

## The practicalities of adapting UK maternity clinical information systems for observational research: Experiences of the POOL study

Fiona Lugg-Widger<sup>1</sup>, Christian Barlow<sup>1</sup>, Rebecca Cannings-John<sup>1</sup>, Chris Gale<sup>2</sup>, Nicola Houlding<sup>3</sup>, Rebecca Milton<sup>1</sup>, Rachel Plachcinski<sup>4</sup>, Michael Robling<sup>1,5</sup>, and Julia Sanders<sup>6,\*</sup>

Submission History	
Submitted:	23/08/2022
Accepted:	16/05/2023
Published:	07/08/2023

<sup>1</sup>Centre for Trials Research, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS, UK

<sup>2</sup>Neonatal Medicine, School of Public Health, Faculty of Medicine, Imperial College London, Chelsea and Westminster Hospital campus, London, SW10 9NH, UK

<sup>3</sup>Wellbeing Software Group, i2 Mansfield, Hamilton Court, Oakham Business Park, Mansfield, NG18 5FB

<sup>4</sup>Parent, patient and public representative, National Childbirth Trust [NCT], Brunel House, Clifton, Bristol BS8 3NG

<sup>5</sup>DECIPHER, School of Social Sciences, Cardiff University, Cardiff, CF10 3WT, UK

<sup>6</sup>School of Healthcare Sciences, Cardiff University, Ty Dewi Sant, Heath Park, Cardiff. CF14 4YS, UK

### Abstract

#### Background

Using routinely collected clinical data for observational research is an increasingly important method for data collection, especially when rare outcomes are being explored. The POOL study was commissioned to evaluate the safety of waterbirth in the UK using routine maternity and neonatal clinical data. This paper describes the design, rationale, set-up and pilot for this data linkage study using bespoke methods.

#### Methods

Clinical maternity information systems hold many data items of value for research purposes, but often lack specific data items required for individual studies. This study used the novel method of amending an existing clinical maternity database for the purpose of collecting additional research data fields. In combination with the extraction of existing data fields, this maximised the potential use of existing routinely collected clinical data for research purposes, whilst reducing NHS staff data collection burden.

Wellbeing Software<sup>®</sup>, provider of the Euroking<sup>®</sup> Maternity Information System, added new study specific data fields to their information system, extracted data from participating NHS sites and transferred data for matching with the National Neonatal Research Database to ascertain outcomes for babies admitted to neonatal units. Study set-up processes were put in place for all sites. The data extraction, linkage and cleaning processes were piloted with one pre-selected NHS site.

#### Results

Twenty-six NHS sites were set-up over 27 months (January 2019 – April 2021). Twenty-four thousand maternity records were extracted from the one NHS site, pertaining to the period January 2015 to March 2019. Data field completeness for maternal and neonatal primary outcomes were mostly acceptable. Neonatal identifiers flowed to the National Neonatal Research Database for successful matching and linkage between maternity and neonatal unit records.

#### Discussion

Piloting the data extraction and linkage highlighted the need for additional governance arrangements, training at NHS sites and new processes for the study team to ensure data quality and confidentiality are upheld during the study. Amending existing NHS electronic information systems and accessing clinical data at scale, is possible, but continues to be a time consuming and a technically challenging exercise.

#### Keywords

clinical information systems; routine data; feasibility; waterbirth; observational research

\*Corresponding Author:

Email Address: [sandersj3@cardiff.ac.uk](mailto:sandersj3@cardiff.ac.uk) (Julia Sanders)

## Background

The use of routinely collected clinical data in observational research has increased substantially over recent years [1]. It has become a cost-effective way to advance research methods and generate evidence to improve the health of the population [2]. In the UK, Electronic Health Records (EHRs) have been successfully used in maternal and neonatal research where rare outcomes of interest, necessitate large study sample sizes [3]. In commissioned calls major funders in the UK encourage, and often dictate, that the use of routine data is maximised to address evidence gaps.

## Waterbirth

The option of immersion in warm water during labour for pain relief has been recommended by the National Institute for Health and Care Excellence (NICE) since 2007 [4]. Whilst immersion in warm water during labour in a birthpool or bath has analgesic benefit, a large study was required to establish whether for women who use water immersion analgesia during labour, there are any benefits or disadvantages, to women or their babies, by remaining in the water for birth.

To address this question, in 2017 The POOL Study was commissioned by the National Institute for Health Research (NIHR) [5]. Commissioning requirements included that the study should be observational and designed to maximise the use of routinely collected clinical data.

A large study was required to build on the existing observational evidence of waterbirth safety [3, 6] and designed on a sufficiently large scale to provide conclusive evidence in relation to infrequent but important maternal and neonatal clinical outcomes.

## Maternity records

At the NHS provider level, detailed data on individual women and their babies are entered into clinical maternity information systems by midwives and other healthcare staff [7]. These data form the electronic clinical record of each maternity episode from early pregnancy to postnatal discharge. Such systems include numerous data items relating to pregnancy, birth and the early neonatal period. The information system itself is supplied to NHS sites by a private company, of which there are many in the UK [8, 9]. The use of water immersion during labour, in a bath or pool, and whether birth occurred into water are fields that have been captured routinely in NHS local maternity systems for many years.

## Designing the POOL study

The primary analysis of the commissioned study was required to include only women with uncomplicated pregnancies, comparing maternal and neonatal outcomes between women who left a birth pool or bath prior to birth, with women who remained in the water for birth.

Local maternity information systems communicate with hospital patient administration systems, and a standardised

minimum NHS maternity data set is passed from local sites to national systems to inform generation of national maternity statistics, for example NHS Digital in England and Digital Health and Care Wales in Wales. Providing such data sets contain all required fields these national datasets can be accessed for research use.

A previous large study conducted in England exploring maternal and neonatal outcomes associated with waterbirths occurring between April 2015 and March 2016 used centralised maternity audit data to compare outcomes for women who gave birth in, or out, of water [3]. The study included all women without labour complexities who gave birth out of water as the control group, as the study team, using centralised records, were unable to identify women who used a pool during labour but left the water prior to birth. Due to our need to identify a cohort of women who used a pool during labour and compare outcomes between those who remained in or left the pool for birth it was essential for our study to identify data sources that captured pool use during labour but not birth. The study team designed the study including establishing routinely collected clinical data sources and method of linkage that could be utilised to answer these specific study objectives.

Mapping data fields within national datasets against the study objectives in 2017 confirmed that some essential data items were not exported from NHS sites to national datasets. This included a field to identify women who used water immersion during labour but not for birth. As this was central to the research question, this confirmed that national maternity datasets could not be used for study purposes. In addition, other data items including some required to inform the neonatal primary outcome, in particular the administration of antibiotics to neonates on postnatal wards, were not captured at either local or national level. Therefore, to design a study to answer the NIHR commissioning brief questions whilst maximising the use of existing routinely collected data, it became apparent that existing routinely data collected at NHS site level could be used but would require to be supplemented with additional data fields.

As local maternity information systems are required to be adaptable to accommodate changing clinical and monitoring requirements, they have the facility for additional data items to be added or adapted. The decision was taken to add study specific data fields to existing local NHS maternity systems to be completed by midwives as part of usual maternity care and for the records to be extracted at regular timepoints during the study.

To our knowledge the POOL study was the first to adapt electronic maternity information systems at individual NHS site level for the purpose of collecting research data. Using and adapting existing data systems in this way we designed an efficient and novel data collection system, that maximised the use of existing routinely collected data fields whilst minimising additional or repetitive data collection for NHS staff [10].

This paper describes the process required to identify data fields available in local site maternity systems, development of new fields added to 26 NHS sites in England and Wales, study set-up processes, and the feasibility assessments used to confirm the data linkage model and data quality were sufficient for study use.

## Methods

### Overview of POOL study

The POOL study is a large cohort study with a nested qualitative component. Full details of the study are published elsewhere [10]. The study aims to establish the safety of waterbirth using linked NHS maternity and neonatal information systems. The primary analysis is required to only include women with pregnancies classified as uncomplicated by NICE [11], comparing maternal and neonatal outcomes between women who left the water prior to birth with women who remained in the water for birth. There are two primary outcomes: The maternal primary outcome is severe perineal trauma of Obstetric Anal Sphincter Injury (OASI). Such trauma is important to women and the NHS as it requires more complex repair and follow-up, and is associated with short term morbidity (pain, infection, incontinence) as well as longer term morbidity; (dyspareunia, urinary and faecal incontinence, future caesarean section) [12]. The infant primary outcome is a composite of 'adverse infant outcomes or treatment' to include: (a) any neonatal unit admission requiring respiratory support; (b) intravenous antibiotic administration within 48 hours of birth (with or without culture proven infection); and (c) intrapartum stillbirth and all deaths prior to neonatal unit/postnatal ward discharge. Such outcomes are important as they cause distress to parents, are associated with potential long-term damage to infants and have added cost to the NHS.

The qualitative component explored factors influencing pool use and waterbirth across six sites in the UK and are reported elsewhere [13, 14].

### Identifying data sources and developing data flows

Although maternity information systems provide similar functionality due to the complexity of amending maternity information systems a decision was taken for the study team to collaborate with a single large maternity information system provider. In 2016, at the time discussions began, the EuroKing<sup>®</sup> maternity information system, provided by Wellbeing Software<sup>®</sup> (WS), was the most commonly used system in the UK [15]. In 2016 EuroKing<sup>®</sup> captured data relating to 96,951 births including 6,037 (6.2%) waterbirths in NHS sites ranging in size from approximately 1,500 to 10,000+ annual births. As such these units could collectively be regarded as representative of UK NHS practice, able to provide the volume of data required for the planned study and the EuroKing<sup>®</sup> maternity information system had the advantage of being familiar to the clinically based investigators.

As the EuroKing<sup>®</sup> maternity information system does not link to neonatal unit information systems attention then focused on how best to obtain clinical data relating to any sick babies admitted to a neonatal unit. Identifiable data including diagnosis and details of clinical care provided to any baby admitted to a neonatal unit in England, Wales, and Scotland, are recorded by neonatal units in an electronic Badgernet system<sup>®</sup> with a sub-set of data transferred to the National Neonatal Research Database (NNRD) based at Imperial College London [16]. Once received at the NNRD these data are subjected to data cleaning and control. This

centralised data source therefore has several advantages over extracting data from neonatal units at individual and multiple study sites.

The POOL study developed the following data flow to be used at all study sites: 1) To retain anonymity potentially identifiable fields were coded, for example 'Mother's age at Delivery' was calculated from 'Mother's Date of Birth' and 'Date of Delivery'; 2) A linking field was added to each maternity record; 3) The linking field and de-identified data from maternity information systems were transferred to Cardiff University; 4) The linking field and identifiers for each baby were sent to the NNRD to identify and match infants who had been admitted into a neonatal unit; 5) Identifiable data were removed and the linked de-identified neonatal clinical records were sent from the NNRD to the POOL study team Cardiff University; 6) Using the linking field, maternity and neonatal unit records were combined. The data flow diagram is available in the protocol paper [6].

### Identifying existing and required data fields for the study outcomes

To answer all research objectives, including reporting the proportion and characteristics of women who use water immersion during labour and comparative analyses, it was estimated that individual computerised maternity records relating to approximately 600,000 births would be required [10].

The *a priori* sample size calculations estimated a required 30,000 women for the maternal primary outcome and 16,200 neonates to inform the neonatal primary outcome. As all the required fields to inform the maternal primary outcome were already being collected at sites, data relating to births prior to the study could be utilised. In contrast as additional new fields were required to inform the neonatal primary outcome, only babies born after site opening could be included in this analysis.

The study started on April 1<sup>st</sup> 2018, and it was anticipated that the first site would open in November 2018. Estimating the number of records already stored within potential site systems indicated that using records relating to births from January 2015 to November 2020 would be sufficient to inform the maternal primary outcome, whilst births occurring between sites being opened and November 2020 would be sufficient to inform the neonatal primary outcome.

### Additional data fields required

To identify which of the existing EuroKing<sup>®</sup> fields would be required for the study, the study statistician, with clinical support, matched the existing EuroKing<sup>®</sup> data dictionary to those required. Required fields included those to define and characterise the study population as well as required confounders and outcomes. Fields required to answer the research questions but not present in the existing data dictionary were identified, defined and developed. New data fields were developed for the following fields: Maternal or pregnancy risk factors present at water entry, the use of continuous electronic fetal monitoring in water; the administration of intravenous oxytocin for labour augmentation in water; births occurring partially in water;

umbilical cord snapping prior to clamping; neonatal antibiotic administration and clinical markers for neonatal sepsis; and management of placental delivery following waterbirth. The cost for the amending of local maternity information systems and all data extraction costs were met by the study. NHS support costs were calculated at a rate of five minutes additional communication and data entry time for each woman using water immersion analgesia during labour. This recognised that clinical time was required to support the additional data entry, albeit far less than would have been needed through conventional paper or electronic data collection methods.

### Study approvals

To maintain integrity of the cohort study it was important that as few women as possible were excluded, and for this reason an opt-out design, without individual consent, was developed. The extraction and access of de-identified medical records for research is a common approach for observational studies. However, matching maternity to neonatal unit data required the transfer of identifiable, neonatal data, outside of the NHS. Section 251 of the NHS Act 2006, as granted by the Confidentiality Advisory Group (CAG), allows the transfer of “*confidential patient information without consent. . . without being in breach of the common law duty of confidentiality*” [17].

It was proposed to CAG that women giving birth in the period January 1<sup>st</sup>, 2015, up to the date of the site opening, would not be able to opt-out due to the impracticalities of contacting thousands of women who gave birth during that period. To uphold data protection regulations the application to CAG included the proposal that women giving birth whilst active study data collection was in progress would be informed of the research activities and given an opportunity to opt-out without any impact on their health care. The intention was for all women giving birth at study sites during site opening, to be informed about the study through methods selected by individual sites including leaflets, posters, and postings on websites, hospital social media or local maternity services Facebook<sup>®</sup> pages. For this purpose, as part of the modification to the local systems, a study opt-out tick-box was added to the local electronic maternity record which could be ticked by any member of the clinical team with access to the system. This flag was visible to the data processor, WS<sup>®</sup>, who would not extract, or transfer, data related to the flagged records to the study team. The team at Cardiff University would not have sight of flagged records but the total number of women who opted out of the study will be reported as part of study findings.

### Site set-up

Potential NHS study sites were identified as those NHS Trusts or Health Boards with an ongoing contract with WS<sup>®</sup> and with waterbirth facilities. Each potential site was contacted by the POOL Study Manager (RM) inviting them to register an interest in study participation. The proposed NHS site principal investigator provided a CV and Good Clinical Practice certificate. Contracts were set-up between Cardiff University and the NHS sites and once signed, a site initiation visit

was undertaken. Site specific study information materials were prepared and issued to site.

For each NHS site WS<sup>®</sup> compared the data dictionary to their central held Euroking<sup>®</sup> data dictionary to ensure all data fields required for the study were available from the site.

Once contracts were in place, WS<sup>®</sup> generated a statement of works describing the required technical interaction between WS<sup>®</sup> and each NHS site, which was subsequently signed off by each NHS site, Cardiff University and WS<sup>®</sup>. Following signing of the statement of works, WS<sup>®</sup> would then approach the local lead for the EuroKing<sup>®</sup> system and agree a timeline for testing and implementation of the 12 additional study specific data fields. A further approval process was required prior to implementation of the new data items, a ‘Request for Change’, which in many sites needed information governance or senior information technology team sign off. All EuroKing<sup>®</sup> maternity information systems are bespoke to individual sites, therefore prior to release into the live clinical system, testing of the new data items in their parallel system was required at each study site. Once the local lead was satisfied the new data items were compatible with their system, and staff were trained in the study, a date for their introduction into the live clinical system was agreed.

### Piloting the data extraction

It was agreed that the first site set-up would be used as a pilot site for data extractions. Other sites were not opened until the pilot data extraction had confirmed feasibility of the planned methods.

The objectives of the pilot were to:

1. Ensure that the requested data fields could be extracted, linked to the NNRD data and were in the expected format.
2. Test and refine the planned steps for data management and cleaning.
3. Assess data completeness for key outcomes and the new data fields.
4. Validate the plan for risk status classification for women (Table 1).

### Management of data

A central component to the study design, and condition under which approvals for study conduct has been approved, was that the POOL study team at Cardiff University would not receive identifiable data from study sites or the NNRD.

A technical specification was developed and agreed with WS<sup>®</sup> and NNRD detailing data fields requested from both organisations and the data fields to be added to the Euroking<sup>®</sup> system. It also detailed the matching ID to be added to each record and how this would be extracted and sent to Cardiff University. It was intended that data would be transferred by Fastfile, (Cardiff University’s secure data transfer system) and stored on the Cardiff University secure server. Data were to be received in a comma separated values file format and imported into SPSS<sup>®</sup> for manipulation and analysis. A detailed data cleaning plan was developed which outlined required steps prior

Table 1: Pilot objectives

Pilot objective	Rationale	Methods of ascertainment
1 Ensure that the requested data fields could be extracted, linked to the NNRD data and were in the expected format.	<ul style="list-style-type: none"> <li>a. To check all requested fields had been received.</li> <li>b. Ensure the study generated linking fields were attached correctly to all mother / baby records.</li> <li>c. Confirm number and nature of identifiers required for data matching between NHS extracted and NNRD records.</li> <li>d. Ensure site records and matched NNRD records can be linked by the POOL study team.</li> </ul>	<p>Cross-checking all requested fields with those received.</p> <p>Assessment of the received dataset.</p> <p>Assessment of match rates using each identifier (NHS Number, Date of Birth, Postcode).</p> <p>Ensure records received from NNRD could be matched onto maternity records.</p>
2 Test and refine the planned steps for data management and cleaning	Ensuring data were received in the right format to enable automated cleaning prior to receiving the required 600,000 records.	Syntax was written to address the data cleaning activities.
3 Assess data completeness for key outcomes and the new data fields.	To identify any problems with existing or new field completion rates.	Analysis of the completeness rates of data fields received, in particular those that would contribute to the study primary outcomes and the new study fields added to the EuroKing <sup>®</sup> system.
4 Validate the plan for risk status classification for women.	To ensure risk factors in pregnancies could be identified in provided fields.	Check available fields against risk factors specified in NICE guidance.

to analysis. Data quality was to be recorded with particular focus on the new and derived fields as well as an assessment of detailed data quality for the data fields required for the primary outcomes. It was appreciated that this would be an extensive process involving several visual checks of each dataset as well as developing syntax to recode the data fields received in string format into numeric values. It was planned that site data would arrive via WS<sup>®</sup> in 3-month batches and be merged, to create datasets for each site by year. To enable management of these large datasets, a specific research data storage space was required on the university server with increased storage capacity that required additional permissions to access, and that could be accessed remotely from home/office.

## Results

### Study approvals

The study started in April 2018 and received NHS ethical approval in September 2018 (18/WA/0291). A three-way agreement between Cardiff University, WS<sup>®</sup> and University College London (for NNRD) was signed in September 2018, enabling data to flow between the three organisations.

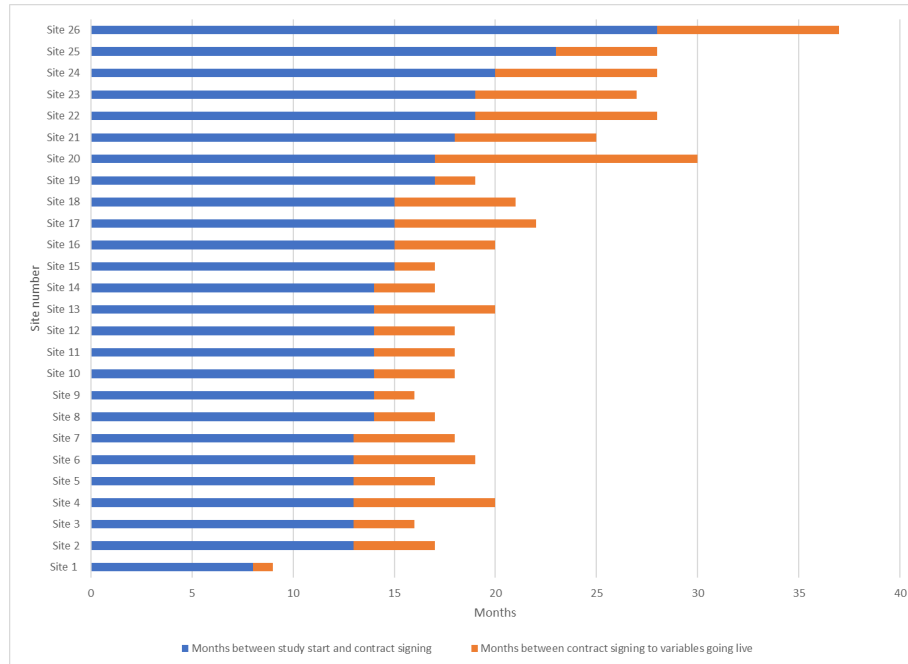
The NHS and CAG approvals committees accepted that women giving birth prior to site opening could not be informed their data would be included. Reviewing the study plans for opt-out CAG were concerned that an individual woman may feel reluctant to indicate her desire to opt-out of the study

to a midwife providing direct clinical care to herself or her baby. It was therefore requested that we include alternative options for women to phone and/or email a contact at the maternity unit if they wished to request opt-out from the study. For this purpose, each site identified an email address and phone number women could contact to opt-out. Approval of the Health Research Authority (HRA) CAG was subsequently granted in November 2018 (18/CAG/0153).

### Site set-up

Emails seeking formal study participation were sent to known prospective sites from August 2018. The first contract with study sites was signed in January 2019, the last in August 2020. The period from the date of a contract being signed between the NHS site and Cardiff University and new data fields being implemented in the NHS sites' maternity software systems ranged from one to 11 months, with most sites (n=17) taking between two and six months (Figure 1). WS<sup>®</sup> required additional authorisations post contract signing, which extended the period between contracts being signed and the new data fields being implemented. Reasons for such delays varied reflecting differing governance arrangements and staff availability at individual NHS sites. The structure and responsibilities of members of each NHS site IT team differed, as did the procedures at each site for agreement of the new data fields, data access permissions and portal opening for data extraction.

Figure 1: Duration of site study setup



## Piloting the data extraction

The first study site was opened in January 2019. In June 2019 a pilot data extraction was undertaken from this single site including data relating to 24,416 babies born during 24,068 births during the period January 2015 to March 2019.

Objective 1. Ensure that the requested data fields could be extracted, linked to the NNRD data and were in the expected format.

The pilot confirmed that the proposed process for data flow and linkage was feasible. Study generated IDs were successfully separately attached to each mother's and baby's records to identify them individually and as dyads/triads. The NHS number, date of birth, gender, postcode and study ID of babies born to the 2,860 women who used a pool during labour between January 1<sup>st</sup> 2015 and March 31<sup>st</sup> 2019 were securely transferred to the NNRD.

The NNRD identified the records of 48 babies (48/2,860, 1.6%) who had been born to women who used a pool during labour and matched all 48 records using the baby's NHS number alone. No additional babies were identified from the other identifiers. Following linkage, the NNRD data were transferred to Cardiff University with records identified only by the baby's study ID. Data held by the NNRD, relating to those 48 babies were subsequently successfully linked to the respective EuroKing<sup>®</sup> mother and infant maternity data that had previously been sent directly to Cardiff University.

Objective 2. Test and refine the planned steps for data management and cleaning.

A data cleaning plan was written by the data manager and approved by the study team, this detailed all the steps taken for each dataset received by WS<sup>®</sup>. Extracts were checked for: i) the correct variables, ii) duplicate records, iii) identifiable data. Datasets were then merged and prepared in the agreed format for the statistician. Each step completed by the data manager was then logged on a data processing document so that the

status of each dataset could easily be identified. Syntax was written to clean each dataset and to make checks on the data, these included checking the data fields derived by WS<sup>®</sup> to ensure they were within feasible ranges.

Although care had been taken not to receive fields of identifiable data, some fields had a free text option box into which NHS staff, on occasions, typed identifiable information. To ensure any free text containing identifiable data was deleted required a manual process of visual checks. This was completed by the data manager who had permission to identify and redact identifiable data.

Objective 3. Assess data completeness for key outcomes and the new data fields.

Maternal demographic characteristics were well completed. The data field relating to the maternal primary outcome (severe perineal trauma) was available in the records for the full period of data extraction (January 2015 to March 2019) and were 99.9% complete (24,044/24,068).

The primary neonatal outcome is a composite of a) intrapartum or neonatal death; b) admission to a neonatal unit requiring respiratory support; or c) antibiotic administration within 48 hours of birth. Data relating to stillbirths or neonatal deaths occurring without NNU admission were 99.9% complete with outcomes provided on 24,415 of the 24,416 babies born during the period.

Data relating to neonatal deaths, respiratory support or antibiotic administration in a NNU were available for all the 48 babies born between 01 January 2015 and 1<sup>st</sup> January 2019 and admitted to a NNU following pool use in labour. These data were 99.0% or above, complete. Antibiotic usage on the postnatal ward, without NNU admission, was only available for the period 02 January 2019 to 31 March 2019 and included all babies regardless of pool use. Data relating to the use and duration of administration of antibiotics were provided on 87 babies with completion rates for the more detailed new data items relating to markers for neonatal sepsis, ranging

from 24% to 100%. Whilst reporting of the attempting or performing of a lumbar puncture was well completed (100% of babies receiving antibiotics) blood results including C Reactive Protein levels and blood culture results were poorly completed (21/87 (24.1%)).

Blood loss at birth was well completed (24,046/24,068, (99.9%)) enabling identification of women who had experienced a postpartum haemorrhage. One requested field, known to be usually completed in clinical records relating to the timing of the first infant feed was empty indicating failure to identify or transfer this field (Table 2).

## Management of data

There were several changes made when data began to flow. One of the key changes related to how the data were to be transferred, the Cardiff University FastFile system involved a member of WS<sup>®</sup> answering a CAPTCHA. This manual interaction needed to transfer each file was time-consuming and inefficient. To ensure data could be sent securely between WS<sup>®</sup> and Cardiff University, access was granted to a member of the WS<sup>®</sup> team. This provided them with a university account and access to a separate space on the secure shared drive. Strict access rights were observed with access to this space being limited to the WS<sup>®</sup> employee and the study data manager. When data was transferred by WS<sup>®</sup> the data manager could then move this across to another secure shared drive which the statistician could also access. This enabled WS<sup>®</sup> to connect securely via a Secure Shell filesystem and automate the upload of the data extracts between their servers and the Cardiff University shared drive folder. Their own in-house system extracted the files securely from the NHS servers to their servers using methods also in use for sending local NHS data extracts to centralised systems.

## Discussion

We have described the detail of the POOL study site set-up and pilot data extraction. Timely approvals were granted, with minimal changes to the proposed opt-out model. NHS ethics and CAG approvals did not delay study progress. Site set-up incurred many, and for some sites, lengthy delays due to the series of steps required for approval and amendment of individual site systems.

The piloting of the data extraction was an important exercise highlighting some issues with existing and new data fields. We found that maternal and neonatal primary outcomes were assessable, although some data fields' completion were poor e.g., blood culture results among babies receiving antibiotics. We also confirmed that the babies identified by the NNRD data could be linked adequately using the identifiers sent by WS<sup>®</sup>, using NHS Number alone. This meant we could reduce the number of identifying fields of data being sent from sites to NNRD to just the NHS Number.

## Lessons learnt

At the end of the pilot we concluded that our planned methods for the amending and use of data held in clinical information systems in multiple NHS sites for research purposes was

feasible. With careful planning and meticulous attention to detail securing NHS ethical and CAG approval was not challenging and should not be regarded as a barrier to future studies with similar planned methods. The study experienced familiar delays in securing signed contracts with study sites. What was unanticipated at study design stage, was the complexity of securing the agreement and involvement of IT departments at each of the study sites, which varied due to site structure and configuration. This process was more complex due to the bespoke nature of the WS system with the requirement to 'test' the new fields in each study site.

Some maternity information systems, such as Badgernet<sup>®</sup> [18] are not localised and collaboration with a different company may have reduced the set-up period at some sites. With plans for a national maternity information system such as that planned for Wales, in future direct linkage between national or regional level data with the NNRD dataset may be possible [19]. Badgernet<sup>®</sup> is in use in nearly every neonatal unit in the UK and therefore, in units where maternity systems are both provided by the same supplier, it is already possible to link maternity and neonatal data 'in house'. However, the advantage of using data from the NNRD is that it is cleaned and validated. For this reason a subsequently funded similarly designed study investigating the safety of induction of labour as an outpatient procedure, continued to use the NNRD as the source of neonatal data despite using only NHS sites with the Badgernet<sup>®</sup> maternity information system [20] For England, with identifiable maternity data held by NHS digital, and neonatal data already held by the NNRD it would be possible to produce a national maternity dataset including maternity and neonatal data, but such an ambition is yet to be realised.

The central IT teams at NHS Trusts were frequently separated from the maternity IT support personnel and local R&D teams who reviewed and agreed the study. This frequently resulted in further delays when requests were made to senior IT staff to provide a portal through which data could be exported, as they challenged the permissions previously given through Trust R&D procedures. Being the first study to modify local maternity information systems for research data collection we were the first to encounter such difficulties.

Following the pilot, the management of the data has changed to facilitate efficient access and processing of large datasets. The technical specification has been a useful document to ensure data extraction and flow are consistent across sites and other documents that document what variables are available/collected by each site (i.e., some sites do not collect postnatal data). Data fields with time/date included have since been removed from the technical specification and have not been extracted from future data extractions due to the potential identifiability of these data fields. Other data (string) fields that contain potentially identifiable data have been retained but a stringent process put in place for these fields to be reviewed and anonymised by one member of staff prior to release for analysis.

One of the challenges of using routine data is the variability of usage between sites. At the site initiation visit this was discussed to ensure data fields with frequently low completion rates are populated moving forward. Some data such as pool use and rates of waterbirth can be checked against the expected range with sites, but as with all anonymised data collected at this scale it will not be possible to validate received

Table 2: Pilot completeness of study confounders, primary and key secondary outcomes

<b>Confounders</b>	<b>Data source</b> ¥	<b>Denominator (N)</b> ∞	<b>N complete</b>	<b>%</b>	<b>Risk to study</b>
Parity	Maternity dataset	All women giving birth 01/01/2015–31/03/2019 (24,068)	24,068	100	Low
Age	Maternity dataset	All women giving birth 01/01/2015–31/03/2019 (24,068)	24,068	100	Low
Gestation at delivery	Maternity dataset	All women giving birth 01/01/2015–31/03/2019 (24,068)	23,961	99.6	Low
<b>Outcomes</b>	<b>Data source</b> ¥	<b>Denominator (N)</b> ∞	<b>N complete</b>	<b>%</b>	<b>Risk to study</b>
<b>Maternal primary outcome</b>					
Perineal trauma, including OASI	Maternity dataset	All women giving birth 01/01/2015–31/03/2019 (24,068)	24,044	99.9	Low
<b>Neonatal primary outcome composite</b>					
Birth outcome live birth/stillbirth or neonatal death without admission to NNU	Maternity dataset	All babies born 01/01/2015–31/03/2019 (24,416)	24,415	99.9	Low
Neonatal death following admission to NNU Yes/No	Neonatal Dataset	Babies admitted to NNU 02/01/2019–31/03/2019 (48)	48	100	Low
Respiratory support provided in NNU Yes/No	Neonatal Dataset	Babies admitted to NNU 02/01/2019–31/03/2019 (48)	48	100	Low
Intravenous antibiotic administration within 48 hours of birth provided in NNU Yes / No	Neonatal Dataset	Babies admitted to NNU 02/01/2019–31/03/2019 (48)	47	99	
Intravenous antibiotic administration within 48 hours of birth provided on Postnatal ward. (Completion indicated antibiotic administration)	Maternity dataset enhanced period	Babies born 02/01/2019–31/03/2019 (1,276)	87	100	
<b>Selected secondary outcomes</b>					
Duration of Intravenous antibiotic administration within 48 hours of birth provided on Postnatal ward	Maternity dataset enhanced period	Babies born 02/01/2019–31/03/2019 who received antibiotics on the postnatal ward (87)	87	100	Low
C Reactive Protein levels and blood culture results for babies receiving antibiotics on the postnatal ward	Maternity dataset enhanced period	Babies born 02/01/2019–31/03/2019 who received antibiotics on the postnatal ward (87)	21	24.1	Moderate
Blood loss at birth	Maternity dataset	All women giving birth 01/01/2015–31/03/2019 (24,068))	24,046	99.9	Low



data against individual records held at NHS sites. Missing variables identified in the pilot extract have since been rectified by WS<sup>®</sup> for all future data extracts. Data quality is being monitored by Cardiff University on an ongoing basis, as data are received from each site, in particular data fields for key outcomes.

## Research using maternity systems data

The approach to local data enhancement, extraction and linkage is bespoke to the POOL study, and at the time of study development, was the best approach to accessing all required data fields including the creation of new data fields to address the commissioned brief. Other options available include datasets or databanks held by individual trusts. Penn et al. 2014 and Oakley et al 2016 report the use of one example of this in England, however key data fields of interest (smoking and body mass index) were incomplete [21, 22]. Another databank of one hospital (maternity and neonatal data) is available in Aberdeen as reported by Bell et al 2001 [23]. The national maternity dataset via the National Maternity and Perinatal Audit holds a wealth of data on this study population and is often utilised for research although its primary function is to evaluate quality in NHS maternity services. It does however have some limitations, for example, Jardine et al. 2020 noted that only 60% of births could be included in their analysis with most exclusions related to data completeness (e.g. severity of medical conditions and timing of stillbirth) [24]. The minimum datasets uploaded from NHS sites to NHS Digital in England (or the equivalent in Wales, Scotland and Northern Ireland) continues to be the most accessible and generalisable dataset for this study population. Hospital Episode Statistics datasets in England (via NHS Digital) is frequently accessed for pregnancy and maternal and infant outcomes and, despite data quality improving over time (from 2002 to 2012 onwards) [25], this can still be a limitation for researchers. For example, Aylin et al 2016 changed study outcome selection due to the quality and availability of data [26]. The Maternity Services Data Set (MSDS) patient level data is now also available from NHS Digital containing information from maternity services first booking appointment to discharge [25]. These national resources are suitable for many research questions and in some instances could have linked with NNRD, however would not be able to incorporate new data fields which was crucial for the delivery of this commissioned work.

Fields held with local and national data sets are updated regularly and future studies will have particular data requirements. At design stage it is important for researchers to consider the precise nature of data requirements and identify the most efficient methods of data collection or extraction. For some studies supplementing routinely collected data may be an option. For large scale observational studies options such as bespoke data collection (e.g., prospective CRFs) to supplement what is routinely collected is not feasible and would add unreasonably to clinician burden. In such circumstances we have demonstrated that the amending of local information systems is a feasible option worthy of consideration.

One of the current limitations of using routine data supplied under Section 251 of the Health and Social Care Act 2006 are the requirements for all study data to be maintained with strict access controls. Even where complete data sets cannot be shared, given the complexity and cost involved, where possible researchers should look to make maximal use of data obtained through appropriately constructed data sharing agreements.

## Impact of the COVID-19 pandemic

The study started in April 2018, sites began collecting data from January 2019 and sites concluded data collection at the end of June 2022. Impact of the COVID-19 pandemic on the delivery of this study was substantially less than for other studies in that births and routine clinical data input continued throughout the pandemic and therefore data collection continued throughout. There will be an impact, however, as many units closed midwifery led facilities and temporarily discontinued water births during the initial phase of the pandemic [27]. Redeployment of staff also impacted the study where site leads for the study were redeployed elsewhere in the trust leading to changes in staff reporting to the study or limited availability of staff to respond to study queries. Additional data fields were added to the Euroking<sup>®</sup> systems relating to COVID-19 during this time, these were not at the request of the POOL study and therefore will not be reported as part of the study.

## Conclusion

The POOL study was the first to adapt local maternity information systems for the purpose of collecting research data. Piloting the data extraction and linkage has been a useful exercise that highlighted the need for additional documentation, training and processes to ensure data quality and confidentiality are upheld for the remainder of the study. Accessing such data on scale, is possible, but continues to be a time consuming and a technically challenging exercise.

## Acknowledgements

We are grateful to Alistair Richards (former WS Product Manager), Andrea Hardy (WS Program Manager), Claire Wright (former WS DBD), Karen Wright (WS DBA), Chris Sewell (former WS Technical Lead), Lee Hallam (WS Integration Specialist), Damian Kay (WS Project manager) and Nicola Houlding (WS POOL Study Lead/Project Manager), Paul Stone (WS Database Services Lead), Reece Percival (WS Support Team Lead), Ryan Brookes (WS Support Team Lead), Imran Hussain (WS Support Technician), Laura Jonathan (WS Support Technician), Jenny Duffin-Bins (WS Application Specialist and Clinical Oversight)

The Centre for Trials Research, Cardiff University, receives funding from Health and Care Research Wales.

## Funding

This project was funded by the National Institute for Health Research 16/149. The views and opinions expressed therein

are those of the authors and do not necessarily reflect those of the NIHR PHR Programme or the Department of Health.

## Availability of data and materials

Routine data supplied to the study from NHS sites and the NNRD are subject to specific data sharing agreements and the study to the arrangements under Section 251 of the Health and Social Care Act 2006. All study data are maintained with strict access controls, which restrict further sharing of data.

## Contributions

JS is the chief investigator for the POOL study. RM is responsible for the study management. FLW is responsible for data governance. RCJ and CB are responsible for data management and analysis. NH was data lead at the time of the pilot work at Wellbeing Software® and throughout the majority of deployment phases. All authors contributed to the content of the manuscript, and all have read and approved the final version.

## Ethics statement

Ethics approval and consent to participate: Ethics approval of the study has been given by the Research Ethics Committee for Wales (18/WA/0291), and the transfer and use of identifiable data has been approved by the Health Research Authority (HRA) Confidentiality Advisory Group (CAG) (18/CAG/0153).

## Consent for publication

The study comprises a pseudonymised dataset, which has been developed via the application of a dissenting model in accordance with the Data Protection Act 1998. No individual will be identified in presented data.

## Statement of conflict of interests

The authors declare that they have no competing interests.

## References

1. D. Agniel, I. S. Kohane, and G. M. Weber, "Biases in electronic health record data due to processes within the healthcare system: retrospective observational study," *Bmj*, vol. 361, p. k1479, 2018. <https://doi.org/10.1136/bmj.k1479>
2. K. M. Lewis and P. Hardelid, "National data opt out programme: consequences for maternity services in England," *Int J Popul Data Sci*, vol. 5, no. 1, 01/30, 2020. <https://doi.org/10.23889/ijpds.v5i1.1126>
3. H. Aughey *et al.*, "Waterbirth: a national retrospective cohort study of factors associated with its use among women in England," *BMC Pregnancy and Childbirth*, vol. 21, no. 1, p. 256, 2021/03/26, 2021. <https://doi.org/10.1186/s12884-021-03724-6>
4. National Institute for Health and Care Excellence, "Intrapartum Care: Care of healthy women and their babies during childbirth," RCOG Press, London, 2007.
5. "National Institute for Health Research, Health Technology Assessment Programme HTA no 16/149 Delivering babies in or out of water." <https://njl-admin.nihr.ac.uk/document/download/2023325> (accessed April, 2023).
6. J. M. Bailey, R. E. Zielinski, C. L. Emeis, and L. Kane Low, "A retrospective comparison of waterbirth outcomes in two United States hospital settings," (in eng), *Birth*, vol. 47, no. 1, pp. 98–104, Mar 2020. <https://doi.org/10.1111/birt.12473>
7. G. Hawley, T. Janamian, C. Jackson, and S. A. Wilkinson, "In a maternity shared-care environment, what do we know about the paper hand-held and electronic health record: a systematic literature review," *BMC Pregnancy and Childbirth*, vol. 14, no. 1, p. 52, 2014/01/30, 2014. <https://doi.org/10.1186/1471-2393-14-52>
8. "Wellbeing Software." <https://www.wellbeingsoftware.com/solutions/sector/maternity/> (accessed April 2023).
9. "Badger Notes." <https://www.badgernotes.net/> (accessed April, 2023).
10. R. Milton *et al.*, "Establishing the safety of waterbirth for mothers and babies: a cohort study with nested qualitative component: the protocol for the POOL study," *BMJ Open*, vol. 11, no. 1, p. e040684, 2021. <https://doi.org/10.1136/bmjopen-2020-040684>
11. National Institute for Health and Care Excellence, "Intrapartum Care: Care of healthy women and their babies during childbirth," RCOG Press, London, 2014.
12. A. H. Sultan, M. A. Kamm, C. N. Hudson, and C. I. Bartram, "Third degree obstetric anal sphincter tears: risk factors and outcome of primary repair," *Bmj*, Research Support, Non-U.S. Gov't vol. 308, no. 6933, pp. 887-91, 1994. [Online]. Available: <https://doi.org/10.1136/bmj.308.6933.887>.
13. S. Milosevic *et al.*, "Factors influencing water immersion during labour: qualitative case studies of six maternity units in the United Kingdom," *BMC Pregnancy and Childbirth*, vol. 20, no. 1, p. 719, 2020/11/23, 2020. <https://doi.org/10.1186/s12884-020-03416-7>
14. S. Milosevic *et al.*, "Factors influencing the use of birth pools in the United Kingdom: Perspectives of women, midwives and medical staff," *Midwifery*, vol. 79, p. 102554, 2019/12/01/2019. <https://doi.org/10.1016/j.midw.2019.102554>
15. NHS England. "Digital Maturity Assessment of Maternity Services in England 2018." (accessed August 7th 2023).

16. "Imperial College London. Utilising the National Neonatal Research Database. <https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/utilising-the-national-neonatal-research-database/>." <https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/utilising-the-national-neonatal-research-database/> (accessed August 7th 2023).
17. "Section 251 of the National Health Service Act 2006." <https://digital.nhs.uk/services/data-access-request-service-dars/how-the-national-data-opt-out-affects-data-released-by-nhs-digital/national-data-opt-out-guidance-for-researchers/appendix-1-section-251-of-the-national-health-service-act-2006> (accessed April, 2023).
18. CleverMed. <https://www.clevermed.com/badgernet-neonatal/> (accessed May 5, 2023).
19. "Women in Wales to benefit from new digital maternity system." NHS Wales. <https://dhcw.nhs.wales/news/latest-news/women-in-wales-to-benefit-from-new-digital-maternity-system/> (accessed May 5, 2023).
20. S. Sarah Jane et al., "Cervical ripening at home or in-hospital—prospective cohort study and process evaluation (CHOICE) study: a protocol," *BMJ Open*, vol. 11, no. 5, p. e050452, 2021. <https://doi.org/10.1136/bmjopen-2021-050452>
21. N. Penn, E. Oteng-Ntim, L. L. Oakley, and P. Doyle, "Ethnic variation in stillbirth risk and the role of maternal obesity: analysis of routine data from a London maternity unit," *BMC Pregnancy and Childbirth*, vol. 14, no. 1, p. 404, 2014/12/07, 2014. <https://doi.org/10.1186/s12884-014-0404-0>
22. L. Oakley, N. Penn, M. Pipi, E. Oteng-Ntim, and P. Doyle, "Risk of Adverse Obstetric and Neonatal Outcomes by Maternal Age: Quantifying Individual and Population Level Risk Using Routine UK Maternity Data," *PloS one*, vol. 11, no. 10, p. e0164462, 2016. <https://doi.org/10.1371/journal.pone.0164462>
23. J. S. Bell, D. M. Campbell, W. J. Graham, G. C. Penney, M. Ryan, and M. H. Hall, "Can obstetric complications explain the high levels of obstetric interventions and maternity service use among older women? A retrospective analysis of routinely collected data," (in eng), *BJOG : an international journal of obstetrics and gynaecology*, vol. 108, no. 9, pp. 910-8, Sep 2001. <https://doi.org/10.1111/j.1471-0528.2001.00214.x>
24. J. Jardine et al., "Risk of complicated birth at term in nulliparous and multiparous women using routinely collected maternity data in England: cohort study," *Bmj*, vol. 371, p. m3377, 2020. <https://doi.org/10.1136/bmj.m3377>
25. "Maternity Services Data Set." NHS Digital. <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/maternity-services-data-set> (accessed April 2023).
26. P. Aylin et al., "Estimating the risk of adverse birth outcomes in pregnant women undergoing non-obstetric surgery using routinely collected NHS data: an observational study". Southampton (UK): NIHR Journals Library. Copyright © Queen's Printer and Controller of HMSO 2016, 2016.
27. J. Jardine et al., "Maternity services in the UK during the coronavirus disease 2019 pandemic: a national survey of modifications to standard care," *BJOG: An International Journal of Obstetrics & Gynaecology*, <https://doi.org/10.1111/1471-0528.16547>, vol. 128, no. 5, pp. 880-889, 2021/04/01, 2021, <https://doi.org/10.1111/1471-0528.16547>.

## Abbreviations

CAG:	Confidentiality Advisory Group
EHR:	Electronic Health Records
HRA:	Health Research Authority
MOH:	Massive Obstetric Haemorrhage
NHS:	National Health Service
NIHR:	National Institute for Health Research
NNRD:	National Neonatal Research Database
OASI:	Obstetric Anal Sphincter Injury
WS:	Wellbeing Software – a Citadel Group Company

