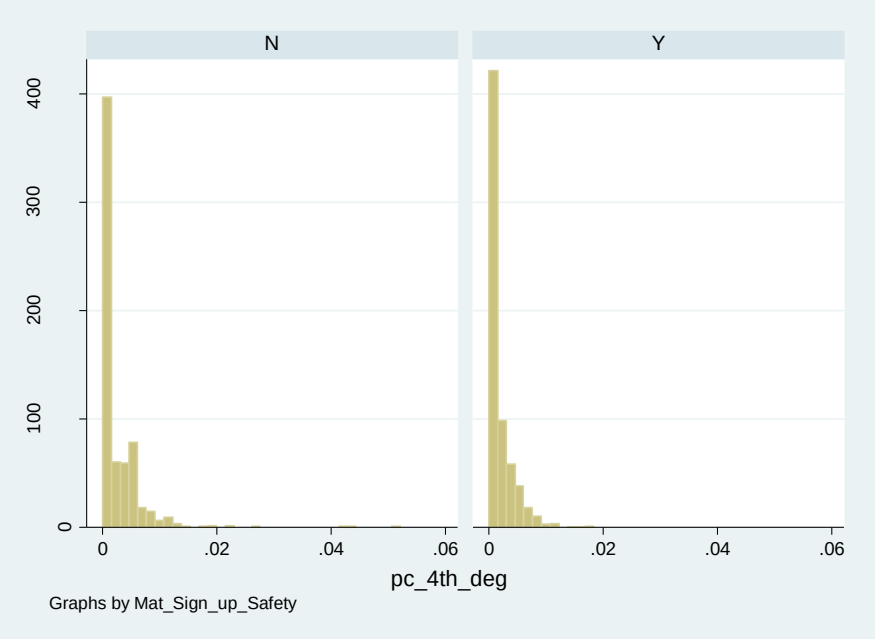
- 3rd degree tears (not including data from the three Trusts that only reported mixed 3rd and 4th degree tears) (Graph 5)
- 4th degree tears (not including data from the three Trusts that only reported mixed 3rd and 4th degree tears) (Graph 6)
- Mixed 3rd and 4th degree tears (including data from all 44 Trusts, i.e. also including the three Trusts that only reported mixed 3rd and 4th degree tears) (Graph 7)

Each pair of histograms shows the unsuccessful Trusts' ("N") values on the left, and those of the successful Trusts ("Y") on the right. The y-axis, called "Density", represents the frequency of each value on the x-axis; for example, in both the N Trusts and the Y Trusts, the most frequent rate of 4th degree tears is close to zero, and the most frequent rate of 3rd degree or mixed 3rd/4th degree tears is around 3% to 4%.

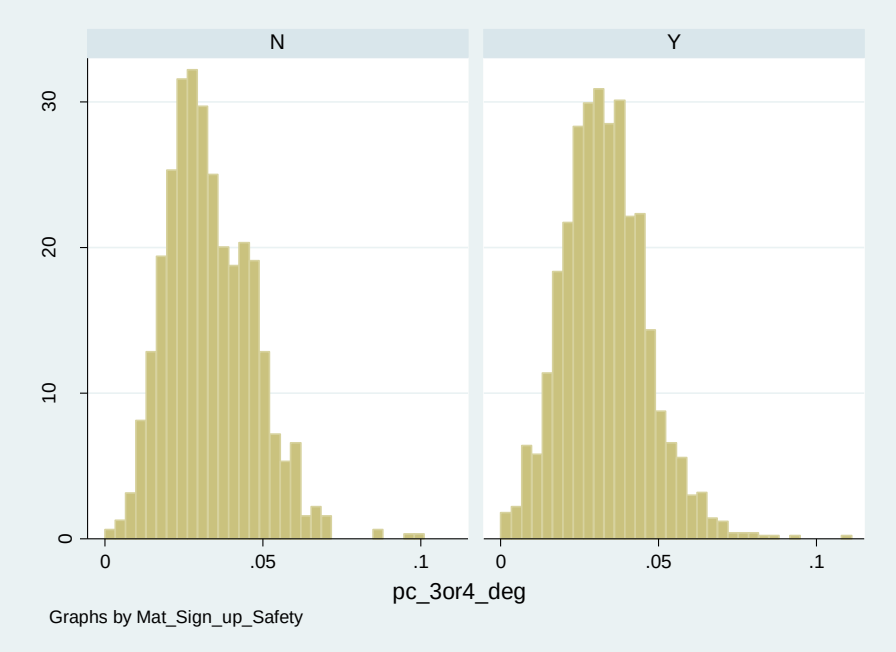
Graph 5. Average proportions of 3rd degree tears in successful and unsuccessful Trusts



Graph 6. Average proportions of 4th degree tears in successful and unsuccessful Trusts



Graph 7. Average proportions of 3rd and 4th degree tears in successful and unsuccessful Trusts



Differences in proportions of mothers' outcomes in our sample

These results are shown in Table 12 (using the date of implementation of the first specific intervention, or the median of that date) and Table 13 (using the date on which the BACS payment arrived at the Trust, or the median of that date). The data are described using the next set of tables (Table 14 and Table 15) as it is easier to discuss when considering numbers of events per 100,000 mothers than when considering the proportions which are given here. The proportions are reported however as they relate to Figures 20 and 21 below.

Table 12. Proportions of outcomes, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the specific intervention implementation date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Proportions	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	0.03253	0.03256	0.00003	0.03376	0.03013	-0.00363	0.00366
lower 95%CI	0.02943	0.02932	-0.00321	0.03027	0.02661	-0.00715	0.00013
upper 95%CI	0.03563	0.03579	0.00326	0.03725	0.03365	-0.00011	0.00718
4th degree tears	0.00160	0.00141	-0.00019	0.00219	0.00186	-0.00033	0.00014
lower 95%CI	0.00110	0.00087	-0.00073	0.00163	0.00129	-0.00090	-0.00043
upper 95%CI	0.00210	0.00194	0.00035	0.00275	0.00243	0.00024	0.00071
Mixed 3rd or 4th degree tears	0.03347	0.03392	0.00045	0.03595	0.03199	-0.00396	0.00441
lower 95%CI	0.03041	0.03071	-0.00276	0.03238	0.02839	-0.00756	0.00082
upper 95%CI	0.03653	0.03714	0.00367	0.03951	0.03558	-0.00037	0.00801

Table 13**.** Proportions of outcomes, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the BACS payment arrival date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Proportions	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	0.03246	0.03276	0.00030	0.03377	0.03158	-0.00219	0.00249
lower 95%CI	0.02955	0.02992	-0.00261	0.03029	0.02814	-0.00568	-0.00100
upper 95%CI	0.03537	0.03560	0.00321	0.03726	0.03502	0.00130	0.00598
4th degree tears	0.00145	0.00153	0.00008	0.00253	0.00193	-0.00059	0.00067
lower 95%CI	0.00096	0.00106	-0.00042	0.00194	0.00136	-0.00118	0.00008
upper 95%CI	0.00195	0.00200	0.00057	0.00311	0.00250	-0.00001	0.00126
Mixed 3rd or 4th degree tears	0.03317	0.03417	0.00100	0.03630	0.03351	-0.00279	0.00379
lower 95%CI	0.03030	0.03132	-0.00187	0.03274	0.03000	-0.00635	0.00023
upper 95%CI	0.03603	0.03701	0.00387	0.03986	0.03703	0.00077	0.00735

Differences in numbers of events per 100,000 mothers

These results are shown in Table 14 (using the date of implementation of the first specific intervention, or the median of that date) and Table 15 (using the date on which the BACS payment arrived at the Trust, or the median of that date).

When considering the Intvn date, the numbers of tears in the Y Trusts undergo small changes, and the numbers of tears in the N Trusts decrease dramatically, at least in the cases of the 3rd degree tears, and the mixed 3rd/4th degree tears. This means that the difference in differences indicates a significant relative increase in the numbers of tears on taking part in SU2S.

It has been suggested that, when the scheme began, Trusts might have begun to recognise and record 3rd and 4th degree tears more than they previously had done, and this could cause an increase in the numbers of tears reported, even if the numbers of tears do not change per se.

The results when using the BACS dates are similar, with significantly higher numbers of mixed 3rd/4th degree tears, and of 4th degree tears (but not now 3rd degree tears) reported.

Table 14. Numbers of events per 100,000 mothers, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the specific intervention implementation date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Events per 100,000 mothers	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	3253	3256	3	3376	3013	-363	366
lower 95%CI	2943	2932	-321	3027	2661	-715	13
upper 95%CI	3563	3579	326	3725	3365	-11	718
4th degree tears	160	141	-19	219	186	-33	14
lower 95%CI	110	87	-73	163	129	-90	-43
upper 95%CI	210	194	35	275	243	24	71
Mixed 3rd or 4th degree tears	3347	3392	45	3595	3199	-396	441
lower 95%CI	3041	3071	-276	3238	2839	-756	82
upper 95%CI	3653	3714	367	3951	3558	-37	801

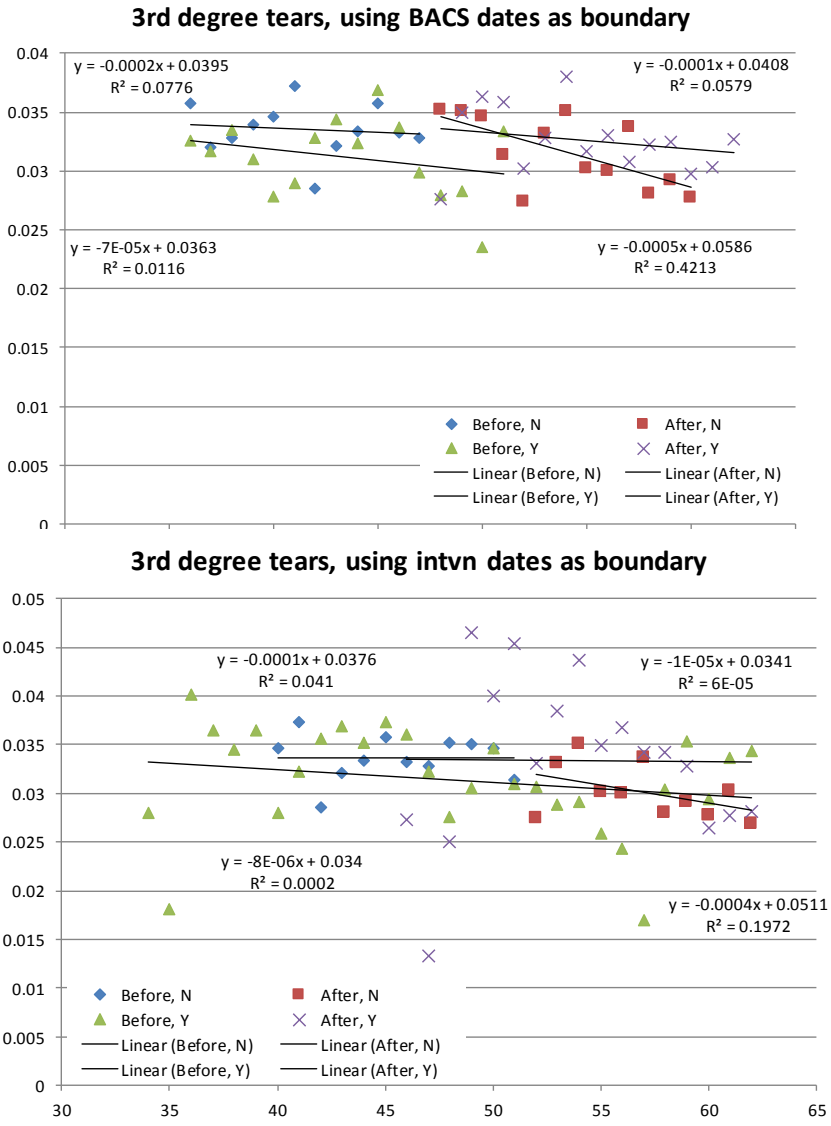
Table 15. Numbers of events per 100,000 mothers, the differences ("Diff" columns) according to whether or not the Trust received SU2S funding or not (using the BACS payment arrival date as the boundary date), and the difference in those differences ("DiD" column) to give the overall effect of the scheme. Values are given as point estimates and 95% confidence intervals.

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Events per 100,000 mothers	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	3246	3276	30	3377	3158	-219	249
lower 95%CI	2955	2992	-261	3029	2814	-568	-100
upper 95%CI	3537	3560	321	3726	3502	130	598
4th degree tears	145	153	8	253	193	-59	67
lower 95%CI	96	106	-42	194	136	-118	8
upper 95%CI	195	200	57	311	250	-1	126
Mixed 3rd or 4th degree tears	3317	3417	100	3630	3351	-279	379
lower 95%CI	3030	3132	-187	3274	3000	-635	23
upper 95%CI	3603	3701	387	3986	3703	77	735

Plots of mean differences in proportions against calendar month, separated by before vs. after and successful vs. unsuccessful Trusts

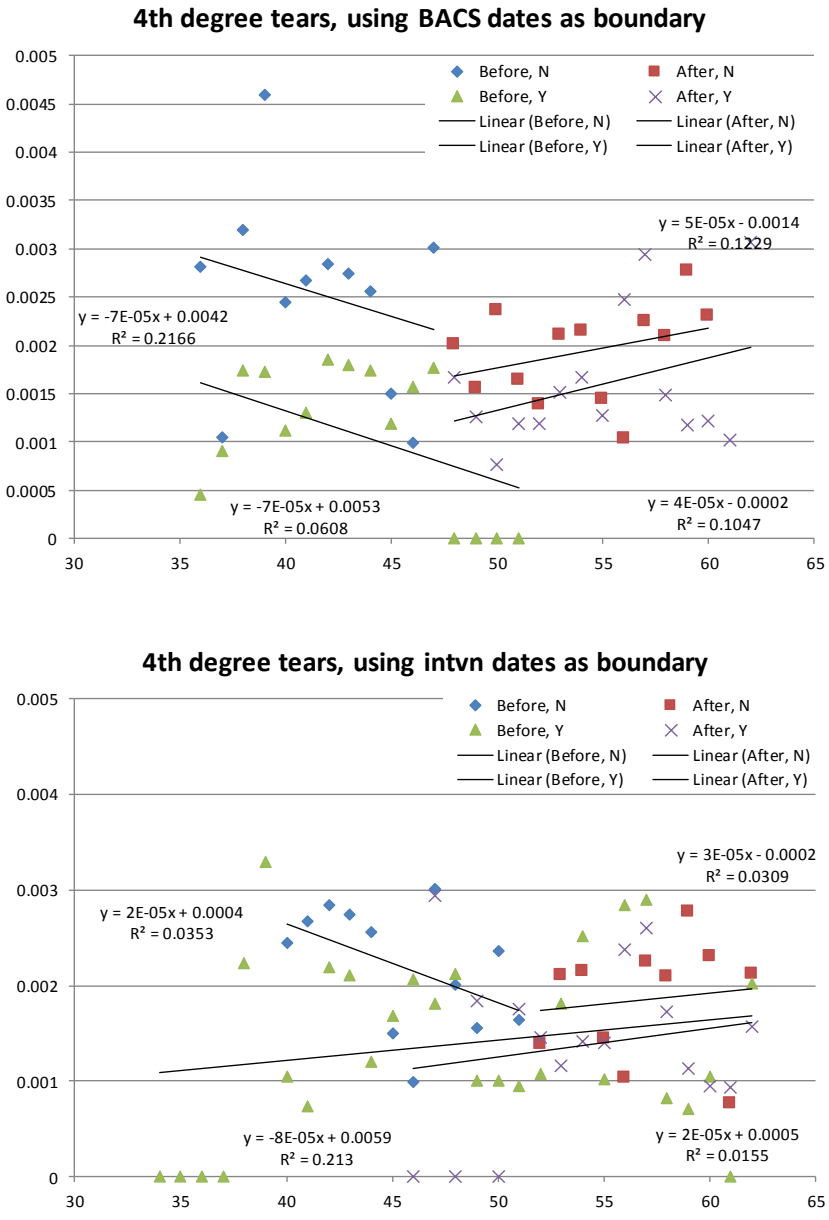
The same plots have been drawn here for the tears data, and they all also suggest that there is a downward (insignificant) trend over time, apart from the 4th degree tears. The numbers of 4th degree tears reported made up approximately 5% of all tears (3rd or 4th degree).

Figure 20. Plots for the 3rd degree tears. The top plot used the BACS dates as the boundary dates, and the bottom plot used the first specific intervention dates as the boundary dates.



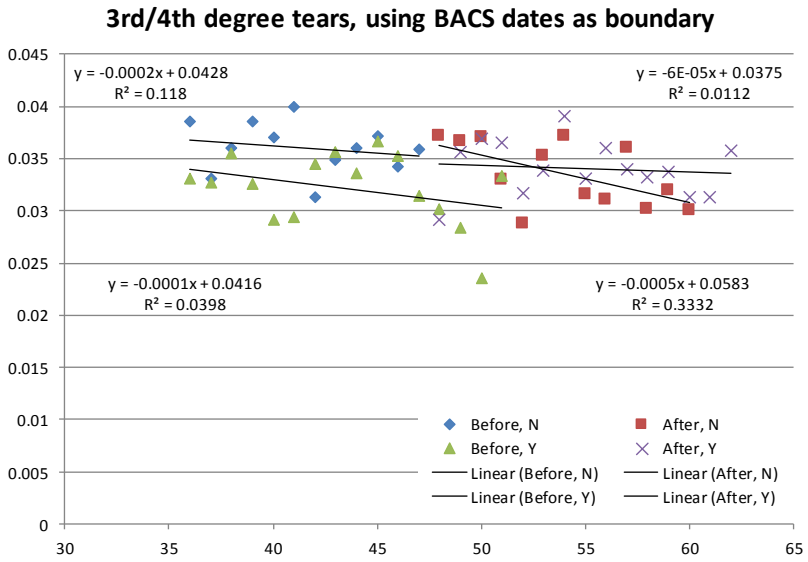
Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Figure 21. Plots for the 4th and 3rd/4th degree tears. The top plot used the BACS dates as the boundary dates, and the bottom plot in each pair used the first specific intervention dates as the boundary dates.

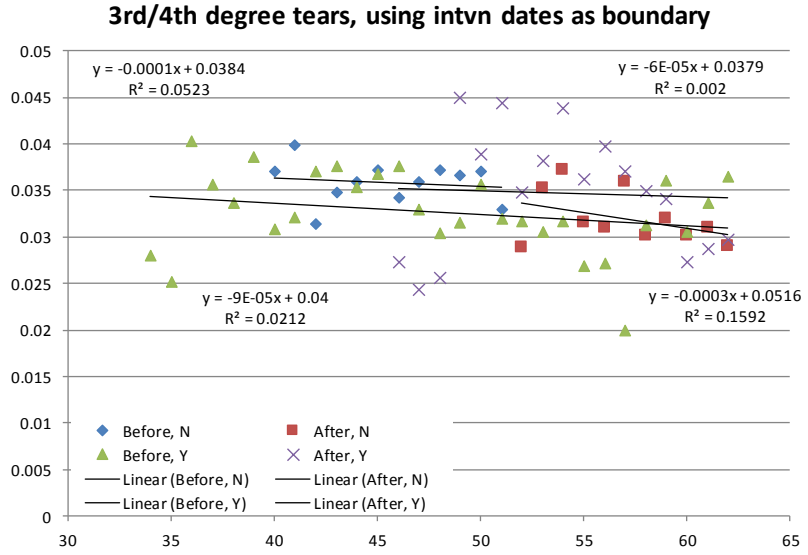


Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

Figure 21. Continued:



.



Note: The y-axis indicates the proportion of events (i.e. 0.01 means 1%), and the x-axis indicates the number of months after the start of data collection, i.e. 1 = June 2011, 30 = November 2013, and 60 = May 2016.

6.6. Sensitivity analysis

We conducted several sensitivity analyses where we: i) varied the boundary data of implementation of interventions; ii) included the whole data period instead of just only one year before and one year after the implementation of interventions; iii) ignored the numbers of "unknown" low Apgar scores rather than removing them from the denominator; iv) included only mothers with instrumental delivery in the analysis of tears; v) ran the analysis for sub-groups of interventions.

Which boundary date?
As discussed briefly above, we have carried out two parallel analyses, firstly looking at the impact of the specific interventions purchased, using the implementation dates. The second analysis considers the date on which funds were received by each Trust via BACS from NHS LA. The results show no difference when using one boundary date instead of the other, so this does not matter.

Over what time period (whole time or 1 year)?
The main analysis compares the average proportion of, for example, low Apgar scores (<7 at 5 minutes) in the 'after' time period (i.e. averaging across all months after the intervention or BACS payment date), with the average proportions with low Apgar scores across the whole 'before' time period. This approach was considered to possibly lead to some bias, as the successful Trusts managed to provide more data in early time periods (i.e. in 2011 and 2012) than unsuccessful Trusts, meaning that there was more missing data for the unsuccessful (control) group in early time periods. A similar issue arose when we considered that the intervention(s) had been implemented less than a year after July 2016, so there was even less information available in the 'after' period compared to the 'before' period, for both sets of Trusts.

To account for this and to minimise any possible bias, we carried out an analysis which only included 12 months before the intervention date, and 12 months after the intervention date as this was felt to be the more appropriate.

The results using the whole time period were also calculated, and the results obtained in each case were similar, and gave the same conclusions for all outcomes, i.e. any changes were not significant, and the broad trends in each outcome were the same regardless of whether 4 (or 3 or 2 depending on missing data) or only one year before the boundary date were included.

Should the proportion of Apgar scores not known at 5 minutes be included?
The results are similar regardless of whether the numbers of unknown Apgar scores are subtracted from the denominator or not. It does not affect the conclusions. The figures given in the tables above are for the analyses where numbers of babies for whom their Apgar score at 5 minutes was not known were excluded from the denominator, i.e. we made no assumptions regarding the unknown Apgar scores. The reason why there was a question over this is that some Trusts reported that there were zero unknown Apgar scores, or they did not have records of this data that were easily accessible without going through all the records by hand, so we did not insist on receiving this.

What denominator should be used for tears?
We tested the results found if only mothers with instrumental deliveries were included as the denominator, and this gave the same conclusions as above – there was a significant increase in tears recorded across both boundary date scenarios, for 3rd degree, 4th degree, and mixed tears (except for 4th degree tears measured with the specific Intvn boundary date – this gave a non-significant increase).

Sub-group analysis of different interventions?
We attempted some sub-group analysis to see if the effects of different interventions could be analysed alone.

We looked at those Trusts that implemented Fetal Monitoring Equipment (intervention B), and at those that implemented Q&S improvement lead / nurse (staff) and/or Training and development (interventions E or H), for stillbirth and for low Apgar rates. The other Y Trusts that had implemented other interventions were omitted from this analysis. Consideration of only these specific intervention types did not lead to any significant results, so the conclusions were the same as the base case analysis.

6.7. How many incidents should be avoided to cover the investment?

Cooling, Apgar scores and cerebral palsy claims
As part of the evaluation, and given the lack of significant effects found in the DiD analysis described above, we have sought evidence from previous studies establishing a relationship between the outcomes collected from the Trusts and cerebral palsy cases and claims. The idea was to determine the minimum number of claims avoided necessary for the Sign Up to Safety Scheme to be good value for money.

The first relationship investigated was between the number of babies cooled and the cerebral palsy cases and claims. After obtaining expert advice and conducting a brief literature review, we have identified one UK-based study that explored the relationship between cooling and poor long-term outcomes in babies (Azzopardi et al., 2009)[73]. According to the authors, 25.71% of the babies cooled will end up being diagnosed with a severe disability, most of which due to cerebral palsy. Although the exact proportion of cerebral palsy cases is not given in the study, one could understand the above-mentioned proportion as the upper limit of cerebral palsy cases.

National level data from the NNRD shows that in the years of 2010 and 2011, there were 1,343 babies cooled in England. We have chosen to focus on those years, given that normally cerebral palsy takes at least 2 years to be diagnosed. Furthermore, the average time gap between the date of incident and the date of claim is on average 3.5 years, therefore we expect most cases referring to incidents that took place in 2010 and 2011 to have been opened already, allowing for a relationship between babies cooled and number of claims to be established.

Using the number of babies cooled in 2010 and 2011 and the proportion calculated by Azzopardi et al, we estimate that 346 infants would end up with some form of severe disability, most of which with cerebral palsy.

Following clinical expert advice, we have also sought to establish a relationship between low Apgar scores and cerebral palsy. The 5-minute Apgar<7 (Apgar $<7^5$) rate in term infants is an important measure of intrapartum care as it is associated with a considerably higher rate of cerebral palsy in later life (OR 62, 95% CI 52-74) as well as lower levels of cognition, lower levels of education and lower incomes (Graham, 2008; Hogan et al., 2007; Thorngren-Jerneck et al., 2001)[74]. There is also

[73] Azzopardi DV et al (2009). Moderate hypothermia to treat perinatal asphyxial encephalopathy. N Engl J Med 361(1):1349-58
[74] Graham (2008). A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy AmJObsGyn; Hogan L, Ingemarsson I et al. (2007) How often is a low 5min Apgar score in term newborns due to asphyxia ? EurJObsGyn 130(2):169-75. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. Obstet Gynecol. 2001;98(1):65–70

evidence of long-term impact of poor birth condition on social and economic outcomes in early adulthood.

Crucially, Apgar scores can be improved by training. Moreover, some pilot work has shown that Apgar scores are reliably collected, with missing data not affecting the robustness of the score as outcome measure.

Finally, a significant majority of claims and also the successful applications are related to reducing asphyxial damage, for which Apgar <7 [5] is a very sensitive marker.
Unfortunately, no UK-based studies were found exploring this relationship. However, a robust population-based cohort study carried in Norway and published in the British Medical Journal (BMJ) has been identified (Lie et al., 2010)[75]. This study has the strength of estimating the relationship between Apgar scores and cerebral palsy using direct observation of the total population born over a long period of time, avoiding important biases due to time lags and sampling.

We have relied on national data (NNRD) to gather information on number of babies born with an Apgar score less than 7 at 5 minutes (Apgar <7[5]) in England between 2010-2011 and applied a rate from the Norway study to estimate the number of babies having cerebral palsy. In 2010 and 2011, there were 5187 born with Apgar scores lower than 7 at 5 minutes. Using the proportion of babies developing cerebral palsy obtained by Lie and other, we would have that 232 babies would be later diagnosed with cerebral palsy, having been born in England between January 2010 and December 2011. The discrepancy between the previous estimate using cooling and the one using Apgar scores may be explained by the fact that the previous did not focus on cerebral palsy cases only, but in any form of severe disability.

Finally, we tried to determine the relationship between cases and claims. According to the NHS Litigation Authority data, 147 cerebral palsy claims were opened referring to incidents occurring in 2010 and 2011. This implies that not all cerebral palsy cases are directly translated into claims, although the majority does become a legal action. Table 16 displays the relationship between cooling and low Apgar and cerebral palsy claims.

Table 16. Cooling, Apgar Scores and Cerebral Palsy Claims

Outcomes	**Cases (2010/11)**	**Proportion that becomes a claim**
Babies cooled	1343	10.95%
Babies with Apgar lower than 7 at 5 minutes	5187	2.83%
Babies with Apgar equal to or lower than 4 at 5 minutes	1783	8.20%

The relationship between cooling, Apgar scores and cerebral palsy claims can be used to estimate the necessary decrease in numbers of babies cooled and with low Apgar necessary for the Sign Up to Safety Scheme to be good value for money.

Considering that on average, a successful cerebral palsy claim has a total cost to the NHS Litigation Authority of £4,745,295 (NHS claims database), and that in total just over £8 million pounds were invested in maternity interventions in the Sign Up to Safety financial incentives Scheme to decrease

[75] Lie KK, Groholt EK, Eskild A. (2010). Association of cerebral palsy with Apgar score in low and normal birthweight infants: population based cohort study. BMJ 341.

both cerebral palsy claims and claims in tears, the scheme could be considered good value for money had it decreased two claims in terms of cerebral palsy. Table 17 displays the necessary reduction in each outcome to reduce the number of cerebral palsy claims by one.

Table 17. Necessary decrease in outcomes

Outcomes	**Necessary decrease**
Babies cooled	18
Babies with Apgar lower than 7 at 5 minutes	70
Babies with Apgar equal to or lower than 4 at 5 minutes	24

Thus, had the Sign up to Safety scheme produced an overall reduction in the number of babies cooled and with low Apgar scores in the amount displayed in Table 17 in the participating Trusts, the scheme could be considered good value for money. This is likely to be possible as the mean Apgar score $<7^{5}$ minutes was 1.3% and there are reported rates in the UK of <0.5% that have been sustained for more than a decade (Draycott et al., 2006)[76].

Similarly, if we think that an admission in NICU cost on average £3,440 (Table A1- Appendix 10) in order to cover the £8 million investment in maternity units would be sufficient to avoid the admission of 2,325 babies.

6.8. Discussion

The results of the analysis don't show a statistically significant effect of the scheme in reducing stillbirth, babies born with a low Apgar score at 5 minutes or receiving cooling, unexpected NICU admissions, or 4^{th} degree tears or instrumental delivery or CSs. The only significant effect is represented by an increase in 3^{rd} degree tears, thought to be due to improved reporting. Hence it was not possible to calculate cost savings associated with the interventions or improvement in health related quality of life and mortality. These results have been confirmed by the sensitivity analysis.

We identified several reasons why the scheme might not have had an impact on reducing intrapartum harm based on our results:

- Choice of interventions. Taking the results of the statistical analysis at face value, the interventions implemented by Trusts may not have been effective in reducing harms. The maternity units invested their funding in a broad range of interventions, and evidence of effectiveness for the interventions they implemented is limited.
- Implementation problems. At the time of the analysis (conducted 12-14 months after the Trusts had received funding from the scheme) not all Trusts had implemented the interventions they had originally proposed to: almost half of the participating maternity units (46% of those awarded funding) had not implemented some or all of their interventions at the time we completed the evaluation.
- Short duration of follow up. There is likely to be a delay before the benefits of the interventions are seen, so even if interventions are fully implemented then it may take

[76] Draycott T, Sibanda T, Owen L, Akande V, Winter C, Reading S et al. (2006) Does training in obstetric emergencies improve neonatal outcome? BJOG: An International Journal of Obstetrics &Gynaecology. 113(2):177-82.

longer than the time horizon of our evaluation (one year) to see a tangible benefit. In at least one unit level intervention with positive outcomes, a year was required to train all of the unit, and outcomes changed in the following year.

- Data quality. The results of our analysis may be limited by the quality of data provided. Maternity datasets are recognised to be some of the most accurate in the NHS, however many Trusts found it difficult to provide data from their IT department for evaluation. Despite the enormous efforts by Trusts, we are aware that the retrospective collection might have negatively impacted on the accuracy, reliability and completeness of the data. In addition, the small sample size (we only had aggregate data for 44 Trusts) and the lack of patient level data may also have affected the results.

Future recommendations

There have been a number of successful maternity improvement programmes and some that have been less successful and it is imperative that we use the learning from both.

Firstly, as recommended by the Trusts, it would be useful in future for the NHS LA to work with experienced academic and clinical collaborators to provide further clinical oversight, help signpost units to potentially effective interventions and also help units with local evaluation.

Secondly, local measurement of care is very important to both prioritise interventions and also measure effect (Macrae, 2016)[77]. We should make measurement of care easier, timely and more understandable to all the actors in the system, from Government to patients themselves. Quality is multi-faceted and we must ensure that measurement is broad enough to include what is important to all stakeholders, not merely what can easily be measured (Draycott et al., 2010)[78].

High quality healthcare systems are those that produce the best outcomes with the fewest interventions, to the satisfaction of their patients, within a cost-effective framework.
Measuring a combination of outcomes and processes is required, with perception of care if possible.
There is probably a sweet spot of best care: the best outcomes, with the least intervention and the best experience.
We should collect and produce a standard, relevant set of quality indicators, ideally from routinely collected data, and present these in a manner that facilitates on-going quality improvement, just as recommended in the recent Better Births report (National Maternity Service Review, 2016)[79].
Ideally, these data could then be employed to focus regulatory visits from bodies like the Care Quality Commission as well as being used for local prioritisation of improvement initiatives and energy.

Thirdly, the evidence base for policy makers is lacking and this should be addressed to help national level bodies in the system to use finite resources most effectively. It is important that we identify which national level levers can be used to incentivise 'bottom up improvement' at a local level.

NHS LA is perfectly placed to align and untangle the 'priority thicket' of national, regional and local actors in the current policy landscape (Dixon-Woods et al., 2013) [80].

---

[77] Macrae C. The problem with incident reporting. BMJ Quality & Safety. 2016; 35(2):71-5

[78] Draycott T, Sibanda T, Laxton C, Winter C, Mahmood T, Fox R. Quality improvement demands quality measurement. BJOG: An International Journal of Obstetrics & Gynaecology. 2010 Nov 15;117(13):1571–4.

[79] National Maternity Services Review. Better Births. Improving outcomes of maternity services in England. London; 2016.

[80] Dixon-Woods M, Baker R, Charles K, Dawson J, Jerzembek G, Martin G, et al. (2013). Culture and behaviour in

Going forward, we propose that a useful framework for insurers and other actors in the Quality Improvement space may be divided into national, regional and local elements:

National

- Establish a network of active academics and clinicians to work with the NHS LA
- Set evidence based standards (proportion staff trained etc.)
- Signpost *evidence based* Interventions
- Monitoring and Evaluation
  - Alignment: Care Quality Commission, NHS England, Professional bodies
- Incentivise good outcomes – Premium modelling

Regional

- Incentivise support networks including buddying
- Support benchmarking

Local

- Consider funding some Implementation research (including General Medical Council and Nursing and Midwifery Council revalidation)
- Prioritisation based on local outcomes/benchmarking

Sustainable improvement in intrapartum outcomes is likely to require an integrated approach of: incentivising best care, local multi-professional training with tools for staff to provide best care and also the measurement of best care. Insurers are very well placed to identify, fund and promote successful models of care.

Finally, there is a Health Foundation funded project investigating a selection of State insurers' approaches to clinician engagement. It is currently at the analysis stage, but there is clearly significant interest in this area from all the participants. Some preliminary observations suggest that Insurers with better clinical insight and partnership can potentially have a more positive influence on patient safety. Whilst there are important learning points from each of the insurers, the single over-arching finding is an aspiration for positive partnerships: embracing and directing research as well as developing collaborations between Insurers, academics and clinical teams, to improve care and thereby reduce both harm and litigation.

---

the English National Health Service: overview of lessons from a large multimethod study. BMJ Quality & Safety.

7. Summary and recommendations

Summary

The aims of this project were threefold:
(1) To examine the feasibility of collecting robust 'cause of harm' data to inform strategies to improve safety, reduce harm and reduce litigation claims.
(2) To evaluate the impact of the SU2S financial incentive scheme on reducing missed fractures in A&E
(3) To evaluate the impact of the SU2S financial incentive scheme on reducing intrapartum harm.

To meet aim 1 we suggested that to be useful a dataset should contain data on the number and type of incidents and their association with health and claims, data on interventions to reduce incidents and claims, and data to routinely evaluate the impact of interventions when they are implemented. We identified several datasets already in existence that could meet these requirements, and rather than collect new data, and incur the associated time and money costs of this, further work would be beneficial to coordinate these data and improve their reliability and reduce under-reporting.

To meet aim 2 we undertook a cost-utility analysis of the interventions put in place to reduce the number of missed fractures in A&E in participating Trusts funded by the Sign up to Safety financial incentives scheme. The interventions were based around hot reporting of imaging results and continuous quality improvement strategies. We approached participating Trusts for data on the effectiveness of these interventions but it was not possible for them to provide it, mainly due to difficulties in identifying and recording missed fractures. We therefore based our analysis on published estimates of the effectiveness of these interventions from published studies, noting the limitations of this approach. Accounting for the costs of the interventions, and their impact on missed fractures, health-related quality of life, health care costs and litigation costs, both interventions were found to reduce overall costs and improve the health outcomes of patients.

To meet aim 3 we undertook a difference-in-differences analysis of the range of interventions put in place to reduce the intrapartum harms in maternity units funded by the Sign up to Safety financial incentive scheme. Our analysis was based on data from units on a range of measures, collected before and after the interventions were implemented, for Trusts who were successful in obtaining funds, and for those who did not receive funds from the scheme. Our detailed analysis showed that the intervention did not have a significant effect on any of the outcomes (other than the number of reported tears, which were higher as a result of the scheme, possibly because with the introduction of the scheme more tears were recognised and reported). Given these results it was not possible to identify cost savings associated with the interventions, or improvements in health-related quality of life and mortality, or a reduction in claims. We discussed a range of possible reasons for our findings, including that the interventions were not effective, that that it was too early to undertake the evaluation, and that the quality of the data precluded a definitive analysis.

We calculated that for the Sign up to Safety financial incentive scheme to be cost saving across the 28 Trusts, then it would need to reduce the number of claims for cerebral palsy by two, which based on published data, would require on average 18 fewer babies to be cooled, 70 fewer babies born with an Apgar score lower than 7 at 5 minutes, or 24 babies born with an Apgar score equal or lower than 4 at 5 minutes across the 28 Trusts who received funding from the scheme.

Another feature of the evaluation was that at the time of the analysis (12-14 months after the Trusts had received funding from the scheme) not all Trusts had implemented the interventions they had originally proposed due to a range of reasons: almost half the participating maternity units (46% of those awarded funding) had not implemented some or all of their interventions at the time we completed the evaluation.

Moving forward more work should be done to make sure that similar schemes are supplemented by an initial evaluation of what are the most critical areas to be improved, what are the main causes of errors and what interventions could potentially be the most effective in reducing errors according to available evidence.

Further analysis at a later date may be beneficial, after a longer period has elapsed since Trusts implemented the interventions. It is difficult to specify what this period might be. We speculate that interventions might take a year to implement fully from the start date and then at least a year to have an effect on incidents. For maternity care, in our data, the average time gap between the date of incident and the date a claim is made was 3.5 years, and so if future research was to measure the impact on claims directly in maternity services a substantial period from the date at which interventions were implemented needs to have elapsed. This time period might vary by the type of intervention and the clinical specialty. Further research on the appropriate time period for measurement would be useful.

At present there is a considerable amount of data available on errors and claims but more effort is required to coordinate and improve its reliability. This could be achieved by improving existing databases so that data are routinely collected, in a standardised way and analysed. For example, the maternity dashboard, now limited to South East could be potentially very useful for building a more comprehensive database to be linked with NHS LA data on claims. This should be accompanied by investments and efforts in improving the IT systems of Trusts so that all the necessary data are collected in a timely, accurate, complete and reliable way.

Further consideration should be given to the trade-off between top-down recommendations on the most cost-effective interventions and interventions that are more responsive to local needs.
If on the one hand decisions on what interventions to invest in should follow clear considerations of cost-effectiveness, on the other hand we should not ignore specific necessities at Trust level.

Recommendations

1. Given the timelines involved, further evaluation of the scheme in the future may be beneficial. This is likely to vary by the type of intervention and the clinical speciality, but is likely to be several years.

2. A balance needs to be made between making top-down recommendations about the interventions Trusts ought to implement, and giving Trusts autonomy regarding which interventions to implement in response to local needs. However, interventions in this and similar schemes should be based on good evidence of effectiveness and cost-effectiveness. We concur with the 2016 NHS England National Maternity review recommendation: *'Most importantly, any training undertaken must have been proven to be effective in improving outcomes or other aspects of quality, and its impact monitored locally'.*

3. Future schemes could include establishing evidence on the main causes of errors, and signposting effective interventions with support for local measurement and regional

benchmarking of clinical outcomes (A&E, maternal and neonatal outcomes), process measures (interventions), and implementation (e.g., proportion of staff trained).

4. Recognition need to be given to the timescale between allocation of funding and implementation time (e.g., implementation of intervention can be delayed by procurement time). Efforts should be made to ensure that Trusts awarded funding from this and similar schemes can act on that funding in a timely manner, so the interventions can be of maximum benefit to patients as soon as possible.

5. At present there is a considerable amount of data available on errors and claims but more effort is required to coordinate these data, improve their reliability and reduce under-reporting.

6. To assist with reporting, investment in providing integrated, flexible, efficient and user-friendly IT systems is needed to bring all Trusts up to a minimum standard. This will allow data to be collected in a timely, accurate, complete and reliable way.

7. The NHS LA could usefully partner with recognised academic, Improvement Science and clinical groups to improve the selection, implementation and evaluation of improvement initiatives in the future.

References

AI Girsen, MB Greenberg, YY El-Sayed, H Lee, B Carvalho and DJ Lyell (2015). Magnesium sulfate exposure and neonatal intensive care unit admission at term; Journal of Perinatology 35:181–185

Anderson A (2013). Ten years of maternity claims: an analysis of the NHS Litigation Authority data – key findings. Clinical Risk;19 (1):24-31.

Azzopardi DV et al (2009). Moderate hypothermia to treat perinatal asphyxia encelopathy. N Engl J Med; 361(1):1349-58

Baker C. Accident and Emergency Statistics: Demand, Performance and Pressure. Briefing paper umber 6964. November 2016. House of Commons Library.

Beckley S, Stenhouse E, Greene K. (2000).The development and evaluation of a computer-assisted teaching programme for intrapartum fetal monitoring. BJOG;107(9):1138–44.

Bergh A-M, Baloyi S, Pattinson RC. (2015). What is the impact of multi-professional emergency obstetric and neonatal care training? Best Practice & Research Clinical Obstetrics and Gynaecology; 29(8):1028-43.

Berglund S, Grunewald C, Pettersson H, Cnattingius S. (2008). Severe asphyxia due to delivery-related malpractice in Sweden 1990–2005. BJOG;115:316–323.

Blotkamp A, Draycott T, Lenguerrand E, Yau C, Winter C, (2016). Report on the feasibility study for the provision of maternity litigation indicators PROMPT. Work in progress.

Braithwaite J, Wears RL, Hollnagel E. From Safety-I to Safety-II: A White Paper. University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia; 2015 Aug pp. 1–43.

Crofts J, Lenguerrand E, Bentham GL, Tawfik S, Claireaux HA, Odd D, et al. (2015). Prevention of brachial plexus injury-12 years of shoulder dystocia training: an interrupted time-series study. BJOG: An International Journal of Obstetrics & Gynaecology;123(1):111–8.

Dixon-Woods M, Baker R, Charles K, Dawson J, Jerzembek G, Martin G, et al. (2014). Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study. BMJ Quality & Safety; 23(2):106-15.

Dixon-Woods M, Martin GP (2016). Does quality improvement improve quality? Future Hosp J. Royal College of Physicians; 2016;3(3):191–4.

Draycott T (2016). Not all training for obstetric emergencies is equal, or effective. BJOG: An International Journal of Obstetrics & Gynaecology (to be published).

Draycott T, Collins K J, Crofts JF, Siassakos D, Winter C, Weiner CP, et al. (2015). Myths and realities of training in obstetric emergencies. Best Pract Research Clin Obstet Gynaecol.;29(8):1067-76

Draycott T, Crofts J, Ash JP, Wilson LV, Yard E, Sibanda T, et al. (2008). Improving neonatal outcome through practical shoulder dystocia training. Obstet Gynecol.;112(1):14–20.

Draycott T, Sibanda T, Laxton C, Winter C, Mahmood T, Fox R. (2010). Quality improvement demands quality measurement. BJOG: An International Journal of Obstetrics & Gynaecology.; 117(13):1571–4.

Draycott T, Sibanda T, Owen L, Akande V, Winter C, Reading S. et al. (2006). Does training in obstetric emergencies improve neonatal outcome? BJOG: An International Journal of Obstetrics &Gynaecology. 113(2):177-82.

Espinosa JA, Nolan TW. (2000). Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. BMJ;320(7237):737-740.

Fawsitt C, BourkeJ, Greene R, Everard C, Murphy A, Lutomski J (2013). At What Price? A Cost-Effectiveness Analysis Comparing Trial of Labour after Previous Caesarean versus Elective Repeat Caesarean Delivery. PLoS ONE 8(3): e58577.

Fransen AF, van de Ven J, Schuit E, van Tetering A, Mol BW, Oei SG. (2016). Simulation-based team training for multi-professional obstetric care teams to improve patient outcome: a multicentre, cluster randomised controlled trial. BJOG: An International Journal of Obstetrics & Gynaecology; 123(11):1753-60

Gerova V, Griffiths P, Jones S, Bick D. (2010). The association between midwifery staffing and outcomes in maternity services in England: observational study using routinely collected data. Preliminary Report and Feasibility.

Graham EM, Ruis KS et al. (2008). A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy AmJObsGyn; 199(6):587-95.

Guly HR. (2001). Diagnostic errors in an accident and emergency department. Emerg Med J.; 18(4):263-269.

Gustafson P, Schultz T, Stefansdtottir A, (2014). Related infections shall be stopped –a national, interdisciplinary collaboration for safer prosthetic knee and hip operations. Ljungbergs Tryckeri AB.

Hamzeh Khazaei, Carolyn McGregor, J Mikael Eklund, Khalil El-Khatib (2015). Real-Time and Retrospective Health-Analytics-as-a-Service: A Novel Framework. JMIR Med Inform;3(4):e36

Hardy M, Hutton J, Snaith B. (2013). Is a radiographer led immediate reporting service for emergency department referrals a cost effective initiative? Radiography;19(1):23-27.

Hogan L, Ingemarsson I et al. (2007). How often is a low 5min Apgar score in term newborns due to asphyxia? EurJObsGyn 130(2):169-75.

Howell AM, Burns EM, Bouras G, Donaldson LJ, Athanasiou T, Darzi A. (2015). Can Patient Safety Incident Reports Be Used to Compare Hospital Safety? Results from a Quantitative Analysis of the English National Reporting and Learning System Data. PLoS One.;10(12):e0144107.

Howell C, Grady K, Cox C. (2007) Managing Obstetric Emergencies and Trauma. The NMOET Course Manual. 2nd edition. Cambridge University Press.

Kind P, Hardman G, and Macran S. UK population norms for EQ-5D. York. Centre for Health Economics, 1999. Available from: http://www.york.ac.uk/che/pdf/DP172.pdf

Knight HE, van der Meulen JH, Gurol-Urganci I, Smith GC, Kiran A, Thornton S, et al. (2016). Birth "Out-of-Hours": An Evaluation of Obstetric Practice and Outcome According to the Presence of Senior Obstetricians on the Labour Ward. Myers JE, editor. PLoS Med.;13(4):e1002000–15.

Lee C, Bleetman A. (2004). Commonly missed injuries in the accident and emergency department. Trauma;6(1):41-51.

Leigh S, Granby P, Turner M, Wieteska S, Haycox A, Collins B. (2014). The incidence and implications of cerebral palsy following potentially avoidable obstetric complications: a preliminary burden of disease study. *BJOG: An International Journal of Obstetrics & Gynaecology*; 121(13):1720–8.

Lie KK, Groholt EK, Eskild A. (2010). Association of cerebral palsy with Apgar score in low and normal birthweight infants: population based cohort study. BMJ 341:c4990.

Luthander CM (2016). The national perinatal patient safety programme: the challenges of implementation and evaluation. BMC Health Serv Res.;12(1):274.

MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH (1998). Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma ;12(8):577-86.

Macrae C. (2016). The problem with incident reporting. BMJ Quality & Safety; 25(2):71-5.

Millde Luthander C, Källén K, Nyström ME, Högberg U, Håkansson S, Härenstam KP, et al. (2016). Results from the National Perinatal Patient Safety Program in Sweden: the challenge of evaluation. Acta Obstet Gynecol Scand.; 95(5):596-603.

National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management Fractures: diagnosis, management and follow-up of fractures NICE Guideline NG38. February 2016

National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management Fractures: diagnosis, management and follow-up of fractures. NICE Guideline NG38. February 2016

National Institute for Health and Care Excellence. Intrapartum Care. NICE Clinical Guideline. 2007;55(1):1–65.

National Maternity Service Review. "Better Births. Improving outcomes of maternity services in England". London, February 2016.

NHS Reference Costs (2014-15)

NHS LA. Ten Years of Maternity Claims. NHS Litigation Authority; 2012 Oct pp. 1–171.

NIHR Imperial Patient safety translational research centre (2016). NRLS Research and Development. www.imperial.ac.uk/media/...of.../IMPJ4219-**NRLS**-report_010316-INTS-WEB.pdf

Olofsson P, Ayres-de-Campos D, Kessler J, Tendal B, Yli BM, Devoe L. (2014). A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labor. Part II: the meta-analyses. Acta Obstet Gynecol Scand.;93(6):571–86.

Pagano E, De Rota B, Ferrando A, Petrinco M, Merletti F and Gregori D, (2010). An economic evaluation of water birth: the cost-effectiveness of mother well-being. Journal of Evaluation in Clinical Practice; 16:916–919.

Preston CA, Marr JJ, Amaraneni KK, Suthar BS. (1998). Reduction of "callbacks" to the ED due to discrepancies in plain radiograph interpretation. Am J Emerg Med.;16(2):160-162.

Regier DA, Petrou S, Henderson J, Eddama O, Patel N, Strohm B, Brocklehurst P, Edwards AD, Azzopardi D, (2010). Cost-Effectiveness of Therapeutic Hypothermia to Treat Neonatal Encephalopathy. Value in Health 13:695-702.

Ricca J. (2016). Limits of "Skills And Drills" Interventions to Improving Obstetric and Newborn Emergency Response: What More Do We Need to Learn? Global Health: Science and Practice; 4(4):518-521.

Royal College of Obstetricians & Gynaecologists. Green-top Guideline No.29. The Management of Third- and Fourth- Degree Perineal Tears. London: RCOG; 2015.

Schiariti V, Klassen AF, JS Hoube, A Synnes, S Lisonkova and SK Lee (2008). Perinatal characteristics and parents' perspective of health status of NICU graduates born at term; Journal of Perinatology 28, 368–376.

Shoushtarian M, Barnett M, McMahon F, Ferris J. (2014). Impact of introducing Practical Obstetric Multi-Professional Training (PROMPT) into maternity units in Victoria, Australia. BJOG: An International Journal of Obstetrics & Gynaecology; 121(13):1710-8.

Sibanda T, Fox R, Draycott T, Mahmod T, Richmond D, Simms RA, (2013). Intrapartum care quality indicators: a systematic approach of achieving consensus. European Journal of Obstetrics & Gynecology and Reproductive Biology 166:23-29.

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Tan JT, Macario A, Carvalho B, Druzi ML, El-Sayed YY, (2010). Cost-effectiveness of external cephalic version for term breech presentation; BMC Pregnancy and Childbirth;10:3; http://www.biomedcentral.com/1471-2393/10/3

Tellesen L (2016). PhD Thesis. Submitted to University of Copenhagen.

The Health and Social Care Information Centre, Hospital Episode Statistics for England, Accident and Emergency (A&E) statistics, 2014-15.

Thomas HG, Mason AC, Smith RM, Fergusson CM. (1992). Value of radiograph audit in an accident service department. Injury;23(1):47-50.

Thorngren-Jerneck K, Herbst A. (2001). Low 5-minute Apgar score: a population-based register study of 1 million term births. Obstet Gynecol.;98(1):65–70.

Turner CE, Young Y, Solomon MJ, Ludlow J, Benness C, Phipps H (2008). Vaginal delivery compared with elective caesarean section: the views of pregnant women and clinicians. BJOG 115 (12):1494-1502.

Yau CWH, Pizzo E, Morris S, Odd DE, Winter C, Draycott TJ. (2016). The cost of local, multi-professional obstetric emergencies training. Acta Obstet Gynecol Scand.; 95(10):111-9.

Websites

http://www.england.nhs.uk/signuptosafety/
http://www.nrls.npsa.nhs.uk/; https://report.nrls.nhs.uk/nrlsreporting/
http://www.clevermed.com/badgernet/badgernet-maternity/
https://www1.imperial.ac.uk/neonataldataanalysis/data/;
http://www.rcpch.ac.uk/improving-child-health/quality-improvement-and-clinical-audit/national-neonatal-audit-programme-nn-3
http://www.datix.co.uk/
http://content.digital.nhs.uk/catalogue/PUB16725
http://content.digital.nhs.uk/hes
https://www.england.nhs.uk/statistics/statistical-work-areas/friends-and-family-test/friends-and-family-test-data/
https://www.npeu.ox.ac.uk/mbrrace-uk
https://www.rcog.org.uk/en/guidelines-research-services/guidelines/good-practice-7/
https://www.rcog.org.uk/eachbabycounts
http://www.gettingitrightfirsttime.com/
https://www.strokeaudit.org/
https://www.rcplondon.ac.uk/projects/national-copd-audit-programme-starting-2013
http://lof.se/other-languages/
http://www.sbu.se/en/publications/sbu-assesses/therapeutic-hypothermia-in-fullterm-infants-affected-by-birth-asphyxia/; 25th Feb 2009
https://www.rcog.org.uk/globalassets/documents/patients/patient-information-leaflets/pregnancy/pi-an-assisted-vaginal-birth-ventouse-or-forceps.pdf;
http://www.publications.parliament.uk/pa/cm200809/cmselect/cmhealth/151/15108.htm

Appendices

1. Description of the SU2S incentivisation scheme process.
2. Questionnaire to capture Trusts impression of the SU2S scheme.
3. What Trusts think of the SU2S incentivisation scheme?
4. Questionnaires sent to Trusts to collect data on problems to be addressed, interventions proposed, funding received and outcome measures (end of year reports).
5. List of interventions and outcome measures: reducing intrapartum harm.
6. List of interventions and outcome measures: missed fractures in A&E.
7. List of interventions and outcome measures: maternity and A&E common interventions.
8. Data collection questionnaire sent to Trusts to measure outcomes in A&E.
9. Data collection questionnaire sent to Trusts to measure outcomes in maternity units.
10. Details of the evaluation of the incentivisation scheme in maternity units.

Appendix 1. Description of the SU2S incentivisation scheme process

NHS LA Bid Incentivisation Process in Support of Sign up to Safety

The following outlines the different steps and timing of these involved in the roll out the Scheme

1. Date of invitation to apply

The Sign up to Safety Campaign was launched on 24 June 2014. The NHS LA advised its members that it was delivering a bid incentivisation scheme on behalf of the Department of Health to support the Sign up to Safety initiative. The deadline for receipt of the bids to be forwarded from the Sign up to Safety Campaign team was 16 January 2015

2. Application process and Time scales

The first 12 Trusts to sign up to Sign up to Safety, known by the campaign as "trailblazers", were informed about the bid scheme in July 2014. The second tranche roll out to the rest of the country took place in September 2014. Members could only apply to the scheme if they had signed up to safety through the campaign. The claims incident scorecards were produced to support the analysis of claims profile for the Trusts and the criteria for the scheme was provided to Trusts to complete (see last page).

3. Judging process

Bids against set criteria were assessed via processes set up by the NHS LA using a cross directorate approach (Safety and Learning, Claims and Finance):

They established the following committees for this process:

- Steering Group for Review of Safety Improvement Plans with bids (SIPs) from Sign up to Safety (cross directorate)
- Approvals Committee (Executive)

4. NHS LA internal governance systems

- All bids were submitted to the Sign up to Safety Campaign office. The NHS LA requested all bids which had been received by 16 January 2015 to be forwarded on to the NHS LA – these were all received by the NHS LA the week beginning 19 Jan 2015
- All bids were reviewed by the Safety and Learning team, supported by the Claims and Finance sections of the NHS LA
- The combined teams reported into a Sign up to Safety steering group
- Outcome of these review scores were presented to the NHS LA Approvals Committee and allocation of the funding to the successful organisations was agreed.

5. Number of bids received and awarded funds

The NHS LA received 249 bids from 114 member Trusts across a wide geographical reach by the time of the closing date on 16 January 2015. The bids covered a range of specialties. The process of assessing bids was undertaken from 16 January- mid March 2015. Compliance with the NHS LA bid criteria was scored (met, partially or fully met) recorded for individual questions as well as additional qualitative information added to determine a threshold of scoring agreed by the Steering

Group. 67 bids were approved for funding from a total of 47 Trusts (some Trusts submitted multiple bids).

The top five focus areas for successful bids were:

1. Maternity- purchase of CTG electronic monitors, recruitment for "second pair of eyes", central CTG monitoring and alert systems, remote access to tracings, training.
2. Safety Culture – a range of human factors and cross cutting areas.
3. Surgical – includes training and equipment, human factors training in a number of surgical specialties particularly orthopaedics and neurological surgery.
4. A&E – improving missed and delayed diagnosis, diagnostics, "hot" radiography reporting in 24 hours, performance feedback on missed diagnoses.
5. Deteriorating patient – early and improved recognition, electronic flags, improved management pathways.

6. Details of verification/review process proposed by NHS LA

The awarding of funding was subject to a number of conditions from the submitting Trusts' CEOs as stated below. The monitoring of the plans was the responsibility for the submitting Trusts, who were required to share learning with the NHS LA Safety and Learning team.

1. Confirmation to be provided that the funds allocated to be used only in relation to the submitted bid;
2. The Trust is asked to publish a summary of their successful bid, including details of the anticipated outcomes, on their public website;
3. The Trust will provide details of their successful bid(s) to their Trust Board and their local commissioners and provide regular updates on the monitoring of their progress;
4. The Trust will provide feedback and share safety and learning themes with external partners and directly with the Safety and Learning team at the NHS LA;
5. The Trust will agree to collaborate with the NHS LA and Royal Colleges in the progress of implementation of the bid and in particular for all maternity bids with relevant Royal Colleges as regards maternity claims and outcomes from the bid. More details will follow.
6. The Trust will agree to 'buddy' with an unsuccessful bidder in terms of sharing best practice to support quality improvements to those requiring additional support;
7. The Trust will agree to coordinate with Trusts requesting the same specific equipment or training to ensure procurement benefits from economies of scale and value for money – NHS LA will be in contact to provide details of those Trusts with shared purposes, equipment and training.

7. Dates Trusts received funding

Payments of bid funding were made to successful Trusts during the period May-September. The different dates of these separate transfers related to when commitments were received from CEOs to the conditions set for receiving the funds as above. Trusts were then responsible for allocating the funds to budget holders for expenditure.

Table 1. Criteria scoring sheet

CRITERIA SCORING SHEET – INCLUDED ON SHAREPOINT FOR CROSS-DIRECORATE SCORING				
Criteria for scoring of Safety Improvement Plans for uploading onto Sharepoint (January 2015) Appendix 1				
	Information for Informatics on Sharepoint areas to include			
	Summary Page	Text boxes	Scoring	
	Full Name of Organisation and T number			
	Date of Bid Submission			
	Summary of Bid to NHS LA			
	Specialties related to the Bid			
	Sum requested and is this within 10%			
	Was External review required			
	Sign up to Safety Steering Group Recommendations to the Approvals Committee			
	Recommend Yes/No			Y/N
	Date Agreed			
	SLT Section to Complete			
1	Provide a summary of comments from Sign up to Safety			
2	Is the Bid signed off by the Executive Sponsor			Y/N
3	Area of Focus: Does the bid adequately describe why the member has selected this area to focus on in relation to claims		out of 5	
4	Goal: Does the Bid articulate how claims could be reduced		out of 5	
5	Timing: Are implementation timelines sufficiently detailed within the bid		out of 5	
6	Actions: Are actions sufficiently detailed in the bid		out of 5	

Table 1. Criteria scoring sheet (continued)

7	Benefits: Do expected outcomes represent a reduction of the identified harm being addressed by this bid		out of 5	
8	measurement: Does the plan state: what will be measured When it will be measured By whom How these measures will be monitored and reported		out of 5	
	Quality and Safety (input by SLT)			
9	Evidence of clinical involvement in the development of the plan	Free text		
10	Cross cutting themes across several specialties	Free text		
11	Impact on patient safety culture for the member	Free text		
12	Evidence of patient engagement in either informing or supporting the plan	Free text		
13	Reference to published evidence base or guidance	Free text		
14	Innovative/reflects an approach from which other members would benefit	Free text		
15	Reference to alignment with other stakeholders and professional bodies e.g. AHSNs	Free text		
	Completion by Claims			
16	Quality: Does the plan describe how it aligns with the member's claims profile			
17	Quality: From your knowledge of this member, does this plan address claims in the way you would expect		0	
18	Number/value Has the member articulated the numbers and values of claims associated with this focus		out of 5	
	For Completion by Finance			
19	How much money has been requested	Figure to be inserted	0	
20	Contribution: is the requested amount within 10% of the member's contribution (general which includes non clinical or maternity)	Free text		Y/N

Table 1. Criteria scoring sheet (continued)

21	Do the figures add up?	Free text		Y/N
22	Does the budget accurately reflect breakdowns of: Staff Equipment Training Any other type of investment		Out of 5	
23	Financial data/timings: Has the member detailed how it would spend the money in the fiscal year 2015/2016 with financial milestones		Out of 5	

Scoring rationale
0=not met
1=partially met by weak
2=partially met and adequate
3=completely met and adequate
4=completely met and good
5=completely met and exceptional/innovative

Did the team have to have to go to member for further information	Date
Date requested information received	

Date of Review by Approvals Committee	
Agreement with SUTS Steering Group Recommendations and Rationale for Decision	
Non Agreement with SUTs Recommendations and Rationale	
Other Qualifying Comments	
Award Agreed	Y/N
Amount of Award agreed	
Rationale for level of funding/(formula)	
Outcome letter to Member	

Appendix 2. Questionnaire to capture Trusts impression of the SU2S incentivisation scheme

Thank you for taking part in the evaluation of the Sign up to Safety Scheme. We think that your views of the scheme are really important and should be taken into account for the future.
Could you please answer the following questions so that we can include a paragraph in the final report with a summary of the Trusts' impressions? (the answers provided here will remain anonymous and we will ensure that your identity will not be recognised). Thank you.

1) Overall, what do you think are the main positive aspects of the Sign up to Safety scheme?

Answer here

2) What do you see as the main limitations of the scheme?

Answer here

3) What would you like to change (either from the NHS LA end, or from your end) if the scheme was going to be run again?

Answer here

4) What are the key messages you would like to send to the DoH regarding the Sign up to Safety scheme?

1. … 2. … 3. ….

Appendix 3. What Trusts think of the SU2S incentivisation scheme?

Positive aspects of the Sign up to Safety Scheme

Funding and monetary aspects

- The scheme financially supported the organisation in taking forward three focussed safety projects. It would not have been possible to take forward the schemes without the additional funding allocated.
- The NHS LA funding has allowed for investment to go directly to front line services to improve patient safety
- Being able to have NHS LA money and thereby resource to make the changes that will make a real difference to patient outcome without having to wrangle with commissioners to make the work a funded CQUIN. Has enabled us to find the amount of money that we would never be able to find within the Trust to do this work.
- Providing cash to enable implementation of steps that can improve safety.
- Funding – receiving external funding was essential in implementation of SU2S. In the current economic climate finding additional resource is extremely difficult. In relation to reducing intrapartum harms in Obstetrics demonstrating a reduction in neonatal brain injury will take years to evaluate as claims come into the Trust years after the birth and the main milestone for longer term predictors of outcome and degree of damage cannot be assessed until 2 years of age in the child. Not being able to demonstrate quick wins and timely improvements/evaluations when presenting business cases to the Trust can be challenging when there are so many demands on resources from all disciplines and whilst it is anticipated the improvements will achieve the desired effect there is no immediate reassurance that can demonstrate its success.

Safety aspects and outcomes

- This scheme has enabled us to get organised around safety, to engage better with staff and to focus the Board to some safety improvement work
- I think having the themes for action clearly identified helps engender a consistent approach and gives organisations a clear direction. In the current financial situation, having safety so clearly at the top of the agenda focuses resources.
- National focus on improving outcomes in conjunction with the ability for Trusts to focus on projects that are specific to their service users thus increasing the potential for improved outcomes'.
- The campaign has supported us to focus on a key area of concern for which improvements will have a significant impact on identifying patient harm in a high flow patient area.
- Having only purchased the software this year we are still waiting for some of the evidence to come in, but the overall impression is that it has been very supportive of changes that were required by our teams to support patient safety. The immediate impact is raising the monitoring and reporting of the initiative to reporting Committees within the Trust on a quarterly basis.
- Raising awareness and focusing teams on steps that can improve safety
- Safety becomes the norm, embedding this in everyone's practise and in everything we do, ensuring the patient (quality and safety) remain the focus
- The focus is on patient safety and staff wellbeing and the emphasis on creating a positive learning culture to reduce avoidable harm and focus on organisational culture
- It recognises that changes and improvements are necessary to reduce avoidable harm and save patients' lives and it has started the debate of the requirements locally and nationally needed to allow this to happen.

- It is creating an environment where learning and reflection on what could be done differently or better can take place safely to reduce avoidable harm which will benefit both patients and staff
- A large majority of incidents and claims are due to human factors – specifically the actions and behaviours of those looking after patients. Being given the opportunity to understand how and why medical errors occur and the contributions made by poor communication/behaviours and having the funding to take targeted focussed action is a step to reinforcing patient and staff experiences and outcomes.

Communication and support

- Webinars good
- Good communications from the central team via webinars and weekly emails
- Overarching support and help with the spread of the campaign via social media (e.g., Twitter).
- Opportunities for publicity and reaching a wider audience.
- The link with the NHS LA was what made it different – and it will be interesting to see whether there is any evidence of impact on claims.
- In addition to supporting the local focus, we have used the signage and branding of the Sign up to Safety campaign as part of our Patient Safety Conference held last month and locally the Patient Safety Lead circulates any relevant Sign up to Safety campaign details to our clinical leads and subject specialists.
- Profile – having the SU2S campaign recognised as a national scheme and the fact that it is supported by the DoH has really helped the staff at ground level understand the investment and commitment to improvements. The fact that they can see, touch, benefit from the investment in the additional resources provides a direct link from them to the DoH. I think it is really valuable for the staff delivering the care to have that recognition and understanding. The midwives and obstetrician providing intrapartum care are passionate about delivering high quality care, they care about their patients and at that time become a part of their birth experience and bring new life into that family. So, when there is a poor outcome for a baby it is devastating for all and knowing that the DoH want to directly help us in reducing that provides a direct connection for them. I think at times the front line staff perceive the DoH to be responsible for 'cuts' in the NHS and this scheme has definitely provided a positive message.
- As a Manager myself it has also allowed me to sell the 'brand' of SU2S and discuss with staff how the campaign is being implemented across the country and also link in with the other SU2S streams in my own Trust and other Trusts, networking ideas and challenges.
- It allows Trusts to develop their own local ideas and plans to support the Scheme goals and aims
- It can give staff and patients the opportunity to have a voice and share their personal experiences in a safe environment

Learning and sharing with other organisations

- Encouraging sharing between organisations is really important
- Being able to learn from and share good practice and improvements nationally via Sign up to Safety updates, webinars etc.
- The scheme also helps promote multi-disciplinary working and shared learning which is always positive.
- The focus is on patient safety and staff wellbeing and the emphasis on creating a positive learning culture to reduce avoidable harm and focus on organisational culture
- It brings together staff at all levels across the NHS all facing similar challenges, issues and concerns

- It enables Trusts to share their ideas and approaches to tackle and make safety improvements
- It brings together people all who want to make a difference to patient and staff safety in a safe manner

Impact on staff

- The NHS LA finding for a Sign up to Safety Matron allowed for ring fenced staff and time to do work that is important (safety culture, Trust wide changes to handover practice) rather than getting tied up doing work that is urgent (fire fighting/quick fixes). It also allowed for a dedicated member of staff rather than the work being an add-on to another person's workload.
- Improved confidence in CTG assessment as shown by Survey Monkey questionnaire.
- Improved knowledge in CTG assessment by an intense program of learning through CTG master classes, weekly CTG meetings and daily bedside teaching.
- Improved multidisciplinary team work between midwives, obstetricians and neonatologists (HIE report, posters presented at the RCOG congress with multidisciplinary contributions)
- The focus is on patient safety and staff wellbeing and the emphasis on creating a positive learning culture to reduce avoidable harm and focus on organisational culture
- It brings together staff at all levels across the NHS all facing similar challenges, issues and concerns
- It can give staff and patients the opportunity to have a voice and share their personal experiences in a safe environment
- Running NHS LA events where staff have been given the chance to speak up and be listened to
- It has served as a reminder to show compassion and care for all those responsible for either delivering or receiving care

Limitations of the scheme

Funding and monetary aspects

- Financial support: the Trust was successful in receiving funds to support developments in three of our four identified areas. The fourth area was not successful in receiving funding and, while the initiative was taken forward, this was limited to what could be achieved within available resource.
- Funding was only made available for a short-time, therefore limiting the potential to observe substantial benefits.
- We were fortunate to receive NHS LA funding for our Sign up to Safety projects, but without this funding finding additional resources from within existing budgets would have been very challenging.
- It is non-recurrent funding from NHS LA bid and there has been no further calls for bids for monetary support.
- The scheme is time and cost-limited so momentum may fall away
- Lack of resources i.e. finance
- Generally there is a short timescale to apply for funds such as these which means that a case has to be pulled together at short notice. This can mean that what at the time seems a good idea is, after reflection, not quite so practical.

Sustainability

- One of the limitations of any external scheme supporting a local initiative is not necessarily maintaining the life of the scheme itself in the early stages, but ensuring that the practice change that has been facilitated by such schemes is embedded into everyday care – in other

words, it becomes 'business as usual'. This is not necessarily a criticism of the campaign itself, nor is it something that the campaign can directly influence with local Trusts, but it is something to be aware of in local implementation of schemes.

- I spent a large proportion of the funding we received on staffing and as the funding was a one of payment I am now going to put a business case into the Trust to retain those posts which may or may not be approved. In the event they are not approved as an increase in my establishment it may be difficult to sustain the improvements long term.

Learning from other organisations

- Headroom and capacity to really understand the learning from other organisations.

Organisation priorities

- Organisations have to ensure that the Sign up to Safety scheme fits in with their strategic objectives otherwise it will not get the Board support and fail.
- SU2S itself as a driver for patient safety was a bit vague. We already identify quality priorities and work relating to improving patient safety that we then had to badge as being part of SU2S. Although the aspirations of the scheme are good, it was the funding from NHS LA that enabled us to maintain a specific focus on the aspects of safety improvement that were included in our safety improvement plan and selected for funding. But alongside this we are progressing other patient safety work streams that are just as important.
- The need to bid for funding before work began meant that the original project scope did not necessary match what was actually needed
- There was no common issue to get behind – it was down to local priorities which we had already identified

Data collection

- I don't really see any limitations but there are challenges. Accurate comparable data being one challenge.
- Limited length for the type of project where main outcomes (HIE) are rare and need a minimum of 3 years to show significant change not explained by normal variation

Staff resources and engagement

- Capacity and engagement from the MDT (multidisciplinary team)….tends to be nurse led
- The limitations are more about the service pressures on the NHS, staffing and capacity issues facing NHS staff and their capacity to remove themselves from the day to day operations and participate in the local interventions being made available to staff and to have the space and capacity to learn and reflect on what could have been done better when harm or medical errors occurs. These are not limitations of the scheme but do have an impact on the scheme and can affect or delay the level of involvement from frontline staff who in reality wants things to get better and support improvements.

Proposed changes

Bidding process

- The Trust was successful in our bid for funds – however – the process felt onerous with detailed project plans being required in a short timeframe
- More support with the bid and project planning
- More of a push on the "Ask and offer" scheme may have helped.

Support and collaboration

- Deliver very focussed on the Trust schemes identified. Centrally – more support with initiatives to support learning – such as Root Cause Analysis training and safety events.
- Locally, running these kind of projects can be isolating, forming an effective team support structure is essential.
- Would be useful to have support from national team and /or academics to evaluate the impact of improvements made
- Clear links with the Regional PS collaboratives might prove helpful – i) in having local expertise to draw on and ii) so that the national team has 'boots on the ground' in all regions to support the scheme.
- A national virtual safety faculty would be a valuable resource.
- We would suggest that each Trust has a linked member of staff from your NHS LA team who can suggest ideas and challenge plans before they are submitted.
- Additional support for IT/ admin/ data collection
- Support for all organisations and consideration of a catalogue of proven safety initiatives to choose and embed
- More time given to Clinicians to be given time and space to focus on preventative actions and less on reactive systems and processes currently in place.

Funding and sustainability

- Investment for a longer period of time of 3 year
- Flexibility of the projects really helps to keep them responsive and effective.
- Funding for more than one year (recurrent funding) for safety improvement work programmes would be beneficial to enable embedding and sustainability of change
- I would have to say funding! There was initial funding and those initiatives which were successful in their bids did benefit.
- Access to funds to enable more projects/schemes
- The scheme to become integral to the way Trusts operate and at the heart of the way Trusts and various departments work more closely together to improve patient safety and staff wellbeing

Learning and sharing with other organisations and stakeholders

- Regarding the NHS LA incentivisation scheme funding, one Trust stated that it would have really helped with collaboration if the successful and unsuccessful Trusts could have been introduced to each other to assist with sharing lessons learnt. Seeking out a Trust' to "buddy" with was perceived as extremely time consuming.
- A further meeting of the successful organisations would have been in inspiring an informative. As far as I know there was only 1 national meet up
- Better collaboration between Trusts, maybe some networking events for those involved.
- Improved networking with other units running the same type of projects
- More patient involvement in the scheme and events run by the NHS LA
- More partnership working between Trusts and collaborative approach to tackling similar areas for improvement within Trusts based on themes identified by NHS LA scorecard data

Type of interventions and their selection

- Our experience has been that very specific clinically based projects are easier to "pin down" and bring about change in the time span. Those that involve broader concepts are much harder to get a grip on.
- More input from NHS LA colleagues to work with the organisations to identify claims themes and required focus of work and triangulate with other data sources.

- It would be useful to focus on a key aspect of patient safety that needs improving across the board and direct support at that – particularly if there is anything that comes out of this scheme that provides evidence of impact on claims.
- If to run again central monitoring is the priority in terms of equipment to improve safety, documentation and as a learning/ teaching tool
- Support for all organisations and consideration of a catalogue of proven safety initiatives to choose and embed
- More identification of interventions which have so far worked, the issues and challenges and how they have overcome them to increase participation and engagement in the scheme

Outcomes, evaluation and impact awareness

- The outcomes of the scheme itself focussed primarily on the process outcome (e.g., the management of missed fractures) and the secondary outcome of litigation prevention in the Trust. For the future, it might be helpful to understand the health economic impact of such interventions which may be larger than just the locally agreed measures, and which might provide some evidence as to why the scheme should continue to be locally sustainable by both providers and commissioners. Too often patient safety initiatives rely solely on the individual desire not to do harm, or the negative connotation of avoiding litigation or prosecution. In the longer term, for work around the avoidance of adverse events to become commonplace, one opinion is that this needs to be linked to productivity and economics.
- I might have liked to have seen the evaluation strategy at the point we put the bid in. My personal thoughts are that every Trust has implemented a project with the same aim but are trying to achieve that in so many different ways and I do not think that can be reflected in arbitrary figures of outcomes within the first 2 years. It will be difficult to use quantitative measures alone to evaluate the impact of individual Trusts success with this scheme.
- I think each Trust should be asked to submit a report at the 3 year point with their individual measures of success which will include other quantitative and qualitative measures which they feel are pertinent in demonstrating the impact of SU2S in their particular Trust for their particular population.

Key messages for the Department of Health regarding the scheme

Positive messages

- Overall, as a Trust that was successful in receiving funds to take forward specific projects – the scheme has been successful
- Great leadership and great engagement via the weekly emails, blogs, twitter feeds, webinars, tool kits etc. Thank you!
- Learning from each other and sharing improvement work nationally is key for the NHS
- Really useful scheme
- This scheme has been very well received across the Trust, the focus on local ownership helps keep the front line staff engaged. It's not about "the management said we should do this", it's about looking at what is really happening, listening to why staff feel it is the way it is and what they would like to do instead then helping to facilitate that change with local ownership and ward level leadership to ensure sustainability.
- Having the opportunity to bid for this type of funding and receiving it boosts morale of a team, allows implementation of change that the team may not have been able to secure through their own Trust due to competing demands for financial investment, and also sends a strong message to frontline staff that the DoH recognise and care about the work they do on a daily basis and the outcomes for their patients. I don't for one minute think that that DoH haven't cared but with so much focus in the media on negative effects of 'cuts in the

NHS' it is really nice for the staff to tangibly feel the investment SU2S has provided and the positive impact that has had on our particular Maternity services, staff and patients alike.
- The scheme is definitely worthwhile as it serves as reminder to put patient and staff safety first and allows Trusts to tackle their local issues with local ideas and interventions
- For organisations and NHS staff to truly learn from patient harm and medical errors it is important to create a safe environment which is less performance based and more values based to allow staff and patients to work more collectively towards creating an open learning service and provide better care and treatment as a result

Funding and sustainability
- Keep going with the scheme
- For unfunded projects, arguably, the scheme was less successful with the need to balance delivery alongside other national safety initiatives within available resources
- A great way of funding different approaches to deliver front line services to improve patient safety, it's a shame the funding only lasts a short period of time
- Good support and publicity but limited effectiveness without funding
- Need more investment to improve outcomes in maternity services as many Maternity Units will not have been successful in their bids for funding.
- Think about how the scheme can influence long term sustainability of changes locally.
- Having the opportunity to apply for external funding is vital in being able to implement changes like these. Every unit knows their own challenges and areas for focus and improvements but very few sustainable changes comes without any additional resource. It is extremely difficult with existing budgets to maintain the status quo and also invest in the future, particularly when trying to demonstrate a reduction in something that will take years to really be able to evaluate its impact.

Challenges and suggestions
- For real national working and avoidance of duplication, NHS organisation need a "push"
- Changing culture is really hard to measure, compared to clinical projects where you can count numbers
- Let organisations decide on what they need to focus on locally instead of producing a set of national standards for us to work to ie NHS LA standards.
- It would be useful to focus on a key aspect of patient safety that needs improving across the board and direct support at that – particularly if there is anything that comes out of this scheme that provides evidence of impact on claims.
- Support wider understanding within Trusts of the health economic impact of adverse events and the intrinsic value of a positive safety culture.
- It would be beneficial to have more collaboration between Trust's and the NHS LA
- Make it easy for teams to share what worked well for them without having to attend meetings (e.g., Webinar)
- Sign up to safety made a bigger impact than the one we can quantify in terms of team work, enthusiasm and a change in culture. We would like to thank you for this opportunity and encourage the scheme to continue
- Although safety is crucial the governance structures and skills are also crucial. Would like to see some specific support for governance leads and some type of benchmarking/gap analysis
- Simplifying the processes and constant request for information, setting targets and monitoring performance which in a way have become more important than the people (receiving & delivering care) and inhibits staff to who just want to do a good job of delivering good quality care and treatment.

- The scheme should continue and be sustained to allow enough time for organisations to grow and become preventative learning hubs, embedding the safety improvements within their culture. It should not just as another scheme but needs to become part and parcel of everyday roles, practice and plans and integral to everyone's way of working in the NHS.

Appendix 4
Questionnaires sent to Trusts to collect data on problems to be addressed, interventions proposed, funding received and outcome measures.

NHS Litigation Authority

NHS Litigation Authority incentivisation scheme within the context of Sign up to Safety

Yearly Sign up to Safety bid report 2016 – completion of a separate form is required for each successful bid

Name of Trust:		
Title of success bid:		
Question 1		
What is the specific problem you are addressing as part of the NHS LA incentivisation scheme?		
Question 2		
How/on what are the funds you have been allocated via incentivisation scheme being spent? For example, what are the key actions and interventions you have implemented with the funds? How much of the funding has been allocated to each activity?	a) £ _____ description of action/intervention	
	b) £ _____ description of action/intervention	
	c) £ _____ description of action/intervention	
	d) £ _____ description of action/intervention	
	(if you do have other actions/interventions please add them here)	

Question 2 (follows)	
How/on what are the funds you have been allocated via incentivisation scheme being spent? For example, what are the key actions and interventions you have implemented with the funds? How much of the funding has been allocated to each activity?	
Question 3	
Have you shared details of the successful bids with your Trust Board and local commissioners?	Yes/No
If yes, please provide detail of information sharing with your Board and when this took place?	
If no, please describe plans for doing this	
Question 4	
Have you been able to buddy up with an unsuccessful trust which bid in the same area and if so, which one?	

Question 5		
When was the intervention /action described in number 2) implemented? (Please state the month and year for each action/intervention if not happening at the same time).	a) month/year	
	b) month/year	
	c) month/year	
	d) month/year	
Question 6		
Is there any intervention that you initially planned to implement but has not started? Please describe the intervention below and state the main reason why it was not possible to implement it.	Intervention not started	
	reason for not implementing it	
	Intervention not started	
	reason for not implementing it	
	Intervention not started	
	reason for not implementing it	

Question 7	
How are you evaluating whether or not the interventions/actions you have described in number 2) have made a difference to patient safety? (describe for each action/intervention using answers in question 2)	a)
	b)
	c)
	d)
	(if you do have other actions/interventions please add them here)

Please use the table 1 below to describe any outcomes measures/data you are collecting to provide evidence of this impact (e.g., reduction in number of incidents and litigation claims, improvement in patient health outcomes, etc.)

Question 8	
What difference have the outcomes made to patient safety?	

Please use the table to answer question 9

Question 9: Are the outcome measures/data described in number 8 routinely collected in your Trust? Are these data available prior to the introduction of the incentivisation scheme? If so, when did you start collecting these outcome measures/data?

Table 1 Outcome measures

Outcome measure (description) *(ex. Reduction of incidents)*	**Are these outcome measure data routinely collected in your Trust?** *(yes or no)*	**Are these data available before the introduction of the scheme?** *(yes or no)*	**When did you start collecting the data? (month and year)** *(ex. January 2003)*

Question 10	
Have you shared these on your public website?	Yes/No
If no, why not and when will you do this?	
Further comments	
Please provide any narrative you wish to add to describe the implementation of the bid e.g. this may be text used in reporting to your Board or in any other Trust communications.	

Appendix 5. List of interventions and outcome measures: reducing intrapartum harm

Trust *	Intervention	Intervention ID	Outcomes Monitored
Trust 2	Fetal monitoring equipment	2	N. of falls, n. of hospital acquired pressure ulcers, Summary Hospital-level Mortality Indicator (SHMI), formal complaints
	Training and development **for fetal monitoring**	8	N. of falls, n. of hospital acquired pressure ulcers, Summary Hospital-level Mortality Indicator (SHMI), formal complaints
Trust 3	Q&S improvement lead / nurse (staff) **Intrapartum clinical practice educator**	5	% babies born with an APGAR less than 7 at 1 minute, % term babies admitted to the NICU
Trust 4	Computer Hardware/Software	1	NA
	Fetal monitoring equipment	2	NA
Trust 5	Fetal monitoring equipment **STAN**	2	Reported incidents and associated harm, n. of falls, emergency CS rate, unexpected term admissions to the NNU
	Q&S improvement lead / nurse (staff) **2 Midwives**	5	Reported incidents and associated harm, n. of falls, emergency CS rate, unexpected term admissions to the NNU
Trust 6	Q&S improvement lead / nurse (staff) **midwife**	5	% 3rd/4th degree tear
	Specialised obstetric equipment **episiotomy scissors**	7	% 3rd/4th degree tear
Trust 7	Q&S improvement lead / nurse (staff)	5	N. of stillbirths and early neonatal deaths, n. of babies transferred for active cooling
	Fetal monitoring equipment **USS capacity and equipment**	2	N. of stillbirths and early neonatal deaths, n. of babies transferred for active cooling
	Training and development **K2 training (fetal monitoring)**	8	N. of stillbirths and early neonatal deaths, n. of babies transferred for active cooling
	Computer Hardware/Software	1	N. of stillbirths and early neonatal deaths, n. of babies transferred for active cooling
Trust 8	Computer Hardware/Software **electronic obs**	1	Sepsis related incidents, % IV antibiotics for sepsis, SHMI, reported incidents
	Training and development CTG **training video/human factors**	8	Sepsis related incidents, % IV antibiotics for sepsis, SHMI, reported incidents
Trust 11	Q&S improvement lead / nurse (staff) **Midwives and consultant time**	5	NNU unexpected admissions, HIE, intrapartum interventions like FBS, emergency caesarean section rates
	Q&S improvement administrator (staff) **Management**	4	NNU unexpected admissions, HIE, intrapartum interventions like FBS, emergency caesarean section rates
	Training and development **CTG masterclass and training**	8	NNU unexpected admissions, HIE, intrapartum interventions like FBS, emergency caesarean section rates

* Trusts have been anonymised and numbered using the coding for the analysis

Appendix 5. List of interventions and outcome measures: reducing intrapartum harm (continued)

Trust *	Intervention	Intervention ID	Outcomes Monitored
Trust 12	Training and development **CTG training package**	8	Stillbirths rate, perinatal morbidity, compliance with guidelines, n. women who stopped smoking
	Fetal monitoring equipment **CTG equipment**	2	Stillbirths rate, perinatal morbidity, compliance with guidelines, n. women who stopped smoking
	Computer Hardware/Software **IT system for ANC and postnatal care**	1	Stillbirths rate, perinatal morbidity, compliance with guidelines, n. women who stopped smoking
	Patient monitoring equipment **More USS machines**	3	Stillbirths rate, perinatal morbidity, compliance with guidelines, n. women who stopped smoking
Trust 13	Q&S improvement lead / nurse (staff) **Midwives and consultant**	5	N. babies diagnosed with HIE, n. babies born with a low cord gas of less than 7.0 (i.e. requiring admission to neonatal unit)
	Q&S improvement administrator (staff)	4	N. babies diagnosed with HIE, n. babies born with a low cord gas of less than 7.0 (i.e. requiring admission to neonatal unit)
	Fetal monitoring equipment **handheld dopplers and central monitoring**	2	N. babies diagnosed with HIE, n. babies born with a low cord gas of less than 7.0 (i.e. requiring admission to neonatal unit)
	Training and development **fetal monitoring training**	8	N. babies diagnosed with HIE, n. babies born with a low cord gas of less than 7.0 (i.e. requiring admission to neonatal unit)
Trust 14	Computer Hardware/Software **K2 software and hardware**	1	NA
Trust 17	Fetal monitoring equipment **central monitoring**	2	Incidence of claims due to HIE, admission temperatures of baby's being transported between hospitals
	Q&S improvement lead / nurse (staff) **midwives**	5	Incidence of claims due to HIE, admission temperatures of baby's being transported between hospitals
	Neonatal transport equipment	9	Incidence of claims due to HIE, admission temperatures of baby's being transported between hospitals
Trust 18	Computer Hardware/Software **IT system**	1	N.of claims, n. 3rd degree tear
	Fetal monitoring equipment **Wifi Sonicaids (CTG)**	2	N.of claims, n. 3rd degree tear
Trust 20	Fetal monitoring equipment **K2 central monitoring**	2	N. intra-partum stillbirths, unexpected admissions of term infants to the NNU, n. of claims
	Computer Hardware/Software **K2 electronic storage of CTG**	1	N. intra-partum stillbirths, unexpected admissions of term infants to the NNU, n. of claims
	Training and development **K2 training**	8	N. intra-partum stillbirths, unexpected admissions of term infants to the NNU, n. of claims
Trust 21	Q&S improvement lead / nurse (staff) **Midwife focus on CTG and sepsis**	5	Reduction in reported incidents, improved training compliance, adherence to sepsis CQUIN

* Trusts have been anonymised and numbered using the coding for the analysis

Appendix 5. List of interventions and outcome measures: reducing intrapartum harm (continued)

Trust *	Intervention	Intervention ID	Outcomes Monitored
Trust 22	Fetal monitoring equipment **K2 central monitoring** Computer Hardware/Software **Pregnancy App** Patient monitoring equipment **More USSscans** Specialised obstetric equipment **fetal pillow / episiotomy scissors** Computer Hardware/Software **IT system for triaging**	2 1 3 7 1	NA NA NA NA NA
Trust 23	Fetal monitoring equipment **Central monitoring and fetal telemetry monitors**	2	Incidents relating to HIE grade 2/3, emergency caesarean sections
Trust 24	Fetal monitoring equipment – **STAN** Training and development **Training for STAN**	2 8	% emergency c-section, % instrumental delivery, HIE, babies into NNU, n. cooling, stillbirths, neonatal deaths (<7days) % emergency c-section, % instrumental delivery, HIE, babies into NNU, n. cooling, stillbirths, neonatal deaths (<7days)
Trust 25	Training and development **Training for central monitoring** Fetal monitoring equipment **Central monitoring**	8 2	N. incidents, N. complaints N. incidents, N. complaints
Trust 26	Patient monitoring equipment **Handheld USS scans to detect breech** Training and development **Training for handheld USS scans**	3 8	N. women scanned, n. breech presentations, birth outcomes N. women scanned, n. breech presentations, birth outcomes
Trust 27	Fetal monitoring equipment - **CTG telemetry** Training and development- electronic **CTG and face-to-face training**	2 8	NA NA
Trust 29	Q&S improvement lead / nurse (staff) **Midwives**	5	Admissions to NICU, Grade 2/3 HIE
Trust 31	Training and development **CTG training**	8	NA
Trust32	Fetal monitoring equipment **K2 central monitoring** Training and development	2 8	NA NA
Trust 34	Q&S improvement lead / nurse (staff) **Midwifes (clinical and CTG)** Training and development **Scanning training** Q&S improvement administrator (staff) Fetal monitoring equipment **K2 monitoring and CTG console**	5 8 4 2	HIE Grades 2 & 3, Babies to NNU >36+6 weeks, APGAR at 5 min <7 at term, Stillbirths, Early Neonatal Deaths, Mat. Serious incidents
Trust 35	Q&S improvement lead / nurse (staff) **Midwifes** Training and development **conference** Computer Hardware/Software **Touch screen TV**	5 8 1	Incident rate
Trust 36	Q&S improvement lead / nurse (staff) **Midwifes** Fetal monitoring equipment - **CTG telemetry**	5 2	babies into NNU babies into NNU

* Trusts have been anonymised and numbered using the coding for the analysis

Appendix 6. List of interventions and outcome measures: A&E missed fractures

Trust *	Intervention	Interv. ID	Outcomes Monitored
Trust 1	Q&S improvement lead / nurse (staff) Q&S improvement administrator (staff) Computer Hardware/Software Training and development	5 4 1 8	Incidence reporting, hospital acquired infections, formal complaints
Trust 2	Q&S improvement lead / nurse (staff) Patient monitoring equipment	5 3	N. of falls, n. of hospital acquired pressure ulcers, Summary Hospital-level Mortality Indicator (SHMI), formal complaints
Trust 4	Radiology equipment	6	NA
Trust 5	Q&S improvement lead / nurse (staff)	5	Reported incidents and associated harm, n. of falls, emergency caesarean section rate, unexpected term admissions to the NNU
Trust 9	Radiology equipment Q&S improvement lead / nurse (staff)	6 5	N. of missed fractures, time from report to ED actions
Trust 10	Computer Hardware/Software Training and development	1 8	NA
Trust 15	Q&S improvement lead / nurse (staff) Q&S improvement administrator (staff)	5 4	% falls, % sepsis, % deteriorating patients, % AKI, % patients with diabetes management problems % falls, % sepsis, % deteriorating patients, % AKI, % patients with diabetes management problems
Trust 16	Q&S improvement lead / nurse (staff) Q&S improvement administrator (staff) Training and development	5 4 8	SHEWS, Sepsis and AKI audits, cardiac arrest rate
Trust 28	Q&S improvement lead / nurse (staff) Q&S improvement administrator (staff)	5 4	N. claims for missed fractures
Trust 30	Training and development Computer Hardware/Software Q&S improvement lead / nurse (staff)	8 1 5	Incidents reported by category, n. falls, % sepsis, % deteriorating patients, missed or delayed diagnosis
Trust 33	Training and development Q&S improvement lead / nurse (staff) Patient monitoring equipment Handheld Profile beds/pressure-relieving heel devices	8 5 3	Freq. pressure ulcers grades 3/4, freq. avoidable harm, Hospital Standardised Mortality Ratio, mortality audits

* Trusts have been anonymised and numbered using the coding for the analysis

Appendix 7. List of interventions and outcome measures: maternity and A&E common interventions

Trust *	Intervention	Intervention ID	Outcomes Monitored
Trust 2	Q&S improvement administrator (staff)	4	N. of falls, n. of hospital acquired pressure ulcers, Summary Hospital-level Mortality Indicator (SHMI), formal complaints
Trust 5	Training and development	8	Reported incidents and associated harm, n. of falls, emergency caesarean section rate, unexpected term admissions to the NNU
Trust 6	Training and development **teaching material for episiotmy scissors**	8	Incidence reporting
	Q&S improvement administrator (staff)	4	Incidence reporting
Trust 19	Q&S improvement lead / nurse (staff)	5	Incidence reporting, formal complaints
	Q&S improvement administrator (staff)	4	Incidence reporting, formal complaints
	Training and development **PROCESS COMMUNICATION MODEL**	8	Incidence reporting, formal complaints

* Trusts have been anonymised and numbered using the coding for the analysis

Appendix 8. Data collection questionnaire sent to Trusts to measure outcomes in maternity units

	jun 2011	.....	sep 2016
1. Total numbers of x-rays taken in A&E			
2. Number of missed fractures in A&E			
2.1. Number of missed fractures in A&E resulting in no harm			
2.2. Number of missed fractures in A&E resulting in low harm			
2.3. Number of missed fractures in A&E resulting in moderate harm			
2.4. Number of missed fractures in A&E resulting in severe harm			
2.5. Alternatively - Number of missed fractures resulting in harm (versus no harm)			
3. Number of claims due to missed fractures in A&E			
3.1. Number of claims due to missed fractures classified as no harm			
3.2. Number of claims due to missed fractures classified as low harm			
3.3. Number of claims due to missed fractures classified as moderate harm			
3.4. Number of claims due to missed fractures classified as severe harm			
3.5. Alternatively - Number of claims due to missed fractures classified resulting in harm (versus no harm)			

Appendix 9. Data collection questionnaire sent to Trusts to measure outcomes in maternity units

	Jun 2011	…	Jul 2016
Total numbers of births			
Numbers of singleton births			
Numbers of term (>=37 weeks) singleton births			
Total numbers of mothers delivered			
Numbers of instrumental vaginal births			
Numbers of caesarean sections			
3rd degree perineal tears: Numbers of mothers with 3rd degree perineal tears			
4th degree perineal tears: Numbers of mothers with 4th degree perineal tears			
APGAR 5-min scores: (1) Numbers of term (>=37 weeks) singleton neonates with APGAR score<7 at 5 minutes			
APGAR 5-min scores: (2) Numbers of term (>=37 weeks) singleton neonates for whom their APGAR score at 5 minutes is not known			
Cooling: Number of [singleton]† babies that were therapeutically cooled after birth			
Unexpected* NICU admissions at term: (1) Number of term (>=37 weeks) singleton neonates admitted unexpectedly to NICU			
Stillbirths: Number of singleton stillbirths			
Stillbirths: Number of term (>=37 weeks) singleton stillbirths			

* **Unexpected** does not include babies who are admitted to the NICU for congenital malformations that require treatment, or for social reasons (e.g., using the NICU as a place of safety).

Questions about the Trust itself - please respond Yes or No	**Y/N**
Are you a level 3 NICU?	
If so, do you provide neonatal surgery?	
Are you a local neonatal unit (LNU)?	
Are you a special care baby unit (SCBU)?	

Appendix 10. Details of the evaluation of the incentivisation scheme in maternity units

Costs

The costs include the following potential components, for successful and unsuccessful Trusts[81]:

- Extra cost of the intervention
- Extra cost to the NHS in terms of resources used as a result of treating harms, i.e. costs of providing maternity care relating to the maternity outcomes that we measured
  Change in consequences in terms of health-related quality of life (HRQOL) or mortality, where values for these can be attached to maternity
- Longer-term savings in costs of claims (assuming there is a reduction)

  The final result here is presented as a cost-consequences analysis (CCA), which is a type of economic analysis where different types of outcomes that can be "bought" by investing in the intervention are listed in their own natural units. We have assigned monetary values where possible, to aid in making comparisons.

We cannot directly measure the numbers of claims resulting from care in the last 5 years, as with maternity claims it can sometimes take a long time for claims to arise and then a further lengthy period until the claim is settled and any financial transfer is made. Costs of claims include not only settlements awarded, but also include legal fees and other costs. Due to this limitation, we are instead measuring short-term outcomes that are thought to possibly have an effect on claims, i.e. they can be used as proxy outcomes. These proxy outcomes have been discussed above according to Sibanda et al 2013[82].

We assume in this analysis that the relationship between short-term outcomes (e.g., low Apgar score) and longer-term outcomes (e.g., cerebral palsy) and from there to numbers of claims, all remain the same over time, and the only thing that changes therefore is the proportion of low Apgar scores or other short-term outcomes, as a result of the interventions introduced.

There are a number of proxy outcomes that could be used to predict numbers of future claims using this approach, and we have focused on those that might be influenced by the specific interventions chosen by the successful maternity units that have been previously discussed, as this then affords a causal pathway by which the interventions could potentially influence the overall outcome. The primary example that we use here is reduction in rates of low (<7) Apgar score at 5 minutes and the relationship of low Apgar scores with cases of cerebral palsy and therefore claims against the NHS for perceived problems with care during labour and birth.

[81] As mentioned before, note that the terms "successful" and "unsuccessful" in the context of this report refer only to whether or not the Trust was successful in being awarded funding for improvements in maternity care via the SU2S (Sign up to Safety) scheme from the NHS LA.

[82] Sibanda T, Fox R, Draycott T, Mahmod T, Richmond D, Simms RA, (2013). Intrapartum care quality indicators: a systematic approach of achieving consensus. European Journal of Obstetrics & Gynecology and Reproductive Biology 166:23-29

Literature review on interventions, impact on claims, QALYs and costs

We have conducted a literature review to obtain information on the linkages from the proxy outcomes to the numbers of claims. We have investigated what information is available on the impact of these interventions (e.g., CTG monitoring) on the claims categories that the NHS LA collects. There is not much evidence that the specific interventions implemented by the Trusts will have any direct impact on maternity indicators that might lead to reductions in claims, which is the overall aim of the scheme.

We have also conducted a literature review to obtain supplementary input data for the model, specifically costs and changes in quality of life relating to the proxy outcomes.

Costs

Unit health care costs for the model were taken from the 2014-15 NHS Reference Costs as described in the Model section below, except for the cost associated with a low Apgar score (Pagano et al., 2010) [83], and that for cooling (Table A.1). It is not clear exactly what extra health care costs are associated with a low Apgar score, as it is not a clinical diagnosis and does not lead to a specific test or admission, so this empirical study where costs of services used during a study were calculated and then different levels of cost associated with different brackets of Apgar scores was the best estimate to use, to our knowledge.

Regarding cooling, a literature search of the Web of Science and the grey literature yielded a report published by the Swedish HTA authority[84] giving a range of costs but not much detail (in English) regarding their provenance. There was also a UK cost-effectiveness analysis (Regier et al., 2010 [85]) which gave a cost for cooling and this lay within the range given in the Swedish report, so we felt that using this point estimate was justified.

Regarding the unexpected NICU admissions, the weighted average cost per day from relevant unit costs in the NHS reference costs was calculated as described in the Model section below, and a literature review was performed to estimate the average length of stay in NICU. This involved searching the Web of Science (Core Collection) for papers mentioning "length of stay", "NICU" and "term", which gave 58 hits, three of which (Khazaei et al.,2015 [86], Girsen et al., 2015 [87], Schiariti et al., 2008 [88]) gave the data we required, and all three of these stated that the mean or median length of stay was 5 days. The cost per unexpected NICU admission therefore was the weighted mean day cost, multiplied by five.

[83] Pagano E, De Rota B, Ferrando A, Petrinco M, Merletti F, Gregori D (2010). An economic evaluation of water birth: the cost-effectiveness of mother well-being. Journal of Evaluation in Clinical Practice 16:916–919

[84] http://www.sbu.se/en/publications/sbu-assesses/therapeutic-hypothermia-in-fullterm-infants-affected-by-birth-asphyxia/; 25th Feb 2009

[85] Regier DA, Petrou S, Henderson J, Eddama O, Patel N, Strohm B, Brocklehurst P, Edwards AD, Azzopardi D, (2010). Cost-Effectiveness of Therapeutic Hypothermia to Treat Neonatal Encephalopathy. Value in Health 13:695-702

[86] Khazaei H, McGregor C, Eklund JM, El-Khatib K (2015). Real-Time and Retrospective Health-Analytics-as-a-Service: A Novel Framework. JMIR Med Inform ;3(4):e36)

[87] Girsen AL, Greenberg MB, El-Sayed YY, Lee H, Carvalho B and Lyell DJ (2015).Magnesium sulfate exposure and neonatal intensive care unit admission at term; Journal of Perinatology 35:181–185

[88] Schiariti V, Klassen AF, Hoube JS, Synnes S, Lisonkova S and Lee SK (2008). Perinatal characteristics and parents' perspective of health status of NICU graduates born at term; Journal of Perinatology 28, 368–376;

No estimate of any increase or reduction in costs has been included for stillbirth as the range is expected to be very wide as parents of a stillborn child can request some or no tests [89]. No empirical studies with usable values were found by our searches, and discussions with clinical colleagues concluded that there is no way of knowing if tests performed in the aftermath a stillbirth would be more costly or numerous than the work done in a similar time period with a live birth, and it is the difference in cost that we would need for our analysis, i.e. how much more (or less) is done compared to a live birth.

For the cost of repairing 3rd or 4th degree tears, we calculated the weighted average cost difference between "Assisted Delivery, with Epidural or Induction" with any CC (complications and comorbidities) score, and "Assisted Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention" with any CC score from the NHS reference costs. We used assisted delivery costs only as it is more common to suffer 3rd or 4th degree tears when forceps or ventouse are used[90], therefore it is more likely that the post-partum surgical intervention is for repairing a tear of this type.

Table A1. Summary of cost data used in the cost-consequences model.

Outcome	Difference in cost	Description	Reference[91]
Stillbirth	£0	We cannot justify a specific difference between the financial cost of a stillbirth and that of a live birth as there is no published data on this.	Authors' assumption
Unexpected NICU admission	£3,440.70	Admission to NICU – weighted average across relevant NHS Reference Costs for a 5-day stay.	NHS Reference Costs (2014-15)[92] Khazaei 2015 Girsen 2015 Schiariti 2008
Cooling	£601.00	Cost of cooling Swedish report 2009 (gives range of £430-860 at 2008 prices)[93]	Regier 2010 Swedish HTA report 2009
Low Apgar score	£289.20	From study looking at water birth, empirical summing of costs split by Apgar score range. (€222.08 at 2010 prices, adjusted to 2014 prices using HCHS P&P index and converted to GBP using 2014 exchange rate)	Pagano 2010
3rd degree tear or 4th degree tear	£582.90	"Post-partum surgical intervention" in NHS Reference Costs. Found by subtracting average costs excluding that factor from average costs including that factor.	NHS Reference Costs (2014-15)

[89] When a baby is stillborn, parents can choose what tests they want to have carried out, if any, to determine or confirm the cause of death. This means that there is a wide range of things that they might want investigated, and a wide range of possible tests, and we don't know the reasons behind any of the stillbirths in our data so can't make any sensible guess. Equally, they might not want to have any tests carried out at all.

[90] https://www.rcog.org.uk/globalassets/documents/patients/patient-information-leaflets/pregnancy/pi-an-assisted-vaginal-birth-ventouse-or-forceps.pdf;Vaginal tears/episiotomy: If you have a vaginal tear or episiotomy, this will be repaired with dissolvable stitches. A third- or fourth-degree tear (a vaginal tear which involves the muscle and/or the wall of the anus or rectum) affects 1 in 100 women who have a normal vaginal birth. It is more common following a ventouse delivery, affecting up to 4 in 100 women (4%). It is also more common following a forceps delivery, affecting between 8 and 12 women in every 100 (8–12%).
Also, if we used the normal delivery figures, it turned out that repairing tears saved £500 per delivery.

[91] See various footnotes for these references

[92] NHS Reference Costs (2014-15)

[93] Should be converted to 2014 prices using HCHS; it was 5000-10000 SEK

Consequences

We searched the York CRD (NIHR Centre for Reviews and Dissemination) database for values for reduction in HRQOL in the event of each of the five outcomes identified above. The consequences for babies and mothers were included as changes in HRQOL for low Apgar, therapeutic cooling, unexpected NICU admission and tears, and as changes in mortality for stillbirth (Table A2). The consequences for babies and mothers were kept separate and independent as there was no published information suggesting how the overlap (e.g., reduction in HRQOL for the mother when her baby is admitted to NICU or is stillborn) might be quantified. It is clear that the additional suffering on the part of the mother will be important, but as we have no figures and cannot guess, it has been omitted. This is a limitation of the analysis.

Some information included values for the reduction in HRQOL when in certain health states or to calculate quality of life and QALYs in patients with cerebral palsy, but did not say for how long the states were thought to last (Leigh et al, 2104 [94]; Turner et al, 2008 [95]). There were other studies where the classification of health states was vague although the time period was defined, for example "operative injury" (Fawsitt et al., 2013 [96]), which would be likely to include 3rd and 4th degree perineal tears, but is unlikely to be made up exclusively of these injuries.

To account for this lack of data, we include only point reduction in HRQOL in the model, and over no specified time period. Therefore, the summing of these point reductions gives only an indication of the direction of travel of quality of life for a particular outcome (i.e. is it positive or negative), and does not reflect on how long this reduced quality of life lasts.

Table A2. Summary of consequences data used in the cost-consequences model.

Outcome	Difference in consequences	Description	Reference
Stillbirth	Mortality (numbers of stillbirths)	We cannot justify a specific difference between the financial cost of a stillbirth and that of a live birth as there is no published data on this.	Authors' assumption
Unexpected NICU admission	0.42 point reduction in HRQOL	This is also called the "disutility". It is the proportion of the baby's quality of life that is lost, on a scale of 0 (no loss of HRQOL) to 1 (total loss of HRQOL).	Tan 2010[97]
Cooling	No change in HRQOL	We found no information on the reduction of a baby's HRQOL on being therapeutically cooled after birth.	Authors' assumption
Low Apgar score	No change in HRQOL	We found no information on the reduction of a baby's HRQOL on being assigned a low Apgar score after birth.	Authors' assumption
3rd degree tear	0.28 point reduction in HRQOL	Mothers suffering from 3rd degree tears are thought to lose 28% of their HRQOL for a short (and undefined) time.	Turner 2008
4th degree tear	0.41 point reduction in HRQOL	Mothers suffering from 3rd degree tears are thought to lose 41% of their HRQOL for a short (and undefined) time.	Turner 2008

[94] Leigh S, Granby P, Turner M, Wieteska S, Haycox A, Collins B. (2014).The incidence and implications of cerebral palsy following potentially avoidable obstetric complications: a preliminary burden of disease study. *BJOG: An International Journal of Obstetrics & Gynaecology*;121(13):1720–8.

[95] Turner CE, Young JM, Solomon MJ, Ludlow J, Benness C, Phipps H. (2008). Vaginal delivery compared with elective caesarean section: the views of pregnant women and clinicians. BJOG 115 (12):1494-1502

[96] Fawsitt CG, Bourke J, Greene RA, Everard CM, Murphy A., Lutomski JE (2013). At What Price? A Cost-Effectiveness Analysis Comparing Trial of Labour after Previous Caesarean versus Elective Repeat Caesarean Delivery. PLoS ONE 8(3): e58577.

[97] Jonathan M Tan, Alex Macario, Brendan Carvalho, Maurice L Druzin, Yasser Y El-Sayed (2010). Cost-effectiveness of external cephalic version for term breech presentation; BMC Pregnancy and Childbirth 2010, 10:3; http://www.biomedcentral.com/1471-2393/10/3

Cost perspective

In terms of the cost perspective taken, we are using that of NHS England (i.e. summing health care costs to the NHS and budgetary costs to NHS LA). This means that unit costs for the various outcomes were taken from NHS reference costs as far as possible then supplemented with information from the literature. The results are presented as differences in costs only, and the total costs for each group are not available as we do not have complete data on all pathways' costs and consequences, only on the comparative costs and consequences for specific outcomes. The results of this model will feed into the longer-term pathway towards the difference in litigation and claims costs, specifically in the cases of low Apgar scores (which is an indicator for cerebral palsy) as this is an important factor in a significant number of expensive claims.

Costs of the intervention(s)

The cost of implementing the specific interventions corresponds to how much has been spent on purchasing the specific interventions that the Trusts planned to buy using the NHS LA funding, and this information, as well as information on dates regarding when interventions were implemented, has been made available to us by the Trusts in the overall SU2S scheme reports that were submitted to us (Appendix 4).

The cost of the overall SU2S scheme corresponds to the funds provided by the NHS LA to Trusts, regardless of whether or not they had been spent immediately. This information, including the date on which the money was sent to Trusts via BACS, has been provided to us by NHS LA.

Costs relating to the outcomes

We have had discussions with clinicians regarding which costs will be the most appropriate ones to include, and how different outcomes should be classified and therefore costed and the conclusions of these discussion are described further below.

We are using NHS reference costs as far as possible, and supplementing this with information found in the published literature for the cost of a low Apgar score (Pagano et al., 2010)[98], unexpected NICU admission and cooling, as described above in the Literature Review section.

Consequences associated with the outcomes

We had anticipated that information on the HRQOL of different health states for newborns would be difficult to find and this has indeed turned out to be the case, as discussed in the Literature review section. We are therefore using point changes in health-related quality of life for all short-term outcomes except stillbirth, for which we are using mortality, and we will not be able to report a full cost-utility analysis with total quality-adjusted life-years based on full utility analysis and associated time periods, as the information is not available and the uncertainty introduced were we to postulate possible values would be enormous.

[98] Pagano E, De Rota B, Ferrando A, Petrinco M, Merletti F and Gregori D, (2010) An economic evaluation of water birth: the cost-effectiveness of mother well-being. Journal of Evaluation in Clinical Practice 16:916–919

Differences in costs, health-related quality of life and mortality – babies

The two halves in Table A3 show the increase or decrease in cost based on these extra outcomes per 100,000 babies, along with the 95%CIs. Similarly, Table A4 shows the change in HRQOL and its 95%CIs. Finally, Table A5 shows the mortality changes per 100,000 babies in stillbirth, as this is a more concrete outcome in this case than HRQOL. Note that the HRQOL of the mothers as a result of suffering a stillbirth are not included in the model.

Table A3**.** Differences in costs per 100,000 babies. These costs are found by multiplying the unit costs by the rates in Table 10-11. The top half relates to the first specific intervention implementation dates being the boundary dates, and the bottom half relates to the BACS payment arrival dates being used as the boundary dates.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Costs	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
lower 95%CI	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
upper 95%CI	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
Low Apgar score	£358,593	£401,594	£43,001	£356,731	£376,304	£19,573	£23,428
lower 95%CI	£285,652	£326,654	-£31,938	£268,424	£287,676	-£69,053	-£65,198
upper 95%CI	£431,534	£476,531	£117,939	£445,037	£464,928	£108,199	£112,054
Cooled	£92,278	£81,399	-£10,878	£114,196	£132,683	£18,487	-£29,365
lower 95%CI	£68,003	£54,258	-£38,016	£83,815	£101,407	-£12,786	-£60,638
upper 95%CI	£116,552	£108,535	£16,260	£144,583	£163,953	£49,760	£1,908
Unexpected NICU	£14,370,939	£14,990,850	£619,911	£14,918,527	£14,747,558	-£170,968	£790,879
lower 95%CI	£11,272,177	£11,866,283	-£2,504,674	£11,178,968	£11,002,839	-£3,915,687	-£2,953,840
upper 95%CI	£17,469,667	£18,115,452	£3,744,495	£18,658,050	£18,492,277	£3,573,751	£4,535,598

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Costs	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
lower 95%CI	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
upper 95%CI	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00	£0.00
Low Apgar score	£344,584	£395,067	£50,482	£328,010	£369,658	£41,647	£8,835
lower 95%CI	£277,731	£328,716	-£16,370	£241,549	£284,047	-£44,814	-£77,626
upper 95%CI	£411,435	£461,414	£117,335	£414,472	£455,272	£128,109	£95,296
Cooled	£104,520	£93,486	-£11,034	£123,554	£127,731	£4,177	-£15,211
lower 95%CI	£77,289	£67,156	-£38,263	£87,061	£92,296	-£32,316	-£51,704
upper 95%CI	£131,745	£119,815	£16,194	£160,046	£163,165	£40,670	£21,281
Unexpected NICU	£14,432,631	£15,443,233	£1,010,602	£15,041,256	£14,669,179	-£372,077	£1,382,679
lower 95%CI	£10,152,195	£11,170,470	-£3,269,817	£9,477,680	£9,112,278	-£5,935,653	-£4,180,897
upper 95%CI	£18,713,033	£19,715,996	£5,291,021	£20,604,832	£20,226,114	£5,191,499	£6,946,255

Table A4. Differences in health-related quality of life per 100,000 babies. These point values are found by multiplying the reduction in health-related quality of life associated with a particular outcome by the rates in Table 10-11.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
HRQOL	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Low Apgar score	0	0	0	0	0	0	0
lower 95%CI	0	0	0	0	0	0	0
upper 95%CI	0	0	0	0	0	0	0
Cooled	0	0	0	0	0	0	0
lower 95%CI	0	0	0	0	0	0	0
upper 95%CI	0	0	0	0	0	0	0
Unexpected NICU admission	1754	1830	76	1821	1800	-21	97
lower 95%CI	1376	1448	-306	1365	1343	-478	-361
upper 95%CI	2132	2211	457	2278	2257	436	554

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
HRQOL	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Low Apgar score	0	0	0	0	0	0	0
lower 95%CI	0	0	0	0	0	0	0
upper 95%CI	0	0	0	0	0	0	0
Cooled	0	0	0	0	0	0	0
lower 95%CI	0	0	0	0	0	0	0
upper 95%CI	0	0	0	0	0	0	0
Unexpected NICU	1762	1885	123	1836	1791	-45	169
lower 95%CI	1239	1364	-399	1157	1112	-725	-510
upper 95%CI	2284	2407	646	2515	2469	634	848

Table A5. Differences in mortality per 100,000 babies for stillbirth.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Mortality	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	133	118	-16	141	160	19	-35
lower 95%CI	102	83	-50	103	122	-19	-73
upper 95%CI	164	152	19	179	199	58	4

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Mortality	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (babies)							
Stillborn	142	130	-13	157	153	-4	-9
lower 95%CI	112	101	-43	118	116	-43	-48
upper 95%CI	172	158	17	196	190	35	31

Differences in costs and health-related quality of life – mothers

The mothers included in this analysis are only those who had vaginal delivery. We subtracted the numbers of mothers delivering via Caesarean section from the total number of mothers delivered, and assumed that all the 3rd or 4th degree tears reported were as a proportion of this group only.

The reductions in numbers of tears lead to reductions in cost, and to improvements in health-related quality of life.

We have not included any uncertainty in the costs themselves - only in the numbers of events.

Table A6. Differences in costs per 100,000 mothers. These costs are found by multiplying the unit costs by the rates in Table 14-15.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Costs	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	£1,896,186	£1,897,725	£1,539	£1,967,790	£1,756,190	-£211,600	£213,139
lower 95%CI	£1,715,422	£1,709,004	-£187,182	£1,764,275	£1,550,833	-£416,957	£7,782
upper 95%CI	£2,076,956	£2,086,446	£190,260	£2,171,305	£1,961,547	-£6,243	£418,496
4th degree tears	£93,043	£81,910	-£11,133	£127,539	£108,269	-£19,271	£8,137
lower 95%CI	£63,857	£50,503	-£42,537	£94,920	£75,084	-£52,458	-£25,050
upper 95%CI	£122,229	£113,311	£20,270	£160,164	£141,459	£13,917	£41,325
Mixed 3rd or 4th degree tears	£1,950,886	£1,977,315	£26,429	£2,095,370	£1,864,459	-£230,912	£257,340
lower 95%CI	£1,772,622	£1,789,911	-£160,972	£1,887,705	£1,654,963	-£440,407	£47,845
upper 95%CI	£2,129,149	£2,164,712	£213,830	£2,303,035	£2,073,954	-£21,416	£466,836

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
Costs	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	£1,892,187	£1,909,517	£17,330	£1,968,699	£1,840,839	-£127,860	£145,190
lower 95%CI	£1,722,696	£1,744,060	-£152,164	£1,765,353	£1,640,414	-£331,206	-£58,156
upper 95%CI	£2,061,684	£2,074,974	£186,824	£2,172,045	£2,041,265	£75,486	£348,536
4th degree tears	£84,737	£89,283	£4,547	£147,259	£112,617	-£34,642	£39,189
lower 95%CI	£55,970	£61,834	-£24,220	£113,025	£79,403	-£68,879	£4,952
upper 95%CI	£113,503	£116,738	£33,313	£181,499	£145,831	-£405	£73,425
Mixed 3rd or 4th degree tears	£1,933,235	£1,991,543	£58,308	£2,116,052	£1,953,456	-£162,595	£220,903
lower 95%CI	£1,766,210	£1,825,876	-£108,717	£1,908,398	£1,748,711	-£370,249	£13,249
upper 95%CI	£2,100,261	£2,157,216	£225,333	£2,323,705	£2,158,201	£45,058	£428,557

Table A7 Differences in level of immediate health-related quality of life per 100,000 mothers. These point values for the reduction in HRQOL are found by multiplying the reduction in HRQOL associated with a particular outcome by the numbers of events in Table 14-15.

	Intvn date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
HRQOL	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	911	912	1	945	844	-102	102
lower 95%CI	824	821	-90	847	745	-200	4
upper 95%CI	998	1002	91	1043	942	-3	201
4th degree tears	65	58	-8	90	76	-14	6
lower 95%CI	45	36	-30	67	53	-37	-18
upper 95%CI	86	80	14	113	99	10	29
Mixed 3rd or 4th degree tears	885	897	12	950	846	-105	117
lower 95%CI	804	812	-73	856	751	-200	22
upper 95%CI	966	982	97	1045	941	-10	212

	BACS date (using only 1 year before and after)						
	Intervention (successful Trusts)			Control (unsuccessful Trusts)			DiD
HRQOL	Before (Y)	After (Y)	Diff (Y)	Before (N)	After (N)	Diff (N)	
Outcome (mothers)							
3rd degree tears	909	917	8	946	884	-61	70
lower 95%CI	828	838	-73	848	788	-159	-28
upper 95%CI	990	997	90	1043	981	36	167
4th degree tears	60	63	3	104	79	-24	28
lower 95%CI	39	43	-17	79	56	-48	3
upper 95%CI	80	82	23	128	103	0	52
Mixed 3rd or 4th degree tears	877	903	26	960	886	-74	100
lower 95%CI	801	828	-49	866	793	-168	6
upper 95%CI	953	978	102	1054	979	20	194

Overall cost-consequences results

Total budget given to Trusts for maternity outcomes (excluding budget that was jointly earmarked for maternity and other things = £8,291,996 in successful Trusts only (zero in unsuccessful Trusts, at least over the dates that we looked at – some reapplied at a later date and were then successful).

Total amount spent so far on specific maternity-related interventions by the successful Trusts = £6,993,749 (over the dates we were looking at).

	No. babies (1 year before + 1 year after)	No. mothers (1 year before + 1 year after)
Intvn	297,731	357,372
BACS	387,702	461,291

Using Trusts' Intvn dates as boundary dates

Therefore, the "items bought" for the money spent by the NHS LA in these 28 Trusts (i.e. per 357,372 mothers or 297,731 term singleton babies) are (cost-consequences analysis):

Non-significant changes, i.e. we cannot say that there is definitely a difference:

- 104 fewer stillbirths
  - 95%CI from 219 fewer to 10 more
- 241 more babies with low Apgar at 5 minutes
  - 95%CI from 671 fewer to 1,154 more
- 145 fewer babies cooled
  - 95%CI from 300 fewer to 9 more
- 684 more NICU admissions
  - 95%CI from 2,556 fewer to 3,925 more
- 50 more 4th degree tears
  - 95%CI from 154 fewer to 253 more

Significant changes, i.e. we can say that there is a statistically significant difference in the numbers recorded:

- 1,307 more 3rd degree tears
  - 95%CI from 48 more to 2,566 more
- 1,578 more mixed 3rd/4th degree tears
  - 95%CI from 293 more to 2,862 more

Using Trusts' BACS payment arrival dates as boundary dates

"Items bought" for the money spent by the NHS LA in these 28 Trusts (i.e. per 461,291 mothers or 387,702 term singleton babies) are (cost-consequences analysis):

Non-significant changes, i.e. we cannot say that there is definitely a difference:

- 33 fewer stillbirths
  - 95%CI from 185 fewer to 118 more
- 118 more babies with low Apgar at 5 minutes

    - 95%CI from 1,041 fewer to 1,278 more
- 98 fewer babies cooled
    - 95%CI from 334 fewer to 137 more
- 1,558 more NICU admissions
    - 95%CI from 4,711 fewer to 7,827 more
- 1,149 more 3rd degree tears
    - 95%CI from 460 fewer to 2,758 more

Significant changes, i.e. we can say that there is a statistically significant difference in the numbers recorded:

- 310 more 4th degree tears
    - 95%CI from 39 more to 581 more
- 1,748 more mixed 3rd/4th degree tears
    - 95%CI from 105 more to 3,391 more

Overall difference in costs for the sample used

TOTAL Costs (using the mixed 3rd/4th degree tears only to avoid double-counting)	**Intvn boundary dates**	**BACS boundary dates**
Point estimate	£3,256,678	£6,354,960
lower 95%CI	-£8,998,167	-£16,649,718
upper 95%CI	£15,511,523	£29,359,637

TOTAL Mortality (including stillbirths only)	**Intvn boundary dates**	**BACS boundary dates**
Point estimate	-104	-33
lower 95%CI	-219	-185
upper 95%CI	10	118

TOTAL Reduction in point HRQOL[99] (including NICU admissions and tears only; using the separate 3rd and 4th degree tears only, to avoid double-counting and allow for the difference in reduction in HRQOL for the two degrees of tear)	**Intvn boundary dates**	**BACS boundary dates**
Point estimate	674	1103
lower 95%CI	-1,123	-2,091
upper 95%CI	2471	4,298

[99] Note that these are not QALYs.