Commission on Human Medicines

REPORT OF THE COMMISSION ON HUMAN MEDICINES EXPERT WORKING GROUP ON OPTIMISING DATA ON MEDICINES USED DURING PREGNANCY

January 2021

MHRA

Table of Contents

Abbreviations

Abbreviation	Definition
AAS	Notification of abortion statistics
ACT	Antenatal Corticosteroid Treatment
ALSPAC	Avon Longitudinal Study of Parents and Children
ASQ	Ages and Stages Questionnaire
BiB	Born in Bradford birth cohort
BMI	Body Mass Index
BPSU	British Paediatric Surveillance Unit
BSO	Business Services Organisation
BUMPS	Best Use of Medicines in Pregnancy
CAG	Confidential Advisory Group
CARDRISS	Congenital Anomalies and Rare Diseases Registration and Information Service for Scotland
CARIS	Congenital Anomaly Register and Information Service for Wales
C-GULL	Children Growing Up in Liverpool
CHESS	COVID-19 Hospitalisation in England Surveillance System
CHI	Community Health Index
CHM	Commission on Human Medicines
CHS	Child Health System
CHSP	Child Health Systems Programme
CHSP-PS	Child Health Systems Programme Pre-School
CLDC	Common Law Duty of Confidentiality
CNST	Clinical Negligence Scheme for Trusts
Co-OPT	Consortium for the study of pregnancy treatments
CPD	Continuing Professional Development
CPRD	Clinical Practice Research Datalink
CSDS	Community Services Dataset
CYPHS	Children and Young People's Health Services (CYPHS)
DAE	Data Access Environment
DARS	Data Access Request Service
DHSC	Department of Health and Social Care
DPA	Data Protection Act

eDRIS	electronic Data Research and Innovation Service
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ENTIS	European Network of Teratology Information Services
EPAU	Early Pregnancy Assessment Unit
EPD	Enhanced Prescribing Dataset
EPMA	Electronic Prescribing and Medicines Administration
EU	European Union
EUROCAT	European surveillance of congenital anomalies
EWG	Expert Working Group
FAIR	Findable, Accessible, Inter-operable and Reusable
FIHR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
GP	General Practice
GRONI	General Register Office for Northern Ireland
GUS	Growing up in Scotland
HBS	Honest Broker Service
HCN	Health and Care Number
HCP	Healthcare Professional
HDR_UK	Health Data Research UK
HDU	High Dependency Unit
HEPMA	Hospital Electronic Prescribing and Medicines Administration
HES	Hospital Episode Statistics
HFEA	Human Fertilisation and Embryology Authority
HPT	Hormone pregnancy test
HQIP	Healthcare Quality Improvement Partnership
HRA	Health Research Authority
HSC	Health and Social Care
HTI	Hospital Treatment Insights
ICNARC	Intensive Care National Audit and Research Centre
ICU	Intensive Care Unit
IGARD	Independent Group Advising on the Release of Data
IMI	Innovative Medicines Initiative
IMRD	IQVIA Medical Research Data
ISAC	Independent Scientific Advisory Committee
ISN	Information Standards Notice

LSHTM	London School of Hygiene and Tropical Medicine
MAS	Minor Ailment Service
MBRRACE-UK	Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK
MCS	Millennium Cohort Study
MHRA	Medicines and Healthcare products Regulatory Agency
MHSDS	Mental Health Services Dataset
MIds	Maternity Indicators Dataset
MSDS	Maternity Services Dataset
NCARDRS	National Congenital Anomaly and Rare Disease Registration Service
NCCHD	National Community Child Health Database
NCMD	National Child Mortality Database
NCMP	National Child Measurement Programme
NHS	National Health Service
NHSBSA	NHS Business Services Authority
NI	Northern Ireland
NICE	The National Institute for Health and Care Excellence
NICPR	Northern Ireland Cerebral Palsy Register
NIHR	National Institute for Health Research
NIMACH	Northern Ireland Maternal and Child Health
NIMATS	Northern Ireland Maternity System
NNRD	National Neonatal Research Database
NPD	National Pupil Database
NRS	National Records of Scotland
NSHPC	National Study of HIV in Pregnancy and Childhood
OMOP	Observational Medical Outcomes Partnership
ONS	Office of National Statistics
OTC	Over the counter
PAS	Patient Administration System
PCCMDS	Paediatric Critical Care Minimum Dataset
PEDW	Patient Episode Database for Wales
PHE	Public Health England
PHR	Personal Health Record
PIS	Prescription Information System
PMRT	Perinatal Mortality Review Tool

POPS	Pregnancy Outcome Prediction Study
PRSB	Professional Record Standards Body
RCGP RSC	Royal College of General Practitioners Research and Surveillance Centre
RCT	Randomised Controlled Trial
RWE	Real World Evidence
SAIL	Secure Anonymised Information Linkage
SBR	Scottish Birth Record
SGSS	Second Generation Surveillance System
SIDS	Sudden Infant Death Syndrome
SIGN	Scottish Intercollegiate Guidelines Network
SMR	Scottish Morbidity Records
SMR01	Scottish Morbidity Record 01
SMR02	Scottish Morbidity Record 02
SPIRE	Scottish Primary Care Information Resource
STROBE	Strengthening the Reporting of Observational studies in Epidemiology
TOPFA	Termination of Pregnancy due to Fetal Anomaly
UK	United Kingdom
UKRC	UK Clinical Research Collaboration
UKTIS	UK Teratology Information Service
UPN	Unique Pupil Number
WHO	World Health Organisation

EXECUTIVE SUMMARY

Introduction

In July 2019 the Commission on Human Medicines (CHM) established a new Expert Working Group (EWG) to advise on better ways to collect and monitor data on the safety of medicines during pregnancy. The EWG was established following a recommendation made by the EWG on Hormonal Pregnancy Tests that concluded in 2017.

The terms of reference of the Group are to advise on:

- Better capturing and linking of existing data on adverse outcomes of pregnancy
- Other ways to capture relevant information on exposure to all medicines during pregnancy
- Improving access to all relevant data on medicines taken during pregnancy to enable studies to be conducted to support pharmacovigilance
- Improving the analytic design of studies examining drug safety in pregnancy
- To make recommendations

The membership of the group was chosen to bring together the expertise relevant to optimising data on the safety of medicines used in pregnancy and is chaired by Professor Jane Norman, Dean of the Faculty of Health Sciences at Bristol University. Members of the Group include key data holders including NHS Digital, the Clinical Practice Research Datalink, Public Health England and individuals with expertise in statistics and epidemiology, data science and artificial intelligence, and current clinical practice with regards to the prescribing of medicines in pregnancy.

Background

Although non-clinical studies are conducted as part of drug development, pregnant women are rarely included in pre-licensure clinical trials, creating a knowledge gap. While work is being done to increase the volume of clinical trial data on the safety of medicines in pregnancy, there will remain a need to further amass evidence on the use, benefits, and risks to both mothers and children exposed to medicines in pregnancy post-licensure. Yellow Card reports of events occurring following exposures during pregnancy in individual women and their children are a key tool for the early detection of potential risks, but this data source has many limitations.

Access to real-world data to facilitate timely and robust identification, quantification, and characterisation of risks associated with exposure to medicines during pregnancy or to provide reassuring evidence of safety, is essential. The UK has a wealth of relevant data, and considerable expertise in the capture, analysis, and interpretation of these data, which has already benefited our understanding of the safety of medicines in pregnancy. With a growing recognition of the need to better monitor long term outcomes including developmental issues in children; increase the availability of data and opportunities for integration of data sources; and develop innovative methods for data capture and analysis, there is potential for strengthening UK data on exposure to medicines during pregnancy. This would improve the evidence base for regulatory and clinical decision making, and the provision of more individual patient-relevant information to allow informed decision-making. These data are also vital for

measuring the impact and effectiveness of actions taken by regulators and healthcare professionals e.g. in monitoring the success of a pregnancy prevention plan for a known teratogen.

Programme of work

The EWG have met five times since July 2019 and have completed the following programme of work:

- Key sources of observational data and those with capabilities for monitoring medicines used in pregnancy have been documented. For the purpose of this EWG “medicines during pregnancy” is defined as any medicinal product (including those taken over-the-counter) taken at any time from 3 months prior to conception to the end of lactation.
- Current proposals or ongoing initiatives for improving data relevant to medicines used in pregnancy and pregnancy/child health outcomes have been ascertained in order to identify opportunities for further optimisation.
- Gaps in available data have been identified with recommendations made for how these might be addressed using existing data sources or using novel approaches to data collection.
- Recommendations have been made on how opportunities for optimisation within current data sources or ongoing developments might be realised.

Gaps in available data relevant to medicines used in pregnancy identified by the EWG related to both exposure and outcome data. Gaps in exposure data include: medicines dispensed and administered in secondary care, use of over-the-counter medicines, medication details such as dosage, duration and indication. Gaps in outcome data include: early miscarriage prior to a maternity booking appointment where the woman may present to either an Early Pregnancy Assessment Unit or to A&E, and long-term physical and neurodevelopmental outcomes in children.

The EWG also discussed a number of other important considerations for data on medicines used in pregnancy including: the importance of data quality, data linkage, the need for public and professional engagement with secondary use datasets and information governance.

Conclusion and recommendations of the EWG

The Expert Working Group on Optimising Data on Medicines used During Pregnancy was set up specifically to identify ways to improve data collection and access to information relevant to research on medicines used during pregnancy and expanded its scope to also include breastfeeding. The UK has a wealth of relevant data on medicines used in pregnancy and breastfeeding but much of it is fragmented across different datasets according to exposure and outcome and UK nation. There are gaps in the available data with some variables either missing completely or poorly recorded and so there are opportunities to optimise data collection and access. Other broader issues around data quality, linkage, public engagement and information governance also present opportunities to facilitate data collection and access for medicines used in pregnancy and breastfeeding but also for all epidemiological research.

The future vision and ultimate aim of this Expert Working Group is to promote and enable complete and accurate data capture of exposure to medications in pregnancy (including the

immediate pre-conception period) and breastfeeding and relevant outcomes and to facilitate access to the data for research purposes. Current barriers include the fragmentation of UK data, variability in data quality and completeness, challenges around information governance and sustainability of linkages and access processes and costs to researchers. The current Covid-19 pandemic has accelerated and improved collaborative working, demonstrating that data can be accessed and shared rapidly whilst maintaining patient confidentiality and there may be lessons to learn from this experience. There are however real opportunities to address some of the current barriers and therefore the Expert Working Group make the following recommendations:

The Group recognises that both motivation and resource are required to implement these recommendations. Our hope is that this report provides the motivation for these recommendations but our fear is that without adequate resource little progress can be made. A cross-sectoral group should therefore be established to oversee the implementation of the following recommendations.

<u>Addressing gaps in data collection:</u>

1. Electronic Prescribing and Medicines Administration (EPMA) systems configured appropriately for use in pregnant women and children should be implemented in maternity and neonatal departments. Data from EPMA needs to be captured in a standardised way, centralised and accessible for data linkage. The project currently underway in England to implement EPMA systems in secondary care is strongly supported.

2. Information on any medication use during pregnancy and breastfeeding including prescribed, over-the-counter or illicit drugs in addition to lifestyle factors should be captured during maternity and health visitor appointments and self-reported directly from women:

    a. Personal Healthcare Records should ensure that women are able to self-report any medication use during pregnancy and breastfeeding.

    b. National maternity datasets should collect any medication data recorded at the booking appointment and other appointments or self-reported.

    c. National clinical guidelines should be updated or developed to include the collection of medication use at the booking antenatal appointment as well as lifestyle factors.

3. Data from Early Pregnancy Assessment Units should be standardised, collated centrally and made available for research purposes

<u>Optimising linkage of existing datasets:</u>

4. Linkage of health and educational datasets is vitally important to study the longer-term impact of medication used in pregnancy and breastfeeding on educational attainment.

a. Progress made to date to link Unique Pupil Numbers (UPN) and NHS numbers in England is welcomed and should be accelerated. Similar linkages should be made in the devolved nations of the UK.
b. The governance approval process needs to be reviewed and clarified regarding the linkage of health and health-related datasets.

5. A mother-baby linkage spine should be set up and maintained in all four devolved nations.

6. Further linkages between datasets relevant to medicines used in pregnancy and breastfeeding are encouraged (e.g. GP, maternity, congenital anomaly and neonatal databases) and a 'task and finish' group of major data custodians should be set up to achieve this.

7. Data controllers are encouraged to facilitate greater access to their data in a timely way for research for the direct benefit of patients, for example, by using the UK Health Data Research Innovation Gateway to ensure datasets can be found and are accessible, inter-operable and reusable (FAIR).

8. Information governance for data linkage of health-related datasets should be optimised to enable greater access to data for the benefit of patients through research:

    a. A sustainable infrastructure for data linkage of datasets is required to ensure that once a linkage has been established it is then available for subsequent projects and improves clinical care.

    b. Researchers should be permitted to archive datasets rather than destroy data to enable future verification of studies or the study of long-term outcomes.

    c. A mutual recognition system for ethics applications, whereby a project approved in one nation is recognised and accepted in the others, should be developed.

<u>Improving data quality:</u>

9. General practice and prescription dispensing data sources should capture indication, dosing, duration and prescription/dispensing dates for all medicines. Information should be structured data rather than free text.

10. Datasets should follow international standards and common coding systems such as SNOMED CT to ensure data interoperability when linking datasets

11. Systems are needed for systematic quality-assurance:
    a. Strategies and processes should be devised to encourage women to review their own data and that of their babies to be able to and report any inaccuracies.
    b. Software used in maternity, neonatal and primary care should implement technical measures that automate data quality through prompts and mandatory fields.

12. The importance and benefits of routinely collected healthcare data for improving patient care both directly and through research into the effects of medicines in

pregnancy and breastfeeding on the child should be emphasised to healthcare professionals and the public:

a. Medicines used in pregnancy should be used as an example to inform the public of the benefits and importance of healthcare data. Specifically examples of how the secondary use of routinely collected healthcare data has helped answer important safety questions should be used.

b. All trainee healthcare professionals should be targeted for training on the importance of routinely collected healthcare data for studying the effects of medicines in pregnancy and breastfeeding on the child.

Encouraging research:

The remit and focus of this Expert Working Group has been on optimising data available for research on medicines used during pregnancy and breastfeeding. The next step is to ensure that these data are used to their full potential to improve the safety of medicines in pregnancy and breastfeeding. In line with the recommendation from the UK Clinical Research Collaboration (UKCRC) the Group also make the following recommendations in relation to use of routine healthcare data in research:

13. Observational research to investigate safety of medicines used in pregnancy and during breastfeeding should be commissioned and a system of ongoing surveillance should be established and funded.

1 INTRODUCTION

In July 2019 the Commission on Human Medicines (CHM) established a new Expert Working Group (EWG) to advise on better ways to collect and monitor data on the safety of medicines during pregnancy.

This report summarises the discussions of the Group on currently available data sources, gaps in the available data and ways in which these gaps might be addressed. The recommendations of the Group are provided at the end of the report.

This chapter outlines the background to the establishment of the EWG and provides information about the membership of the Group, its remit and the scope of its work.

1.1 Background

1.1.1 The need for real-world data

Although non-clinical studies are conducted as part of drug development, pregnant women are rarely included in pre-licensure clinical trials, creating a knowledge gap. While work is being done to increase the volume of clinical trial data on the safety of medicines in pregnancy, there will remain a need to further amass evidence on the use, benefits, and risks to both mothers and children exposed to medicines in pregnancy post-licensure. Yellow Card reports of events occurring following exposures during pregnancy in individual women and their children are a key tool for the early detection of potential risks, but this data source has many limitations.

Access to real-world data to facilitate timely and robust identification, quantification, and characterisation of risks associated with exposure to medicines during pregnancy or to provide reassuring evidence of safety, is essential. The UK has a wealth of relevant data, and considerable expertise in the capture, analysis, and interpretation of these data, which has already benefited our understanding of the safety of medicines in pregnancy. With a growing recognition of the need to better monitor long term outcomes including developmental issues in children; increase the availability of data and opportunities for integration of data sources; and develop innovative methods for data capture and analysis, there is potential for strengthening UK data on exposure to medicines during pregnancy. This would improve the evidence base for regulatory and clinical decision making, and the provision of more individual patient-relevant information to allow informed decision-making. These data are also vital for measuring the impact and effectiveness of actions taken by regulators and healthcare professionals e.g. in monitoring the success of a pregnancy prevention plan for a known teratogen.

1.1.2 Recommendation to establish an Expert Working Group on Optimising Data on Medicines used in Pregnancy

The Expert Working Group (EWG) on Optimising Data on Medicines used in Pregnancy was established following a recommendation made by the EWG on Hormonal Pregnancy Tests that concluded in 2017. Hormonal Pregnancy Tests (HPTs) were medicines available between the 1950-70s used to diagnose pregnancy. HPTs were widely used in the UK and in the 1960s

concerns arose over a possible risk of congenital anomalies associated with their use. Many studies were conducted that showed mixed results and HPTs were withdrawn in the 1970s due to ongoing safety concerns and the availability of alternative methods of diagnosing pregnancy.

Interest in the possible association between HPTs and congenital anomalies has continued however and in 2015, an EWG of the Commission on Human Medicines (CHM) was established to review all the available evidence on a possible causal association between HPTs and adverse outcomes of pregnancy including congenital anomalies and miscarriage or stillbirth. The evidence considered by the EWG included non-clinical studies, mechanistic and pharmacological data and epidemiological studies. A large number of epidemiological studies were identified in the published literature but these often had many methodological limitations including a lack of adjustment for any potential confounding factors. Largely conducted 40-50 years ago, the design and methodological rigour of many of the studies identified for the review was not consistent with today's standards. The studies were therefore considered likely affected by bias, making the results difficult to interpret.

Epidemiological research has developed greatly since HPTs were on the UK market, both methodologically and in terms of data quality. Techniques for handling different types of bias have been developed together with statistical software packages to easily run the analyses. Substantial improvements have also been made with respect to the recording and linking of patient records facilitated by the use of electronic patient healthcare records. Other improvements in the quality of epidemiology studies may be due in part to international initiatives to improve the design, conduct and reporting of observational studies, through initiatives such as those implemented by STROBE (Strengthening the Reporting of Observational studies in Epidemiology) and ENCePP (European Network of Centres for Pharmacoepidemiology and Pharmacovigilance).

Despite these advances in epidemiology over the last 40-50 years, there are clear opportunities to strengthen further how safety concerns about drugs used in pregnancy are detected and evaluated. The UK has several national datasets/programmes for collecting information on exposure to medicines in pregnancy and on neonatal outcomes. A key limitation is the inability to perform more extensive linkage between the datasets and this makes it difficult to perform pharmacoepidemiology studies to detect or evaluate potential safety signals. A lack of standardisation between datasets also makes data harmonisation difficult.

The CHM EWG on HPTs agreed that while considerable improvements have taken place within the fields of pharmacovigilance and pharmacoepidemiology since use of HPTs in the UK, more could be done to safeguard future generations. Accordingly, the CHM EWG on HPTs made a number of recommendations to strengthen the systems in place for detecting, evaluating, managing and communicating risk associated with medicines used in pregnancy. One of the recommendations (Recommendation 2) related to data available for pharmacoepidemiological research stated that:

Optimising collection of, access to and use of data on medicines in pregnancy

A new Working Group should be set up to advise on better ways to collect and monitor data on the safety of medicines during pregnancy. The Working Group's remit should, in particular, explore the potential for:

- *better capturing and linking of existing data on adverse outcomes of pregnancy, including congenital anomalies identified prenatally and neonatally, and developmental disorders that take longer to become apparent, to facilitate regular surveillance*
- *other ways to capture relevant information from, amongst others, midwives and pregnant women on exposure to all medicines, including prescription and over the-counter, during a pregnancy*
- *improving access to all relevant data on medicines taken during pregnancy to enable studies to be conducted to support pharmacovigilance*
- *improving the analytic design of studies examining drug safety in pregnancy*
- *a system for the early sharing and expert review of possible signals or concerns regarding teratogenicity of a drug*
- *systematic, detailed clinical and genetic evaluation of patients in whom a teratogenic effect is being queried*

The current Expert Working Group on Optimising Data on Medicines used in Pregnancy was set up to address the first four points captured within this recommendation. The final two points, to consider a system for the early sharing and expert review of possible signals and the systematic clinical and genetic evaluation of patients in whom a teratogenic effect is being queried, is being followed up in separate strands of a wider programme of work within the MHRA.

1.2 The Expert Working Group on Optimising Data on Medicines used in Pregnancy

1.2.1 Remit of the Group

The terms of reference were finalised and adopted by the Group at the first meeting on 4th July 2019 as follows:

To advise the Commission on Human Medicines on better ways to collect and monitor data on the safety of medicines during pregnancy, in particular:

- *better capturing and linking of existing data on adverse outcomes of pregnancy, including congenital anomalies identified prenatally and neonatally, and developmental disorders that take longer to become apparent, to facilitate regular surveillance*

- *other ways to capture relevant information from, amongst others, midwives and pregnant women on exposure to all medicines, including prescription and over the-counter, during a pregnancy*
- *improving access to all relevant data on medicines taken during pregnancy to enable studies to be conducted to support pharmacovigilance*
- *improving the analytic design of studies examining drug safety in pregnancy*
- *to make recommendations.*

The Group also made some points of clarification for the terms of reference:

- For the purpose of this EWG "medicines during pregnancy" is defined as any medicinal product (including those taken over-the-counter) taken at any time from 3 months prior to conception to the end of lactation.
- Developmental disorders will include developmental, cognitive and metabolic disorders.
- Consideration will also be given to improving the capture and quality of data on potentially important confounding factors e.g. wider maternal health information such as diet and ensuring the availability of data on unexposed women to be used as control groups in epidemiological studies.
- The Group will make recommendations on what datasets are currently available, who could access these, how these could best be linked and best practice for access to linked datasets to a variety of users.
- The Group also agreed that it should make recommendations to improve the quality of the available data.
- The Group considered that while data from industry-led RCTs were out of scope and investigator-led trials and data from observational cohorts were within scope, the main focus of the group would be on secondary use of routinely collected healthcare data.
- The Group considered that the way in which data on medicines used in pregnancy should be reported was also out of scope.

The EWG met three times between July and December 2019 to fulfil the terms of reference and a further meeting was held in June 2020 to finalise its report.

1.2.2 Programme of work

The Group considered the following programme of work to fulfil the terms of reference:

- Documenting all currently known key sources of observational data and capabilities for monitoring medicines used in pregnancy to enable gaps in the available data to be identified.
- Ascertaining current proposals for improving data relevant to medicines used in pregnancy and pregnancy/child health outcomes in order to identify opportunities for further optimisation.
- Identifying gaps in currently available data and how these might be addressed using either existing data sources or novel approaches.

- Considering how the analytic design of observational studies investigating drug safety during pregnancy might be improved.
- How the opportunities for further optimisation within current data sources or developments and opportunities to address gaps in available data could be realised, including barriers and challenges and how these can be overcome.

1.3 Membership of the Group

1.3.1 Conflicts of interest

Expert Working Groups of the Commission on Human Medicines are usually established to consider issues relating to the safety, efficacy or quality of a medicinal product or class of products. As such, to ensure the integrity and impartiality of the review, a conflict of interest policy is applied which precludes members from participating in the EWG if they have interests in any pharmaceutical companies that market any of the products under review.

Since this new EWG does not concern any specific medicines or companies, a conflict of interest policy as described above was not deemed appropriate. In the interest of complete transparency however members were asked to declare any personal and financial interests held in any pharmaceutical companies or any other interests relevant to the scope and remit of the EWG (see Annex 1).

1.3.2 Membership

The membership of the group was chosen to bring together the expertise relevant to optimising data on the safety of medicines used in pregnancy. At its November 2018 meeting the Commission on Human Medicines (CHM) endorsed the formation of the new Expert Working Group (EWG) and subsequently appointed Professor Jane Norman, Dean of the Faculty of Health Sciences at Bristol University to chair the group. The CHM considered that Professor Norman's experience in the field of maternal and fetal medicine in addition to her experience with data science made her an ideal choice of chair for the group.

Relevant stakeholders in maternity data and organisations with prescribing data and/or maternity or child health outcome data were invited as potential members of the group as well as a patient representative. Participation of key data holders including NHS Digital, the Clinical Practice Research Datalink, Public Health England and individuals with expertise in statistics and epidemiology, data science and artificial intelligence, and current clinical practice with regards to the prescribing of medicines in pregnancy was considered essential.

In addition to the full members of the EWG, five experts were invited to the December 2019 meeting to provide additional input and information regarding data collected in Scotland, Wales and Northern Ireland that might differ from that collected in England. Representatives from the Independent Medicines and Medical Devices Safety Review were invited to attend all meetings as observers.

Chair

Professor Jane Norman MBChB MD FRCOG F MedSci FRCP Edin, FRSE
Dean, Faculty of Health Sciences, University of Bristol

Members

Professor Peter Brocklehurst MBChB, MSc, FRCOG, FFPH, FMedSci
Professor of Women's Health, Director for Birmingham Clinical Trials Unit (BCTU)

Mr Paul Brown
Clinical Lead NHS Digital, Prescribing, Medicines and Pharmacy

Ms Caroline Cake
Chief Executive Officer at Health Data Research UK

Dr Rachel Charlton
Research Fellow in Epidemiology, Dept of Pharmacy and Pharmacology, University of Bath

Mr Chris Dickson
Senior Clinical Lead Platforms and Infrastructure, Paediatric Nurse, Digital Child Health and Digital Maternity

Professor Helen Dolk
Professor of Epidemiology & Health Services Research
School of Nursing Institute of Nursing and Health Research Jordanstown campus
University of Ulster

Professor Elizabeth Draper BSc (Hons), MPhil, PhD, FFPH, FRCOG ad Eundem
Professor of Perinatal & Paediatric Epidemiology, University of Leicester

Dr Kenneth Hodson MBChB, MD, MRCP(UK), MRCOG
Head of UK Teratology Information Service, UK Teratology Information Service, Consultant in Obstetrics and Maternal Medicine

Dr Matthew Jolly
National Clinical Director for the Maternity & Women's Health, Acute Medical Directorate, NHS England & NHS Improvement

Professor Neena Modi MBChB MD FRCP FRCPCH FFPM (Hon) FMedSci
Professor of Neonatal Medicine, Consultant, Imperial College London and Chelsea and Westminster NHS Foundation Trust; Chief Investigator, National Neonatal Research Database

Professor Joan Morris
Professor of Medical Statistics, Population Health Research Institute, St George’s, University of London

Dr Puja Myles
Head of Observational Research, Clinical Practice Research Datalink (CPRD)

Ms Katharine Robbins
Information Analysis Lead Manager, Maternity, Child Health and Community at NHS Digital Leeds, West Yorkshire

Mrs Sarah Stevens
Public Health Consultant, Service Lead, The National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)
Public Health England

Dr Sarah Stock
Reader and Consultant in Maternal and Fetal Medicine, University of Edinburgh

Mrs Madeleine Wang BA (Hons)
Lay Representative. Patient Advocate

Invited Experts

Dr Llion Davies MRCS MD MFPH
Public Health Consultant, Public Health Wales

Ms Adele Graham
Senior Health Intelligence Manager, HSC Public Health Agency

Ms Alison Little
Midwife Consultant, HSC Public Health Agency

Mr Dave Roberts
Head of Product for Arms Length Bodies, NHS Digital

Dr Rachael Wood
Consultant in Public Health Medicine, NHS National Services Scotland, Information Service Division

Observers

Dr Sonia Macleod
Independent Medicines and Medical Devices Safety Review Representative

Dr Mel Ramasawmy (December 2019 meeting only)
Independent Medicines and Medical Devices Safety Review Representative

2 CURRENTLY AVAILABLE KEY DATA SOURCES

2.1 Background

The four devolved regions of the UK have a wealth of relevant data sources which have contributed to the better understanding of medicine exposure in pregnancy and breastfeeding, nevertheless, there are areas which could be further enhanced to facilitate increased availability of data and opportunities for better integration of data sources. Overall, the Group considered there is huge potential for strengthening UK data on exposure to medicines during pregnancy and breastfeeding, including better monitoring of long-term outcomes, such as developmental issues in children. The overall impact of this would be to improve the evidence base for regulatory and clinical decision making and allow women to make better informed choices. These data are also important for measuring the impact and effectiveness of actions taken by regulators and healthcare professionals.

The four UK devolved nations have electronic healthcare systems that collect data related to prescribing, maternity, clinical outcomes and other potentially important variables, although not all these elements would necessarily be available in one data source. Although individual differences exist between these datasets in terms of quality and completeness, the core data that are collected is similar across the regions.

A summary of existing data sources in the United Kingdom, which can provide information on medication exposure during pregnancy and breastfeeding and/or relevant outcomes, is provided in the table in Annex 2 (up to October 2020). The table presents key characteristics, where available, of the data source: time period of data collection, population covered, geographical coverage, data captured, availability of maternity data, data linkage, data controller and access mechanisms and any known ongoing initiatives. The table is structured according to data sources from primary care, secondary care, community based, registries, observational research study cohorts, national statistics and others which includes data banks/platforms.

The aim of the table is to provide a comprehensive list of relevant UK data sources with data on medicine exposure during pregnancy and breastfeeding and associated outcomes. Whilst the Group endeavoured to make the table as complete as possible it is not an exhaustive list of data sources, with a focus on larger datasets. Of note, data sources from all the four devolved nations of the UK have been included in the table and experts from the four UK nations have provided valuable input into making this table as complete and up to date as possible. While some data sources incorporate data from all 4 nations, there are others which are specific to a UK nation and the nations are indicated in the table. In addition, some data sources comprise data on both exposure of medicines during pregnancy and outcomes of interest, whereas others might have information on only one of these and all with varying quality, completeness and coverage.

2.2 Type of data source

As the table in Annex 2 shows, a range of data sources are available in each of the four UK regions, giving a wealth of information relating to medicines exposure in pregnancy and relevant outcomes. These can be divided into different categories depending on the type of information contained in each data source.

Table 2.1 shows the type of datasets available in all four devolved nations as well as some datasets that cover either the whole or a sample of the UK population. Examples of some of the main data sources in each category are summarised below.

2.2.1 General practice datasets

The Clinical Practice Research Datalink (CPRD), IQVIA Medical Research Data (IMRD), QResearch, ResearchOne, Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC), Welsh primary care dataset, Scottish Primary Care Information Resource (SPIRE) are general practice (GP) datasets which contain information on prescribing in primary care, diagnoses, test results and some lifestyle factors. There are also regional collections through local health & care record linkage, for example via Discover-NOW which is a health data research hub for real world evidence and led by Imperial College Health partners. Linkages to other data sources are available for some of the GP datasets.

CPRD is one of the most widely used GP datasets for research in the UK, with 25% of the UK population, and studies have shown that the representativeness of the data is good. CPRD contains routinely collected de-identified patient data from a network of 1 in every 5 GP practices across the UK.

Data linkages are currently possible for practices in England and these have had ethics and Confidentiality Advisory Group (CAG) approval. Seventeen standard linkages are available and refreshed on a quarterly basis and these include: hospital episodes statistics (HES) data, maternity data in HES [better coverage than the Maternity Services Dataset (MSDS) but not as detailed] that covers all maternity appointments – see section 2.2.4), Office for National Statistics (ONS) death registry, cancer registry, and deprivation data. Defining pregnancy periods in electronic GP record data is challenging, therefore CPRD has developed an algorithm to create a CPRD pregnancy register with all relevant information related to pregnancy in one place. Data on mothers are not automatically linked to their children, and so CPRD has also developed a mother-baby link that links information from the maternal record to that of the child by using an algorithm to identify pairs of mothers and babies in this dataset. The pregnancy register and mother-baby link currently include data primarily from GP practices that use Vision software but CPRD is in the process of additionally developing these links for practices using EMIS software. There have also been discussions about linking CPRD data to other datasets in England including: the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS), Maternity Services Dataset (MSDS) and dispensing data; and in the UK, namely the National Neonatal Research Database (NNRD). Recently, CPRD has also established linkage to Public Health England (PHE) Second Generation Surveillance System (SGSS) COVID-19 positive virology test data and COVID-19 Hospitalisation in England Surveillance System (CHESS) which include epidemiological data on COVID-19 infection in persons requiring hospitalisation and ICU/HDU admission, to facilitate COVID-19 public health research and surveillance. CPRD has also obtained approvals for linkage to Intensive Care National Audit and Research Centre (ICNARC) data on COVID-19 intensive care admissions and will be making these COVID-19 linkages available for research shortly.

The Welsh primary care dataset contains data for around 78% of the Welsh population. The data captures symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients' home environment. Access to this dataset is available via the Secure Anonymised Information Linkage (SAIL) databank (see section 2.2.9). In Scotland, SPIRE is a service by which information can be requested from GP practice records and centrally collected to generate statistics for Scotland. There is no central dataset with all of GP information. Practices are sent requests for some information relating to specific topics and these need to be approved by the SPIRE Strategy and Oversight Group

prior to being sent out to GP practices. The practices can choose whether or not they want to opt-in to the request and patients can also ask their GP for exclusion of their data from any patient identifiable extracts.

2.2.2 Prescribing/dispensing datasets

Most data sources in this section contain detailed information on all primary care prescriptions dispensed in the community, and hence will include any prescriptions dispensed to women during pregnancy.

In England, the NHS Business Services Authority (NHSBSA) processes prescriptions, issued primarily by GPs in England and also other settings such as from community clinics, dentists, hospitals, and dispensed by NHS contractors, predominantly community pharmacy. Prescription details of medicines such as cost and quantity is collected but indication, dose and frequency are not available. However, for some medicines dose and frequency can be inferred from strength/quantity. Data is collected monthly but the exact date the prescription was prescribed or dispensed is unknown. Data for any particular month is available around 6-7 weeks after the end of the month. Since April 2015 NHS number, age, and gender are also available, for over 95% of the prescriptions, making it possible to link the data to other datasets.

In Scotland, data on medicines prescribed and dispensed in the community and their costs are collected in the Prescription Information System (PIS). The data also includes information on prescriptions written in Scotland that were dispensed elsewhere in the UK. Data includes the Community Health Index (CHI) number (unique patient identifier throughout NHS in Scotland), prescriber and dispenser details, costs and medicine details (e.g. manufacturer, pack, formulation code, strength) as well as data on practices, organisational structures. The Hospital Electronic Prescribing and Medicines Administration (HEPMA) in Scotland is an electronic system for the prescribing and administration of medicines for inpatients that is used in some hospitals in Scotland.

The Enhanced Prescribing Dataset (EPD) collates information on prescriptions issued in general practice and dispensed in a community pharmacy in Northern Ireland. Prescriptions dispensed by community pharmacies are forwarded to the Business Services Organisation (BSO), each month for reimbursement, where the information is held in a secure database i.e. EPD. Data from approximately 85%-90% of all prescriptions are added to the EPD database and includes date of prescription. Details such as drug name, strength, form, prescribed quantity, pack size and price, gross cost and prescriber/practice information are available. Other details, such as gender, age, practice information, residential status, and demographics, can be retrieved by linking the patients' HCN (patient identifier) to other datasets in the NHS BSO.

In Wales the General Practice prescribing data extract includes prescriptions that are prescribed in general practice in Wales and then dispensed in the community within Wales or England. Hence, only data on prescriptions that have been dispensed will be collated in the dataset. The month the prescription was submitted for reimbursement is included in the data extract, however the date the prescription was issued or dispensed is not available. All prescribed medicines, dressings and appliances that are dispensed each month are included in this dataset.

2.2.3 Maternity datasets

This section includes datasets with patient level data on activities carried out by maternity services relating to mother and baby, from the point of the first booking appointment until the mother and baby are discharged from maternity services. These datasets can obtain data solely from the maternity system i.e. electronic data recording system used by midwives to record information about all women who present to maternity services, or they can collect some data from the maternity system along with other hospital systems. Hospital activity datasets with information on delivery and birth episodes are also included in this section. 'Episodes' refer to period of care for a patient under a single consultant at a single hospital.

In England, NHS Digital is the data controller for the Maternity Services Dataset (MSDS), a secondary uses dataset, which comprises some information from the maternity system including: pregnancy booking appointment, clinical details from all appointments and scans, labour and birth details, baby's details and also includes data on inpatient stays in hospital during pregnancy from other hospital systems such as patient administration system (PAS). The data collection started in 2015 and in the last year every maternity provider was submitting electronic data and also records relating to booking appointments and birth (e.g. pregnancy outcome, birth date, delivery method). There has been work going on with NHS resolution for a financial incentive to ameliorate the quality of information submitted and there has been 2 years of the incentive scheme which has led to improvements to MSDS v1.5. Version 2 of MSDS is now available and is an update to the original MSDS (v1.5) and introduces a new structure and content. It also mandates the submission of all maternity records in scope of the dataset, including that records held on paper. As MSDS v2 contains more information than its predecessor there are greater opportunities for linkage with this dataset and there are regular discussions with the Public Health England (PHE) team with regards to health visiting data. However, data quality and completeness of MSDS v2 needs to be improved and the incentive scheme seeking improvements to MSDS v2 is currently paused due to COVID-19.

Another source of maternity information in England is the maternity section of the Hospital Episodes Statistics (HES). HES maternity collects delivery and birth episodes for all births occurring in England, excluding those at home and in non-NHS hospitals where there is no NHS funding. This is part of HES which collects information on all hospital admissions, outpatients' appointments and A&E attendances at NHS hospitals in England (further details in section 2.2.4). HES maternity includes details such as method of delivery, method of onset of labour, complications (such as eclampsia, complications of labour and delivery), miscarriages which resulted in hospital stay, gestation and ethnicity as well as information about the baby, such as date of birth, gender, birthweight and geographical information on where the baby was born.

In Scotland maternity data is captured in the Scottish Morbidity Record 02 (SMR02). This data source, with data collected since the 1980s, has near complete coverage of the population and is quality assured. The SMR02 is similar to the hospital episodes statistics (HES) maternity data for England in terms of clinical detail, however there is funding from the Scottish Government to develop an enhanced maternity return. All health boards, except one, in Scotland use a single maternity record system and there is potential to supplement data from SMR02. SMR02 gives information about pregnancies and collects episode level data every time a mother goes for an obstetric event (antenatal, delivery or postnatal episode). The basic data set for the mother includes demographic information, episode management information, maternity and clinical information. The basic data set for the baby includes birthweight, sex, Apgar score and neonatal indicator (referring to whether baby has been admitted and for how long). Data items including drug misuse during current pregnancy and alcohol consumption have been added in the SMR02 dataset since October 2002.

The Northern Ireland Maternity System (NIMATS) is the maternity system used in Northern Ireland. NIMATS has full coverage from 2011 on all sites in Northern Ireland and holds data on both mother and child including demographic, clinical information (including past medical and obstetric history), current medication (including non-prescription medicines) recorded by midwives and medication administered to the mother during pregnancy or delivery. Ongoing initiatives in Northern Ireland are trying to improve the standardisation of clinical data, however standardising the recording of current medicines exposure is not yet included. Encompass is a 5-year initiative that will introduce a digital integrated care record to Northern Ireland to ensure all patient data are in one place; this presents an ideal opportunity to ensure how maternity data is collected is included in the new programme.

In Wales, the Maternity Indicators dataset (MIds) was established in 2016 and combines records from a mother's booking appointment with a child's birth record in Welsh maternity units. This data enables the Welsh Government to monitor its Welsh maternity indicators which were established to measure the effectiveness and quality of Welsh maternity services. It includes babies born at home, although they might not be identifiable in all health boards but does not include women whose pregnancy did not lead to a birth delivered in a Welsh maternity unit even if the mother resides in Wales. In 2018, MIds covered 94.1% of births to mothers resident in Wales when compared to Office for National Statistics (ONS) data.

The Patient Episodes Database Wales (PEDW) is similar to HES and also encompasses data on maternity patients which comprise pregnant or recently pregnant women being admitted to a maternity ward (including delivery facilities). Every hospital birth in Wales should have a PEDW record and the record consists of general information, diagnosis and procedure information for the mother. The delivery record is attached to each general record for each baby born and contains fields with data relating to the relevant mother and babies.

2.2.4 Data from hospital

There are several data sources which hold data on hospital activity e.g. hospital admissions, A&E attendances [Hospital Episodes Statistics (HES) in England, Scottish Morbidity Record01 and 02 (SMR01 and SMR02) in Scotland, Patient Episodes Database Wales (PEDW) in Wales, Patient Administration System (PAS) in Northern Ireland]. The National Neonatal Research Database (NNRD) collects data specifically from neonatal units in England, Scotland and Wales. Details of these datasets are summarised below. Of note, 'episodes' refer to period of care for a patient under a single consultant at a single hospital.

Hospital Episode Statistics (HES) is a database containing details of all hospital admissions, A & E attendances and outpatient appointments at NHS hospitals in England including private patients treated in hospitals in England and also care funded by the NHS and delivered by other treatment centres e.g. independent sector. Each HES record contains a wide range of information about an individual patient admitted to an NHS hospital, including clinical information about diagnoses and operations, patient information, such as age group, gender and ethnicity, and administrative information. HES also includes data on maternity, as summarised in section 2.2.3.

In Scotland electronic healthcare data for individual patients is collected as a series of Scottish Morbidity Records (SMR). The SMR01 contains information relating to all inpatients and day cases discharged from non-obstetric and non-psychiatric specialties but captures information on early miscarriages that require day case or inpatient care. Information about SMR02, which

captures maternity data, is discussed in section 2.2.3 Scotland also has the Scottish Birth Record (SBR), a web-based system to ensure that every baby born in Scotland has one record, which will act as the foundation for future information collection. All Scottish hospitals which provide midwifery and/or neonatal care have had this system implemented since 2002. The aim of the system is to record all of a baby's neonatal care in Scotland, including readmissions and transfers in one electronic record. The SBR will be completed for all births in Scotland including still births and home births. The SBR includes data on neonatal care as well as some maternal information and has potential to link mothers and babies' data.

The Patient Episode Database for Wales (PEDW) contains all inpatient and day case activity undertaken in NHS hospitals in Wales plus data on Welsh residents treated in English Trusts. The dataset contains demographic, clinical and administrative detail and operative procedures undertaken during the episode of care in hospital. Patients whose episodes are captured by the PEDW database are classified as in-patients, day cases, maternity patients and regular attendees. Information on the maternity part of PEDW is stated in section 2.2.3.

In Northern Ireland, data on admitted patient care delivered by health and social care hospitals is included in the patient administration system (PAS). This administrative data source records information on patients admitted to acute hospitals as inpatients or day cases. However, it does not collect information on patients attending hospital as outpatients, those who attended an accident and emergency department but were not admitted or patients admitted under psychiatric or mental health specialties. The data includes information such as patient demographics, diagnoses, operations, and type of episode, which includes births and delivery.

In terms of neonatal data, the National Neonatal Research Database (NNRD) includes data from every admission to a neonatal unit in England, Wales and Scotland. There are 200 neonatal units in the UK and every baby who requires above normal care will come into a neonatal unit and almost exclusively an NHS neonatal unit. NNRD has also been investigating linkage of the dataset to sample collection (genomic analysis). NNRD is also a part of the new real-world evidence (RWE) data hub: Discover-NOW, recently announced by Health Data Research UK (HDR UK) to link data with other organisations and is also in collaboration with NHS Digital to move the NNRD within the NHS Digital safe haven. Other ongoing and future plans include potential linkage with MSDS and discussions regarding the European neonatal database in order to obtain data from neonatal units across Europe and internationally aiming to create a global network real world evidence (RWE) data from neonatal units around the world.

2.2.5 Congenital anomaly registers

England and Wales have established congenital anomaly registers. The National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) records those people with congenital abnormalities and rare diseases across the whole of England. Currently coverage is 100% of all births in England. NCARDRS is part of EUROCAT, a European network of population-based registries for the epidemiological surveillance of congenital anomalies. Data are collected on all suspected and confirmed congenital anomalies and rare diseases identified in utero, at birth, in childhood or as an adult. In addition to live births and stillbirths affected by congenital anomalies, information about terminations of pregnancy with a diagnosed fetal anomaly at any gestation and late miscarriages (20–23 weeks' gestation) where an anomaly is present is also collected. Data are collected from a number of multiple independent sources (e.g. maternity units, biochemistry and cytogenetic labs, neonatal units, paediatrics, specialist clinics) to maximise and validate the detail of each case. NCARDRS do not rely on individual case reports. Providers send any information that they have readily available, even if it is partial, registration staff then link this information together. This proactive

multiple-source capture enables NCARDRS to achieve the highest possible ascertainment and completeness of cases in the population. NCARDRS is also working with NHS England to establish data flows from the new genomics medicines service. In terms of linkage to other routine data sets currently data is linked to Office for National Statistics (ONS) births and deaths and HES, with discussions with CPRD, MSDS and HQIP on further data linkage. A proof of concept piece of work has also been undertaken in conjunction with UK Teratology Information Service (UKTIS) to link NCARDRS data with NHS Business Services Authority (NHSBSA) prescribing data.

In Wales, CARIS (Congenital Anomaly Register and Information Service) collects information about any fetus or baby who has or is suspected of having a congenital anomaly and whose mother is normally resident in Wales at time of birth. It includes babies in whom anomalies are diagnosed at any time from conception to the end of the first year of life. CARIS is also a member of EUROCAT. Data collection commenced on 1st January 1998 and includes any baby where pregnancy ended after this date. CARIS uses a wide range of sources within the NHS which include antenatal ultrasound, clinical letters, post-mortems and laboratory results. In addition, CARIS also accesses a number of databases including SHIRE (Medical Genetics database), PEDW (Patient Episode Database for Wales), NCCHD (National Community Child Health Database), and the Paediatric Cardiology database. Validation of the data is done by accessing medical records to confirm the information already collected as well adding further relevant information if needed. Due to the nature of this dataset linkage based on the babies is challenging since many pregnancies (approximately 20% of the total) end in fetal loss or termination. However, data linkage based on the mother can be performed.

In Scotland Congenital anomaly records is currently based on linkage of existing records and a national congenital anomaly registration service CARDRISS (Congenital Anomalies and Rare Diseases Registration and Information Service for Scotland) will launch in 2021. The Scottish Morbidity Record 02 (SMR02) also contains a congenital anomaly indicator.

Northern Ireland does not have a congenital anomaly register.

2.2.6 Community-based

There are several community-based datasets which can give important information on relevant developmental outcomes in the children and provide information on breastfeeding status.

The Community Services Data Set (CSDS) is a patient-level dataset which allows community service providers and commissioners to view local and national information from community services, to improve patient care in England. It supersedes the ‘Children and Young People’s Health Services (CYPHS) dataset’ which only focussed on 0-18 year olds. Since October 2017 the CYPHS was renamed CSDS which is an all ages dataset, hence allowing submission of adult community data. Since inclusion of data on adults is a fairly recent development, the dataset has more established data for children. Data extracted includes demographics, social and personal circumstances, and the dataset has the capability to record data on breastfeeding and nutrition, screening, diagnoses including long-term conditions and disabilities and scored assessments. Healthcare activities could take place in settings such as Sure Start centres, day care facilities, schools or community centres, mobile facilities or the patient’s own home. The completeness rate varies, but there is a substantial amount of data on many areas of health visiting e.g. 6-8 week breastfeeding, and the Ages and Stages questionnaire (ASQ) at age 2 to 2.5 years.

The National Child Measurement Programme (NCMP) aims to assess overweight and obesity levels in children within primary schools in England. It measures the height and weight of children in Reception class and year 6. Data are collected by local authorities from all state schools within their area and this data is submitted to NHS Digital. The data can be linked with previous height and weight measurements and data such as dental survey and HES.

The Child Health Systems Programme (CHSP) Pre-School system in Scotland supports the delivery of child health reviews for pre-school children. The child health reviews incorporate assessment of children's health, development, and wider wellbeing and include the neonatal hearing screening, first health visitor visit at around 10 days and subsequent reviews until age 5 years. The CHSP Pre-School system allows consistent recording of the findings and outcomes of child health reviews.

The Child Health System (CHS) records information on all children in Northern Ireland from birth to early adulthood and has historic data going back to the 1980s. It is a community based operational system comprising seven modules: child register, preschool vaccination and immunisation, preschool developmental surveillance, school health, special needs, new-born hearing and influenza.

The National Community Child Health Database (NCCHD) in Wales includes details relating to maternal and child health related indicators such as births, immunisation screening and safeguarding of children. Anonymised records for all children born, resident or treated in Wales and born after 1987 are included in the NCCHD.

2.2.7 National statistics

This section includes some examples of national statistics data that might be useful in providing information on relevant outcomes. The Office for National Statistics (ONS) produces birth registration data which provides information such as birthweight, live or stillborn, gestational age at stillbirth for babies born in the England and Wales. In England, data on pupils' demographics, test, exam results and progression at each key stage is held in the National Pupil Database (NPD). This is for pupils in state funded education, non-maintained special schools, sixth form and further education colleges and (where available) independent schools' further education. In Scotland national statistics databases include the National Records of Scotland (NRS) statutory registers of live births, stillbirths and deaths. Statutory notification of abortions carried out in Scotland are held by Public Health Scotland on behalf of the Chief Medical Officer (AAS database). Data on all pupils in state funded schools in Scotland (for example additional support needs and underlying reasons) is held by the Scottish Government. The statutory register of births contains detailed information about every child born in Scotland since 01/01/1855 and most records include name, place of birth, date and time of birth, sex and details of parents. The Statutory Register of Deaths contains detailed information about each person who has died in Scotland since 1 January 1855 and information includes age at death and cause of death. The General register office for NI (GRONI) stores records of births, deaths, marriages, civil partnerships, stillbirths and adoptions in Northern Ireland.

2.2.8 Birth cohorts

There have been several birth cohort studies carried out over the lives of groups of participants from birth onwards. Table2.1 and the table in Annex 2 list a few of these, however this list is

not exhaustive and includes only larger cohorts. These birth cohorts provide valuable information on wide-ranging factors including exposures during pregnancy, such as medication exposure, and relevant outcomes. Some examples include: The Avon Longitudinal Study of Parents and Children (ALSPAC) which recruited around 14,500 pregnant women from the Bristol area in the 1990's and has been recording their and their children's health since then. It considers multiple genetic, epigenetic, biological, psychological, social and other environmental exposures and a diverse range of health, social and developmental outcome. The variables collected during pregnancy included medication exposure for a range of indications such as nausea, anxiety, infection, allergies, skin condition, depression as well as frequency of aspirin and use, amongst others.

Another example of an established birth cohort study is the Millennium cohort study (MCS), which is currently following the lives of around 19,000 young people born across England, Scotland, Wales and Northern Ireland in 2000-02. This study also provides measures of cohort members physical, socio-emotional, cognitive and behavioural development over time as well as detailed information on their daily life, behaviour and experiences, economic circumstances, relationships. Data on pregnancy, labour of the mother has also been collected, such as illnesses or complications during pregnancy which required treatment, however medication exposure as such was not recorded. Records from the National Pupil database in England have also been linked to the MCS data.

2.2.9 Data/Research platforms

In Scotland, Wales and Northern Ireland, data/research platforms exist that are secure environments for data access/collation of multiple data sources that facilitate data management, data linkage, analysis and dissemination of anonymised person level data. In Scotland, the electronic Data Research and Innovation Service (eDRIS) provides a single entry point and assists researchers in data access in a secure environment, as well as provides support with study design, approvals and data linkage process. In Wales, the Secure Anonymised Information Linkage (SAIL) is a databank of anonymised person-based data and data sources can be linked together to address important research questions. Datasets with information on pregnancy as well as child health outcomes are available via these data platforms e.g. Scottish Morbidity Record 02 (SMR02), Child health programme in Scotland; National Community Child Health Database (NCCHD), Congenital Anomaly and Register Information Service (CARIS) in Wales. In Northern Ireland the Business Services Organization (BSO) has established the Honest Broker Service (HBS) for Health and Social Care (HSC), which enables access to anonymised and pseudonymised, aggregated health and social care data to the DHSC, HSC organisations and for health and social care related research. Examples of data sources included in the HBS are Northern Ireland Maternity Service (NIMATS), data from Child Health System and Enhanced Prescribing Database (EPD).

2.2.10 Other sources of mortality data

This section includes three other datasets which are important sources of information when looking at factors leading to mortality in pregnant women and infants.

The 'Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK' (MBRRACE-UK) is a surveillance data collection system which applies to England, Wales and Scotland, with modified arrangement for Northern Ireland in terms of anonymisation

of data and method of data collection which is centralised at the Northern Ireland Maternal and Child Health (NIMACH). Deaths are reported to the system by nominated staff is every unit and this includes all deaths of pregnant women and women up to one year following end of pregnancy, late fetal losses, stillbirths, early neonatal deaths, late neonatal deaths.

In England, data for all child deaths, up to 18 years of age, is collected via the National Child Mortality Database (NCMD). This work has been commissioned by the Health Quality Improvement Partnership (HQIP) on behalf of NHS England and is being led by University of Bristol. This started in April 2019 and prior to this NHS digital collected aggregate child death review data for two years as an interim measure. The information collected will aid in learning to reduce preventable child mortality.

In Wales, data from the 'Child Death Review Programme' helps in assessing common factors contributing to child deaths, with the aim to reduce preventable child deaths in Wales. The programme covers deaths of children who are live born, where the death occurred after 1 October 2009 and before the child's 18th birthday and where the child is either normally resident in Wales or dies within Wales.

2.3 Other relevant organisations and initiatives

This section provides further information on the work of other organisations and initiatives involved with data collection relevant to medicines used in pregnancy and breastfeeding. These include organisations such as the UK Teratology Information Service (UKTIS) and Health Data Research UK (HDR UK), cross-disciplinary groups established to investigate issues associated with data on medicine exposure during pregnancy and relevant outcomes [Consortium for the study of pregnancy treatments (Co_OPT), the Innovative Medicines Initiative ConcepTION (IMI ConcepTION] or groups involved in establishing evidence base on UK pregnancy research needs (Pregnancy research review). These are summarised below.

2.3.1 The UK Teratology Information Service (UKTIS)

This national service was established in London in 1983 and is hosted in Newcastle by the Newcastle foundation trust. UKTIS is one of the founder members of the European Network of Teratology Information Services (ENTIS), a group of teratology information services across Europe and worldwide. UKTIS aims to support appropriate use of medicines in pregnancy and to advise on management after in utero exposure to potentially harmful substances. They run an enquiry phone line for healthcare professionals and prepare monographs which comprise systematic reviews of drug exposure in pregnancy from literature. UKTIS are commissioned by Public Health England (PHE) to perform national surveillance of known and emerging human teratogens across UK by routinely following-up pregnancies reported voluntarily to the service. Standardised procedures are used to collect pregnancy and fetal outcome information following reporting of maternal or paternal perinatal environmental exposures. Unexposed control data as well as disease matched control data (e.g. diabetes) are available and data can also be linked to other data sources.

In 2014 UKTIS launched a new public facing website: Bumps (Best use of medicines in pregnancy), which collects information about the mother and pregnancy outcomes and women can add information about their pregnancy at any time. Bumps offers a wide range of patient-focussed information sheets to compliment the scientific reviews that UKTIS produces for health care providers.

In terms of ongoing and future developments, UKTIS plans to commission more data linkages especially with larger outcome databases e.g. maternity services dataset (MSDS) as well as the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS). UKTIS is currently working with NCARDRS to assess whether congenital anomaly data can be linked with prescribing data in the first trimester, focusing on neural tube defects as the outcome.

2.3.2 Health Data Research UK (HDR UK)

Health Data Research UK is the national institute for health data that includes England, Wales, Scotland and Northern Ireland. Its mission is to unite the UK's health data to enable discoveries that improve people's lives. By working in partnership with the NHS, industry, academia and patients, and providing safe and secure access to rich health data, it aims to better understand diseases and discover new ways to prevent, treat and cure them. It is a not-for-profit public benefit company funded by UK funded by UK Research and Innovation, the Department of Health and Social Care in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities.

Health Data Research UK is leading a £37.5million government investment on behalf of UK Research and Innovation to create a UK-wide capability for the safe and responsible use of health-related data at scale for research and innovation (the Digital Innovation Hub Programme). This four-year programme is funded by the Industrial Strategy Challenge Fund and includes three essential functions: (i) the UK Health Data Research Alliance; (ii) Health Data Research Hubs; and, (iii) the Health Data Research Innovation Gateway.

The UK Health Data Research Alliance is an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale. Convened by Health Data Research UK, the Alliance develops and co-ordinates the adoption of tools, techniques, conventions, technologies, and designs that help researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK. The National Neonatal Research Database (NNRD), Avon Longitudinal Study of Parents and Children (ALSPAC) and the Human Fertilisation and Embryology Authority (HFEA) are all members of the Alliance.

The Health Data Research Hubs are centres of expertise dedicated to making data available, curating data, and providing expert research services. The Hubs enable researchers and innovators to collaborate and co-create with colleagues, clinicians and patients to draw insight and understanding from raw and fragmented data.

The Health Data Research Innovation Gateway is a portal to find and request access to UK health datasets controlled by members of the UK Health Data Research Alliance. The first phase of the Gateway provides detailed descriptions (metadata) of these datasets, which researchers can search, browse and request access to health data, plus projects, tools and a community forum. It does not hold or store any patient or health data. It aims to increase transparency around accessible datasets and processes associated with their access.

2.3.3 Consortium for the study of pregnancy treatments (Co_OPT)

Co_OPT is an international collaboration established to examine the effect of treatments given during pregnancy and is formed of expertise from a variety of relevant disciplines. Data are

being contributed from Canada, Iceland, Finland, England, Scotland, The Netherlands, Sweden and Aarhus. The aim of Co_OPT is to address knowledge gaps through individual patient data level meta-analysis (IPD) and data linkage studies with data extracted from electronic health records and from birth registries, cohort studies and randomized control trials (RCTs).

The emphasis of Co_OPT is on treatments given in secondary care for acute obstetric conditions which are often old and off-label. In the first instance focus will be on antenatal corticosteroid treatment (ACT), which are usually given to pregnant women between 22 to 39 weeks of gestation for elective caesarean section and unnecessarily treat more than 50% of women who do not end up delivering within 7 days. Scottish health records and a range of relevant anonymised routinely collected observational data sources will be used to assess short-term and long-term benefits and harms of ACT in pregnant women who are at risk of preterm birth. Several data sources from many countries (e.g. Finland, Iceland, Canada, Netherlands, Sweden etc) will provide relevant data fields and further linkages will be done to ascertain outcomes, including long-term outcomes.

2.3.4 The Innovative Medicines Initiative (IMI) ConcePTION Project

The aim of this European initiative is to build and test a pan-European ecosystem for generating, monitoring and providing robust and rapid real-world evidence on medication safety during pregnancy and breastfeeding in a collaborative and standardised manner to inform labels, women, families and heath care professionals (HCPs). This 5-year project, which started in April 2019, is funded by the EU IMI public-private partnership, co-led by the university medical centre in Utrecht and Novartis and includes over 200 researchers from 88 institutions including the regulatory authorities, academia, pharmaceutical companies, public health authorities, women's health organisations and teratology networks.

IMI ConcePTION comprises 8 work packages one of which has a focus on the reuse of existing population data. A data catalogue is being created which must be updateable and searchable. Initially this is being done using the existing EUROmediSAFE inventory that includes UK and European data sources, but this catalogue will also have non-European sources. Details of the UK's data sources, as outlined in annex 2 can be provided to IMI ConcePTION for inclusion in this data catalogue. In the first instance the aim is to find out the characteristics of the different data sources and convert to a common data model.

Another focus of IMI ConcePTION is performing five demonstration projects on selected medication classes each with a drug utilisation component, maternal disease risk component and a medication risk component.

2.3.5 EUROmediCAT

EUROmediCAT is a European research network dedicated to improving medication safety in pregnancy, with 27 members in 17 countries. Data sources are EUROCAT congenital anomaly registers which record data on medication exposure in pregnancy, and healthcare databases including prescription databases, primary care databases and maternity databases, which may be linked to congenital anomaly registers. EUROmediCAT also studies other adverse pregnancy outcomes and medication utilisation. UK data sources currently included are the Secure Anonymised Information Linkage (SAIL), including the Congenital Anomaly Register and Information Service (CARIS), in Wales and the Clinical Practice Research Datalink (CPRD).

2.3.6 Pregnancy research review

In the 2014 annual report on women's health, England's chief medical officer had recommended a review of research needs and expenditure in pregnancy in the UK. As a result of this, the UK Clinical Research Collaboration (UKRC) had commissioned a study, funded by NIHR and the Wellcome Trust, with the aim to establish a comprehensive evidence base on UK pregnancy research needs and how this relates to the current funding for this population. The methodology included three main components: 1) review of current level of spending on pregnancy research in the UK, for which data were gathered from bodies that had funded research related to pregnancy 2) review of healthcare costs associated with pregnancy compared to other areas of health 3) identifying the main research priorities from key stakeholders.

The findings showed that less than 1p is spent on research in pregnancy for every £1 spent on pregnancy care in the UK, which is significantly less than for other conditions. Of the pregnancy related topics, mental health during and after pregnancy is top priority for all stakeholders and is likely underfunded. Other priority topics with varying levels of funding include stillbirth, preterm birth, postnatal support and safety of medicines during pregnancy. An understanding of which medications are safe to take during pregnancy was considered high priority by the stakeholders, due to the challenge in evidence-based prescribing in this population. Overall, the report highlighted the need for further investment in pregnancy research and areas to target the funding.

Table 2.1: Type of datasets available in the UK

Dataset type	UK	England	Scotland	Northern Ireland	Wales
General practice	The Clinical Practice Research Datalink (CPRD) IQVIA Medical Research data (IMRD)	ResearchOne QResearch Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC)	Scottish Primary Care Information Resource (SPIRE)		Primary care GP dataset (via SAIL)
Prescribing		NHSBSA prescribing data Hospital Treatment Insights (HTI)	Prescribing Information system (PIS) Hospital Electronic Prescribing and Medicines Administration (HEPMA)	Enhanced prescribing database (EPD)	
Maternity datasets		Maternity Services dataset (MSDS) Hospital episodes statistics (HES) maternity	Scottish Morbidity Record 02 (SMR02)	Northern Ireland Maternity System (NIMATS)	Maternity Indicators dataset (MIDS) Patient Episode Database for Wales (PEDW) maternity
Hospital data	Paediatric Critical Care Minimum dataset (PCCMDS)	Hospital episodes statistics (HES) National Neonatal Research Database (NNRD) Neonatal critical care minimum dataset Mental health services dataset (MHSDS)	Scottish Morbidity record 01(SMR01) Scottish birth record (SBR) National Neonatal Research Database (NNRD)	Patient Administration System (PAS)	Patient Episode Database for Wales (PEDW) National Neonatal Research Database (NNRD)
Congenital anomaly registers		National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)	Congenital anomalies and rare disease registration and information service for Scotland (CARDRISS) – In development		Congenital Anomaly Register & Information Service (CARIS)
Community based		Community Services dataset (CSDS)	Child Health Systems Programme – Pre-school (CHSP pre-school)	Child health system database (CHS)	National Community Child Health Database (NCCHD)

Dataset type	UK	England	Scotland	Northern Ireland	Wales
		National Child Measurement Programme (NCMP)	Minor Ailment Service (MAS) – in development		
National statistics	Office for National Statistics (ONS) (births and deaths in England and Wales only)	ONS births and deaths National pupil database (NPD)	National records of Scotland statutory register of births National records of Scotland (NRS) statutory register of deaths School education data Notification of abortion statistics (AAS)	General register office for NI (GRONI)	ONS births and deaths
Birth cohorts	Millennium Cohort Study (MCS)	Born in Bradford birth cohort (BiB) Pregnancy Outcome Prediction Study (POPS) Avon Longitudinal Study of Parents and Children (ALSPAC) Children Growing Up in Liverpool (C-GULL) – in development	Growing up in Scotland (GUS)		
Registries (apart from congenital anomaly registries)	Human Fertilisation and Embryology Authority (HFEA) National Study of HIV in Pregnancy and Childhood (NSHPC) UK epilepsy and pregnancy register			The Northern Ireland Cerebral Palsy Register (NICPR)	
Data/Research platforms			electronic Data Research Innovation Service in Scotland (eDRIS)	Honest Broker Service (HBS)	The Secure Anonymised

Dataset type	UK	England	Scotland	Northern Ireland	Wales
					Information Linkage Databank (SAIL)
Other sources of mortality data	Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK (MBRRACE)	National Child Mortality Database (NCMD)			Child death review programme
Other sources	UK Teratology Information Service (UKTIS)				

3 CURRENT DATA CAPABILITIES AND GAPS IN AVAILABLE DATA

High quality observational research on the safety of medicines (including vaccines) taken during pregnancy and breastfeeding requires good quality data on exposures during pregnancy and breastfeeding (including timing of pregnancy), health outcomes and other important covariates. This chapter summarises how and where relevant data are captured within the healthcare system across the UK and where the gaps in available data are. The data sources considered include any that capture relevant data regardless of their current ability to be linked and used to study the use of medicines during pregnancy and breastfeeding. Consideration is also given to issues regarding data quality, public trust and confidence, professional engagement and information governance that affect current ability to conduct high quality research.

3.1 Exposure data on medicines used in pregnancy and breastfeeding

Exposure can be defined into several distinct categories according to stage of pregnancy or during breastfeeding. With the exception of the peri-conceptional period during artificial reproductive therapy and the peripartum period, exposure to medicines in all categories could occur in any of primary care, secondary care or over-the-counter (OTC) settings.

Table 3.1 below summarises the currently known capability of existing data sources as detailed in Annex 2 to identify exposure to medicines at different stages of pregnancy and breastfeeding. The data sources require the ability to capture medication use and identify pregnancy either within the same data source or through linkage with another dataset.

Table 3.1. Availability of data on medicines used at different stages during pregnancy and during breastfeeding

	Exposure		
	Primary care	Secondary care	OTC
Pre-conception*	Y	N	N
Peri-conception**	n/a		n/a
First Trimester (up to 12 weeks)	Y	N	Some
Second Trimester (13-27 weeks)	Y	N	?
Third Trimester (28 weeks onwards)	Y	N	?
Peripartum†	n/a	Y?	n/a
Breastfeeding	Some	N	N

* up to 3 months prior to conception
** refers to embryonic exposure during artificial reproductive therapies
† covers exposures immediately prior to and during labour

3.1.1 Primary care

Primary care exposure data is generally well captured within General Practice record databases such as the UK Clinical Practice Research Datalink (CPRD) and the Secure Anonymised Information Linkage (SAIL) in Wales. Data may also become available through

an NHS Digital GP dataset in the future that is currently under development. GP databases such as the CPRD are however a sample of the UK population and therefore rare exposures may be difficult to study. In this respect the CPRD is expanding its coverage and is developing the technical capability to enable coverage of a large proportion of the UK population while the NHS Digital GP dataset will cover the whole of England.

Prescription data sources such as the NHSBSA (NHS Business Services Authority) in England capture data on all prescriptions dispensed in a community setting and have the advantage of being nation-wide and therefore larger than GP datasets; however, they also tend to lack the completeness and richness of data that are available from GP datasets. In particular, the NHSBSA data does not currently collect the date the prescription was issued or dispensed, which is critical for determining exposure during pregnancy or breastfeeding. The data is only collated and reported by month. The exact date the prescription was issued and submitted to the NHSBSA are known but not collected. However, the latter may not correlate to the dispensing/collection. If available, the date of issue and dispensing would make a critical difference in terms of exposure during early pregnancy.

When using GP prescription data assumptions have to be made that the prescription has been dispensed and then that the patient has actually taken the medicine. Community pharmacy prescription data from sources have an advantage that it is known that the medication was dispensed but an assumption still has to be made that the patient then took it.

3.1.2 Secondary care

Medicines dispensed and administered in some secondary care settings are not currently well captured. Few datasets exist with any routinely entered prescribing information in secondary care and these tend to have incomplete information as only named-patient prescriptions are captured on the hospital system and not those administered on the ward which tend to be written on paper/cardex and not captured electronically. This likely results in incomplete exposure information for drugs such as antibiotics and steroids that are often administered on the ward as well as through the hospital pharmacy.

Maternity information systems have the capability to capture some information on current medication that is recorded by midwives. This information is captured at the booking appointment but may not be updated again and is unlikely to be as complete and accurate as GP record data for prescriptions. It is also not currently in the National Institute for Health and Care Excellence (NICE) guidance on Antenatal care for uncomplicated pregnancies[1] that information on medication should be recorded so this may not always be done. This guidance details what information should be given to and asked of women at routine antenatal appointments. NICE guidance is followed in England and also in Wales and Northern Ireland where appropriate. In Scotland the Scottish Intercollegiate Guidelines Network (SIGN) produces national clinical guidelines but currently there isn't one covering antenatal care. Furthermore, data on medication use recorded in maternity information systems is collected directly from patients and may therefore suffer from issues with maternal recall. Some medication information may therefore be available from Northern Ireland in the Northern Ireland Maternity System (NIMATS) but medication data are not routinely collated in the current version of the Maternity Services Data Set (MSDS) in England.

[1] Antenatal care for uncomplicated pregnancies. https://www.nice.org.uk/guidance/cg62

During the peripartum period, hospital prescribing information may be captured in maternity information systems but as medicines may be used urgently these will be used from a stock on the ward and written in the notes on paper/cardex and hence not captured in electronic records. Neonatal units generally do not use a fully integrated electronic prescribing system either and a clinician will use a menu for a dropdown item or use free text to record a prescription which results in medicines being captured in a non-standardised manner.

NHS Digital in England are working on a long-term project to create a dataset to capture data on medicines prescribed and administered in a secondary care setting that uses an EPMA (Electronic Prescribing and Medicines Administration) system. This will greatly improve data collection and access to data on prescribing in secondary care. However complete coverage won't be available initially as not all hospitals currently have EPMA systems and there is variation in the use of EPMA within hospitals that do. The current project aims to capture by early 2021 EPMA data from 45 hospital trusts in England and 4 NHS Health Boards in Scotland. The long-term ambition is to capture EPMA data from all hospitals and create a dataset for secondary care similar to the primary care dataset. Timescales for the completion of this project is in the region of 3 to 5 years but the program of work has been accelerated under covid-19.

3.1.3 Over-the-counter medicines

Data on over-the-counter (OTC) medicines used during pregnancy and breastfeeding are also not currently well captured. In England, the maternity information system has the capability to record what drugs a woman is taking, and a field has been created for OTC medicines. This information is captured in a free text field at the maternity booking appointment at around 8-10 weeks gestation but may not be routinely captured any further data at time points after this. Aspirin use is of particular clinical interest and is more likely to be captured than other OTC medicines. Medicines that are taken regularly are more likely to be captured than short-term exposures. The Maternity Services Data Set (MSDS) does not currently flow free-text information from the maternity information system but relies on SNOMED CT being used to capture information. SNOMED CT is a structured clinical vocabulary used in electronic health records.

In England and Wales congenital anomaly registers (NCARDRS and CARIS) capture information on medicines prescribed during the first trimester for cases of congenital anomalies and will also collect information on OTC medicines. However, this information is obtained from the booking appointment and will therefore be the same as that held in the maternity system data.

In Scotland, Wales and Northern Ireland prescriptions are free and therefore some OTC medicines may be prescribed. Prescribing data are unlikely to be a complete record of OTC use, however. In Scotland the NHS Minor Ailments Service gives certain patients (including those with a valid maternity exemption certificate) free medication for minor illnesses from a pharmacy. A national dataset to collect these data is under development and some OTC data may therefore be available from this. Early pregnancy exposure will not be captured in this data source however as the certificate will only be issued following the booking appointment.

3.1.4 Medication details

As with any pharmacoepidemiological research, high quality data on exposure to medicines during pregnancy or during breastfeeding requires accurate and detailed exposure records to include information on timing of exposure in relation to gestational age, dosage, duration of treatment and indication. Drugs are known to have different effects at different gestations and so information on the gestational age at exposure is needed. Timing of exposure can usually be inferred from prescription dates (which are well recorded in GP datasets) and gestational age can also be inferred from birth and delivery details where available but this is an imprecise method. Data on dose, duration and indication is often lacking. Dose in particular is often poorly recorded and not standardised resulting in free text entries that make data analysis challenging. Similarly, data on duration of treatment are often lacking or recorded as a free text. Another issue is when prescriptions state 'take as directed' or similar which is appropriate for clinical care but not sufficiently informative for research purposes. In community pharmacy datasets such as the NHS Business Services Authority (NHSBSA) dosage information is not available, only quantity of drugs which then requires researchers to make assumptions about how long the patient might have been taking the medicine for to work out the dose.

Data on indication is frequently not specified and in GP data researchers have to rely on making assumptions based on clinical diagnoses made at the same appointment or from previous appointments as GPs are only required to enter a diagnosis once and not with every repeat prescription. Indication data may be more easily inferred from maternity datasets as whilst individual medications are not linked to individual indications these datasets do contain a lot of data on current medical conditions women have during pregnancy.

3.1.5 Identifying pregnancy and breastfeeding

High quality research requires not only accurate records of the exposure and timing of exposure but also the capability to accurately identify pregnancy and breastfeeding. Particularly in early pregnancy during the period of organogenesis, an inaccuracy of dating the pregnancy of a few weeks can make a huge difference in terms of biological plausibility for many congenital anomalies.

For maternity datasets such as the Maternity Services Data Set (MSDS) in England pregnancy status is implicit and so this is not an issue. In GP records a pregnancy can be identified using data elements relating to maternity appointments or the last menstrual period date but this is challenging. To address this challenge the Clinical Practice Research Datalink (CPRD) have developed a pregnancy register to identify pregnancies from primary care data, using an algorithm developed in collaboration with the London School of Hygiene and Tropical Medicine. The algorithm is based on a number of different maternity data elements to ascertain pregnancy status and dates and have demonstrated high validity when compared with Hospital Episode Statistics (HES) maternity data.[2]

Exposure through breastfeeding presents a challenge as the information on whether a mother is currently breastfeeding may not be well captured, particularly longer-term breastfeeding. The World Health Organisation (WHO) recommends that mothers should exclusively breastfeed babies for the first six months until the baby is eating solids and then

[2] Minassian C., Williams R., Meeraus WH. *et al.* Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. *Pharmacoepidemiol Drug Saf.* 2019; 28(7): 923-33

supplementary breastfeeding is encouraged up to two years or beyond. Breastfeeding rates from a recent maternity survey suggests that around 80% of mothers start breastfeeding, between 62% are still breastfeeding at 6-8 weeks old and 45% at 6 months old[3]. Breastfeeding status is captured at birth through the maternity system datasets and during routine infant health reviews at 1-2 weeks, 6-8 weeks and 9-12 months in community services datasets (e.g. Community Services Data Set (CSDS) in England, National Community Child Health Database (NCCHD) in Wales) that are subsequently shared with the digital Red Book. Breastfeeding status should be ascertained at all health visitor appointments whether routine or ad hoc and these data where captured will be available. NHS Digital have recently published a national data standard that will standardise the recording of breastfeeding status in community datasets to include details on feeding status and method of the baby and approximate date breastfeeding was started and stopped.[4] The implementation of the standard will improve the quality of recording of breastfeeding status but these details will only be captured during health visitor appointments. For time periods in between health visitor appointments, breastfeeding status should be ascertained prior to the initiation of a new medication by the prescribing clinician (primary or secondary care) but it is uncertain how complete this information would be. This might occur for initiation of a new treatment but for existing medication it is less likely that breastfeeding status would be ascertained prior to issuing a repeat prescription. GP datasets may therefore contain limited information on breastfeeding status.

3.2 Pregnancy and child development outcomes of interest

Outcomes of interest include any adverse fetal outcomes during pregnancy and labour, and adverse neonatal outcomes including neonatal mortality and longer-term cognitive, developmental and other outcomes. Table 3.2 below summarises the current UK data sources with pregnancy and child developmental outcome data for all four UK regions.

3.2.1 Adverse fetal outcomes

3.2.1.1 Miscarriage and stillbirth

Information on early miscarriage (< 13 weeks gestation as defined by The National Institute for Health and Care Excellence (NICE)[5]), particularly those that occur prior to the maternity booking appointment at 8-12 weeks gestation is likely to be incomplete. For miscarriages that occur in very early pregnancy (< 6 weeks gestation) the woman may not even be aware that she is pregnant or has miscarried. Most women will not have informed their GP or made a booking appointment prior to 6 weeks gestation. Those that do seek medical care early in a pregnancy are likely to be women with a past history of adverse pregnancy outcomes or women with an underlying medical condition. Data on very early miscarriages are therefore

[3] You & Your Baby. A national survey of health and care 2018. NPEU https://www.npeu.ox.ac.uk/maternity-surveys

[4] DCB Healthy Child Programme: An information standard for child health services, to allow everyone involved in caring for children to share information by using standardised digital records. NHS Digital https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb3009-healthy-child-programme

[5] https://cks.nice.org.uk/miscarriage#!backgroundSub

not likely to be well recorded or captured. Women experiencing an early miscarriage prior to a booking appointment are likely to present to an Early Pregnancy Assessment Unit (EPAU) which are dedicated units that provide specialist care to women in the first three months of pregnancy. EPAUs operate outside of the standard hospital day case, outpatient clinic or hospital admission environment and the data from EPAUs are not currently collated into a dataset and may also not be well captured at the point of entry. Some EPAUs capture data using BadgerNet and Concerto but not all do. Many EPAU are only open on weekdays however and some only in the mornings and therefore women experiencing a miscarriage may also contact either their GP or present to Accident & Emergency so data will also be available from Hospital Episode Statistics (HES) in England or GP record data. Data on early miscarriages that require hospital inpatient care or day cases may be available from the standard general hospital discharge dataset (SMR01) in Scotland. These data will not cover miscarriages that occur outside a hospital setting however. Information on miscarriage that occurs after the maternity booking appointment is likely to be well recorded in maternity clinical systems (e.g. Maternity Services Data Set (MSDS) in England and SMR02 in Scotland) as it is mandatory to record how the pregnancy ends.

Information on stillbirths (> 24 weeks gestation) will be much more complete than for miscarriage as it is a legal requirement for stillborn babies to be formally registered. In England and Wales this is completed within 42 days, in Scotland within 21 days and in Northern Ireland a stillbirth may be registered within 1 year. Data on the stillbirth may be available from several sources including HES, GP record data, maternity clinical systems and Office for National Statistics (ONS). In addition any stillbirth will have a detailed investigation using the Perinatal Mortality Review Tool (PMRT) where a structured review is conducted with healthcare professionals and the parents to ascertain what has happened. Data from the review is stored with MBRRACE (Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK) perinatal surveillance data and could potentially be linked to the MSDS through NHS Digital. MBRRACE also investigates late fetal losses that occur between 22-23 weeks gestation and neonatal deaths.

3.2.1.2 Congenital anomalies

Congenital anomalies may be detected at different stages of pregnancy. A congenital anomaly may be detected after a woman has a miscarriage or stillbirth either from post-mortem results or from visual inspection. Routine antenatal scans and blood tests at 12 weeks may detect Down's, Edwards' and Patau's syndromes and the detailed structural anomaly scan at 20 weeks specifically checks for fetal anomalies. Congenital anomalies may also be detected at birth if not previously identified and some, such as hearing defects or cardiac defects, may only be detected later in infancy. Many congenital anomalies may not be detected if these occur in fetuses that are miscarried in the first trimester. As discussed, very early miscarriages may be undetected in women who are unaware they are pregnant. In addition, miscarriages in the first trimester for which medical attention is sought are not routinely investigated for the presence of anomalies unless a woman has experienced three or more consecutive losses. The rationale for this is that most women who have one or two miscarriages will go on to have a successful pregnancy next time. Many pregnancy losses in the first trimester may be due to congenital anomalies (particularly chromosomal anomalies) and therefore go unidentified or recorded. Investigations are offered to women who have a miscarriage in the second trimester onwards but these are optional and therefore congenital anomalies may be missed if women decline an investigation into their pregnancy loss.

Where a congenital anomaly has been identified, the data should be well recorded within national congenital anomaly registers. NCARDRS (National Congenital Anomaly and Rare Disease Registration Service) in England and CARIS (Congenital Anomaly Register and Information Service) in Wales are likely to have the most complete data on congenital anomalies as information is collected at delivery and from antenatal scans, with the aim to follow-up throughout the life course of a child. In Scotland a congenital anomaly register is currently being set up (CARDRISS – Congenital Anomalies and Rare Diseases Registration and Information Service for Scotland) which will capture data from 2021 onwards. Congenital anomaly registers also record terminations of pregnancy for fetal anomaly (TOPFA).

Additionally some data on congenital anomalies is also available from maternity system data and GP record data but the completeness and detail will not be as good as in the congenital anomaly registers. Furthermore, the UK teratology information service (UKTIS) systematically follows up all pregnancies reported to the service with the aim of detecting congenital anomalies associated with medicines used during pregnancy. This is a voluntary system of reporting pregnancies that are exposed to medicines and chemicals during pregnancy and therefore the data are not captured for all cases of congenital anomalies.

3.2.1.3 Other adverse outcomes during pregnancy

Other adverse outcomes during pregnancy such as fetal growth issues may be captured within maternity information systems. Collection of growth data has recently been incentivised in England by Clinical Negligence Scheme for Trusts (CNST) but the data will not flow into the Maternity Services Data Set (MSDS) unless it is recorded using SNOMED codes. Fetal growth is captured through good recording of birthweight and gestation at delivery on SMR02 in Scotland (allowing birthweight centile to be calculated). Estimated fetal growth during ongoing pregnancies (i.e. at antenatal care appointments) is not captured in SMR02 however as returns are only triggered by inpatient care.

3.2.2 Adverse neonatal outcomes

Data on neonatal death and other neonatal outcomes are well recorded within a number of data sources across the UK. Outcomes such as preterm delivery, low birth weight will be captured at delivery with the majority of deliveries occurring in the hospital setting. Information on these outcomes may be available from maternity clinical systems, hospital episode data (e.g. HES, SMR02) and also the National Neonatal Research Database (NNRD). Additional neonatal outcomes occurring during the neonatal inpatient stay are also available from the NNRD in addition to data on the newborn physical examination, blood spot test for rare diseases and hearing test. Some data will be available on these outcomes within GP datasets such as the Clinical Practice Research Datalink (CPRD). Any neonatal outcomes occurring outside the hospital setting are likely to be managed either within General Practice or by a community midwife or health visitor. These details should be captured in GP record data, maternity data for outcomes prior to discharge from maternity services or community services data such as the Community Services Data Set (CSDS) in England.

3.2.3 Physical, neurodevelopmental and educational outcomes in children

Data on short and longer-term physical and child neurodevelopmental outcomes are available from a variety of sources. Data from the first 6 weeks after birth are available from midwife appointments and are captured in maternity datasets such as Maternity Services Data Set (MSDS) as care up to 6 weeks is part of maternity care. The majority of women are discharged from maternity services after 10 days post-delivery, however. The National Neonatal Research Database (NNRD) will also capture data on babies during the first few weeks after birth as well as health and neurodevelopmental status at age 2-years for very preterm babies. All babies aged 6-8 weeks are also invited for a thorough examination usually by their GP and information on this examination will be available from GP record data. Information from this 6-8 week review is also captured in the Community Services Data Set (CSDS) in England.

Data from health visitor appointments are captured in community datasets such as the CSDS in England, Child Health Systems Programme Pre-school (CHSP-PS) in Scotland and National Community Child Health Database (NCCHD) in Wales. Any special educational needs in children up to 5 years of age that are identified within community services in England will be captured in the CSDS. Babies aged 9–15 months should also be offered another review to include an assessment of language, learning, diet and behaviour. This review is usually carried out by health visitors but is not offered in all areas of the UK and therefore data will not be complete across the UK. All children aged 27-30 months are currently offered another health and development review which is usually carried out by a health visitor. This review (The Ages and Stages Questionnaire – ASQ-3) includes an assessment of general development including movement, speech, social skills, behaviour, hearing and vision. Data from the ASQ-3 is captured in data sources such as the CSDS in England and also in the NNRD in England, Scotland and Wales. In Scotland a further assessment is conducted at 4-5 years. After this, further child development outcomes will be captured in an ad hoc fashion when they are diagnosed in either a GP or hospital setting. GP data and hospital data such as HES should therefore provide a comprehensive range of data from childhood onwards.

Educational record data such as the National Pupil Database in England could also be a useful source of data on neurodevelopmental delay in older children but there are issues with the accessibility these data and the ability to link them to other data sources. Every pupil in England has a unique pupil number (UPN) assigned but this is different to the NHS number and there is currently no direct linkage between the two numbers. Linkage between health and educational datasets in England is possible and has been performed for individual studies but is a cumbersome process based on matching date of birth and postcode. Progress has been made recently, however, towards linking NHS and UPN numbers, with approval given by the Department for Education and the Department of Health and Social Care for this. In Scotland patients are allocated a CHI (Community Health Index) number at birth which is used in place of the NHS number. It is also technically possible to link health records with educational records in Scotland but there are similar challenges.

3.2.4 Other long-term health outcomes

Data to assess the impact of intrauterine exposure on long term health outcomes is also important. For example, there has been interest in the causal factors resulting in the substantially increased risk of diabetes in offspring of women with diabetes with possibilities (not mutually exclusive) including genetic, drug-related and intrauterine environmental (e.g. hyperglycaemia) determinants. A further example is with the increased risk of vaginal adenocarcinoma associated with diethylstilbestrol in children and women exposed in utero

with evidence suggesting that the risk is transgenerational and affects the grandchildren of the women exposed during pregnancy.

There is therefore a need for data on maternal exposure to medicines during pregnancy and breastfeeding that is linked to the health records for the children throughout the course of their lives. GP system data with a mother-baby link such as that developed by the Clinical Practice Research Datalink (CPRD) provides an important data source to investigate these types of outcomes but only if the exposure is prescribed in primary care.

3.2.5 Normal outcomes in pregnancy and breastfeeding

Whilst data on adverse outcomes in pregnancy and during breastfeeding are essential for safety monitoring and research, data on normal pregnancies/outcomes are also vitally important as a control group. Maternity datasets contain data on all pregnancies and will therefore have good data on normal pregnancies. Data should be available on all babies up until discharge from hospital and afterwards child development for all children is monitored through the statutory health visitor and child review appointments discussed above. There should therefore be basic information on normal pregnancies and child development up until the review at 2 years old. GP data will also capture information on normal outcomes in pregnancy and breastfeeding although there may be some misclassification if some have had adverse outcomes that have not been recorded/captured with the GP system.

Table 3.2. UK data sources with pregnancy and child developmental outcome data

	UK- wide	England	Scotland	Wales	Northern Ireland
Very early miscarriage (prior to booking)	GP record data	HES A&E	SMR01, SMR02		PAS
Early miscarriage (<13 weeks but after booking)	GP record data	MSDS, HES A&E	SMR01, SMR02		NIMATS, PAS
Late miscarriage (13-24 weeks)	GP record data	MSDS, HES A&E	SMR02		NIMATS, PAS
Stillbirth	GP record data, ONS, MBRRACE-UK	MSDS, HES A&E	SMR02, NRS stillbirth	MIds	NIMATS, PAS, General register office
Congenital anomalies	GP record data, UKTIS	NCARDRS, MSDS	CARDRISS	CARIS	NIMATS, CHS, BadgerNet
Other adverse outcomes during pregnancy e.g. fetal growth issues	GP record data	MSDS, HES	SMR02	MIds?	NIMATS, BadgerNet
Neonatal death	ONS, MBRRACE-UK	NNRD, MSDS, CSDS	NNRD, NRS infant deaths	NNRD	General register office, NIMATS
Other neonatal outcomes e.g. preterm, low birth weight	GP record data	MSDS, NNRD, HES maternity	NNRD, SMR02, SBR	NNRD, NCCHD, MIds	NIMATS, CHS

CARDRISS – Congenital Anomaly and Rare Diseases Registration & Information Service for Scotland, CARIS – Congenital Anomaly Register and Information Service, CHS – Child Health System, CHSP – Child Health Systems Programme, CSDS – Community Services Data Set, HES – Hospital Episode Statistics, MBRRACE-UK – Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK, MIds – Maternity Indicators data set, MSDS – Maternity Services Data Set, ONS – Office for National Statistics, NCARDRS – National Congenital Anomaly and Rare Disease Registration Service, NCCHD – National Community Child Health Database, NIMATS – Northern Ireland Maternity System, NNRD – National Neonatal Research Database, NRS – National Records of Scotland, PAS – Patient Administration System, SBR – Scottish Birth Record, SMR01 – Scottish Morbidity Record 01, SMR02 – Scottish Morbidity Record 02, UKTIS – UK Teratology Information Service

3.3 Data elements other than exposure and outcome

The development of a robust observational study investigating a possible association between an exposure and an event will account for potential confounding factors in the statistical analysis. However, in order to do this accurate and reliable data must exist on the relevant confounding factors. Some potentially important factors for exposures during pregnancy and breastfeeding include maternal age, maternal medical history including past obstetric history, use of folic acid, certain medications, smoking and alcohol use, recreational drug use and Body Mass Index (BMI). Genetic factors may also be important to consider but this is not routinely collected or available. Some genetic data is collected by Genomics England on a sample of 100,000 patients with rare diseases and from UK Biobank from 500,000 patients but the possibility to link these to other datasets is currently unknown and sample sizes are small.

Maternal medical and obstetric history will be available through GP records and need only be recorded once. Key medical and obstetric history information will also be recorded in the maternity system datasets at the booking appointment.

Use of folic acid, smoking and alcohol use and to a lesser extent BMI can vary during pregnancy and therefore require data capture ideally at multiple times during the pregnancy. Data on these factors are currently captured at the maternity booking appointment and will therefore be available through maternity information systems. Social factors are also recorded at the start of labour and again at discharge from maternity services. Data on smoking status is generally well recorded at every maternity appointment and women who smoke should also have carbon monoxide testing performed at every appointment. Alcohol use may be less well-recorded, and it is unclear if women are questioned about their alcohol consumption at appointments after the booking appointment or how accurate the responses are. Information on exposure to infectious diseases could also be important covariate for medicines prescribed in pregnancy and this data would likely be recorded in GP datasets and possibly maternity system datasets.

Capture of important covariates during breastfeeding such as smoking and alcohol consumption may be collected during health visiting appointments. With smoking an established risk factor for SIDS (Sudden Infant Death Syndrome) smoking status may be recorded. These data will then be available in community services datasets such as the CSDS in England.

Even with high quality data on known confounding factors, confounding by indication and residual confounding are two notoriously difficult issues to fully account for in observational studies. Instrumental variable methods have been proposed as potential approaches to reduce confounding in epidemiology that involve additional non-individual level data to be collected e.g. prescriber thresholds for prescribing particular medicines. This further highlights the need for rich datasets that contain data on a wide variety of variables in addition to the exposure and outcome of interest. In the example of using prescriber preference as an instrumental variable data is required on the individual prescriber. This information is readily available from GP datasets but not from community prescription data.

Ethnicity, or race, is another factor with potential implications for the safety of medicines taken during pregnancy and breastfeeding. “Race” is usually associated with biology and can give rise to differences in disease susceptibility, genetic determinants of disease and most importantly in the context of the efficacy and safety of medicines, difference in drug metabolism. “Ethnicity” is more a marker of cultural identification. The difficulty lies in that both

terms are largely social constructs that defy precise definition. In the UK, the standard NHS categories of “ethnicity” are confusing as these are variably based on skin colour, or country or continent of origin and there are around 18 or so categories. Making their use even more problematic is the growing proportion of the population that identify themselves as “mixed”. In addition, in the absence of more specific data, “ethnicity” is often used as a proxy for socio-economic status. A better understanding of the effect of medicines in different groups will require i) knowledge of genetic markers of disease susceptibility and drug metabolism (i.e. population genomics), ii) the individual variables that are societal/environmental confounders of disease risk (e.g. socioeconomic status, support status, occupation, urban/rural location,) and iii) phenotypic variable (e.g. co-morbidities, age, sex, BMI). Accurate and consistent collection of ethnicity data in routine clinical datasets will help to provide a better understanding of the effect of medicines in different groups as well as identify potential barriers to accessing therapies experienced by different parts of our population and can support efforts to improve inclusion. Improving data on ethnicity is challenging and not restricted to data relevant to medicines used in pregnancy and breastfeeding and is therefore beyond the scope of this working group to make recommendations on this. There are however ongoing initiatives on this topic including The Health Data Research UK (HDR-UK) Alliance of data custodians who are aiming to increase the representativeness of datasets across the board with an initial focus on ethnicity.

3.4 Data Quality

High quality data are essential to conduct robust research. Data quality can be defined in different ways and in this report it is considered in terms of completeness and accuracy of the individual data elements and the extent of availability of relevant data elements. The availability of relevant data elements has already been described in the preceding sections so this section will consider data quality relating to completeness and accuracy.

The completeness and accuracy of data from secondary-use sources such as electronic patient records or healthcare administrative sources such as the Hospital Episode Statistics (HES) in England is reliant on the diligence of the person entering the data (either a healthcare professional or a clerical assistant). An element of human error may therefore be present which will be greater if the person responsible for data entry i) has little or no subject knowledge; ii) is insufficiently motivated; or iii) does not appreciate the importance of the task. Training of personnel responsible for recording data is therefore essential.

Data entered into a clinical system in real-time may be more accurate than data captured retrospectively. This is an issue across all healthcare specialities and not just in pregnancy. An example is data from hospital appointments and treatments that are sent (often in letter form) to the patient’s GP to be stored with their primary care record. The information is not always entered into the structured fields of the GP record however and may simply be attached as a scanned version of the letter which is not useable for research. Furthermore, as the hospital/secondary care data are entered retrospectively onto the GP system there will be a delay between the event occurring and being captured on the system. This may result in outcome misclassification and patients being wrongly included in a control group. Another example is retrospective data collected directly from the patient such as medication use which relies on the patient being able to accurately recall the information and may be subject to recall bias if an outcome has already occurred.

The lack of electronic prescribing in secondary care creates an issue with data quality as clinicians may end up using free text to capture medication use or it may not be captured

electronically at all. This is a current issue in neonatal units where data on medication use are available through the National Neonatal Research Database (NNRD) but the quality is variable.

3.4.1 Improving data quality

A number of initiatives currently in place to increase the quality of routinely collected healthcare data related to maternity and neonatal care are described below.

The Clinical Practice Research Datalink (CPRD) offers a GP validation service whereby clinical diagnoses and outcomes recorded in the coded data can be verified by the GP. Linkage to other healthcare and health-related datasets can enrich data quality by providing data validation. CPRD also undertakes a number of data quality checks including collection level validation to ensure all data elements are of correct type, length and format with no duplicates; transformation level validation to ensure there are no orphan records; and research quality validation including practice and patient level quality flags.

Technical approaches to data quality assurance are available in some healthcare settings that create a pop-up information box if for example a value is outside a plausible range. Additionally, clinical software can mandate that certain data fields are not left blank.

NHS Digital have published two important data standard notices relating to maternity care – one primary maternity record standard on how clinical events are recorded and an Information Standards Notice (ISN) that legally obliges trusts to procure electronic maternity information systems that operate to the standard.[6] The maternity record standard will ensure that maternity record information is consistently captured to facilitate sharing of information across healthcare systems. A further ISN for the Maternity Service Data Set (MSDS) version 2 states that data must be submitted to NHS Digital. Work is also ongoing with Clinical Negligence Scheme for Trusts (CNST) to further drive up the quality of data flowing to MSDS. The standards have been developed to enhance data collection for direct patient care but a benefit is enhanced secondary use data for research. The standard is to be implemented by February 2021.

NHS Digital have also recently published a national data standard, the Healthy Child Record Standard (also mentioned in section 3,1,5 under breastfeeding) that will facilitate uniform data for children under the Healthy Child Programme and ensure all clinical providers and systems use the same data. This standard is part of a digital strategy, Healthy Children: Transforming Child Health Information[7], published by NHS England in 2016 that aims to make child health information more accessible.

Issuing a standard does not however guarantee that data quality will improve. This requires the systematic application of quality-assurance methods.

[6] Digital Maternity Record Standard. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb3066-digital-maternity-record-standard

[7] Healthy Children: Transforming Child Health Information, NHS England, Nov 2016. https://www.england.nhs.uk/wp-content/uploads/2016/11/healthy-children-transforming-child-health-info.pdf

3.4.2 Evaluating data quality

For the benefits of routine data use to be realised, systems are needed for systematic quality-assurance. This should cover quantification of missing data, and identification of potentially erroneous entries through e.g. identification of out-of-range values, internal inconsistencies in logic and temporal sequences, and duplicate entries. Completion of quality-assurance cycles requires feed-back of potentially erroneous entries to clinical teams and/or patients and subsequent correction.

The National Neonatal Research Database (NNRD) have had a strong focus on data quality with multiple quality assurance mechanisms. A range of checks are performed on the data using standard algorithms that look for issues such as missing data temporal inconsistencies and duplicate entries. Any errors are fed back to clinical teams and the baby's record is corrected. Additionally, research is ongoing to involve parents are in quality assuring entries in their baby's record.

3.5 Data linkage

As highlighted above, data relevant to studying the safety of medicines used during pregnancy are currently fragmented across many different data sources. Much data on exposure, outcomes of interest and other important factors is captured electronically somewhere in the UK and some data sources capture both exposure and outcome data but with variable quality. In order to make optimal use of the best available data on exposure and outcomes data sources must be linked and intergenerational linkage of mother and child records is critical.

The majority of data captured during pregnancy are from maternity information systems as these capture all hospital-based maternity appointments and ultrasound scans. As discussed above, these also have the capability to record medication but the data are likely to be incomplete and suffer from issues with recall. Furthermore, secondary use datasets such as the Maternity Services Data Set (MSDS) in England may not flow all the information captured in the system. GP datasets such as the Clinical Practice Research Datalink (CPRD) have much higher quality of GP prescription data as the prescriptions are issued during GP appointments and automatically entered into the GP system. Whilst GP record data does contain information on hospital appointments and events as letters are sent from the hospital to a patient's GP, these data are unlikely to be complete or entered in a timely fashion. Similarly for neonatal admission data, sources such as the National Neonatal Research Database (NNRD) capture very high quality data from all neonatal units in England, Scotland and Wales and although some neonatal outcome data are captured in maternity systems, it would be preferable to use data from the NNRD for data relating to babies admitted to a neonatal unit. Some data elements such as educational data are only available from educational datasets with no data currently available in other datasets. Data from national congenital anomaly registers such as NCARDRS and CARIS contain the most complete and accurate data on congenital anomalies including those that result in Termination of Pregnancy due to Fetal Anomaly (TOPFA), miscarriage and live birth.

There is a critical need therefore to link data sources with the best quality data together to ensure the highest quality of research and there are a number of developments in this area. CPRD currently links to HES A&E (Hospital Episode Statistics Accident and Emergency) and maternity data and ONS death registration data and are in dialogue to link to NCARDRS, MSDS and primary care dispensing data. NHS Digital currently collect over 200 datasets including national mental health, social care, primary care, community and maternity and work

is currently being undertaken to develop the capability to link primary care prescribing to other datasets such as HES, maternity and mental health. The aim within 12-18 months is to build a platform that will enable data to sit and be accessible and be linked. Also within NHS Digital is the capability for researchers to bring in smaller datasets and link these to a national dataset.

As described in section 2.3.2, the Health Data Research UK is also working to link different NHS datasets. One of the seven hubs is dedicated to real world evidence (Discover-NOW) and involves linking primary care, secondary care and non-health data covering North-West London. A comprehensive linked dataset will be available for a geographically defined population. The Health Data Research Innovation Gateway acts as a common portal through which data can be searched for and accessed and with the potential to host 'collections' of data. The Gateway supports the use of data, facilitates interoperability and provides analytical capability.

Intergenerational linkage of mother and baby records can be challenging. As described in section 2.2.1 above, the CPRD have developed a pregnancy register and mother-baby link to address this challenge, but a study published in 2019 found that 31% of pregnancy register deliveries (excluding stillbirths and deliveries based on late pregnancy or third trimester codes) had no linked infant.[8] Work is ongoing in collaboration with the London School of Hygiene and Tropical Medicine (LSHTM) to characterise pregnancies with no recorded outcomes with a view to refining the CPRD pregnancy register. Public Health Scotland routinely maintains an intergenerational linkage spine by seeding both the mother and baby's CHI (Community Health Index) number onto the National Records of Scotland (NRS) statutory live birth registrations. A similar system exists in Northern Ireland where babies are given a Health and Social Care (HSC) number at birth and this is recorded in the Northern Ireland Maternity System (NIMATS) that also holds the mother's HSC number. There is no mother-baby linkage spine currently in England or Wales.

An important aspect to linkage of data sets is sustainability. There are examples of routine data linkages that are ongoing and maintained such as NCARDRS linkages to HES and ONS (Office for National Statistics) and also CPRD linkages to HES and ONS. These linkages enable monitoring and surveillance and for the NCARDRS linkages also inform high quality clinical care in addition to research. However, many different data sets are linked on an ad hoc basis for individual research projects with a requirement to destroy the linkage at the end of the project. This results in a non-standardised approach to the linkage and duplication of effort to create the same linkage either for a different project by the same research group or for a different organisation. A large amount of time and resource can therefore be invested in linking datasets together accurately for individual projects which is lost if it cannot then be reused. Furthermore, if the same linkage is duplicated by different research groups/organisations each time it is required errors are likely to be introduced. Routine linkages also benefit registries and clinical care by avoiding duplication of data requests to the NHS and by improving the quality and completeness of the data. This issue is further discussed below under information governance (section 3.7).

Another important aspect to data linkage is interoperability to ensure that data from different data sets are compatible with each other. International common data models (e.g. OMOP – Observational Medical Outcomes Partnership) exist that transform data into a common format using common terminologies and coding systems to facilitate linking of data sets. Data

[8] Minassian C., Williams R., Meeraus WH. *et al.* Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. *Pharmacoepidemiol Drug Saf.* 2019; 28(7): 923-33

processing standards exist such as FHIR (Fast Healthcare Interoperability Resources) that enables the exchange of data between healthcare applications. Common coding systems such as SNOMED CT are also being widely adopted which allows improved interoperability between data sets.

3.6 Public trust and confidence

The availability of high-quality electronic healthcare data for research purposes requires both the engagement of healthcare professionals to record high quality data and the trust of patients to consent to their data being used for secondary purposes. The recording of high-quality data directly benefits the patient, through improved patient care but also indirectly benefits patients and the public by improving the quality of research studies and the evidence on which policy and regulatory decisions are made. Without the engagement of healthcare professionals and the trust of the public the quality, completeness and availability of data suffers.

Data protection and patient confidentiality are important and issues with these have featured prominently in the media in recent years. The public's views on using linked administrative data for research purposes were examined by Ipsos MORI in 2014.[9] The results showed low initial awareness of research and methods which drove scepticism about the value of the research. Furthermore, the public were more uncertain around commercial access to data. It is therefore vitally important that patients and healthcare professionals are aware of the benefits of capturing data and allowing access to it to ensure that research that benefits patients is not impeded.

3.7 Information governance

In the UK, the legal frameworks that cover patient healthcare record data are the Data Protection Act (DPA) 2018 (which brought the EU General Data Protection Regulation (GDPR) into law) and the Common Law Duty of Confidentiality (CLDC). The data protection legislation requires a valid, legal reason for the collection and processing of data and the data cannot normally be disclosed without patient consent. For the benefit of important medical research however confidential patient information is required but it is not always practical to obtain consent in which case patient consent is implied. In England and Wales, confidential patient data may be used in accordance with Section 251 of the NHS Act 2006 to support medical purposes that are in the interest of patients or the wider public. Separate arrangements are in place for Scotland and Northern Ireland. If an implied patient consent approach is used, a robust transparency agreement should be in place to ensure that members of the public are fully informed appropriately.

In England and Wales, Section 251 approval is considered by the Confidentiality Advisory Group (CAG) who advise the Health Research Authority (HRA). In Scotland advice is sought from the Public Benefits and Privacy Panel for access to patient information and in Northern Ireland from the Privacy Advisory Committee.

Linking existing datasets together requires the use of patient identifiable information and therefore requires approval from the bodies described above. In addition, individual data controllers will have their own information governance processes that researchers must adhere to and apply for approval to use the data. Examples are IGARD (Independent Group

[9] Dialogue on Data. Ipsos MORI. https://www.ipsos.com/ipsos-mori/en-uk/dialogue-data

Advising on the Release of Data) who review all applications for NHS Digital data and ISAC (Independent Scientific Advisory Committee) for applications to use CPRD (Clinical Practice Research Datalink) data. Individual universities and grant-giving bodies may also have separate requirements. It can be very time-consuming to obtain permissions to analyse the data even for one nation.

Not just for research relating to medicines used during pregnancy and breastfeeding but in planning any observational study using electronic patient healthcare record data, approvals from all the relevant ethical bodies must be sought and this can be an extremely complex and time-consuming process especially if data from different devolved nations is required.

Data access may also be different across the devolved nations, for example Northern Ireland provides a secure setting for data access, the “Honest Broker Service” which is a supervised room, with disclosure control procedures for outputs. Linked healthcare data are only available in this setting.

A further information governance requirement is that datasets and data linkages must be destroyed at the end of the project in order that national patient opt-outs are respected. As these may change over time data may not be repurposed. There is some inconsistency in this with the accepted good practice that a dataset should be archived so that it can be re-interrogated should there be a need for verification in the future. CPRD permits researchers to archive analysis-ready datasets in line with their institutional data retention and archiving policies so that they can be used for verification purposes but cannot be repurposed for another project. There is a need however to keep data in order that long-term issues might be studied.

Approvals for linking health datasets to non-health related datasets such as educational data have previously been difficult to obtain as the HRA and CAG advise on health datasets that have medical purposes. Educational data is not always been viewed as being related to health and therefore does not fall within remit of the HRA.

4 ADDRESSING GAPS IN AVAILABLE DATA AND REALISING OPPORTUNITIES FOR OPTIMISATION WITHIN CURRENT DATA SOURCES

This chapter details the discussion of the expert working group on ways to address the gaps in currently available data relevant to medicines used during pregnancy and breastfeeding discussed in the previous chapter. This chapter also focuses on how opportunities for optimisation of current data sources might be realised.

4.1 Exposure data on medicines used in pregnancy and breastfeeding

Exposure data from primary care is generally well captured in GP datasets and although currently these datasets cover a sample of the UK population there are ongoing developments to expand coverage across the UK within existing datasets such as the Clinical Practice Research Datalink (CPRD) and with novel datasets such as that under development by NHS Digital. There is an opportunity however to improve currently available prescription dispensing data from community pharmacists to ensure that the date the prescription was issued is captured within the dataset. Without this data element it is impossible to accurately determine the timing of exposure in relation to pregnancy, but also breastfeeding.

Exposure data from secondary care is currently not well captured and considered a gap. This is true across all secondary care and not just for medicines prescribed in pregnancy and breastfeeding. Electronic Prescribing and Administration (EPMA) is not yet available in all hospital departments and there are issues with the recording of medicines administered on the ward being captured on paper/cardex. The project underway by NHS Digital to create a dataset to capture medication prescribed and administered within hospitals with EPMA will largely address the current gap in secondary care data within the next 5 years. This initiative is very much supported and encouraged by this Working Group with an aim of being able to link maternity datasets such as the MSDS (Maternity Services Data Set) to secondary care prescribing data. In order to maximise the potential of this dataset for capturing medication use during pregnancy and breastfeeding, maternity and neonatal departments should be encouraged to implement EPMA systems. There has however recently been issues with the roll-out of such systems for paediatrics and neonates with systems predicated on adult algorithms that are not safe for use in children. EPMA systems for maternity and neonatal departments therefore need to be configured appropriately to enable recording of medication use in infants with appropriate data cleaning steps put in place.

Over-the-counter (OTC) medicines are another exposure that are currently not well captured. Maternity datasets such as the MSDS in England and NIMATS (Northern Ireland Maternity System) in Northern Ireland have the capability to capture information on medication use including OTC medicines, although the MSDS does not currently routinely collate these data and any data captured in the maternity system would need to be entered as SNOMED CT codes in order to flow to the MSDS. Information on medication use including OTC should be captured by midwives at the booking appointment and ideally at subsequent maternity appointments as well. As women self-declare the information to a midwife, this approach would allow information to be captured on actual medication use rather than medicines prescribed but perhaps not taken. There is also an opportunity to collect information on illicit drug use through this route. Medication use captured should cover the pre-conception period (3 months prior to conception) and beyond, and the timing of use should also be recorded.

Currently the National Institute of Health and Care Excellence (NICE) guidance on Antenatal Care for Uncomplicated Pregnancies[10] does not currently include a recommendation to capture any medication during routine antenatal appointments. The guidance is currently being updated however and there is therefore an opportunity to request that medication use (including OTC, prescribed and illicit drug use) is included in the recommended information to be collected. The use of SNOMED CT codes to record the information should be encouraged rather than free text. The updated guidance is due for consultation in October 2020 and to be published in April 2021. NICE guidelines are implemented in Wales and Northern Ireland if relevant. In Scotland the Scottish Intercollegiate Guidelines Network (SIGN) produces national clinical guidelines and also adopts NICE guidelines if relevant. While there is not one currently that covers antenatal care, groups or individuals are able to propose guideline topics. It is recommended that Scotland either adopts the NICE guidance or proposes antenatal care as a guideline for development by SIGN to include the collection of medication.

A further possible option to better capture any medication use during pregnancy and breastfeeding is to ask women to self-report use. This could be achieved through patient access to their electronic Personal Health Record (PHR). Several electronic health record suppliers have products in use that allow patient self-reporting. One example is BadgerNet Maternity Notes which is an online portal and app that allows women to access their maternity records online with the facility for women to add information on topics such as allergies, birth preferences and feedback. In some hospital trusts BadgerNet Maternity Notes is replacing the paper-based handheld notes. There is therefore an opportunity to include the recording of any medication use within products such as BadgerNet Maternity Notes for women to update themselves.

A potential source of OTC data might also be available through commercial loyalty cards for pharmacies such as Boots. Issues with such an approach would be knowing who the medication was for (as people can buy for others or to keep at home for possible future use) and how to identify pregnancy from the data which may not be patient-level data. Furthermore, people may not always buy all their OTC medication from the same pharmacy but alongside their grocery shopping or from different pharmacies. This approach would likely be highly challenging to implement and unlikely to result in high quality data. Another potential approach would be to mandate community pharmacists to record OTC supply to pregnant or breastfeeding women. This would similarly have issues for people buying medicines for others or to keep at home.

Data on medication dosage, duration and indication is often poorly recorded or lacking. A drive towards better capture of medication details would significantly improve the quality of data on medication use not just in pregnancy and breastfeeding but for all patients. A move away from free text capture of these prescription details towards more structured data capture would be beneficial and reduce the need for data cleansing and the potential introduction of errors. The recording of indication is currently lacking and is generally inferred from clinical diagnoses made at the same or previous appointments, specific recording of indication would again increase the quality of medication data and improve the ability to investigate potential confounding by indication.

The implementation of a national data standard that will standardise the capture of breastfeeding information at routine health appointments will increase the quality of data captured. Data will only be captured at these appointments however, and therefore breastfeeding information between appointments or after these have stopped will be missing.

[10] NICE guidance on Antenatal care for uncomplicated pregnancies - https://www.nice.org.uk/guidance/cg62

To obtain more frequent and complete information on breastfeeding, a possible option could be for women to self-report breastfeeding information through their PHR as mentioned above as a way of capturing data on OTC medication use.

4.2 Pregnancy and child development outcomes of interest

Data on stillbirth and miscarriage that occurs after the maternity booking appointment is well recorded and should be complete. Data on early miscarriage that occurs prior to the booking appointment is currently considered to be a gap, however. Many women experiencing an early miscarriage that seek medical care now present to an Early Pregnancy Assessment Unit (EPAU) and there is therefore a potential to access data from these units. Data from EPAUs are not currently contributing to a dataset and some do not currently collect data but there is an opportunity to address this so that the data are captured in a standardised way and made available for research.

High quality congenital anomaly data are captured in congenital anomaly registries but also to some extent in maternity datasets. Data on congenital anomalies in fetuses that are lost through miscarriage in the first trimester are however incomplete as only women having recurrent miscarriage will be investigated. Other adverse outcomes during pregnancy such as fetal growth issues should be captured within maternity datasets and neonatal outcomes for babies admitted to neonatal units are well recorded in the National Neonatal Research Database (NNRD) in addition to other datasets.

Data on physical and neurodevelopmental outcomes in children are recorded in a number of data sources from routine health visitor and GP appointments up to the age 2 years. However, data on longer-term child development outcomes may be incomplete as the ASQ-3 (Ages and Stages Questionnaire) is one of the only routine reviews conducted on child development with no further routine assessment beyond 2 years of age. GP record data or hospital data may capture some information after this on an ad hoc basis. Educational record data may provide information on neurodevelopmental delay in older children using educational attainment as an outcome measure but as discussed in the previous section the data are difficult to link due to the unique identifier in the educational data being different to the NHS/CHI (National Health Service/Community Health Index) number with no direct automated way to link the two. Linkage between educational datasets and health datasets has been achieved in England, Scotland and Wales but has been challenging. Some progress has been made towards linking the NHS number and UPN (Unique Pupil Number) in England and this development is very much welcomed and supported by the Working Group. The importance of establishing a link between health and educational datasets to study the longer-term impact of medication during pregnancy and breastfeeding needs to be emphasized with encouragement for similar linkage to be enabled in the devolved regions of the UK.

4.3 Data elements other than exposure and outcomes

The accurate capture of ancillary data to exposure and outcome data is critically important to ensure high quality observational studies. A broad dataset is therefore necessary that accurately captures factors such as medical and obstetric history, concomitant medications and lifestyle factors including smoking and alcohol consumption, recreational drug use and folic acid and other vitamin supplementation.

Some data on potentially important confounding factors is available within different datasets and maternity systems have the capability to record information on lifestyle factors such as smoking, alcohol and the use of vitamins and folic acid however these factors are unlikely to be recorded routinely at every appointment. The National Institute of Health and Care Excellence (NICE) guidance on Antenatal care for uncomplicated pregnancies recommends giving advice on lifestyle factors at the first contact and at the booking appointment. With the guidance due to be updated shortly an opportunity potentially exists to recommend capture of these factors at regular time points. As mentioned above Scotland could adopt the NICE guideline or an antenatal care topic could be proposed for a new guideline in Scotland by Scottish Intercollegiate Guidelines Network (SIGN).

4.4 Data quality

High quality data capture is essential to produce high quality research and inform high quality clinical care. As detailed in chapter 3, there are current initiatives within the Maternity Service Data Set (MSDS) and the National Neonatal Research Database (NNRD) to drive up the quality of the routinely captured data but there is a need to go further and creating standards could help. Clinical standards are needed to recommend that pregnant women are asked about medicines and other factors and that this information is recorded. Information standards are needed to ensure that the information recorded is linked to an appropriate dataset and quality standards are needed to ensure that information is recorded in way that is accurate and structured. Healthcare professionals have a professional responsibility to record clinical data accurately but there needs to be a central responsibility to ensure that any data used for secondary purposes are quality-assured. NICE (National Institute of Health and Care Excellence) guidance could help in making recommendations on recording quality data but also the Professional Record Standards Body (PRSB) which was established in 2013 to ensure that there are consistent standards for care records and the PRSB were involved in the development of the maternity record standard. As established by the National Neonatal Research Database (NNRD), there is considerable value in having mothers reviewing their babies' data and making clinicians aware of any errors. Strategies and processes should be devised to encourage women to review their own data and that of their babies and report any inaccuracies.

Data quality could also be improved through clinical software prompts. Software used in maternity, neonatal and primary care should implement technical measures that either flag up out of range values entered or mandate that certain fields are completed.

Though not specific to medicines used in pregnancy, a better understanding of the wider scope of the data healthcare professionals enter may help to increase the quality of data entry and these aspects should be part of Continuing Professional Development (CPD). Trainee healthcare professionals should be targeted for training on the importance of healthcare data in medical research. Furthermore, growing use of drop-down menus, forced choice entries, and speciality-specific data standards incorporating unambiguous definitions will help drive data quality.

4.5 Data linkage

A large amount of data relevant to medicines used in pregnancy and breastfeeding currently exists and is available for research. However, as discussed previously the data is fragmented

across many different datasets. To ensure the best quality data is used for exposure, outcomes and other ancillary data, the most accurate and comprehensive datasets must be linked.

A number of initiatives are currently ongoing to link datasets relevant for medicines used in pregnancy and breastfeeding which are welcomed but further linkages between key datasets are required if data on medicines used during pregnancy and breastfeeding are to be fully optimised. Utilising the best available data on exposure, outcomes and ancillary data linked data from GP records, prescription dispensing records, maternity system datasets, congenital anomaly registers, birth and death registries, MBBRACE (Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK), neonatal datasets, community health datasets and educational datasets should be prioritised. A 'task and finish' group including major data custodians could help with this.

The remit of the Health Data Research UK (HDR-UK) is to bring together datasets and facilitate data linkage and access for research purposes. Data controllers are encouraged to facilitate greater access to their data in a timely way for research for the direct benefit of patients, for example, by using the UK Health Data Research Innovation Gateway to ensure datasets can be found and are accessible, inter-operable and reusable (FAIR). A maternal and child health-specific HDR-UK hub would also be highly beneficial.

A mother-baby linkage spine is also required in all four nations to facilitate the linkage of maternal exposures to outcomes in the baby or child. This currently exists for Scotland and Northern Ireland but not for England or Wales.

A sustainable infrastructure for linkage of datasets is required to ensure that once a linkage has been established it is then available for subsequent projects and researchers to make use of. This reduces duplication of effort to create the linkage and promotes a more standardised approach and a reduction in errors that might be introduced from repeated work.

Consideration also needs to be given to the interoperability of different data sets, with a need to ensure they follow international data interoperability standards where possible and that data are mapped to common terminologies to facilitate data linkage.

4.6 Public trust and confidence

Public engagement with secondary use of electronic healthcare record data is vitally important to maintain patient trust in what their data are being used for and to gain better appreciation of the benefits of allowing access to them. This is not specific to medicines used in pregnancy but for all electronic healthcare data for all medical research purposes. Medicines used in pregnancy could however be an excellent example to inform the public of the importance of healthcare data for research. The field of pharmacovigilance was set up due to exposure to thalidomide during pregnancy and phocomelia in the child, and there are public concerns about drug use in pregnancy. Communication of examples of how the secondary use of routinely collected healthcare data has helped answer important safety questions on medicines used in pregnancy may help to maintain public trust.

Currently a distinction is made between research and clinical need which is not helpful. Using routine healthcare data for research purposes directly improves patient care and patient outcomes and it is therefore an artificial construct to separate the two. It is considered good practice that data is shared as widely as possible for patient benefit and data controllers are encouraged to make their data available in a timely way.

4.7 Information governance

Information governance for data linkage is currently a fragmented and consequently resource intensive and time-consuming process, particularly for studies involving data from across the UK. Researchers often opt to only go through one approval process and choose England as the biggest nation. Patients in Scotland, Wales and Northern Ireland are then not able to benefit from the research taking place and the UK's ability to lead and participate in international research is also being heavily impeded. A mutual recognition system where if a project is approved in one nation it is recognised and accepted in the others would greatly simplify the process and encourage more research in the devolved UK nations. There has been a precedent for this type of approach with a study conducted by the British Paediatric Surveillance Unit (BPSU) with collaboration between Scotland and England on ethics approval.

A lack of clarity exists over the requirements to destroy data at the end of a research project. There is a need to archive datasets so that they can be used for verification purposes in the future and it seems that this need is not always reflected in ethics application processes. Furthermore, there is a need to archive data so that longer-term outcomes may be studied in the future. The consequences of medicines used in pregnancy may not become apparent for decades or may even have generational impact such as diethylstilbestrol and therefore it is imperative that data are not destroyed. The implications of destroying data in these cases are that a new study may take decades before results are available.

The information governance approvals process for linking health datasets to health-related datasets such as educational data where the relevance of the data to health may not be immediately obvious needs to be clarified and reviewed to facilitate linkage of these data.

5 CONCLUSION AND RECOMMENDATIONS

The Expert Working Group on Optimising Data on Medicines used During Pregnancy was set up specifically to identify ways to improve data collection and access to information relevant to research on medicines used during pregnancy and expanded its scope to also include breastfeeding. The UK has a wealth of relevant data on medicines used in pregnancy and breastfeeding but much of it is fragmented across different datasets according to exposure and outcome and UK nation. There are gaps in the available data with some variables either missing completely or poorly recorded and so there are opportunities to optimise data collection and access. Other broader issues around data quality, linkage, public engagement and information governance also present opportunities to facilitate data collection and access for medicines used in pregnancy and breastfeeding but also for all epidemiological research.

The future vision and ultimate aim of this Expert Working Group is to promote and enable complete and accurate data capture of exposure to medications in pregnancy (including the immediate pre-conception period) and breastfeeding and relevant outcomes and to facilitate access to the data for research purposes. Current barriers include the fragmentation of UK data, variability in data quality and completeness, challenges around information governance and sustainability of linkages and access processes and costs to researchers. The current Covid-19 pandemic has accelerated and improved collaborative working, demonstrating that data can be accessed and shared rapidly whilst maintaining patient confidentiality and there may be lessons to learn from this experience. There are however real opportunities to address some of the current barriers and therefore the Expert Working Group make the following recommendations:

The Group recognises that both motivation and resource are required to implement these recommendations. Our hope is that this report provides the motivation for these recommendations but our fear is that without adequate resource little progress can be made.

The EWG recommends that a new group should be established to oversee implementation of its recommendations. It was considered that this group should be distinct from the EWG but should include representatives from stakeholders in all sectors concerned with the capture of data on the use of medicines during pregnancy and breastfeeding. It was noted that funding and administrative support would need to be identified to ensure that the recommendations of the EWG can be implemented effectively.

<u>Addressing gaps in data collection:</u>

1. Electronic Prescribing and Medicines Administration (EPMA) systems configured appropriately for use in pregnant women and children should be implemented in maternity and neonatal departments. Data from EPMA needs to be captured in a standardised way, centralised and accessible for data linkage. The project currently underway in England to implement EPMA systems in secondary care is strongly supported.

2. Information on any medication use during pregnancy and breastfeeding including prescribed, over-the-counter or illicit drugs in addition to lifestyle factors should be captured during maternity and health visitor appointments and self-reported directly from women:

a. Personal Healthcare Records should ensure that women are able to self-report any medication use during pregnancy and breastfeeding.

b. National maternity datasets should collect any medication data recorded at the booking appointment and other appointments or self-reported.

c. National clinical guidelines should be updated or developed to include the collection of medication use at the booking antenatal appointment as well as lifestyle factors.

3. Data from Early Pregnancy Assessment Units should be standardised, collated centrally and made available for research purposes

<u>Optimising linkage of existing datasets:</u>

4. Linkage of health and educational datasets is vitally important to study the longer-term impact of medication used in pregnancy and breastfeeding on educational attainment.
    a. Progress made to date to link Unique Pupil Numbers (UPN) and NHS numbers in England is welcomed and should be accelerated. Similar linkages should be made in the devolved nations of the UK.
    b. The governance approval process needs to be reviewed and clarified regarding the linkage of health and health-related datasets.

5. A mother-baby linkage spine should be set up and maintained in all four devolved nations.

6. Further linkages between datasets relevant to medicines used in pregnancy and breastfeeding are encouraged (e.g. GP, maternity, congenital anomaly and neonatal databases) and a 'task and finish' group of major data custodians should be set up to achieve this.

7. Data controllers are encouraged to facilitate greater access to their data in a timely way for research for the direct benefit of patients, for example, by using the UK Health Data Research Innovation Gateway to ensure datasets can be found and are accessible, inter-operable and reusable (FAIR).

8. Information governance for data linkage of health-related datasets should be optimised to enable greater access to data for the benefit of patients through research:

    a. A sustainable infrastructure for data linkage of datasets is required to ensure that once a linkage has been established it is then available for subsequent projects and improves clinical care.

    b. Researchers should be permitted to archive datasets rather than destroy data to enable future verification of studies or the study of long-term outcomes.

    c. A mutual recognition system for ethics applications, whereby a project approved in one nation is recognised and accepted in the others, should be developed.

Improving data quality:

9. General practice and prescription dispensing data sources should capture indication, dosing, duration and prescription/dispensing dates for all medicines. Information should be structured data rather than free text.

10. Datasets should follow international standards and common coding systems such as SNOMED CT to ensure data interoperability when linking datasets

11. Systems are needed for systematic quality-assurance:
    a. Strategies and processes should be devised to encourage women to review their own data and that of their babies to be able to and report any inaccuracies.
    b. Software used in maternity, neonatal and primary care should implement technical measures that automate data quality through prompts and mandatory fields.

12. The importance and benefits of routinely collected healthcare data for improving patient care both directly and through research into the effects of medicines in pregnancy and breastfeeding on the child should be emphasised to healthcare professionals and the public:

    a. Medicines used in pregnancy should be used as an example to inform the public of the benefits and importance of healthcare data. Specifically examples of how the secondary use of routinely collected healthcare data has helped answer important safety questions should be used.

    b. All trainee healthcare professionals should be targeted for training on the importance of routinely collected healthcare data for studying the effects of medicines in pregnancy and breastfeeding on the child.

Encouraging research:

The remit and focus of this Expert Working Group has been on optimising data available for research on medicines used during pregnancy and breastfeeding. The next step is to ensure that these data are used to their full potential to improve the safety of medicines in pregnancy and breastfeeding. In line with the recommendation from the UK Clinical Research Collaboration (UKCRC) the Group also make the following recommendations in relation to use of routine healthcare data in research:

13. Observational research to investigate safety of medicines used in pregnancy and during breastfeeding should be commissioned and a system of ongoing surveillance should be established and funded.

Annex 1 – Interests held by members of the EWG

	Interest
Chair	
Professor Jane Norman MBChB MD FRCOG F MedSci FRCP Edin FRSE Dean, Faculty of Health Sciences, University of Bristol (previously University of Edinburgh)	I have no shares nor received any lecture fees from the pharmaceutical industry for the last three years. I spoke at a European Society of Diabetes and Endocrinology at Merck sponsored symposium: Edinburgh University paid my travel expenses, registration and accommodation, Merck then reimbursed Edinburgh University for this. In the last three years I have done consultancy for Dilafor, LifeArc and GlaxoSmithKline. All fees were paid to Edinburgh university and not directly to me. I received no personal income from any of these organisations. I have an MRC funded research grant which is termed a "MICA" (industry collaboration grant). For this research grant, Merck are donating active and placebo metformin for a randomised clinical trial. Again, the "donation" goes to the University of Edinburgh and not to me.
Members	
Professor Peter Brocklehurst MBChB MSc FRCOG FFPH FMedSci Professor of Women's Health, Director for Birmingham Clinical Trials Unit (BCTU)	Personal Interests: Biotest AG. Consultancy. 12th February 2019.
Mr Paul Brown Clinical Lead NHS Digital, Prescribing, Medicines and Pharmacy	None
Ms Caroline Cake Chief Operating Officer at Health Data Research UK	None
Dr Rachel Charlton Research Fellow in Epidemiology, Dept of Pharmacy and Pharmacology, University of Bath	Personal Interests: I own a small number of shares in GlaxoSmithKline. Non Personal Interests: In the past 12 months the Group within which I work at the University of Bath has held research grants from pharmaceutical companies for research on topics outside of those covered by this working group.
Mr Chris Dickson Senior Clinical Lead Platforms and Infrastructure, Paediatric Nurse, Digital Child Health and Digital Maternity	None
Professor Helen Dolk Professor of Epidemiology & Health Services Research, School of Nursing Institute of Nursing and Health Research Jordanstown campus, University of Ulster	None. Note that I work within an EU Innovative Medicines Initiative funded project CONCEPTION, which involves collaboration between public and industry partners, but we (public partners) are not funded by industry, and we work under the EncEPP Code of Conduct for Scientific Independence and Transparency.

Professor Elizabeth Draper BSc (Hons) MPhil PhD FFPH FRCOG ad Eundem Professor of Perinatal & Paediatric Epidemiology, University of Leicester	None
Dr Kenneth Hodson MBChB MD MRCP(UK) MRCOG Head of UK Teratology Information Service, UK Teratology Information Service, Consultant in Obstetrics and Maternal Medicine	I have received honoraria from UCB Pharma to present at two sponsored meetings (discussing treatment options for women with arthritis and psoriasis in pregnancy) 20th June 2018 and 9th May 2019.
Dr Matthew Jolly National Clinical Director for the Maternity & Women's Health, Acute Medical Directorate, NHS England & NHS Improvement	None
Professor Neena Modi MBChB MD FRCP FRCPCH FFPM (Hon) FMedSci Professor of Neonatal Medicine, Consultant, Imperial College London	Personal Financial Interests: • I am employed by Imperial College London and have an honorary consultant contract with Chelsea and Westminster NHS Foundation Trust; I undertake no personal private practice • I receive travel and accommodation reimbursements for lecturing at academic conferences • In the last five years I also received conference travel and accommodation costs from Chiesi, Nestle and Prolacta Life Sciences, speaker honoraria from Chiesi Pharmaceuticals, and travel costs and an honorarium as member of the jury for the Belgian Anton Faes Foundation Award • I am member of the Independent Scientific Advisory Board of the Singapore National Medical Research Council, Flagship Translational and Clinical Research Programme on Developmental Pathways to Health and Disease, for which I receive travel and accommodation, and an honorarium Non-Personal Financial Interests: • I am Director of the Neonatal Data Analysis Unit at Imperial College London that undertakes commissioned work for NHS England, NHS Litigation Authority, Department of Health, Royal College of Obstetrics and Gynaecology and Royal College of Paediatrics and Child Health • In the last 5 years I have received research grants from the UK Medical Research Council, National Institute for Health Research, March of Dimes, British Heart Foundation, Westminster Medical School Research Trust, HCA International, Chiesi, Nestle, Prolacta Life Sciences, Shire Pharmaceuticals, Collaboration for Leadership in Applied Health Research and Care for Northwest London, Healthcare Quality Improvement Partnership, Bliss, NHS England and UK Department of Health • I am a member of the Nestle Scientific Advisory Board; I receive no personal remuneration; payment is made to my employer, Imperial College London Other Interests: Unpaid positions

	• Trustee (and President), Medical Women's Federation • Trustee, David Harvey Trust • Trustee, Action Cerebral Palsy • Trustee, TheirWorld • Member, UK Health Research Authority National Research Ethics Advisory Panel • President-Elect, British Medical Association
Professor Joan Morris Professor of Medical Statistics, Population Health Research Institute, St George's, University of London	My husband is an employee of GSK.
Dr Puja Myles Head of Observational Research, Clinical Practice Research Datalink (CPRD)	None
Ms Katharine Robbins Information Analysis Lead Manager, Maternity, Child Health and Community at NHS Digital Leeds, West Yorkshire	None
Mrs Sarah Stevens Public Health Consultant, Service Lead, The National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) Public Health England	None
Dr Sarah Stock Reader and Consultant in Maternal and Fetal Medicine, University of Edinburgh	None
Mrs Madeleine Wang BA (Hons) Lay Representative. Patient Advocate	None

Annex 2 - Table 1 Summary of data sources with data on medication exposure during pregnancy and/or relevant outcomes

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
Primary care								
Clinical Practice Research Datalink (CPRD)	Since 1987	25% of UK population from 1 in 5 GP practices. Includes data on currently registered patients as well as historic records on patients who are deceased and transferred out.	UK-wide.	Prescriptions, medical events Immunisations, referrals, laboratory investigations requested in primary care and other procedures undertaken in primary care	Data sources including secondary care (HES), deprivation and national death registration	Maternal medical history, prescriptions, delivery-related data from linked hospital maternity services data Pregnancy register based on algorithm to ascertain all pregnancies in CPRD data Pregnancy outcomes and child health via mother-baby link	Data controller: Department of Health and Social Care (DHSC) Data access: via a licence; restricted to public health research and surveillance More details available at: www.cprd.com	Discussions about expanding linkage to other PHE and NHS Digital datasets Plans to refine and extend pregnancy register and mother-baby link
IQVIA Medical Research Data (IMRD) (formerly known as THIN)	Since 1994	~ 5% of UK population currently registered	Sample of General practices using the Vision software in the UK	Prescriptions, medical events, tests and referrals.	currently no linkages available	Maternal medical history, prescriptions Data during pregnancy Pregnancy outcomes and longer-term follow-up (Mother-baby linkage)	Data controller: IQVIA. Data access: through a sublicense, or subset for a bespoke study.	
QResearch	Since 1989		General practices using EMIS	Data from patient's primary care records	Entire database linked to cause of	Pregnancy records from hospital data?	Data controller: The University of Oxford for datasets linked to QResearch and point of access to data;	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
			largely in England		death data, cancer and hospital data		Data access: Database open to academic researchers. More details available at: https://www.qresearch.org/information/information-for-researchers/	
Oxford - Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC)	Started systematic national surveillance for influenza and other infections in 1967	1764 practices covering over 15 million patients in England and Wales	England and Wales	Mainly capture primary care data: demographics, diagnoses, symptoms, prescriptions, vaccination history, lab tests, referrals to hospitals, specialist care. Also include RSC population configuration (based on age, ethnicity, index of multiple deprivation and rurality), clinical ontologies, covariates and outcomes. Practices also collect data such as serology, virology	Have capability to link to other datasets, researchers must obtain other datasets separately. "Fixed" linkage applications in progress with Public Health England (PHE) and NHS Digital's DARS	Ontologies to identify pregnant women, links to babies and family. Maternity and child health data linkage pending current data application	Data controller: All data handling and management follows the current Oxford-RCGP RSC data management SOP. Data access: Data only be used for SQUIRE (surveillance, quality improvement, research and education) purposes. https://clininf.eu/index.php/orchid-data/. Can request data via data access application form. Also accelerating access through provision of themed data sets (cardiovascular, cancer, surveillance and acute pathways of care, and mental health). More details available at: https://www.rcgp.org.uk/rsc	ORCHID hub https://www.phc.ox.ac.uk/publications/1108889 Extended Surveillance to include COVID-19 containment strategy, separately and together with PHE surveillance systems. Current plans: https://preprints.jmir.org/preprint/24341 Implementation of SNOMED CT coding
ResearchOne*			Electronic patient records currently held on the TPP	Prescriptions and medical records captured in the SystmOne record system – GP, Community Care, Child Health, Palliative Hospital,		Mother and child outcomes	Data controller: Data controller at each provider organization must opt-in to ResearchOne.	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
			SystmOne clinical system in England	Acute Hospital, Urgent Care, Accident & Emergency and Out-of-Hours.			Data access: Can request data from ResearchOne More details available at: http://www.researchone.org/research-faqs/	
NHS Business Services Authority (NHSBSA) prescribing data		All prescription dispensed in the community in England	England	Prescribed and dispensed medicines; Age and gender for patients identified via NHS number captured from prescription; Since 04/2015 started to capture patient ids			Data controller: NHSBSA Data access: Aggregated / anonymised data via ePACT2 (registered NHS users) and monthly Practice Level Prescribing Report (open data). Patient level data (currently from April 2018 but intention to backdate to April 2015) available via NHS Digital Data Access Request Service (DARS) See: Medicines dispensed in Primary Care data More details available at: https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data	
Scottish Primary Care Information Resource (SPIRE)		Dependent on consent of GP practices?	Scotland	Data from patient's primary care records: Prescriptions, medical events.			Data controller: National Services Scotland (NSS)? Data access: via SPIRE Strategy & Oversight Group? More details available at: https://spire.scot/professional/data-request/	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
Prescribing Information System (PIS)	Since April 1993	Every prescription dispensed in the community	Scotland	Prescriber, dispenser details, cost and drug information. Data on practices, organisational structures, prescribable items (manufacturer, formulation code, strength)			Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	
Primary care GP dataset	Since 2000	Around 78% of the Welsh population	Wales	Symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patient's home environment.			Data controller: Swansea University Data access: Can be requested via SAIL, via the information governance panel (IGRP) More details available at: https://saildatabank.com/application-process/	
General practice prescribing data		All primary care prescription dispensed in Wales	Wales	All prescribed medicines, dressings and appliances that are dispensed each month			Data controller: NHS Wales Shared Services Partnership Data access: Available for use and re-use under the Open Government Licence (OGL) More details available at: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ https://nwssp.nhs.wales/ourservices/primary-care-services/general-information/data-and-	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
							publications/general-practice-prescribing-data-extract/	
Enhanced prescribing database (EPD)	Since 2010	All primary care prescriptions dispensed in NI	Northern Ireland	Primary care prescriptions dispensed	Possible to link to other datasets in NHSBSO to get gender, age, demographics; maternity, clinical?		Data controller: Business Services Organisation (BSO), Honest broker services (HBS) Data access: via HBS More details available at: http://www.hscbusiness.hscni.net/services/2454.htm	Possible linkage to primary care
GP Dataset	In development						Data controller: NHS Digital	
Secondary care								
Hospital episodes statistics (HES)		Hospital patients in England	England	Details of hospital admissions, outpatient appointments, A&E attendances	Linked to some primary care databases (see CPRD, THIN)	HES – delivery episodes	Data controller: NHS Digital Data access: via Data Access Environment (DAE), Data Access Request Service (DARS) More details available at: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/users-uses-and-access-to-hospital-episode-statistics	
Maternity Services Data Set (MSDS)	Since 2015	Around 90% of deliveries recorded in HES (England)	England	Key information at each stage of the maternity care pathway: mother's demographics, booking appointments, admissions and re-admissions, screening tests, labour and delivery along with baby's demographics, admissions,	Data may be linked to data held by NHS Digital from various other data sets and collections and external data sources	See 'Data captured'	Data controller: NHS Digital Data access: via Data Access Request Service (DARS) More details available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/maternity-services-data-set	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
				diagnoses and screening tests				
Hospital Treatment Insights (HTI)	Hospital pharmacy dispensing data from January 2010; linked by patient to HES data from 2005	Around 7 million patients from 43 participating NHS trusts. Patients representative by age and gender of all hospitalised patients	27% of hospital trusts in England	All hospital claims activity from HES and drugs dispensed including inpatient, outpatient, discharge medications, accident and emergency (except ward stock drugs)	Data routinely linked by NHS Digital to HES every quarter	Pregnancy information from hospitalisation including maternity information HES, drugs dispensed specifically to patients with a pregnancy ICD-10 code	Data controller: IQVIA Data access: record level data restricted to academic researchers and employees of IQVIA. Other non-commercial researchers by request. More details available at: https://www.iqvia.com/locations/united-kingdom/information-for-members-of-the-public/hospital-treatment-insights-service	Linkage to primary care Expansion of the database
Neonatal critical care minimum dataset	Since 2007	Neonatal intensive care units, maternity wards?	England	Critical care and discharge details, weight, gestational length at delivery (subset of HES but more details)		See 'Data captured'	Data controller: NHS Digital Data access: via Data Access Request Service (DARS) More details available at: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/scci0075-neonatal-critical-care-minimum-data-set-version-2	
Mental Health Services Dataset (MHSDS)		Specialist secondary mental health care	England	Demographics, patient indicators, various elements captured re mental health service use	MSDS	Children's mental health, learning disabilities, ASD	Data controller: NHS Digital Data access: via Data Access Request Service (DARS)	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
		services, specialists secondary learning disabilities					More details available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set/access-data	
Electronic prescribing and medicines administration (EPMA)				Hospital prescribing from EPMA system				In development
Scottish birth record (SBR)			Scotland	Birth and delivery related, smoking, alcohol, illicit drugs, neonatal care discharge details			Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	Improved national neonatal data return is under development
Scottish Morbidity Record (SMR01)	Since 1960	All patients admitted as general patient day case and inpatient care	Scotland	Admission type, patients' condition, operations, waiting times etc		Identifies early pregnancy losses	Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access. More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	
Scottish Morbidity Record (SMR02)	Since 1975	All women admitted as inpatients or day cases	Scotland	Clinical and demographic characteristics and outcomes		Pregnancy related: e.g. LMP, estimated gestation, antenatal steroids, smoking history, weight (booking), drug, alcohol misuse;	Data controller: Public Health Scotland	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
		to Scottish maternity units				Labour related data; Baby related e.g. mode of delivery, outcome (later pregnancy losses) of pregnancy, birthweight, gestational age Apgar score	Data access: Contact Public health Scotland's research support team for data access: More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	
Hospital Electronic Prescribing and Medicines Administration (HEPMA)		Inpatients in some hospitals in Scotland?	Scotland	Inpatient prescribing			Data controller: NHS Scotland? Data access?	
Patient Episode Database for Wales (PEDW)	Since 1997	All NHS hospital patients in Wales and Welsh resident patients treated in English NHS hospitals	Wales	In-patient and day case activity, including hospital stays for giving birth.		Delivery record attached to each general record for each baby born, data relating to mother and babies	Data controller: NHS Wales Informatics Service (NWIS) Data Access: NWIS to supply information on application process. PDIT.Requests@wales.nhs.uk Can be requested via SAIL, via the information governance panel (IGRP) More details available at: https://saildatabank.com/application-process/	
Maternity indicators dataset (MIds)	Since 2016	Antenatal care provided in Welsh maternity units and	Wales	The dataset combines records from a mother's initial assessment with a child's birth record.		Data items related to pregnancy/initial assessment: e.g. gestation, parity, weight, height, smoking (booking); Labour and birth related data e.g.	Data controller: NHS Wales Informatics Service (NWIS) Data access: request via NWIS to supply information on application process	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
		care of deliveries in Wales.				birth mode, outcome of birth, Apgar, birthweight	PDIT.Requests@wales.nhs.uk More details available at: https://nwis.nhs.wales/contact-us/	
Patient Administration System (PAS)		Patients in Northern Ireland	Northern Ireland	Details about admissions, appointments, ward attendances, day ward.		Includes A&E miscarriage data	Data controller: Business Services Organisation (BSO), Honest broker services (HBS) Data access: via HBS More details available at: http://www.hscbusiness.hscni.net/services/2454.htm	
Northern Ireland Maternity System (NIMATS)	from 2009; Full coverage from 2011	All 5 health trusts in NI; Data coverage and completeness varying	Northern Ireland	Information about all women who present to maternity services in NI: demographic, health, past medical/obstetric and family history; medication recorded by midwives; breastfeeding on discharge; Some information on infants (e.g. birthweight)	Linkable to other NI health datasets, including prescribing datasets	See 'Data captured'	Data controller: Business Services Organisation (BSO), Honest broker services (HBS) Data access: via HBS More details available at: http://www.hscbusiness.hscni.net/services/2454.htm	
Paediatric Critical Care Minimum dataset (PCCMDS)	Since 2007	All paediatric intensive care units in UK	United Kingdom	Data for whole episode of critical care; Data relating to each day within episode of care (e.g. weight, interventions, drugs etc). Collected as part of PICANet (paediatric intensive care audit network)		Child health data	Data controller: HQIP Data access: ?	Possibly be expanded to all critical care levels Possible agreement with NHS Digital
National Neonatal Research	Since 2007	All neonatal units in England,	England, Wales and Scotland	Demographic details, daily records of interventions and treatments throughout		See 'Data captured'	Established at Neonatal Data Analysis Unit, Imperial College London/ Chelsea and	Work in progress – refresh of Neonatal Data Set (NHS

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
Database (NNRD)		Wales and Scotland submit data (currently >1 million babies included)		the neonatal inpatient stay, information on diagnoses and outcomes, and follow-up health status at age two years			Westminster NHS Foundation Trust by chief investigator Professor N Modi Data access: submit protocol, data request form; Associated costs More details available at: https://www.imperial.ac.uk/neonatal-data-analysis-unit/	Information Standard comprising items included in the NNRD) Access jointly via Imperial and NHS Digital expected in the future
Community based								
Community Services dataset (CSDS)	2015 (preceding dataset – CYPHS mandated for national flow from 2015)	England; Completeness rate varying	England	Demographic, social and personal circumstances, breastfeeding and nutrition, care event and screening, diagnoses, long term conditions, childhood disabilities, scored assessments)		Health visiting data-breastfeeding, ages and stages questionnaire	Data controller: NHS digital Data access: via Data Access Request Service (DARS) More details available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/community-services-data-set	
National Child Measurement Programme (NCMP)	Since 2005	Primary school children aged 4-5 (reception) and aged 10-11 (year 6)	England	Height, weight (BMI)	Can be linked to data sources such as dental survey, HES	See 'Data captured'	Data controller: NHS Digital Data access: Details available at: https://digital.nhs.uk/services/national-child-measurement-programme	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
Child Health Systems Programme – Pre-school (CHSP Pre-school)	Since 2011	All pre-school children in Scotland	Scotland	Assessment of children's health, development, wider well-being		Health visitor 10-14 day, 6-8 week, 13-15 month child health review records - breastfeeding	Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/?	
Minor Ailment Service (MAS) data		Patients registered with GP in Scotland with some eligibility criteria e.g. aged <16 or <19 and in full-time education, aged >60 years, on benefits etc.	Scotland	Information on medication provided directly by pharmacists			Data controller: NHS National Services Scotland? Data access: Database under development	Database under development
National Community Child Health Database (NCCHD)	Since 1987	All children born, resident or treated in Wales; approx.. 35,000 births a year	Wales	Details of child health examinations and immunisations	Can be linked to other data sources via SAIL.	See 'Data captured'	Data controller: NHS Wales Informatics Service (NWIS) **PDIT.Requests@wales.nhs.uk** Data Access: NWIS to supply information on application process. Can be requested via SAIL, via the information governance panel (IGRP)	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
							More details available at: https://saildatabank.com/application-process/	
Child Health System (CHS) database		Coverage of all children	Northern Ireland	Community based operational system comprising 7 modules – child register (breastfeeding?), preschool vaccination and immunisation, preschool developmental surveillance, school health, special needs, new-born hearing, influenza)		See 'Data captured'	Data controller: ? Data access: ?	
Registries								
Human Fertilisation and Embryology Authority (HFEA) Register	Since 1991	Licensed UK fertility clinics	United Kingdom	Some past pregnancy details, demographics, Fertility treatment, birth outcomes	Linkage possible on ad-hoc basis	See 'Data captured'	Data controller: HFEA Data access: Anonymised register data available on website; Contact HFEA if linkage with another dataset More details available at: https://www.hfea.gov.uk/about-us/our-data/	
National Study of HIV in Pregnancy and Childhood (NSHPC)	Population-based surveillance of obstetric and paediatric HIV since 1990	United Kingdom and Ireland	United Kingdom and Ireland	Active reporting scheme for pregnancies in women living with HIV, babies born to women living with HIV and other children diagnosed with HIV		Pregnancy and outcome of pregnancy details, infection details, antiretroviral drugs, paediatric details	Data controller: UCL Data access: Summarized data from the NSHPC available; Contact If require data which is not available on website.	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
							More details available at: https://www.ucl.ac.uk/nshpc/resources/quarterly-data-update	
UK epilepsy and pregnancy register	Established in 1996	Sample size: end of 2014 >8,000	United Kingdom	Completed registrations either through a healthcare provider or pregnant woman		Details about pregnancy; health details of baby	Data controller: Royal Victoria hospital, Belfast. Data access: Contact the register More details available at: http://www.epilepsyandpregnancy.co.uk/pages/contact.htm	
National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)	Since 2015	Births with congenital anomalies and rare disease across England; since 2017/18 100% coverage	England	Antenatal, Delivery and postnatal data	Births/deaths from ONS, HES	See 'Data captured'	Data controller: Public Health England (PHE) Data access: Researcher to contact NCARDRS More details available at: https://www.gov.uk/guidance/the-national-congenital-anomaly-and-rare-disease-registration-service-ncardrs	Discussions with CPRD, MSDS, HQIP; Possibility of linking with prescribing data from NHSBSA
Congenital anomalies and rare disease registration and information service for Scotland (CARDRISS)		All babies affected by major structural or chromosomal anomaly, or recognised syndrome in Scotland. Also babies affected by inherited endocrine,	Scotland	Data on mother, pregnancy, baby and anomaly		See 'Data captured'	Data controller: Public Health Scotland Data access: Database in development	In development – will launch in 2021 (currently congenital anomaly records based on linkage of existing records)

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
		metabolic and haematological conditions.						
Congenital Anomaly Register & Information Service (CARIS)	Since 1998	All cases of congenital anomaly born to mothers' resident in Wales; Approx.. 1500 babies per year in Wales	Wales	Antenatal, Delivery and postnatal data	Linkages possible (through SAIL)	See 'Data captured'	Data controller: Public Health Wales Data access: via SAIL More details available at: https://saildatabank.com/application-process/	
The Northern Ireland Cerebral Palsy Register (NICPR)	Born since 1977 or living in the area since 1992		Northern Ireland	Child's demographics, type and severity of movement problems, presence of any other problems (seizures, learning, speech and language, vision and hearing), birthweight, gestation, if the child was a singleton or multiple birth (e.g. twin, triplet), schools attended, professionals seen and child's G.P.		See 'Data captured'	Data controller: Queen's University Belfast Data access: https://www.qub.ac.uk/research-centres/NorthernIrelandCerebralPalsyRegister/ContactUs/	
Cohorts								
Millennium cohort study (MCS)	Babies born between 09/2000-01/2002	~ 19,000 children in UK (selected through child benefit records)	UK	A range of healthcare, socio-emotional, cognitive and behavioural development data collected at different timepoints: family composition, parental education,	Linked health data from birth registration, maternity records, National pupil	Antenatal care, pregnancy related and birth outcomes	Data controller: UK Data Service Data access: Freely available to researchers under standard conditions More details available at:	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
				employment, parental health, ante-natal care, labour and delivery, breastfeeding, health, childcare, hospital visits, social & cognitive development, education record	database (England), KS1 & 2(Wales, Scotland); other linked administrative data		https://cls.ucl.ac.uk/data-access-training/	
Born in Bradford birth (BiB) cohort	Babies born between March 2007 – December 2010 with new phase starting January 2019.	Births at the Bradford Royal infirmary; 13,858 babies; 12453 mothers, 3353 fathers	Bradford (UK)	Mothers: sociodemographics, pregnancy measures, birth outcomes; blood and urine measurement at recruitment; pregnancy biomarkers, genomics, GP records – prescriptions etc, hospital admissions, Babies: cord blood sample, genomics, child follow-ups, GP records – prescriptions etc, Hospital admissions etc, Education record	Linkage to Yorkshire and Humber Congenital anomalies register, GP records, Education records, Maternity hospital records, hospital records	Pregnancy measures, birth outcomes, pregnancy biomarkers, prescriptions	Data controller: BiB executive group? Data access: Complete proforma for proposal More details available at: https://borninbradford.nhs.uk/research/how-to-access-data/	
Children Growing Up in Liverpool (C-GULL)	Will start in 2021?	Birth cohort - 10,000 Liverpudlians	Liverpool	Health and well-being of children and their families		Health and well-being of children and their families	Data controller: University of Liverpool Data access: Contact research team More details available at: https://www.liverpool.ac.uk/research/research-themes/living-well/c-gull/	
Pregnancy Outcome Prediction Study (POPS)	2008-2013	4,512 women with first	Cambridge	Blood obtained at recruitment, 20, 28 and 36 weeks of gestational age. After delivery, samples		Generated a simple screening test which is strongly predictive for preeclampsia and FGR at term.	Data controller: Cambridge University? Data access:?	POPs 2 study in progress

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
		singleton pregnancies		from the placenta, umbilical cord, and cord blood.				
Avon Longitudinal Study of Parents and Children (ALSPAC)	All pregnant women resident in a defined area in the South West of England, with an expected date of delivery between 1st April 1991 and 31st December 1992, were eligible and 13 761 women (contributing 13 867 pregnancies) were recruited.			Mother's demographics, life course, health behaviours, obstetric, clinic assessments, biological samples, Outcomes in offspring	Cancer registry, deaths from ONS, NHS Digital, CPRD, geo-spatial linkages, National pupil database	See 'Data Captured'	Hosted by University of Bristol Data access: Requests from all researchers - Charges applied for data request More details available at: http://www.bristol.ac.uk/alspac/researchers/access/	Linkages under development – health (GP, HES, community care), education (HESA)
Growing up in Scotland (GUS)	Since 2005;	Birth cohort 1: 5217 children (born 2004/05) Child cohort: 2,858 children, born in 2002/03 Birth cohort 2: 6,127 children born in 2010/11	Scotland	Cognitive, social, emotional, behavioural development; Physical and mental health well-being, Childcare, education and employment, Home, parenting, family, community, social networks, involvement in offending and risk behaviour.	Some data has been linked to administrative data held by health and education authorities	Pregnancy and birth related topics, child development, child health	Data controller:? Data access: via UK Data Service More details available at: https://growingupinscotland.org.uk/using-gus-data/accessing-gus-data/	
National statistics								
Office for national statistics (ONS)		UK	UK; although births and deaths data is for England	ONS birth registration: e.g. Birth date, gestational age, live/stillbirth		See 'Data captured'	Data controller: ONS Data access: via ONS More details available at:	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
			and Wales only				https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy	
National Pupil Database (NPD)	Since 2002	Pupils in state funded education, non-maintained special schools, sixth form and further education colleges and (where available) independent schools' further education	England	Pupil's demographics, detailed information on pupils' test and exam results, prior attainment and progression at each key stage		See 'Data Captured'	Data controller: Dept for Education Data access: ONS secure research service (ONS SRS) More details available at: https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts	
Scottish school education data		School children in Scotland	Scotland	Attainment, achievement, pupil statistics		See 'Data Captured'	Data controller: Scottish Government Data access: ?	
National Records of Scotland (NRS) Statutory Register of Deaths	Since 1855	Deaths in Scotland	Scotland	Date of death, cause of death, occupation, marital status, sex, age		Date and cause of death	Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access More details available at:	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
							https://www.isdscotland.org/Products-and-Services/eDRIS/	
Notification of abortion statistics (AAS)	Since 1968	Women who have had termination of pregnancy	Scotland	Data related to termination of pregnancy – place, age, deprivation, gestation, parity, repeat terminations		See 'Data Captured'	Data controller: Scotland's Chief Medical Officer (CMO), ISD, Scotland Data access: CMO More details available at: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=64	
National Records of Scotland (NRS) statutory register of births	Since 1855	Births in Scotland	Scotland	Birth date, time of birth, place of birth, occupation of father		See 'Data Captured'	Data controller: Public Health Scotland Data access: Contact Public Health Scotland's research support team for data access More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	
General Register Office for Northern Ireland (GRONI)			Northern Ireland	Records of births, deaths, marriages, stillbirths, adoptions		Neonatal deaths, stillbirths	Data controller: Northern Ireland Statistics & Research Agency Data access: ?	
Data/Research platforms								
eDRIS (electronic data research innovation service in Scotland) **	Dependent on database	Scotland – dependent on database	Scotland	Research platform comprising electronic patient records; Range of datasets	Scottish birth record can be linked to community prescription data	Maternity patient data of all births	Data controller: Public Health Scotland Data access: Public Health Scotland's data access team.	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
							Charges for Research Co-ordination services More details available at: https://www.isdscotland.org/Products-and-Services/eDRIS/	
The Secure Anonymised Information Linkage Databank (SAIL)**	Dependent on database	Wales - Dependent on database	Wales	Primary care prescriptions medical records, Secondary care (e.g. hospital episodes, A&E)	Linkage done for various data sources.	Patient episode database Wales (PEDW) captures pregnancy outcome; National Community Child Health Database-NCCHD; Congenital anomaly register and information service (CARIS)	Data controller: Swansea university Data access: Core datasets (e.g. NCCHD) in accordance with SAIL's standard info governance process – Information Governance Review Panel (IGRP); Core restricted e.g. CARIS – also need agreement of data provider More details available at: https://saildatabank.com/application-process/	
Honest Broker Service (HBS)	Dependent on database	Northern Ireland – Dependent on database	Northern Ireland	Primary care prescriptions, medical records, Secondary care, maternity system		Northern Ireland Maternity System (NIMATS); Child Health System, Enhanced Prescribing Database (EPD)	Data controller: Business Services Organisation (BSO) Data access: via the Honest Broker Governance Board (HBGB). More details available at: http://www.hscbusiness.hscni.net/services/2454.htm	
Other sources of mortality data								
Mothers and Babies: Reducing Risk	Since 2018	>6,300 reviews started;	England, Scotland and Wales;	Reviews of circumstances, care leading up to and surrounding each stillbirth		Details (antenatal, intrapartum, neonatal issues) of all perinatal deaths from 22+0 weeks	Data controller: HQIP+NHS England for England data; HQIP+NHS National Services	

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
through Audits and Confidential Enquiries across the UK (MBRRACE-UK)		~1,500 completed	Modified arrangements for Northern Ireland	and neonatal death, and deaths of babies in the post-neonatal period having received neonatal care		gestation until 28 days after birth (excluding termination, and birthweight<500g)	Scotland for Scotland; HQIP – rest of UK Data access: researchers are able to apply to the Healthcare Quality Improvement Partnership (HQIP), the commissioners of the MBRRACE-UK programme who are also the data controllers More details available at: https://www.npeu.ox.ac.uk/mbrrace-uk/research	
National Child Mortality Database (NCMD)		Children who died before their 18th birthday in England	England	Information about children who die before 18 years of age, including medical conditions, any factors intrinsic to the child (including mother's pregnancy), social factors		See 'Data captured'	Data controller: HQIP Data access: via University of Bristol? More details available at: https://www.ncmd.info/	
Child Death Review Programme	Since Oct 2009	Children who died after Oct 2009 before their 18th birthday and who were normally resident in Wales or died with Wales	Wales	Patterns and causes of child deaths		See 'Data captured'	Data controller: Public Health Wales Data access: ChildDeath.Review@wales.nhs.uk More details available at: https://phw.nhs.wales/services-and-teams/child-death-review/	
Other sources								

Name of data source	Time period of data collection	Population covered	Geographical coverage	Data captured	Data linkage	Maternity and child health data	Data controller and access mechanisms	Ongoing developments
UK Teratology Information Service	From 1983	UK	UK	Systematic follow-up of reported pregnancies to ascertain foetal outcomes		See 'Data Captured'	Hosted by Newcastle upon Tyne Hospitals NHS Trust Data access: via application from researchers More details available at: http://www.uktis.org/index.html	Currently collaborating with NCARDRS linking congenital abnormality data with prescribing data.

*Includes data from primary and secondary care

**These databases/data banks can include many data sources hence could be primary or secondary sources