Investigation and Application of a Pregnancy Register based on Electronic Primary Care Data

Jennifer Campbell

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

June 2023

Department of Non-Communicable Disease Epidemiology
Faculty of Epidemiology and Population Health
London School of Hygiene & Tropical Medicine
University of London

Funded by The Clinical Practice Research Datalink,
part of the Medicines and Healthcare Products Regulatory Agency

LSHTM Research group affiliation: Electronic Health Records Group

Declaration

I, Jennifer Campbell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Jennifer Campbell (28/03/2023)

Abstract

Background:

Electronic Health Records (EHR) are useful for studying pregnancy exposures and the outcomes for mother and child. However, as these data are not collected primarily for research accurately identifying the timing of pregnancies is challenging. Pregnancy episodes can be uncertain either because they have no recorded outcome or they overlap with another pregnancy episode. The aims of this work were to better understand how to handle uncertain pregnancy episodes in EHR data and to apply this knowledge to the question: does having COVID-19 during pregnancy increase the risk of pregnancy loss?

Methods:

Scenarios were identified potentially explaining why uncertain pregnancy episodes occur. Criteria were established and systematically applied to determine whether episodes had evidence of each scenario. Recommendations on how to handle these episodes were generated. A matched cohort study was conducted using EHR data to examine whether COVID-19 infection during pregnancy is associated with pregnancy loss and to test the implementation of developed recommendations.

Results:

Evidence found suggests that most uncertain pregnancy episodes are true and current pregnancies for which the data contain valuable information. Utilising EHR data found evidence that women who had COVID-19 during pregnancy had an 18% higher risk of pregnancy loss compared to contemporary controls and a 39% higher risk compared to pre-pandemic controls. Adjustments to the study population to include uncertain pregnancy episodes highlighted the potential risk of exposure misclassification associated with including all uncertain episodes. Blanket decisions to include or exclude uncertain episodes may lead to under ascertainment of pregnancies, biased study populations and errors in analysis such as exposure misclassification.

Conclusions:

Researchers should consider a tailored approach to utilising uncertain pregnancy episodes dependent on the design and purpose of the study. COVID-19 during pregnancy may increase the risk of pregnancy loss supporting the use of vaccination campaigns to protect pregnant women.

Acknowledgments

I would like to offer my heartfelt thanks to the following people without whom I could not have got here…

First and foremost to my supervisors Dr Caroline Minassian, Professor Krishnan Bhaskaran, and Professor Sara Thomas for all of your patience, wisdom, and encouragement. And to Dr Rachael Williams, for your unwavering support not only as a supervisor but as a line manager. It has been an absolute privilege to learn from each of you and I could not have asked for a better supervisory relay team, thank-you.

To the Clinical Practice Research Datalink for funding me, and to CPRD's Directors Dr Janet Valentine and Dr Puja Myles thank-you for granting me this opportunity. To my friends and colleagues at CPRD, especially to my fellow Senior Researchers past and present: Eleanor Axson, Helen Booth, Sonia Coton, Dan Dedman, Arlene Gallagher, Rebecca Ghosh, Susan Hodgson, Mike Longeron, Tarita Murray-Thomas, Jessie Oyinlola, Helen Strongman and Ellie Yelland. Thank-you for all of your kindness and support through-out the juggling act that the last six years have been.

To my co-authors Dr Helen McDonald and Mia Harley thank-you for your valuable input.

To all of the women who gave their data enabling this research to be possible. The trauma and loss that the outcomes in this analysis represent does not go unacknowledged. I hope that the methodological aspect of this work can play a small part in improving research aimed at reducing the risk of pregnancy loss.

Love and thanks to my parents for the many opportunities you worked so hard to give me which have led me here. And to my husband George, your unquestioning support of me and my pursuit of this PhD has meant everything, thank-you.

Finally I would like to dedicate this work to my children. Jude, Albie, you were both so little when I began this journey, I hope I have shown you that anything can be possible if you work hard enough.

Contents

Abstract……3

Acknowledgments……4

List of Tables……8

List of Figures……11

Abbreviations……12

Chapter 1: Introduction……13

1.1 Rationale for this Research……13

1.2 Thesis Structure……14

1.3 Background……16

1.3.1 Electronic Health data in the UK……16

1.3.2 Pregnancy Research……16

1.3.3 The importance of EHR data in pregnancy research……17

1.3.4 The challenges of using EHR data to study pregnancy……18

1.3.5 COVID-19……19

1.3.5.1 The impact of the COVID-19 pandemic in the United Kingdom……19

1.3.5.2 COVID-19 in Pregnancy……19

1.4 Aims……22

1.5 Objectives……22

1.6 Chapter Summary……23

Chapter 2: Data Sources……24

2.1 Clinical Practice Research Datalink (CPRD)……24

2.1.1 Data Structure, Coding and Key Variables in CPRD……24

2.2 The CPRD Pregnancy Registers……25

2.3 Linked Data……28

2.3.1 Hospital Episode Statistics……28

2.3.2 Office for National Statistics (ONS) Mortality Data……28

2.3.3 Second Generation Surveillance System (SGSS)……28

2.3.4 Index of Multiple Deprivation (IMD)……29

2.3.5 Utilisation of Linked Data……29

2.4 Chapter Summary……31
Chapter 3: Uncertain Pregnancies in the CPRD Pregnancy Registers……32
3.1 Introduction……32
3.2 Published Paper: Examining Uncertain Pregnancy Episodes in the CPRD Pregnancy Registers……33
3.3 Published Supplementary Files……49
3.4 Chapter Summary……90
Chapter 4: A Systematic Review of COVID-19 and Pregnancy Loss……91
4.1 Introduction……91
4.2 Published Paper: Systematic Review Protocol……92
4.3 Published Supplementary Files……99
4.4 Research Paper Submitted for Publication……110
4.5 Research Paper Supplementary Files……141
4.6 Chapter Summary……146
Chapter 5: COVID-19 and Risk of Pregnancy Loss: An Applied Example……147
5.1 Introduction……147
5.2 Research Paper Intended for Publication……148
5.3 Research Paper Supplementary Files……166
5.4 Investigating the Potential Impact of Excluding Uncertain Pregnancy Episodes……174
5.4.1 Obtaining Additional Pregnancy Outcomes from HES linked data……174
5.4.2 Pregnancy Episodes which have been split by the algorithm into conflicting episode……178
5.4.3 Conclusions……180
5.5 Chapter Summary……185
Chapter 6: Discussion……186
6.1 Introduction……186
6.2 Overview of Key Findings……186
6.2.1 Uncertain Pregnancy Episodes in the CPRD Pregnancy Registers……186
6.2.2 A Systematic Review of COVID-19 and Pregnancy Loss……188
6.2.3 COVID-19 and risk of Pregnancy Loss: An Applied Example……189
6.3 Summary of the strengths of this work……190
6.4 Summary of the potential limitations of this work……191

6.5 Implications for policy and future research……..192
6.5. 1 Implications for EHR researchers and data providers……193
6.5.2 Implications for clinicians and public health……193
6.5.3 Recommendations for future research……193
6.6 Conclusions……194
Bibliography……195

List of Tables

Chapter 2		
Chapter 2, Table 1	*Key variables to consider when utilising CPRD data.*	25
Chapter 2, Table 2	*Utilisation of linked data in this thesis*	30
Chapter 3		
Chapter 3, Table P1	*Description of potential scenarios leading to pregnancy episodes with no recorded outcome and scenario criteria applied*	39
Chapter 3, Table P2	*Description of potential scenarios leading to conflicting episodes and scenario criteria applied*	41
Chapter 3, Table P3	*Baseline characteristics of the pregnancy episodes in the February 2018 Pregnancy Register*	43
Chapter 3, Table P4	*Numbers of pregnancy episodes with recorded outcome missing which were consistent with applied criteria of each scenario*	44
Chapter 3, Table P5	*Numbers of conflicting pregnancy episodes which were consistent with applied criteria for each scenario*	46
Chapter 3, Table P6	*Issues with different approaches to dealing with uncertain episodes and recommendations*	47
Chapter 3, Table S1	*Key CPRD GOLD variables*	49
Chapter 3, Table S2	*CPRD Pregnancy Register Variables*	50
Chapter 3, Table S3	*ICD codes indicating end of pregnancy*	53
Chapter 3, Table S4	*OPCS codes indicating end of pregnancy*	62
Chapter 3, Table S5	*HES Maternity Values to indicate delivery*	65
Chapter 3, Table S6	*Pregnancy Read codes identified as likely to be recorded as useful pregnancy history*	66
Chapter 3, Table S7	*Antenatal Read codes identified as pregnancy advice codes*	70
Chapter 3, Table S8	*Read codes potentially misclassified as antenatal rather than outcomes*	71
Chapter 3, Table S9	*Outcome Groupings*	73
Chapter 3, Table S10	*Read Codes identified as likely to only be recorded during current pregnancy*	74

Chapter 3, Table S11	*Outcome Group Combinations*	78
Chapter 3, Table S12	*Read codes for Antenatal scan*	79
Chapter 3, Table S13	*DID Snomed fetal scan codes*	81
Chapter 3, Table S14	*Number of episodes with a suitably timed outcome in linked HES data*	82
Chapter 3, Table S15	*Numbers of pregnancy episodes with recorded outcome missing which were within practice UTS follow-up and patient's current registration period that were consistent with applied criteria for each scenario*	83
Chapter 3, Table S16	*Numbers of conflicting pregnancy episodes which were within practice UTS follow-up and patient's current registration period that were consistent with applied criteria for each scenario*	86
Chapter 3, Table S17	*Number of conflicting episode pairs by outcome combination*	89
Chapter 4 Paper 1		
Chapter 4, Table P1	*Inclusion and exclusion criteria*	95
Chapter 4, Table P2	*Database search strategy*	95
Chapter 4, Table P3	*Example of data collection form*	97
Chapter 4, Table S1	*Prisma-P checklist*	99
Chapter 4, Table S2	*NICE Quality appraisal checklist for quantitative studies reporting correlations and associations*	101
Chapter 4 Paper 2		
Chapter 4, Table P1a	*Characteristics of included studies which looked at miscarriage as the outcome of interest*	121
Chapter 4, Table P1b	*Characteristics of included studies which looked at stillbirth as the outcome of interest*	122
Chapter 4, Table P1c	*Characteristics of included studies which looked at all types of pregnancy loss*	125
Chapter 4, Table S1	*Search Strategy*	141
Chapter 4, Table S2	*NICE Quality appraisal checklist for quantitative studies reporting correlations and associations*	142

Chapter 4, Table S3	*Quality Assessment Results in Full*	143
Chapter 5		
Chapter 5, Table P1a	*Distribution of Outcomes and Covariates among the exposed women and the matched contemporary controls.*	156
Chapter 5, Table P1b	*Distribution of Outcomes and Covariates among the exposed women and the matched historical controls.*	157
Chapter 5, Table P2	*Distribution of pregnancy outcomes*	159
Chapter 5, Table P3	*Hazard ratios by type of pregnancy loss for exposed vs contemporary and historical controls*	159
Chapter 5, Table P4	*Hazard ratios for all pregnancy loss by pregnancy trimester for exposed vs contemporary and historical controls*	160
Chapter 5, Table P5	*Descriptive statistics of the number GP surgery consultations during pregnancy for women in each cohort.*	160
Chapter 5, Table S1	*COVID-19 ICD 10 codes used in HES APC*	166
Chapter 5, Table S2	*COVID-19 Codes used in CPRD Aurum*	166
Chapter 5, Table 1	*The distribution of pregnancy outcomes before and after the inclusion of additional pregnancy outcomes from HES (before matching).*	176
Chapter 5, Table 2	*The distribution of pregnancy outcomes after matching*	176
Chapter 5, Table 3	*Crude and adjusted hazard ratios for pregnancies in the shortened cohort before HES outcomes were included*	177
Chapter 5, Table 4	*Crude and adjusted hazard ratios for pregnancies in the shortened cohort after additional HES outcomes were included and exposure status and follow-up time were adjusted*	177
Chapter 5, Table 5	*Conflicting pregnancy episode pairs which have evidence of scenarios indicating a pregnancy has been split into two episodes.*	181

List of Figures

Chapter 2		
Chapter 2, Figure 1	*Pregnancy Register Algorithm*	27
Chapter 3		
Chapter 3, Figure S1	*Example of how a pregnancy may appear in the Register vs GOLD data vs reality*	52
Chapter 4, Paper 1		
Chapter 4, Figure P1	*Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of study selection process*	96
Chapter 4, Paper 2		
Chapter 4, Figure P1	*Systematic Review PRISMA flow chart of included studies*	117
Chapter 4, Figure P2	*Summary Data Quality scores and reported Risk/Odds ratios for studies which included miscarriage as an outcome*	126
Chapter 4, Figure P3	*Summary Data Quality scores and reported Risk/Odds ratios for studies which included stillbirth as an outcome*	127
Chapter 4, Figure P4	*Meta Analysis of the risk of stillbirth among women with COVID-19 at point of delivery*	128
Chapter 4, Figure S1	*Funnel plot of studies included in Meta Analysis of the risk of stillbirth among women with COVID-19 at point of delivery*	144
Chapter 4, Figure S2	*Comparison of adjusted vs crude ratios for studies which presented both*	145
Chapter 5		
Chapter 5, Figure S1	*Cohort Selection*	173

Abbreviations

BMI	Body Mass Index
CPRD	Clinical Practice Research Datalink
DM+D	Dictionary of Medicines and Devices
EDC	Estimated date of conception
EDD	Estimated date of delivery
EHR	Electronic Health Record
FRD	First Registration Date
GOLD	GP OnLine Database
GPRD	General Practice Research Database
HES	Hospital Episode Statistics
HES A&E	Hospital Episode Statistics Accident and Emergency
HES APC	Hospital Episode Statistics Admitted Patient Care
HES DID	Hospital Episode Statistics Diagnostic Imaging Data
HES OP	Hospital Episode Statistics Outpatient
HR	Hazard Ratio
ICD	International Classification of Disease
IMD	Index of Multiple Deprivation
ISAC	Independent Scientific Advisory Committee
LEO	London School of Hygiene and Tropical Medicine Ethics Online
LSHTM	London School of Hygiene and Tropical Medicine
MBL	Mother Baby Link
MHRA	Medicines and Healthcare Products Regulatory Agency
NIHR	National Institute for Health Research
NHS	National Health Service
ONS	Office for National Statistics
OPCS	Office of Population Census and Surveys
RDG	Research Data Governance
SES	Socio-economic Status
SGSS	Second Generation Surveillance System
TOP	Termination of pregnancy
UK	United Kingdom
UKHSA	UK Health Security Agency
UTS	Up-to- standard
WAPM	World Association of Perinatal Medicine
WHO	World Health Organisation

Chapter 1: Introduction

1.1 Rationale for this Research

The use of longitudinal electronic health records data (EHR), such as the Clinical Practice Research Datalink (CPRD) primary care data sources, is an expanding area of public health research. Such data allow for the study of exposures to drugs and diseases during pregnancy and the outcomes for mother and child. This is particularly useful given that pregnant women are generally excluded from clinical trials. However, as these data are not collected primarily for research, they can sometimes be sub-optimal and accurately estimating the timing and outcome of pregnancy episodes is challenging (sees section 1.3.4). There have been several attempts to develop methodology to assist researchers with identifying pregnancies in CPRD. However, these methods have typically ignored incomplete pregnancy episodes despite evidence the woman may have been pregnant at that time point. A Pregnancy Register was recently developed from a collaboration between the London School of Hygiene and Tropical Medicine (LSHTM) and CPRD which attempts to capture all (including incomplete) pregnancy episodes in CPRD. The Pregnancy Register differs from previous approaches by including pregnancies with no outcomes and pregnancies that appear to overlap with other pregnancies in the database. However, further work was needed to investigate and characterise these pregnancies and to attempt to understand the implications of excluding or including them in public health research studies. This is the focus of the first main study of my thesis, and the motivation for this work was to provide useful insights for researchers conducting pregnancy research using EHR data.

Coronavirus disease (COVID-19) emerged during the course of this PhD and is a new exposure in pregnancy for which the implications on pregnancy outcomes are not fully understood. This presented an opportunity to conduct a proof-of-concept study and apply the methodological recommendations developed from this work to a question of public health importance. This is presented in the later chapters of my thesis; the work goes beyond what other researchers have attempted to do by aiming to investigate and utilise uncertain pregnancy data and will provide a valuable tool for researchers wishing to study pregnancy in EHR data.

1.2 Thesis Structure

This is a research paper style thesis comprising chapters formatted in the style of a journal article and more traditional thesis style chapters.

Chapter 1 introduces the motivation, aims and objectives of the research as well as contextual background information on themes throughout the thesis, beginning with an overview of pregnancy research using EHR data. This overview includes an outline of why EHR data are a useful tool for pregnancy research, the challenges, and limitations of using it for research and previous approaches to utilising pregnancy EHR data. I then present an introduction to COVID-19 and what is known about its impact on pregnancy. This is followed by a description of each of the EHR data sources which were utilised in this work in Chapter 2.

Chapter 3 introduces the methodological work carried out to investigate why uncertain pregnancy episodes exist in the CPRD Pregnancy Registers. This work is based on the original CPRD GOLD Pregnancy Register and was conducted as two separate studies: one which looked at why pregnancy episodes with no recorded outcome might occur and another which looked at why pregnancy episodes which conflict with another episode for the same woman might occur. This work was combined as one research paper and published in *the BMJ Open* in February 2022. (1) In the paper I describe a series of detailed scenarios as to why uncertain pregnancy episodes may occur along with criteria which researchers can apply to ascertain which episodes may fit each scenario in order to maximise their usefulness for research.

In Chapter 4 I introduce the public health research question: Is having COVID-19 in pregnancy associated with an increased risk of pregnancy loss (either miscarriage or stillbirth)? This chapter outlines a systematic literature review of this question and consists of two research papers. The first paper was published in *the BMJ Open* in October 2022 and outlines in detail the protocol and methodology for the review (2). The second paper presents the findings of the systematic review including a subgroup meta-analysis looking at whether having COVID-19 at point of delivery increase the risk of stillbirth. This paper is currently under review by *the BMJ Open.*

Chapter 5 describes a matched cohort study which was conducted to examine the previously outlined question: Does having COVID-19 during pregnancy increase the risk of pregnancy loss? This cohort study utilises the CPRD Aurum Pregnancy Register and the findings are presented as a paper intended for publication. I then present some exploratory analysis examining the impact of excluding uncertain pregnancy episodes on the results of the study.

Finally in Chapter 6 I discuss the findings of the PhD and how they fit in the context of previous work. The strengths and limitations of the presented approach are also discussed along with implications for policy and future research.

1.3 Background

1.3.1 Electronic Health data in the UK

The United Kingdom (UK) has a publicly funded National Health Service (NHS) which is free at the point of care. The NHS works on a gate-keeper system for non-emergency care where patients typically see a General Practitioner (GP) as the first point of contact to access the healthcare system. Patients are then referred to secondary care services by the GP when required. Over 98% of people in England are registered at a GP practice (3). GPs use specialist software to record events pertaining to the care of their patients. Each patient registered at an NHS provider is given a unique patient identifier known as an NHS number. NHS numbers allow events recorded by different healthcare providers including secondary care settings to be linked. This healthcare structure makes the UK a rich source of longitudinal electronic health data. A number of organisations provide access to anonymised UK electronic health care data for research of public health benefit.

1.3.2 Pregnancy Research

The study of drug exposure and co-morbidities in pregnancy is an important and challenging area of research of increasing public interest. Safety of medications during pregnancy is difficult to study in traditional randomised clinical trial (RCT) settings. Pregnant women are excluded from many trials due to potential risks to the woman and her unborn child thus creating a knowledge gap. Therefore, while RCTs have traditionally been considered the gold-standard for evidence-based medicine, the generalisability of outcomes to excluded populations, including pregnant women, is questioned.

In the real-life setting pregnant women are exposed to various drugs and medical interventions, including inadvertent exposure in the first trimester when the woman may not realise, she is pregnant. Women with chronic illnesses (e.g., epilepsy, diabetes) face difficult decisions about whether to continue, switch, reduce or stop medication whilst pregnant. Furthermore, disease exposures can carry risks for both pregnant women and their unborn babies. It is important that these risks are fully understood so that vaccination campaigns can be targeted correctly. Observational research using EHR data has thus become a well-established vital tool for research into disease prevalence, risk factors and post-market pharmacovigilance.

1.3.3 The importance of EHR data in pregnancy research

Large EHR datasets contain a wealth of information on pregnancies, exposures, and outcomes for both mother and child. In the UK a woman's GP is the main point of contact for antenatal care; therefore, most pregnancy events are captured in primary care data (4). The use of such datasets for safety monitoring following drug or vaccine exposure in pregnancy is increasing.

A key example of important post market pregnancy research has been the safety of anti-epileptic drugs in pregnancy. Numerous studies have shown that the anti-epileptic medication sodium valproate increases the risk of congenital malformations and developmental delays (5). However, for some patients it remains the most effective drug to manage their epilepsy. Furthermore, there remain concerns as to whether other anti-epileptics could be associated with developmental delays (6). EHR data have been utilised extensively to assess the safety of anti-epileptics and the effectiveness of risk minimisation programmes such as pregnancy prevention strategies for women who are prescribed sodium valproate (7). For this type of research, it is vital that all evidence of pregnancy is identified from the data regardless of completeness in order to gain an accurate picture of whether the pregnancy prevention strategy is effective.

As another example, in recent years vaccination campaigns for pregnant women have become a key public health strategy for illnesses such as influenza, pertussis, and COVID-19 (8). These vaccines sometimes need to be rolled out due to a pressing need to protect pregnant women where it is known that the disease itself can cause harm to them and their infants. However, lack of trial data can lead to understandably high levels of vaccine hesitancy in pregnant populations (9). Post-license monitoring of vaccine safety in pregnancy is therefore essential, to continue to assess benefits and risks and to offer reassurance to pregnant women. In this case it is critically important that the timings of pregnancies included in studies are correct in order to ascertain that the vaccine was indeed given whilst the woman was pregnant and avoid exposure misclassification but also to establish in which trimester she was vaccinated as exposures may have different effects depending on the stage of pregnancy. Furthermore, EHR data can be useful in measuring vaccine uptake in pregnancy to monitor the impact of public health campaigns. For this type of research it is again important to be able to detect all pregnancies to get a true picture of the campaigns impact.

Electronic health records are also useful for assessing the risk of disease exposures during pregnancy. A recent example of this by Minassian et al utilised EHR data to assess the risk of pre-eclampsia associated with acute maternal infections during pregnancy (10) . For this study the timing of pregnancy was key to ensure that exposures early in pregnancy were not missed.

1.3.4 The challenges of using EHR data to study pregnancy

Whilst EHR data are extremely useful for pregnancy research incomplete data capture can make it difficult to identify accurately the start and end of pregnancies, and in turn pinpoint exposures timings in relation to gestational age. In 2015 Margulis et al published a review of methods used to identify pregnancies in electronic health data (11). These methods can be grouped into three broad categories. The first of these, simple imputation, involves identifying the end of a pregnancy and applying a default duration to allocate a start date. An example of this method in UK data was developed by Devine et al using data from the General Practice Research Database primary care data (GPRD now known as CPRD) (12). The advantage of this method is that it is quick and easy to implement however, setting a standard duration for all pregnancies results in an overestimation of pregnancy length for pre-term pregnancies or those which end in miscarriage or termination. This may in turn result in exposure status misclassification for early pregnancy exposures. This is particularly problematic given that exposure in early pregnancy is of particular importance, being the time of organogenesis, and incurs the highest risk of congenital malformations(13)

The second approach involves mapping timings of pregnancy events in the data: identifying records of gestational age or timed antenatal events and using these to map timepoints throughout the pregnancy to estimate a start date. An example of this method was developed by Hardy et al using GPRD data(14). Whilst this slightly more tailored approach potentially offers more accurate timings it ignores all incomplete pregnancy data and any pregnancy without a recorded outcome in the data. Thirdly some studies have attempted a combination of methods utilising both imputation and mapping methods (15,16). Furthermore, these approaches sometimes fail to differentiate between spontaneous pregnancy loss and induced abortion which limits their usefulness for studies in which these pregnancy outcomes are key (15).

The main limitation of all previous approaches has been the exclusion of pregnancies without identified outcomes and pregnancy records which do not fit chronologically into an identified pregnancy. Ignoring these records potentially excludes periods when women may have been pregnant. If these pregnancies systematically differ from those captured more completely, their exclusion may lead to biased effect estimates in studies. For example, pregnancies ending in miscarriage may be less likely to have the outcome recorded in the data than pregnancies ending in live birth. Ignoring pregnancy data which are challenging to interpret may therefore underestimate adverse outcomes. Incomplete capture of pregnancies also impacts descriptive studies that need pregnancies as denominator data. Furthermore, the existence of seemingly overlapping pregnancy

episodes in the data highlight that estimated timings of some pregnancies may be suboptimal and/or some pregnancy episodes may not be true pregnancies. Such episodes, if unresolved, may also lead to misclassification of exposure timings.

1.3.5 COVID-19

1.3.5.1 The impact of the COVID-19 pandemic in the United Kingdom

COVID-19, a respiratory illness caused by the coronavirus SARS-CoV2 is a new disease. The first cases were detected in China in December 2019 and on the 11th of March 2020 the World Health Organisation (WHO) characterised the outbreak as a global pandemic (17). To date there have been >640 million confirmed COVID-19 cases worldwide and >6.5 million confirmed deaths (18). The UK has reported >24 million cases and ~200,00 deaths (19). Initial public health measures attempting to control the virus began in the UK in March 2020 and included national lockdowns and restrictions on people's movement and daily lives. These limitations laid out by the UK government fluctuated in severity over time dependent on the level of virus currently circulating in the population(20). National restrictions were in place at some level until February 2022 when all legal restrictions related to COVID-19 were removed in the UK.

In December 2020 the first vaccine for COVID-19 was approved for supply by the UK's Medicines and Healthcare Products Regulatory Agency (MHRA) (21). Vaccines were rolled out rapidly across the population in a stepwise fashion based on risk with age being the main deciding risk factor (22) From March 2021 pregnant women in the UK began receiving the COVID-19 vaccination in line with their age group (22). However, a press release by the UK Health Security Agency (UKHSA) published in May 2022 stated that uptake of COVID-19 vaccination among pregnant women was only 60% (23).

1.3.5.2 COVID-19 in Pregnancy

Pregnant women are more vulnerable to respiratory illnesses due to physiological changes in their immune and respiratory systems. It has been established that influenza infection can be dangerous for pregnant women in terms of increased morbidity and mortality (24). A systematic review by Mosby et al found that during the 2009 influenza pandemic pregnant women had a higher risk of hospitalisation, intensive care admission and death compared to non-pregnant women (25). Early

evidence has also suggested that infection with coronaviruses SARS-CoV and MERS-CoV puts pregnant women at an increased risk of needing intensive care (26).

Since the start of the COVID-19 pandemic it has been proposed that pregnant women infected with SARS-CoV2 also have a high risk of serious illness (27). Multiple studies have demonstrated an increased risk of morbidity and mortality amongst pregnant women compared to their non-pregnant counterparts. A review of studies looking at adverse outcomes in pregnant women by Pathirana et al concluded that several adverse maternal outcomes were significantly higher among pregnant women with COVID-19 than non-infected pregnant women (28). A UK national surveillance study by Knight et al showed that 1 in 10 pregnant women admitted to hospital with SARS-CoV-2 infection needed respiratory support and one in 100 died (29) Furthermore, a multinational retrospective cohort study conducted by the World Association of Perinatal Medicine (WAPM) Working Group on COVID-19 concluded that SARS-CoV-2 infection in pregnant women is associated with a 0.8% rate of maternal mortality, and an 11.1% rate of admission to the ICU (30).

Maternal viral infections can also be harmful to unborn fetuses (31), and it has been hypothesised that SARS-CoV2 infection may be associated with an increased risk of miscarriage or stillbirth potentially mediated by placental damage (27,32). Rapid research has been conducted as the world attempts to understand this new virus better however, due to the urgency, many studies have been hospital focused with small sample size and ways of determining COVID-19 exposure have varied. An early systematic review by Kazemi et al concluded an increased risk of pregnancy loss among mothers with a positive test for COVID-19 however, this was based purely on case reports and case series analyses which, without a robust comparison group, have limited value (33). Further systematic reviews of these types of studies have reached similar conclusions (34,35).

Epidemiological studies to date have drawn mixed conclusions around COVID-19 and risk of stillbirth. A large cohort study using English hospital data conducted by Gurol-Urganci et al concluded an increased risk of fetal death when mothers were infected with SARS-CoV2 during delivery (36). A USA based cohort study using hospital based administrative data for over 1 million women reached similar conclusions however, they stated limitations around the study including the potential for misclassification due to missing data and difficulties around ascertaining the timing of COVID-19 records (37). A nationwide cohort study conducted in Sweden assessed stillbirth as an outcome and SARS-CoV2 exposure across the whole pregnancy using registry data (38). They found no evidence of an increased risk of stillbirth when universal COVID-19 testing was conducted and suggested that perhaps studies where testing was not universal were only capturing the more severe cases in their exposed population.

Studies assessing risk of miscarriage have been less frequent presumably due to the difficulties around accurately measuring an outcome that is notoriously underreported. Those which have looked at miscarriage have also reached differing conclusions. Cavalcante et al conducted a review of the existing evidence in 2021 and concluded an urgent need for robust studies examining whether SARS-CoV-2 infection increases the risk of miscarriage (39). Often studies have looked only at COVID-19 status at the point at which the miscarriage occurred. One such study by Cardona-Pérez, of women presenting at a hospital in Mexico City, concluded no increased risk of pregnancy loss (40). However, their sample size was small, and they excluded women with severe COVID-19 symptoms who were sent to a different hospital for care.

A UK based prospective cohort study conducted by Balachandran et al in 2020 found an increased risk of miscarriage was associated with early pregnancy exposure to COVID-19 (41). However, they relied on self-reporting of both exposure status and outcome. The number of women in the unexposed group was relatively low and therefore the study was potentially underpowered. Most recently a study using Scottish EHR data found no evidence of an increased risk of miscarriage among women who had COVID-19 in early pregnancy. Whilst this study benefited from a large sample size and robust methodology there were concerns about the ability to estimate gestation at exposure for pregnancies ending in loss and they were unable to adjust for key potential confounders (smoking and body mass index) due to the level of missing information (42). Chapter 4 of this thesis describes a systematic review of epidemiological studies to date which have examined the association between COVID-19 and the risk of miscarriage and stillbirth.

1.4 Aims

The first aim of this thesis was to conduct a detailed investigation into uncertain pregnancy episodes in the CPRD Pregnancy Register and generate appropriate recommendations for handling these data in research. The second linked aim was to utilise the CPRD Pregnancy Register to answer an important and topical clinical question about whether COVID-19 infection during pregnancy is associated with pregnancy loss, and as part of this study, to implement of my recommendations on the handling of uncertain pregnancy episodes.

1.5 Objectives

The aims will be addressed by the following objectives:

1. Investigate possible reasons why pregnancy episodes without a recorded outcome occur in the CPRD Pregnancy Register and whether linked data can help understand these episodes.
2. Investigate possible reasons why some pregnancy episodes in the CPRD Pregnancy Register are overlapping and whether linked data can help understand these episodes.
3. Systematically review all of the evidence to date as to whether COVID-19 during pregnancy increases the risk of pregnancy loss (miscarriage and stillbirth).
4. Evaluate the potential impact of having COVID-19 during pregnancy on the risk of miscarriage or stillbirth utilising the CPRD Pregnancy Register including the impact of changes in the definition of the pregnant study population.

1.6 Chapter Summary

Introduction

- Whilst pregnant woman are often excluded from clinical trials in the real-world setting, they are often exposed to medication and vaccines.

- UK electronic health data provides a rich source of information on pregnancies and their outcomes as well as disease and drug exposures which may have occurred. However, completeness in these data sources can sometimes be sub-optimal making it difficult to accurately determine pregnancy timings and outcomes.

- Researchers have utilised differing methodologies to try to maximise EHR data as a tool for pregnancy research. However, methods tend to ignore uncertain pregnancy data which may lead to errors such as exposure misclassification or underestimation of vaccines uptake.

- COVID-19 is a new disease and questions remain as to whether exposure during pregnancy increases the risk of pregnancy loss.

- The first aim of this PhD was to conduct a detailed investigation into uncertain pregnancy episodes in the CPRD Pregnancy Register and generate appropriate recommendations for handling these data in research.

- The second linked aim was to utilise the CPRD Pregnancy Register to answer an important and topical clinical question about whether COVID-19 infection during pregnancy is associated with pregnancy loss, and as part of this study, to implement of my recommendations on the handling of uncertain pregnancy episodes.

Chapter 2: Data Sources

2.1 Clinical Practice Research Datalink (CPRD)

Clinical Practice Research Datalink is a government research service collecting de-identified and fully coded patient-level EHR from GP practices across the UK. CPRD primary care data include >62 million patients with >16 million currently registered(43). Data are representative of the UK population with respect to age, gender, and ethnicity (44,45). Data are currently collected from two GP software systems, Vision which feeds in to the CPRD GOLD database and EMIS which feeds into the CPRD Aurum database.

CPRD data contain registration information and all care events that general practice staff record to support clinical care and management. This includes demographic information (birthyear, sex, etc.), clinical events (signs, symptoms, medical diagnoses), referrals to specialists and secondary care, prescriptions issued in primary care, vaccinations, test results, lifestyle information (e.g. smoking and alcohol status), and other care administered as part of GP practice. These data also include pregnancy related events recorded by the GP.

Linkage of CPRD primary care data with other patient-level datasets is available for English practices who have consented to participate in the linkage scheme. These linkages cover approximately ~44% of contributing CPRD GOLD patients and ~ 93% of CPRD Aurum patients (43).

2.1.1 Data Structure, Coding and Key Variables in CPRD

CPRD data are structured as tables which can be linked to one another using the encrypted patient identifier (patid). Events relating to a medical observations are recorded in the CPRD GOLD Clinical, Referral and Immunisation tables using Read version 2 codes, and in the Observation table in CPRD Aurum using a combination of SNOMED-CT, Read and local EMIS codes. Medication prescriptions are coded using Gemscript product codes in the CPRD GOLD Therapy table and Dictionary of Medicines and Devices (DM+D) codes in the CPRD Aurum Drug Issue table (46,47). Event information can be accessed by developing a list of all relevant codes and applying it to the data. Code-list development is a crucial aspect of any observational study utilising these data as missed or incorrect codes can result in missed events and misclassified patients.

The CPRD databases also contain indicators of data quality at the patient and practice level (*acceptability flag* and *up to standard date*, respectively; Table 1). As CPRD databases are longitudinal, updated monthly, they also contain variables indicating whether the patient and practice are still contributing data. Patients may leave and then re-join a practice later causing gaps in patient follow-up, hence current registration date is often used rather than first registration date when considering start of follow up.

Table 1: Key variables to consider when utilising CPRD data.

Variable	**Description**
Last Collection Date	Last date on which CPRD received data from a practice.
First Registration Date (FRD)	Date a patient first registered at the practice
Current Registration Date	Date a patient last registered at the practice (may be the same as FRD if the patient has never left)
Transferred Out Date	Date a patient left a practice.
Up to Standard Date (UTS)	A practice-based quality marker indicating from when a practice is considered to have continuous and complete recording of patient data.
Acceptability Flag	Indicates whether a patient's data are acceptable for research based on patient-level data quality markers.

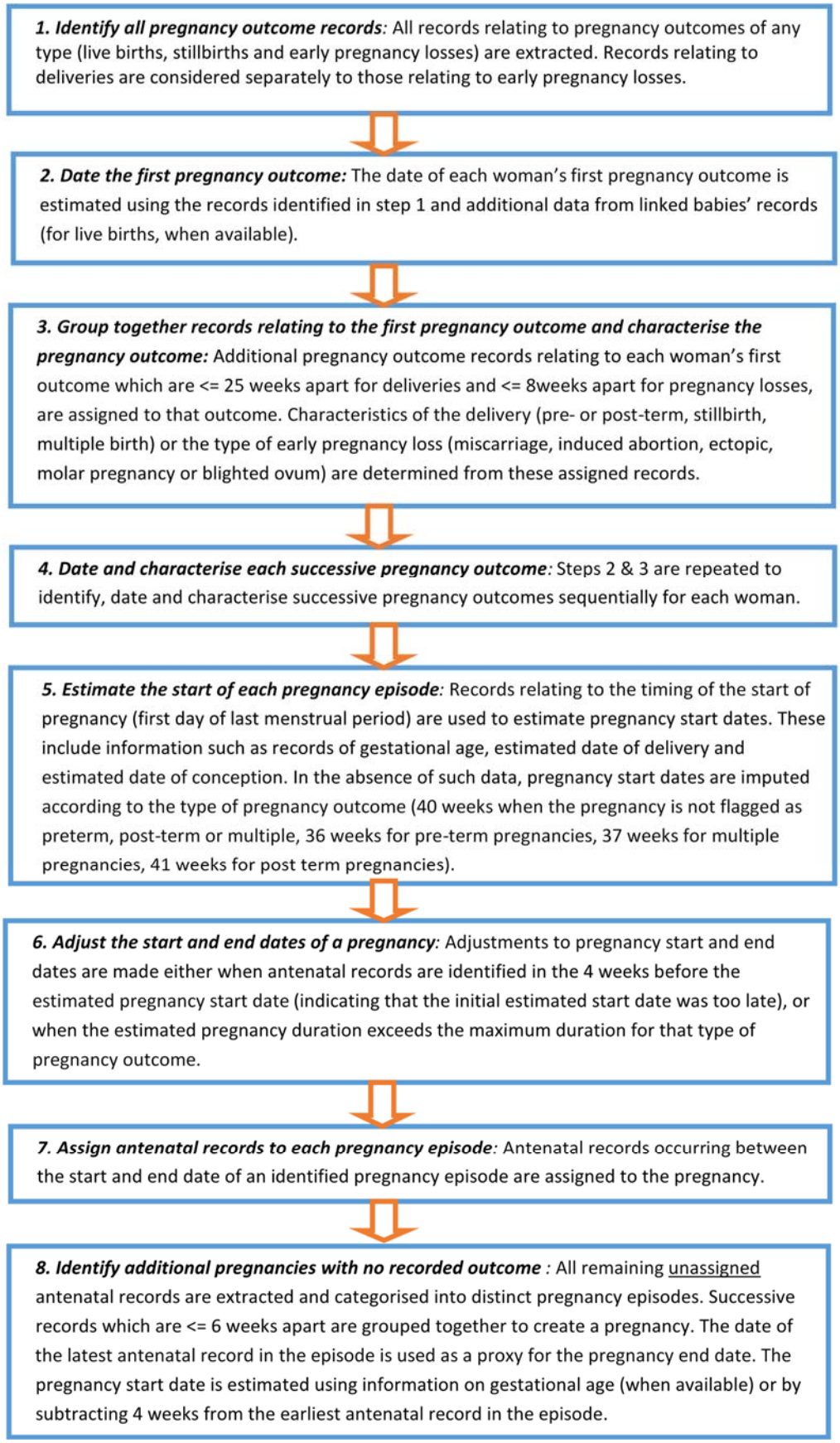
2.2 The CPRD Pregnancy Registers

The CPRD Pregnancy Registers list and characterise all pregnancies identified in CPRD primary care data. A single record represents a unique pregnancy episode. Women may have multiple episodes within the register. Information includes the start and end of pregnancy, its outcome and whether it was a singleton or multi-fetal pregnancy. For live birth pregnancies, patient identifiers of linked babies identified through the CPRD Mother-Baby-Link (MBL)(48) are provided in the CPRD GOLD

register. The Pregnancy Registers allow CPRD data users to study exposures and outcomes of pregnancy more easily.

The CPRD GOLD Pregnancy Register was created in 2016 by a partnership between CPRD and LSHTM (49). In 2022 an equivalent register was created in the CPRD Aurum database (50). Both registers are based on a categorised list of pregnancy-related codes (~4,000 codes for CPRD GOLD and ~ 6,000 codes for CPRD Aurum) which are used to ascertain patients who had a pregnancy recorded in the databases (Pregnancy code lists are provided in the corresponding publications (49,50) Example codes in each category are shown in Appendix 1. All pregnancy data are then extracted from CPRD primary care data, and an eight-step algorithm is applied to the data (Figure 1). The algorithm attempts to ascertain the start and end of each pregnancy episode as well as the trimester dates and what the outcome was. The registers are designed to be sensitive rather than specific and to capture all indications of pregnancy regardless of completeness.

Figure 1: Pregnancy Register Algorithm (figure from (1) *)*



2.3 Linked Data

2.3.1 Hospital Episode Statistics (HES)

CPRD data are linked to HES which include details of admissions to hospitals and outpatient appointments at NHS hospitals in England (51,52). Diagnostic data in HES are coded using ICD10 codes; procedure information is coded using the UK Office of Population, Census and Surveys classification (OPCS).
Linked HES Admitted Patient Care (APC) data includes information on admission and discharge dates, diagnoses, specialists seen and procedures undertaken for linked patients with a hospitalisation record(51). HES APC also includes maternity data for deliveries which took place in NHS hospitals. However, the completeness and quality of these data vary greatly depending on the hospital recording them (53,54).

HES Outpatient (HES OP) data are records of outpatient appointments. Diagnostic information in these data are mainly incomplete. However, information include the type and date of consultation, the main specialty under which the patient was treated, and referral source.

The HES Diagnostic Imaging Dataset (DID) provides detailed information about diagnostic imaging tests carried out from April 2012, including x-rays, MRI scans and fetal growth scans, taken from NHS providers' radiological information systems. However, scan images and test results, including estimated gestation age, are not included. Rather the data just provides an indication that a scan has been carried out.

2.3.2 Office for National Statistics (ONS) Mortality Data

Deaths in the UK are recorded by compulsory registration. Mortality data collected and provided by the ONS include the official date of death, place and causes of death (using ICD codes). ONS mortality data are linked to CPRD primary care data for patients in consenting practices in England.

2.3.3 Second Generation Surveillance System (SGSS)

SGSS is the English national laboratory reporting system used to capture routine laboratory data on infectious diseases and antimicrobial resistance. During the early stages COVID-19 pandemic data on

COVID-19 testing from SGSS was linked to CPRD in order to capture PCR COVID-19 test results for swab tests offered to those in hospital and NHS workers. The CPRD-SGSS linked data contains positive test results only (43).

2.3.4 Index of Multiple Deprivation (IMD)

The English Indices of Deprivation measure relative levels of deprivation based on small neighbourhoods called Lower-layer Super Output Areas (LSOA) (55). Deprivation is measured using a combination of seven domains: Income; Employment; Education Skills and Training; Health Deprivation and Disability; Crime; Barriers to Housing and Services; Living Environment Deprivation. IMD data are linked to CPRD at the patient level based on postcode and can be used as an indicator or socioeconomic status.

2.3.5 Utilisation of Linked Data

Table 2 outlines where each of the linked data sources described in this chapter were utilised in the thesis. Further details on how they were utilised are given in the corresponding chapters.

Table 2: Utilisation of linked data in this thesis

Linked Data Source	**Where it is utilised**
HES APC	• To ascertain additional pregnancy outcomes in the exploration of uncertain pregnancy episodes (Chapter 3 and Chapter 5) • To look for records of COVID-19 in the COVID-19 and pregnancy loss study (Chapter 5)
HES OP	• To ascertain additional pregnancy outcomes in the exploration of uncertain pregnancy episodes (Chapter 3)
HES DID	• To look for records of fetal scans in the exploration of uncertain pregnancy episodes (Chapter 3)
ONS Mortality	• To check for maternal death records which may be missing from CPRD in the exploration of uncertain pregnancy episodes (Chapter 3)
SGSS	• To look for records of COVID-19 in the COVID-19 and pregnancy loss study (Chapter 5)
IMD	• To adjust for socioeconomic status in the study of COVID-19 and pregnancy loss.

2.4 Chapter Summary

Data Sources Utilised

- The work presented in this thesis centres around the use of EHR data provided by the Clinical Practice Research Datalink.
- Data consists of structured tables and coded event information which requires the development and application of carefully considered code lists to access it.
- CPRD primary care data consists of two databases CPRD GOLD and CPRD Aurum both of which have corresponding pregnancy registers available.
- CPRD Pregnancy Registers are generated using an algorithmic approach which aims to capture all pregnancy information regardless of completeness.
- CPRD primary care data are also linked to a number of other data sources which were utilised in this work including secondary care data, lab test information, mortality recording and socioeconomic data.

Chapter 3: Uncertain Pregnancies in the CPRD Pregnancy Registers

3.1 Introduction

As outlined in Chapter 2 the algorithm which generates the CPRD Pregnancy Registers was designed to be sensitive rather than specific (49). All women who have any kind of record indicating pregnancy in the CPRD data are picked up by the algorithm and included in the Register regardless of how complete the recording of their pregnancy is. Initial validation by the creators of the original CPRD GOLD Pregnancy Register against linked Hospital Episodes Statistics Data indicated overall good agreement between the two data sources, suggesting that most pregnancies are well captured in CPRD GOLD (49). However, there are recognised issues with the Register including approximately 950,000 pregnancies (16% of all pregnancies) with no outcome recorded. Furthermore, there are nearly 500,000 (8.5%) pregnancy episodes within the Register which overlap with another episode belonging to the same woman.

Whilst the approach taken to develop the Register ensures that potential pregnancies are not dropped it can leave researchers with questions around how to handle uncertain pregnancy episodes when designing their studies. This chapter outlines my investigations into the potential reasons why uncertain pregnancy episodes exist in the data. This work was carried out as two separate studies: 1. Investigating why pregnancy episodes with no recorded outcome in the data may occur; 2. Investigating why pregnancy episodes which overlap with another episode for the same woman may occur. These two studies were amalgamated into one published paper in order to make the work more accessible for researchers, this is presented in section 3.2. The work outlines a series of proposed scenarios which may result in these uncertain episodes. Criteria to identify evidence of these scenarios were developed and applied to the database and linked secondary data to ascertain how many pregnancy episodes had evidence of each scenario. The paper presents these results along with advice to researchers on the implications of each scenario and how they may wish to handle them. This work was conducted using the CPRD GOLD Pregnancy Register however, the scenarios can also be applied to the CPRD Aurum Register which was created using the same algorithm and many of the scenarios are applicable to other EHR data sources (50)

3.2 Published Paper: Examining Uncertain Pregnancy Episodes in the CPRD Pregnancy Registers

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	71130	**Title**	Mrs
First Name(s)	Jennifer		
Surname/Family Name	Campbell		
Thesis Title	Investigation and Application of a Pregnancy Register Based on Electronic Primary Care Data		
Primary Supervisor	Krishnan Bhaskaran		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open (Open Access please see the following link for copright approval https://www.bmj.com/company/wp-content/uploads/2019/03/Author-Permissions-Policy.pdf)		
When was the work published?	22/02/2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	JC contributed to the initiation, planning and design of the studies included in this paper. JC performed all of the analysis for both studies . JC wrote the manuscript with critical revision from supervisors.

SECTION E

Student Signature	[redacted]
Date	01/03/23

Supervisor Signature	[redacted]
Date	7/3/23

BMJ Open

Investigating the optimal handling of uncertain pregnancy episodes in the CPRD GOLD Pregnancy Register: a methodological study using UK primary care data

Jennifer Campbell ,[1,2] Krishnan Bhaskaran,[2] Sara Thomas,[3] Rachael Williams,[1] Helen I McDonald ,[3] Caroline Minassian[2]

To cite: Campbell J, Bhaskaran K, Thomas S, *et al*. Investigating the optimal handling of uncertain pregnancy episodes in the CPRD GOLD Pregnancy Register: a methodological study using UK primary care data. *BMJ Open* 2022;**12**:e055773. doi:10.1136/bmjopen-2021-055773

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-055773).

Received 16 August 2021
Accepted 02 February 2022

[1]Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK
[2]Faculty of Epidemiology and Population Health, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[3]Faculty of Epidemiology and Population Health, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

Correspondence to
Jennifer Campbell;
jennifer.campbell@mhra.gov.uk

ABSTRACT

Objectives To investigate why episodes of pregnancy identified from electronic health records may be incomplete or conflicting (overlapping), and provide guidance on how to handle them.

Setting Pregnancy Register generated from the Clinical Practice Research Datalink (CPRD) GOLD UK primary care database.

Participants Female patients with at least one pregnancy episode in the Register (01 January 1937–31 December 2017) which had no recorded outcome or conflicted with another episode.

Design We identified multiple scenarios potentially explaining why uncertain episodes occur. Criteria were established and systematically applied to determine whether episodes had evidence of each scenario. Linked Hospital Episode Statistics were used to identify pregnancy events not captured in primary care.

Results Of 5.8 million pregnancy episodes in the Register, 932 604 (16%) had no recorded outcome, and 478 341 (8.5%) conflicted with another episode (251 026 distinct conflicting pairs of episodes among 210 593 women). 826 146 (89%) of the episodes without outcome recorded in primary care and 215 577 (86%) of the conflicting pairs were consistent with one or more of our proposed scenarios. For 689 737 (74%) episodes with recorded outcome missing and 215 544 (86%) of the conflicting pairs (at least one episode), supportive evidence (eg, antenatal records, linked hospital records) suggested they were true and current pregnancies. Furthermore, 516 818 (55 %) and 160 936 (64%), respectively, were during research quality follow-up time. For a sizeable proportion of uncertain episode, there is evidence to suggest that historical outcomes being recorded by the general practitioner during an ongoing pregnancy may offer explanation (73 208 (29.2%) and 349 874 (37.5%)).

Conclusions This work provides insight to users of the CPRD Pregnancy Register on why uncertain pregnancy episodes exist and indicates that most of these episodes are likely to be real pregnancies. Guidance is given to help researchers consider whether to include/exclude uncertain pregnancies from their studies, and how to tailor approaches to minimise underestimation and bias.

Strengths and limitations of this study

- ► This work carefully examines the way in which pregnancies are recorded in electronic health data in order to maximise its usefulness for pregnancy research.
- ► Detailed scenarios were developed as to why uncertain pregnancy episodes may occur along with criteria which researchers can apply to ascertain which episodes may fit each scenario.
- ► Clinician advice and clinical guidelines were used to generate assumptions as to why and when clinicians may record information relating to pregnancy; however, these may not be correct in every case.
- ► Electronic health data are not collected for the purposes of research and can be messy for a variety of reasons, some of which may not have been captured in this study.

INTRODUCTION

Understanding how diseases, drugs and other exposures affect pregnant women and their children is an important public health priority. However, pregnant women are excluded from many trials due to potential risks to the woman and her unborn child. Observational research using electronic healthcare records (EHRs) has thus become a well-established vital tool for investigating disease prevalence, risk factors and pharmacovigilance in pregnant women. UK primary care databases are particularly useful due to the gate-keeper healthcare system meaning all antenatal care is overseen by a general practitioner (GP).[1] One example of such a database is CPRD GOLD. This database is produced and maintained by the Clinical Practice Research Datalink (CPRD), a government research service collecting de-identified and fully coded patient-level

EHR from primary care practices across the UK.[2] However, challenges such as incomplete data capture in EHR data can make it difficult to identify accurately the start and end of pregnancies. Recently, a collaboration between CPRD and the London School of Hygiene and Tropical Medicine established a Pregnancy Register of all pregnancies in CPRD GOLD[3] which includes approximately 6 million estimated pregnancies (henceforth, pregnancies in the Register will be referred to as pregnancy episodes).

Previous approaches to generating pregnancy registers have been limited by the exclusion of pregnancies without identified outcomes and pregnancy records which do not fit chronologically into an identified pregnancy episode.[4] Ignoring these records potentially excludes periods when women were pregnant. If these pregnancies systematically differ from those captured more completely, their exclusion may lead to bias. For example, pregnancies ending in miscarriage may be less likely to have the outcome recorded than pregnancies ending in live birth.[3] Ignoring pregnancy data which are challenging to interpret may therefore underestimate adverse outcomes. Incomplete capture of pregnancies also impacts descriptive studies that need pregnancies as denominator data, such as vaccine uptake studies. A further limitation of previous approaches is that some women have pregnancies that seemingly overlap in the data, and these are not addressed. These conflicting pregnancies highlight that estimated timings of some pregnancies may be suboptimal and/or some pregnancy episodes may not be true pregnancies. Approaches which exclude incongruent or incomplete pregnancy data may lead to misclassification of exposure timings.

The unique advantage of the CPRD Pregnancy Register is that it uses all pregnancy data in CPRD GOLD, thereby capturing all documented pregnancies regardless of completeness. However, this also presents interpretational challenges: approximately 950 000 pregnancy episodes (16% of all pregnancy episodes) have no outcome recorded and approximately 500 000 pregnancy episodes conflict with another episode for the same woman (episodes identified by the algorithm with at least 1 day of overlap). These episodes are flagged in the Register enabling researchers to identify them when designing their study. However, there may be multiple reasons for the occurrence of uncertain episodes and therefore absolute rules on whether to include or exclude them from a study may be inappropriate.

We therefore aimed to investigate possible reasons why the algorithm used to generate the CPRD Pregnancy Register identifies uncertain episodes and thus generate information to guide future use of this important resource. Our specific objectives were:

1. To identify potential scenarios which may result in pregnancy episodes without a recorded outcome or those which conflict with another episode for the same woman.
2. To use available data (including linked data) to investigate these potential scenarios and flag pregnancy episodes which are consistent with each one.
3. To provide information to researchers using the Register to help inform their decisions on how to handle these uncertain episodes when designing studies.

METHODS

Data sources

CPRD primary care data and the Pregnancy Register

The CPRD GOLD UK primary care database contains registration information and all care events that general practice staff record to support clinical care. This includes demographic information (birth year, sex, etc), clinical events (signs, symptoms, medical diagnoses), referrals to specialists and secondary care, prescriptions issued in primary care, vaccinations, test results, lifestyle information (eg, smoking status) and other care administered as part of GP practice.[5] CPRD data also contain indicators of data quality at the patient level (known as the acceptability flag; online supplemental appendix 1) and at the practice level (known as the practice up-to-standard (UTS) date; online supplemental appendix 1). As CPRD GOLD is a longitudinal database, updated monthly, it contains variables indicating whether the patient and practice are still contributing data.

The Pregnancy Register lists and characterises all pregnancies identified in CPRD GOLD based on an algorithm.[3] A single record represents a unique pregnancy episode. Each woman may have multiple episodes. Information includes the estimated start and end of pregnancy, its outcome (when recorded) and whether it was a singleton or multiple pregnancy. For live birth pregnancies, patient identifiers of linked babies identified through the CPRD Mother-Baby-Link[6] are provided. Figure 1 gives an overview of the algorithm steps, including how gestational ages were applied, and online supplemental appendix 2 gives a list of the variables provided in the Register. Figure in online supplemental appendix 3 shows an example of how a real pregnancy might manifest in (a) raw CPRD gold data and (b) the processed Pregnancy Register dataset.

Linked data

Person-level linkage of CPRD primary care data with other datasets (eg, Hospital Episode Statistics HES) is available for English practices who have consented to participate in the linkage scheme.[7] These linkages cover approximately ~56% of contributing CPRD GOLD practices in the UK. Where available, we used linked data to look for further information about the pregnancy episodes within the Register. HES APC (Admitted Patient Care) data include information on admission and discharge dates, diagnoses, specialists seen and procedures undertaken for linked patients with a hospitalisation record.[8] We searched HES APC data for records of pregnancy outcomes using International Classification of Diseases (ICD-10) and

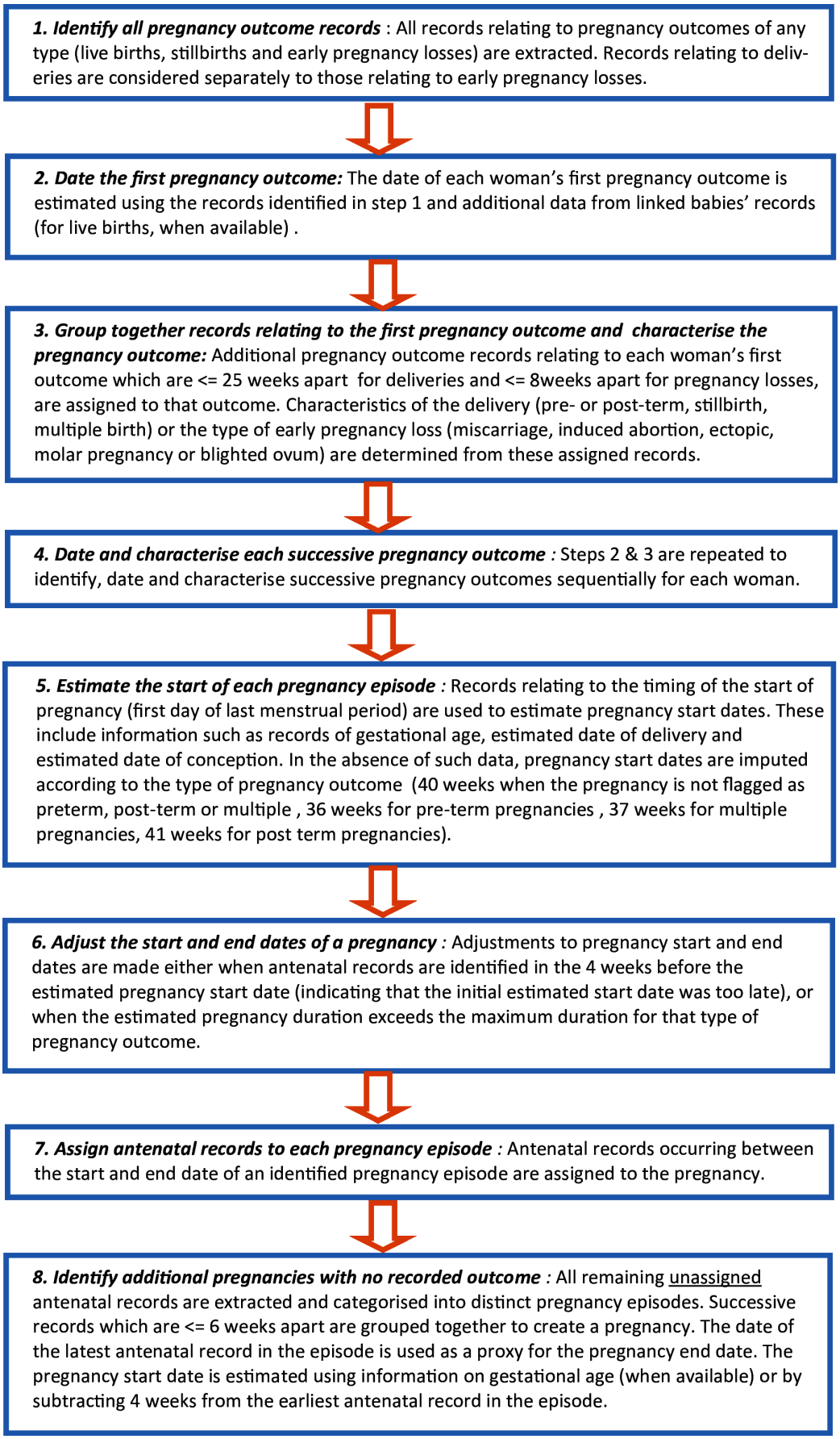


Figure 1 Pregnancy register algorithm steps used to create the CPRD Pregnancy Register. CPRD, Clinical Practice Research Datalink.

Operating Procedure Codes (OPCS) (online supplemental appendices 4 and 5). HES APC maternity records were also used: a recording of an acceptable value in any of the variables identified as relating to delivery (online supplemental appendix 6) was taken as evidence that a delivery had taken place.

The HES Diagnostic Imaging Dataset (DID) provides detailed information about diagnostic imaging tests, including X-rays, MRI scans and fetal growth scans, taken from National Health Service (NHS) providers' radiological information systems. This was used for records of fetal scans. Office for National Statistics (ONS) mortality data were also used to ascertain additional death records which may have been missing from CPRD.

We used set 17 of the CPRD linked data for which the coverage periods were: HES APC 01 April 1997–31 July 2017; HES DID 01 April 2012–31 July 2017; ONS Mortality Data 02 January 1998–19 September 2017.

Study population

This study included all individuals who had at least one pregnancy episode without a recorded outcome or at least one conflicting pregnancy episode in the February 2018 version of the Pregnancy Register. All pregnancy records for these patients were extracted from the CPRD GOLD database using the pregnancy code-list upon which the pregnancy algorithm is based,[3] thereby creating a dataset which included all pregnancy records and the summary Pregnancy Register information for these women. Women were followed up until the minimum of leaving the practice, death or practice last collection date. In the linked data analysis, women with HES records beyond this point were followed up until the end of linked data coverage.

Identifying scenarios to explain the occurrence of uncertain episodes

Potential scenarios which may result in uncertain pregnancy episodes, including those without recorded outcomes and those which conflicted with another episode, were identified through discussions with the creators of the Register (CM, ST, RW), clinicians and CPRD data experts. The scenarios are based on the structure of the CPRD GOLD data and the Pregnancy Register algorithm (figure 1, steps 1–8). The scenarios are not mutually exclusive; thus, episodes may be consistent with more than one scenario.

Pregnancy episodes with recorded outcome missing

Scenarios with the potential to result in episodes with missing outcomes were identified. There are four overarching problems with various specific scenarios within them: the pregnancies are true and current, but the outcome was not captured in CPRD primary care data; the pregnancies are true and current, but the pregnancy was still ongoing at the end of follow-up in the database; the patient was not pregnant at the time of the database record; the pregnancy is really part of another pregnancy episode in the Register. The 12 scenarios which fall under these problems are described in table 1.

Conflicting pregnancy episodes

Scenarios with the potential to result in conflicting episodes were proposed and are described in detail in table 2. Identifying the scenarios was an iterative process, after applying initial scenarios we took a sample of 50 conflicting pregnancy episodes and reviewed the patient data. This allowed us to validate existing scenarios and identify further scenarios. Scenarios can be grouped under four overarching problems: both pregnancies are true but one is a historical pregnancy; both pregnancies are historical; both pregnancies are true and current but the gestation of the second pregnancy estimated by the algorithm is too long; the woman was pregnant, but one pregnancy has been split into multiple episodes by the rules of the algorithm (online supplemental appendix 3).

Applying criteria to identify evidence of each scenario

Evidence in HES

For each episode, it was ascertained whether the woman was eligible for linkage to other data and whether the episode occurred within the coverage period of each linked data source. For pregnancy episodes occurring within the linkage coverage period, the linked HES data were examined for evidence of pregnancy outcomes. The period for which outcomes were searched was from the episode start date to 9 months after the episode end date; we excluded from this analysis pregnancies where this period was entirely outside the coverage dates for linked HES data.

ICD-10 and OPCS code lists were used to look for evidence of outcomes in the HES APC Episodes, Diagnosis and Procedures tables (online supplemental appendices 4 and 5). In the HES APC maternity data, a recording of an acceptable value in any of the variables identified as relating to delivery (online supplemental appendix 6) was flagged as evidence that a delivery had taken place. In the HES outpatient data, an ICD-10 code list for evidence of delivery, termination or early pregnancy loss was used. Snomed codes (online supplemental appendix 14) were used to identify all fetal scan records in the HES DID data.

Pregnancy episodes with recorded outcome missing

All episodes coded as outcome unknown ('13' in the outcome field) were extracted from the Pregnancy Register. For each episode, we extracted information on the timing of the episode in relation to the start and end of patient follow-up and the period of research standard (UTS) data recording in CPRD, and we also searched for relevant codes in the patient's record, namely: early pregnancy codes which were likely to be recorded in the patient's first antenatal visits to the GP; codes which are likely to be recorded by the GP as clinically important in the patient's medical history even when the patient was not pregnant; codes which may indicate an outcome but were originally classified by the Register as antenatal;

Table 1 Description of potential scenarios leading to pregnancy episodes with no recorded outcome and scenario criteria applied

Scenario	How does this appear in the data?	Criteria used to determine if there is evidence in the data that an episode is consistent with the scenario in question
Problem 1: The women was pregnant at the time of the database record, but the outcome *was not captured in CPRD primary care data.*		
1a. The woman was pregnant. She had a delivery, miscarriage or termination of pregnancy (TOP) in hospital or elsewhere and information either was not fed back to the general practice, or was fed back but not coded in the woman's records.	There will be no evidence of an outcome in CPRD data up to 38 weeks* (for delivery) or up to 20 weeks (for miscarriage or TOP) after the first antenatal record for the pregnancy. However, there may be evidence of delivery/miscarriage/TOP in one of the linked HES APC data.	▶ The woman must be eligible for linkage. ▶ There must be at least 1 day of overlap between the data coverage for each HES source and the pregstart+294 days (42 weeks) to give a maximum potential end date. ▶ There must be a record in HES of delivery or loss within 294 days (42 weeks).
1b. The pregnancy outcome was recorded in the primary care data but has no event date recorded alongside it and is therefore not picked up by the algorithm.	There will be an outcome code with missing eventdate† within 38 weeks after the first antenatal record of the pregnancy episode (using the systemdate† as a proxy for the event date).	▶ There must be an antenatal code with missing eventdate† recorded with a systemdate† ≥294 days after pregnancy episode starts.
1c. The pregnancy outcome occurred before the patient was registered at their current practice or before the start of the practice up-to-standard follow-up (UTS). When the patient joined the practice, information was recorded about the pregnancy but not the outcome.	The pregnancy episode will occur before the start of the patient's current registration and/or UTS.	▶ Pregnancy episode end date must be <UTS date† OR ≤current registration date.
Problem 2: The women was pregnant at the time of the database record*, but the pregnancy was still ongoing at the end of available follow-up in the database.*		
2a. The woman moved practices before the end of her pregnancy. If a patient transfers out of a CPRD practice, then follow-up is lost. OR The woman died before the end of her pregnancy.	There will be a transfer out date or death date (in either CPRD or the ONS mortality data) less than 38 weeks after the earliest antenatal record for the pregnancy episode.	▶ The earliest of the woman's transfer out date† or death date (in either CPRD or the ONS mortality data) minus pregnancy episode start date must be ≤294 days.
2b. The last collection of data from the practice was before the pregnancy outcome.	There will be a last collection date less than 42 weeks after the start of the pregnancy episode.	▶ The woman's last collection date minus pregnancy episode start date must be ≤294 days.
Problem 3: *The patient was not pregnant at the time of the database record.*		
3a. A historical pregnancy was recorded retrospectively in the first few months after patient joins the practice. In this scenario, information about the pregnancy is recorded with the current date (by GP software default) rather than the date it occurred (different from scenario 1c). This is more likely to occur when a woman joins a practice and the GP may wish to record past pregnancy events which are relevant to her current clinical care.	The pregnancy episode will occur less than 1 year after the women's current registration date. There will be a record of a pregnancy event which may be clinically useful for future care between the start and end of the pregnancy episode.	▶ Pregnancy episode start date is <365 days after current registration date. ▶ There is a record of a pregnancy code from a list identified as likely to be recorded as useful pregnancy history information (online supplemental appendix 7). ▶ This must have an eventdate ≥pregstart† & ≤pregend.†
3b. The woman was not pregnant but was planning a pregnancy and discussed this with the GP, for example, due to other medical conditions which may complicate pregnancy.	The pregnancy episode will include a pregnancy advice code, for example, '67AF.00 Pregnancy advice for patients with epilepsy'.	▶ The woman has antenatal codes identified as pregnancy advice codes (online supplemental appendix 8) with an eventdate† ≥pregstart† & ≤pregend.†
Problem 4: *The pregnancy record belongs to another pregnancy episode in the Register.*		
4a. There was a delay in recording the outcome of a pregnancy by the practice. Thus, the outcome code has an eventdate† which is later than the true outcome date. The algorithm then calculates the Last Menstrual Period (LMP) date as being later than it was (figure 1, steps 5 and 6). Records which occurred early in pregnancy are then left unassigned to the pregnancy episode and appear as if belonging to a previous pregnancy episode which has no outcome recorded (figure 1, step 8).	As the pregnancy episode without outcome has been created from unassigned records at the beginning of the pregnancy, it will be followed by another pregnancy episode. There is unlikely to be more than a 3-month delay in outcome recording due to the mother attending the practice for postnatal checks and/or infant vaccinations. Therefore, there will be less than 12 weeks between the end of the episode with no recorded outcome and the start of the next pregnancy episode.	▶ The woman must have >1 episode in the Pregnancy Register. ▶ Episodes with recorded outcome missing were eligible if they were not the last pregnancy episode for that woman. ▶ There must be ≤84 days (12 weeks) between the pregend† of the episode without outcome and the pregstart† of the woman's next episode.
4b. The LMP is derived from information in the data and is estimated by the algorithm to have occurred later than reality (figure 1, steps 5). This may lead to a short pregnancy episode and unassigned codes before the estimated start of pregnancy. These are then grouped to form a pregnancy episode with no recorded outcome (figure 1, step 8).	The pregnancy episode without outcome will be followed by another pregnancy episode which will be less than 40 weeks long.	▶ The woman must have >1 episode in the Pregnancy Register. ▶ The episode after the episode with missing outcome must have a startsource†=2, 4, 5 or 6 (online supplemental appendix 2). The length (gestdays) of the episode must be <280 days.

Continued

Table 1 Continued

Scenario	How does this appear in the data?	Criteria used to determine if there is evidence in the data that an episode is consistent with the scenario in question
4c. If there are pregnancy records within 4 weeks before the estimated LMP, the identified pregnancy episode is shifted earlier in time by the algorithm (within plausible limits) to encompass those records (figure 1, step 6). This may leave unassigned pregnancy records which occurred shortly after the new estimated delivery date which will then be grouped to form a pregnancy episode with no recorded outcome (figure 1, step 8).	The pregnancy episode must not be the only pregnancy for this to apply. There will be another pregnancy episode which ends <8 weeks before the first antenatal record of the pregnancy episode without outcome for which the end has been adjusted by the algorithm.	▶ The woman must have >1 episode in the Pregnancy Register. ▶ The episode before the one with recorded outcome missing must have an endadj†=2 (online supplemental appendix 2). ▶ The pregend† date for the episode with missing outcome must be ≤56 days (8 weeks) after the pregend† for that previous episode.
4d. The GP records a code relating to the patient's pregnancy outcome history while the patient is pregnant. This is incorrectly identified by the algorithm as the outcome of the current pregnancy (figure 1, step 3). If the actual outcome is ≤25 weeks after for delivery or ≤12 weeks after for pregnancy losses, they will be grouped together as the same outcome. Subsequent antenatal records may then be grouped together to form a new pregnancy episode with no recorded outcome (figure 1, step 8).	The pregnancy episode must not be the patient's first pregnancy. The pregnancy episode would be within 25 weeks after the previous outcome.	▶ The woman must have >1 episode in the Pregnancy Register. ▶ The pregend† date for the episode with missing outcome had to be ≤175 days (25 weeks) after the pregend† for the previous episode.
4e. The outcome of the pregnancy episode has been misclassified as an antenatal event, for example, 'Failed abortion', 'refer to TOP counselling', 'premature labour', etc.	There will be an antenatal code which should have been an outcome code within 38 weeks after the first antenatal record of the pregnancy episode with recorded outcome missing.	▶ There must be an antenatal record from a code list of potentially misclassified outcomes (online supplemental appendix 9) 266 days (38 weeks) of the firstantenatal† record.

*The first antenatal record is assumed to be recorded ≥4 weeks after the LMP as the woman is unlikely to know she is pregnant before then.
†Refers to a CPRD GOLD-specific variable, for example: pregend=the end of episode as defined by the algorithm; pregstart=the start of episode as defined by the algorithm; endadj=an indication that the end of the episode has been adjusted and how; startsource=which data were used to generate the start of the episode. These variables and others are defined in more detail in online supplemental appendix 2.
APC, Admitted Patient Care; CPRD, Clinical Practice Research Datalink; GP, general practitioner; HES, Hospital Episode Statistics; ONS, Office for National Statistics.

codes which are likely to be recorded by the GP as part of a consultation about the potential health impacts on a patient of becoming pregnant (code lists in online supplemental appendices 7–9).

For each scenario, a set of criteria based on how these should appear in the data were established (described in detail in table 1). Criteria were systematically applied to the data to establish which episodes were consistent with each scenario.

Conflicting pregnancy episodes

All conflicting episodes (those with at least 1 day of overlap with another episode for the same woman) were ascertained using the conflict flag in the Register. Pregnancy episodes may conflict with more than one other episode. Each conflicting pair was treated separately and therefore an individual pregnancy episode could appear in the analysis multiple times. A dataset was created which contained one row per pair of conflicting pregnancy episodes.

Episodes were ordered by start date with episode one being the earlier start date of the two. Descriptive variables were added to the dataset from the CPRD GOLD data to indicate if the episodes were during current registration and UTS follow-up. Pregnancy episode outcomes were grouped into three categories: delivery, loss or missing, and a variable was generated to indicate the combination of outcomes in each conflicting pair (online supplemental appendix 12).

For each scenario, a set of criteria based on how these should appear in the data were established (described in detail in table 2). Criteria were systematically applied to the data to establish which conflicting pairs were consistent with each scenario.

PATIENT AND PUBLIC INVOLVEMENT

There was no patient or public involvement in this methodological work.

RESULTS

There were 2 438 493 women with a pregnancy episode in the February 2018 version of the Pregnancy Register; of these patients, 731 368 (30%) had at least one uncertain episode. Mean patient follow-up time for all women was 4720 days, this was slightly lower for women with a missing outcome record (4349 days) (table 2). Women with an uncertain episode were more likely to be over 30 years of age. Uncertain pregnancy episodes were also more likely to be recent (after 2000) (table 2).

Pregnancy episodes with recorded outcome missing

Of the 5.8 million pregnancy episodes in the Pregnancy Register, there were 932 604 (16%) episodes with no recorded outcome of which over half (516 818, 55.4%) were during UTS follow-up and current registration (table 3). A total of 826 146 (89%) had evidence

Table 2 Description of potential scenarios leading to conflicting episodes and scenario criteria applied

Scenario	How does this appear in the data?	Criteria applied to pairs of conflicting episodes to determine if there is evidence in the data that the pair is consistent with the scenario in question
Problem 1: *Both pregnancies are true, but one is a current pregnancy and one is a historical pregnancy.*		
1a. The GP records a past delivery during a current pregnancy >25 weeks before the true delivery of that pregnancy. OR a past pregnancy loss >12 weeks before the actual loss of that pregnancy.	Both pregnancies will have the same outcome type. Evidence of current pregnancy codes would be expected to fall within the second pregnancy.	▶ The outcome combination of the two episodes must be delivery/delivery or loss/loss (see online supplemental appendix 10 for outcome classifications). ▶ The second episode had an antenatal code from a list deemed likely to only be recorded if the patient was currently pregnant (online supplemental appendix 11) OR a scan record in the HES DID data between firstantenatal* and pregend*.
1b. If a patient has a record relating to a previous loss recorded during a pregnancy ending in delivery or vice-versa, then conflicting episodes will be created by the algorithm. The algorithm first generates episodes for consecutive deliveries; it then does the same thing for pregnancy losses. There is no step in the algorithm to check that the loss episodes do not coincide with the delivery episodes (figure 1, steps 1–6).	The conflicting pregnancies must consist of one loss and one delivery. Evidence of current pregnancy codes would be expected to fall within the second pregnancy.	▶ The outcome combination of the two episodes must be delivery/loss or loss/delivery (see online supplemental appendix 10 for outcome classifications). ▶ The second episode had an antenatal code from a list deemed likely to only be recorded if the patient was currently pregnant (online supplemental appendix 11) OR an antenatal scan record in the HES DID data between firstantenatal* and pregend.*
Problem 2: *Both pregnancies are historical.*		
2a. A patient joins a new practice (or has another reason for a full obstetric history to be taken) and has information on historical pregnancies recorded with the current date rather than the actual date of the event. Losses and deliveries recorded on the same date will result in conflicting episodes in the Register as different outcome types are generated separately by the algorithm (figure 1, steps 1–5).	The conflicting pregnancies must consist of one loss and one delivery. The pregnancy end dates will be the same for both pregnancies. Both pregnancies are likely to be <1 year after the patient's current registration date. We would not expect to find codes indicating current pregnancy.	▶ The outcome combination of the two episodes must be a delivery and a loss. ▶ The pregend* dates must be the same. ▶ There must be no antenatal codes relating to current pregnancy (online supplemental appendix 11) or HES DID antenatal scan recorded between the firstantenatal* date and the pregend* date of either episode.
Problem 3: *Both pregnancies are true and current but the gestation of the second pregnancy estimated by the algorithm is too long.*		
3a. The woman has two pregnancy losses which are >8 weeks and <12 weeks apart. The second pregnancy has no information about gestation recorded so the algorithm applies a default of 12 weeks and the episodes overlap.	Both conflicting pregnancies must be losses. The maximum overlap between the two pregnancies must be 4 weeks. Evidence of current pregnancy codes could be found in either pregnancy.	▶ The outcome combination of the two episodes must be two losses. The pregend* for the first episode must be ≤28 days after the pregstart* of the second episode.
3b. The woman has two pregnancies close together and the second pregnancy ends in delivery. If the information on the Last Menstrual Period date (LMP) in the data of the second pregnancy is wrong, then the algorithm may generate the start too early resulting in an overlap.	The second pregnancy must be a delivery and have no information about gestation in the data. The overlap must be <15 weeks (otherwise the two outcomes would be <25 weeks apart and would have been grouped as one; see figure 1, step 3). There may be evidence of current pregnancy codes in either pregnancy.	▶ The outcome of the second episode must be a delivery. ▶ The startsource* of the second episode must not be equal to 4 or 5 (online supplemental appendix 2). ▶ The pregstart* of the second episode must be <105 days (15 weeks) before the pregend* of the first episode.
Problem 4: *The pregnancy is true and current but is split into separate episodes by the rules of the algorithm.*		
4a. The GP records further information about a pregnancy outcome >25 weeks after the delivery date for pregnancies ending in delivery OR >8 weeks but <12 weeks for pregnancies ending in loss. The algorithm assumes this further information is a different pregnancy and generates a new episode, which may overlap with the 'true' episode.	Both pregnancies must be of the same outcome type. Evidence of current pregnancy codes would be expected to fall within the first pregnancy.	▶ The outcome combination of the two episodes must be delivery/delivery or loss/loss (online supplemental appendix 12). ▶ The first episode had an antenatal code from a list deemed likely to only be recorded if the patient was currently pregnant (online supplemental appendix 11) OR a scan record in the HES DID data between firstantenatal* and pregend*.

Continued

Table 2 Continued

Scenario	How does this appear in the data?	Criteria applied to pairs of conflicting episodes to determine if there is evidence in the data that the pair is consistent with the scenario in question
4b. The GP records further antenatal information about a pregnancy after delivery or pregnancy loss. This will then be used to generate a new pregnancy without outcome episode by the algorithm. If the code is within 4 weeks of the end of the true pregnancy episode, the two will overlap.	The first pregnancy must be a pregnancy with an outcome recorded in the data. The second pregnancy must be a pregnancy without outcome which consists of one antenatal code not related to a scan.	► The first episode must have outcome=1–10 in the Register (online supplemental appendix 2) and must have endadj*=0. ► The second episode must have no recorded outcome (outcome=13). ► The second episode must have a gestdays*=28 (likely to consist of one code) and there must NOT be a scan code (online supplemental appendix 13) with an eventdate*=pregend* of the second episode.
4c. The patient has a follow-up scan after a pregnancy loss. This is recorded in the data by the GP as an antenatal scan. The algorithm then creates a second pregnancy episode based on the antenatal scan code which becomes a pregnancy without outcome in the Register.	The first pregnancy must be a pregnancy loss. The second pregnancy must be a pregnancy without outcome which consists of one antenatal code related to a scan.	► The outcome combination of the two episodes must be loss/missing. ► The second episode must have a gestdays*=28 (likely to consist of one code) and there must be a scan code (online supplemental appendix 13) with an eventdate*=pregend* of the second episode.
4d. The GP records information about a pregnancy but no information about the outcome. If records relating to this pregnancy are more than 6 weeks apart, they will be turned into multiple episodes. Once estimated start dates are generated for these episodes based on the data recorded (figure 1, step 8), episodes may overlap. For example, if there is gestational information included in the second episode, the start of this episode will be assigned before the start of the previous episode resulting in a nested pregnancy episode.	Both pregnancies must be pregnancies without outcome in the Register. The end of the first pregnancy must be greater than 6 weeks before the first antenatal of the second.	► The outcome combination of the two episodes must be missing/missing. ► The pregend* of the first episode is >42 days before the firstantenatal* date of the second episode.
4e. The first pregnancy episode ended in delivery and has been shifted backwards by the rules of the algorithm leaving unassigned late pregnancy or third trimester records. These records will then be identified by the algorithm as end of pregnancies (figure 1, step 6) and new conflicting episodes will be created.	The first pregnancy must be a pregnancy with a delivery outcome recorded in the data. The end of the first pregnancy must have been adjusted. The second pregnancy must be a pregnancy where the outcome is based on a late pregnancy or third trimester record.	► The first episode must have a delivery outcome code and endadj* variable not=0. ► The second episode must have outcome=11, 12 or 13.

*Refers to a CPRD GOLD-specific variable, for example: pregend=the end of episode as defined by the algorithm; pregstart=the start of episode as defined by the algorithm; endadj=an indication that the end of the episode has been adjusted and how; startsource=which data were used to generate the start of the episode. These variables and others are defined in more detail in online supplemental appendix 2.
CPRD, Clinical Practice Research Datalink; DID, Diagnostic Imaging Dataset; GP, general practitioner; HES, Hospital Episode Statistics.

consistent with at least one of the identified scenarios (table 4). On the other hand, 689 737 (74%) had evidence of a scenario indicating they were true (either current or historical) pregnancies (scenarios 1a, 1b, 1c, 2a, 2b or 4e). The largest proportion of pregnancy episodes occurred before the patient registered at their current practice which contributed the data to CPRD or before that practice was deemed to be contributing research standard data (415 807, 44.6% scenario 1c). A total of 211 070 (22.6%) episodes had data in HES consistent with the outcome occurring in hospital and not being fed back to the GP (scenario 1a), representing approximately 50% of episodes with recorded outcome missing which were eligible for linkage. HES APC data were the most useful linked data source for ascertaining pregnancy outcomes with a small number found in HES outpatient (online supplemental appendix 15).

The second most common potential explanation for pregnancies without outcome was scenario 4d, where a code relating to the patient's pregnancy history may have been recorded by the GP while the patient was pregnant. A total of 349 874 (37.5%) episodes without outcome were consistent with this scenario. Relatively fewer episodes were consistent with scenario 4a, 4b and 4e, none were consistent with 4c. For 242 698 (26%) episodes, follow-up ended before the predicted end of the pregnancy (scenario 2a and 2b) for 822 episodes (<0.1%) of these episodes follow-up ended due to death. Only small proportions of episodes were consistent with other scenarios. The distribution of scenarios that occurred during the period left censored by the practice UTS date and patient current registration date was similar to that of the Pregnancy Register as a whole (table 4, online supplemental appendix 16).

Table 3 Baseline characteristics of the pregnancy episodes in the February 2018 Pregnancy Register

	Episodes with recorded outcome missing N (%)	Conflicting episodes N (%)	All episodes in the Pregnancy Register N (%)
Number of patients	*643 689 (26.4)*	*210 593 (8.6)*	2 438 493
Mean patient follow-up time (years)	*11.92*	*12.92*	*12.93*
Mean number of pregnancy episodes per patient	*3.63*	*4.66*	*3.44*
Pregnancy end was during UTS follow-up and current registration	516 818 (55.4)	160 936 (64.1)	1 926 077 (33.1)
Age group of the patient at the end of the pregnancy episode			
11–14	1344 (0.1)	76 (0.0)	7867 (0.1)
15–19	72 543 (7.8)	15 420 (6.1)	551 025 (9.5)
20–24	196 979 (21.1)	48 273 (19.2)	1 397 717 (24.0)
25–29	254 352 (27.3)	65 601 (26.1)	1 624 350 (27.9)
30–34	235 995 (25.3)	69 236 (27.6)	1 339 439 (23.0)
35–39	126 369 (13.6)	40 079 (16.0)	685 421 (11.8)
40–44	37 640 (4.0)	11 355 (4.5)	194 354 (3.3)
45–49	7382 (0.8)	953 (0.4)	24 208 (0.4)
Year pregnancy episode ended			
pre-1950	1417 (0.2)	41 (0.0)	16 695 (0.3)
1950–1959	8061 (0.9)	522 (0.2)	98 436 (1.7)
1960–1969	19 312 (2.1)	1887 (0.8)	283 757 (4.9)
1970–1979	24 296 (2.6)	3882 (1.5)	493 217 (8.5)
1980–1989	38 768 (4.2)	9135 (3.6)	803 380 (13.8)
1990–1999	248 016 (26.6)	54 254 (21.6)	1 530 212 (26.3)
2000–2009	336 523 (36.1)	116 429 (46.4)	1 705 380 (29.3)
2010–2018	256 211 (27.5)	64 843 (25.8)	893 304 (15.3)
Total pregnancies	932 604	251 026	5 824 381

UTS, up-to-standard.

Conflicting pregnancy episodes

There were 478 341 (8.5%) pregnancy episodes with a conflict recorded in the February 2018 Pregnancy Register, amounting to 251 026 conflicting pregnancy pairs. Over half of the pairs (160 936, 64%) were during UTS follow-up and current registration. There were 215 577 (88.6%) pairs which were consistent with at least one identified scenario. Of the remaining 106 458 (11.4%), less than half were during UTS follow-up and current registration (table showing these pregnancies by scenario is given in online supplemental appendix 17). Across all scenarios, at least 40% were during UTS follow-up and current registration. Of the pregnancy pairs, 215 544 (86%) had evidence of a scenario indicating that at least one episode was a true and current pregnancy (scenarios 1a, 1b, 3a, 3b and 4a–e). Most conflicting pairs had at least one pregnancy episode ending in loss (201 783, 80.3%) (online supplemental appendix 18). Furthermore, 41% (101 760) of pairs included at least one pregnancy with no outcome recorded.

A total of 75 672 (30%) of all conflicting pairs were shown to have evidence that they were consistent with problem 1, that a patient had a record relating to the outcome of a previous pregnancy recorded during a current pregnancy. This includes scenario 1b: a record of a previous loss recorded during a pregnancy ending in delivery or vice-versa, one of the most common scenarios (29% of conflicting pairs) (table 5).

A total of 73 191 (29%) of pairs were consistent with scenario 4e: that adjusting of pregnancy dates by the algorithm had led to unassigned records. Of these, over 96% (70 472) were consistent with this scenario only, and 73% (53 464) of these pairs had a linked baby identified. A total of 43 581 (17.4%) of episodes had evidence that they were consistent with further antenatal information having been recorded after the end of pregnancy (scenario 4b).

For approximately 16% (39,373) of conflicting pairs, there was evidence to suggest that the gestation of the second pregnancy episode specified by the algorithm may

Table 4 Numbers of pregnancy episodes with recorded outcome missing which were consistent with applied criteria for each scenario*

Scenario	Description	N of pregnancy episodes with evidence of this scenario (% of total episodes with missing outcome)	N of pregnancy episodes with evidence of this scenario only (% of total episodes with missing outcome)	N of pregnancy episodes with evidence of an outcome in linked HES (% of linkage eligible episodes with recorded outcome missing†)	N of episodes during current registration and UTS follow-up (% of total episodes with missing outcome)*
Denominator		932 604	932 604	424 375†	932 604
Problem 1: The women was pregnant at the time of the database record, but the outcome was not captured in CPRD primary care data.					
Scenario 1a	The pregnancy outcome occurred in hospital or elsewhere and information wasn't fed back to the practice.	211 070 (22.6)	1934 (0.2)	211 070 (49.7)	139 084 (14.9)
Scenario 1b	The outcome of the pregnancy is recorded in the primary care data but has no event date associated with it.	1595 (0.2)	48 (0.0)	523 (0.1)	475 (0.1)
Scenario 1c	The pregnancy occurred before the patient was registered at the practice or before UTS.	415 807 (44.6)	204 176 (21.9)	60 423 (14.2)	0 (0.0)
Problem 2: The women was pregnant at the time of the database record, but the pregnancy was still ongoing at the end of available follow-up in the database.					
Scenario 2a	The patient transferred out or died before the putative end of pregnancy.	177 557 (19.0)	40 191 (4.3)	71 012 (16.7)	117 571 (12.6)
Scenario 2b	The last collection date of the practice was before the putative end of pregnancy.	65 141 (7.0)	22 039 (2.4)	24 091 (5.7)	58 698 (6.3)
Problem 3: The patient was not pregnant at the time of the database record.					
Scenario 3a	Episode is derived from historical pregnancy information recorded in the first few months after the patient joined the practice.	10 235 (1.1)	588 (0.1)	3058 (0.7)	3875 (0.4)
Scenario 3b	Patient asks for advice while planning a pregnancy.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Problem 4: The pregnancy record belongs to another pregnancy episode in the Register.					
Scenario 4a	Delay in recording the outcome of a pregnancy, algorithm calculates the last menstrual period date (LMP) too late and uncovers records at the beginning of pregnancy creating this pregnancy with recorded outcome missing.	61 662 (6.6)	9299 (1.0)	23 099 (5.4)	35 255 (3.8)
Scenario 4b	The LMP is derived from the data and is wrong resulting in early codes being uncovered creating this episode.	29 057 (3.1)	4022 (0.4)	11 304 (2.7)	17 110 (1.8)
Scenario 4c	The LMP has been shifted earlier in time uncovering records at the end of the pregnancy.	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Scenario 4d	A code recorded relating to the patient's delivery history is incorrectly identified by the algorithm as a delivery uncovering records at the end.	349 874 (37.5)	113 688 (12.2)	90 274 (21.3)	219 505 (23.5)
Scenario 4e	The outcome of the pregnancy episode has been misclassified as antenatal	38 848 (4.2)	8000 (0.9)	6611 (1.6)	18 222 (2.0)
None	These pregnancy episodes did not meet the criteria for any identified scenarios.	106 458 (11.4)	–	–	94 769 (10.2)

Continued

Table 4 Continued

Scenario	Description	N of pregnancy episodes with evidence of this scenario (% of total episodes with missing outcome)	N of pregnancy episodes with evidence of this scenario only (% of total episodes with missing outcome)	N of pregnancy episodes with evidence of an outcome in linked HES (% of linkage eligible episodes with recorded outcome missing†)	N of episodes during current registration and UTS follow-up (% of total episodes with missing outcome)*

*A version of this table restricted to episodes which occurred during practice UTS follow-up and patient's current registration is given in the appendices (online supplemental appendix 16).
†Denominator=pregnancy episodes which had at least 1-day overlap with the available HES follow-up period and where the woman was eligible for linkage.
CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; UTS, up-to-standard.

have been too long leading to an overlap (scenario 3a and 3b).

Ten per cent of conflicting pairs had a loss and delivery recorded on the same date and no 'current pregnancy' antenatal codes suggesting they may have been recorded as part of an obstetric history (scenario 2a). Only small percentages of episodes were consistent with other scenarios. Proportional distribution of the scenarios was similar when restricted to those recorded during UTS and current registration to that of the whole Pregnancy Register.

DISCUSSION

This work has shown that uncertain pregnancy episodes in the CPRD Pregnancy Register can contain valuable information about a woman's pregnancy. A high proportion of the uncertain episodes were during research quality follow-up time and therefore comprise data which would usually be included in study designs.[9] We have systematically identified potential reasons for the existence of uncertain episodes within the pregnancy register to allow researchers to consider in more detail whether inclusion is appropriate for their study. This work adds further value to the CPRD Pregnancy Register which is already unique in its inclusion of all pregnancy data regardless of completion.[3 4] To our knowledge, no previous studies have attempted to examine uncertain pregnancies in EHR data in this way and many of the scenarios we have described will also be applicable to other EHR data sources.

We found that most episodes with a missing outcome could be explained by the outcomes not being captured in the CPRD GOLD primary care database; either the patient was not registered at the time of the pregnancy, the outcome was not recorded by the GP but could be found in linked data, or follow-up ended before the outcome. These are likely to be genuine and contemporaneous pregnancies which would be missed if episodes with recorded outcome missing were excluded from the Register. In fact, most of the scenarios we identified are consistent with the episodes being true and current pregnancies. When conducting drug utilisation or vaccine uptake studies, researchers may wish to include episodes where the database follow-up ended before the outcome to avoid underestimation especially for new drugs or vaccination programmes. Further to our objective to provide guidance, table 6 outlines potential considerations for researchers deciding whether to include or exclude uncertain episodes from their study.

There is evidence to suggest that historical outcomes being recorded by the GP during an ongoing pregnancy may explain a sizeable proportion of the uncertain episodes generated by the algorithm. This can lead to true pregnancies being split by the algorithm and depending on the timing, this will either generate an additional episode with outcome missing or two separate episodes with outcomes (figure 1, step 3). In either case, the resulting episodes may conflict with one another. Based on our findings, this appears to be something that happens fairly frequently. One concern is that these episodes are likely to appear more frequently for women with a history of complicated pregnancy outcomes. For example, previous caesarean sections may be likely to be noted by the GP during current care as would outcomes such as ectopic pregnancies. Researchers should be aware that exclusion of women who have overlapping pregnancies for this reason might therefore systematically exclude those with a history of pregnancy complications, introducing bias.

It is also possible that current pregnancies with serious complications are more likely to have an uncertain episode in the Register. For example, women with pre-eclampsia are more likely to have consultant-led antenatal care carried out in hospital, increasing the chances that their primary care record is incomplete and has no recorded outcome.[10] This data pattern is likely to result in the pregnancy being split into multiple episodes without outcome (figure 1, step 8). Dropping all uncertain episodes at the study design stage may mean that these patients are missed. Researchers who are interested in specific pregnancy complications should take this into consideration and use a tailored approach when selecting a study population.

While some conflicting episodes may be caused by poor quality data, there are many conflicting episodes for which it may be possible to clarify which time period is likely to be the true pregnancy. We found that episode conflicts were more likely to occur for pregnancies ending in loss; this is of little surprise given the wider variation around the true gestation of such pregnancies.[11] There

Table 5 Numbers of conflicting pregnancy episodes which were consistent with applied criteria for each scenario*

Scenario	Description	N of pregnancy pairs with evidence of this scenario (% of total conflicting pregnancy pairs)	N of pairs with evidence of only this scenario (% of total conflicting pregnancy pairs)	N of pairs with a linked baby in the MBL (% of total conflicting pregnancy pairs)	N of pairs with evidence of pregnancy in linked HES (% of pairs eligible for HES linkage†)	N of pairs during current registration and UTS follow-up MBL (% of total conflicting pregnancy pairs)
Denominator		251 026	251 026	251 026	160 461†	251 026
Problem 1: Both pregnancies are true but one is a current pregnancy and one is a historical pregnancy.						
Scenario 1a	The GP records a past delivery or loss during a current pregnancy with the same outcome resulting in another episode being created.	2464 (1.0)	413 (0.2)	2164 (0.9)	2332 (1.5)	1981 (0.8)
Scenario 1b	A patient has a record relating to a loss recorded during a pregnancy ending in delivery or vice-versa. Conflicting episodes are generated by the algorithm.	73 208 (29.2)	35 026 (14.0)	11 388 (4.5)	19 900 (12.4)	31 526 (12.6)
Problem 2: Both pregnancies are historical.						
Scenario 2a	A patient has information on historical pregnancies recorded with the current date rather than the actual date.	27 250 (10.9)	0 (0.0)	175 (0.1)	6835 (4.3)	12 557 (5.0)
Problem 3: Both pregnancies are true and current but the gestation of the second pregnancy estimated by the algorithm is too long.						
Scenario 3a	The woman has two losses which are >8 weeks and <12 weeks apart.	6425 (2.6)	12 (0.0)	0 (0.0)	1336 (0.8)	2284 (0.9)
Scenario 3b	The woman has two pregnancies close together and the second ends in delivery. If the last menstrual period date (LMP) information is wrong for this pregnancy, then algorithm episodes may overlap.	32 948 (13.1)	3705 (1.5)	1564 (0.6)	7833 (4.9)	13 464 (5.4)
Problem 4: The pregnancy is real but is split into separate episodes by the rules of the algorithm.						
Scenario 4a	The GP records further information about a pregnancy outcome >25 weeks later for deliveries or >8 weeks or <12 weeks later for losses.	2939 (1.2)	251 (0.1)	2646 (1.1)	2824 (1.8)	2347 (0.9)
Scenario 4b	The GP records further antenatal information after the end of a pregnancy. Conflicting episodes are generated by the algorithm	43 581 (17.4)	40 928 (16.3)	13 531 (5.4)	16 718 (10.4)	27 131 (10.8)
Scenario 4c	The patient has a follow-up scan after a pregnancy loss. The scan is recorded in the data as an antenatal scan, a conflicting episode is then generated by the algorithm.	2734 (1.1)	0 (0.0)	0 (0.0)	744 (0.5)	2088 (0.8)
Scenario 4d	The GP records information about a pregnancy but no outcome with >6 weeks between records. If the second episode has gestational information, the start may be assigned before the start of the first episode.	14 695 (5.9)	14 695 (5.9)	0 (0.0)	7392 (4.6)	9911 (3.9)
Scenario 4e	The pregnancy dates have been shifted backwards by the rules of the algorithm leaving uncovered records. Conflicting episodes are generated by the algorithm.	73 191 (29.2)	70 472 (28.1)	53 464 (21.3)	42 785 (26.7)	55 205 (22.0)
None	These pairs of pregnancies did not meet the criteria for any identified scenarios.	35 449 (14.1)	–	13 241 (5.3)	14 173 (8.8)	15 650 (6.2)

*A version of this table restricted to episodes which occurred during practice UTS follow-up and patient's current registration is given in the appendices (online supplemental appendix 17).
†Denominator=pregnancy episodes which had at least 1-day overlap with the available HES follow-up period and where the woman was eligible for linkage.
GP, general practitioner; HES, Hospital Episode Statistics; MBL, Mother-Baby-Link; UTS, up-to-standard.

Table 6 Issues with different approaches to dealing with uncertain episodes and recommendations

Example uses	Issues with a highly specific approach: excluding all uncertain episodes	Issues with a highly sensitive approach: including all uncertain episodes	Recommended tailored approach: including or excluding uncertain episodes based on scenario criteria
Vaccine uptake study	▶ Underestimate of uptake during pregnancy	▶ Overestimate of uptake during pregnancy where historical episodes are included	▶ Consider using episodes without recorded outcome which continue after data follow-up to maximise the capture of exposure events. ▶ Consider using linked data to obtain additional outcomes. ▶ Exclude episodes which are likely to be derived from historical data based on our described scenarios.
Drug/vaccine safety study	▶ Underestimation of pregnancies ending in loss ▶ Underestimation of pregnancy complications	▶ Misclassification of exposure status ▶ Overestimation of outcomes	▶ Consider using linked data to obtain additional outcomes restricting the study population to those patients eligible for linkage. ▶ Exclude episodes which are likely to be derived from historical data based on our described scenarios. ▶ Consider merging conflicting episodes which are consistent with problem 4 and adjusting the timing accordingly (deciding which of the outcomes is likely to be the true outcome based on the scenarios we have described and then estimating a start date. This should be based on a combination of the patient's antenatal records and default duration dependent on outcome type[3]). ▶ Consider ensuring pregnancy start is at least 9 months before the last data collection date to allow for attainment of outcomes.
Ascertaining pregnancy history	▶ Underestimation of parity ▶ Underestimation of certain pregnancy events ▶ Underestimation of pregnancies ending in loss	▶ Overestimation of parity	▶ Consider using linked data to obtain additional outcomes restricting the study population to those patients eligible for linkage. ▶ Exclude episodes which are likely to be derived from historical data based on our described scenarios. ▶ Consider ensuring pregnancy start is at least 9 months before the last data collection date to allow for attainment of outcomes.
Excluding pregnant women from a study cohort	▶ Reduction in potential study population	▶ Potential misclassification of pregnancy status ▶ Potential errors in pregnancy timing	▶ Consider merging conflicting episodes which are consistent with problem 4 and adjusting the timing accordingly (deciding which of the outcomes is likely to be the true outcome based on the scenarios we have described and then estimating a start date. This should be based on a combination of the patient's antenatal records and a default duration dependent on outcome type[3]). ▶ Consider using linked data to obtain additional outcomes, restricting the study population to those patients eligible for linkage. ▶ Exclude episodes which are likely to be derived from historical data based on our described scenarios.

was also a large overlap between the conflicting episodes and those that were missing an outcome. Again, this is not surprising as the start and end dates for the missing outcome episodes have large margins of error, given they are often estimated based on one or two antenatal codes (figure 1, step 8).[3] Not including uncertain episodes may lead to underascertainment of miscarriage as an outcome. However, including them all may lead to exposure status misclassification due to mistimed start and end dates or past pregnancy outcomes being counted.

Researchers may consider using multiple imputation to handle missing outcomes. However, there is a strong likelihood that the pattern of missing pregnancy outcomes is not missing at random and both multiple imputation and listwise deletion could result in biased results. Investigation of the linked HES data has shown that using these additional data alongside the Register could help users to identify many missing outcomes.[7 8 12] Potentially useful pregnancy outcome data were found in multiple places across the HES APC database (NHS Digital, 2021). Identifying outcomes in HES could allow users of the Register to adjust the dates of the pregnancy episodes. While HES data are useful as a complementary source of information, it is also an EHR database derived from data that were not collected for research purposes and there may be gaps in recording. It is, however, less likely that pregnancy outcome events which happen in hospital will be recorded retrospectively and therefore dates of recorded outcomes may be considered more reliable.

Furthermore, using the HES DID data to access antenatal scan records offers a useful way to validate the dates of primary care pregnancy episodes as patients are unlikely to have an antenatal scan when they are not currently pregnant.[13] When using linked data, we recommend that the study population be restricted to those patients in the Pregnancy Register who are eligible for linkage.

The main limitation of this work is that it relies on the assumption that real-life scenarios will consistently result in the same data patterns. EHR data such as CPRD GOLD are not collected for the purposes of research and can be messy for a variety of reasons. As the criteria we applied to identify our proposed scenarios may not have been a true fit to each pregnancy episode, this may have resulted in misclassification of the true underlying cause. While we did validate a random sample of pregnancy episodes by looking at the individual Read codes recorded, it was

not possible to look at every episode in detail. Furthermore, some of our scenarios relied on assumptions as to why and when GPs may record clinical information relating to pregnancy. While this was informed by clinician advice and clinical guidelines, it may not be correct in every case. There is also the possibility that there are other scenarios which we did not identify, and special cases of scenarios that we could not test. For example, since 2007, women in the UK have been given the option of accessing midwife-led care directly. While information about the pregnancy should be fed to their GP, this may not always be the case. A survey report by the Quality Care Commission published in 2020 estimated that in 2018, 47% of women accessed antenatal care directly through a midwife.[14] As yet, no routinely linked data allow for investigation of this special case of scenario 1a.

We have described in detail reasons why uncertain pregnancy episodes may occur in the CPRD Pregnancy Register and criteria which researchers can apply to ascertain which episodes may fit each scenario. This work offers researchers the opportunity to tailor their study to accommodate these episodes where appropriate (table 6).

CONCLUSIONS

This work has shown evidence that most uncertain pregnancy episodes are consistent with true and current pregnancies for which the data contain valuable information. It is important that researchers carefully consider the impact of including or excluding these episodes from their study. We have demonstrated that examining patterns of events within the primary care data or looking for further evidence in linked data can help to identify possible explanations. Here we offer users of the Pregnancy Register an insight into why these episodes exist and guidance on how to tailor their study population accordingly.

Acknowledgements This work uses data provided by patients and collected by the NHS as part of their care and support.

Contributors JC, KB, ST, RW, HIM and CM contributed to the initiation, planning and design of the study. JC performed the analysis. KB, ST, RW and CM conducted study supervision. HIM and ST provided clinical input. JC wrote the manuscript with KB, ST, RW, HIM and CM performing critical revision. JC is acting as guarantor for this work

Funding This work forms part of JC's PhD which is funded by CPRD (grant number N/A). KB is funded by a Wellcome Senior Research Fellowship (220283/Z/20/Z). CM was supported by a UKRI Innovation Fellowship at Health Data Research UK London (MR/S003932/1). HIM and ST were funded by the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Immunisation (IS-HPU1112-10096) at the London School of Hygiene and Tropical Medicine in partnership with Public Health England (PHE).

Disclaimer The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, or PHE.

Competing interests JC and RW are employees of CPRD. There are no other conflicts of interest to report.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the Independent Scientific Advisory Committee (ISAC) for Medicines and Healthcare Products Regulatory Agency Database Research (protocol no: 17_285R2 and 19_140) and the London School of Hygiene and Tropical Medicine Ethics Committee. This study uses de-identified electronic health records only.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data used for this study were obtained from the Clinical Practice Research Datalink (CPRD). All data are available via an application to CPRD's Research Data Governance (RDG) Process (see https://www.cprd.com/research-applications). Data acquisition is associated with a fee and subject to ethics approval.

ORCID iDs

Jennifer Campbell http://orcid.org/0000-0002-0684-4437
Helen I McDonald http://orcid.org/0000-0003-0576-2015

REFERENCES

1 National Institute for Health and Care Excellence. Antenatal care for uncomplicated pregnancies. *Clin Guidel* 2008:1–55 https://www.nice.org.uk/guidance/cg62/resources/antenatal-care-for-uncomplicated-pregnancies-975564597445
2 CPRD. CPRD: clinical practice research Datalink, 2021. Available: www.cprd.com
3 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;28:923–33. doi:10.1002/pds.4811
4 Margulis AV, Palmsten K, Andrade SE, *et al*. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiol Drug Saf* 2015;24:335–42.
5 Padmanabhan S. CPRD GOLD Data Specification [Internet], 2017. Available: https://cprd.com/sites/default/files/CPRD GOLD Full Data Specification v2.0_0.pdf [Accessed cited 2020 Jul 8].
6 Boggon R, Gallagher A, Williams T, *et al*. Creating a mother baby link and pregnancy register for a UK population. *Pharmacoepidemiol Drug Saf* 2011;20.
7 Padmanabhan S, Carty L, Cameron E, *et al*. Approach to record linkage of primary care data from clinical practice research Datalink to other health-related patient data: overview and implications. *Eur J Epidemiol* 2019;34:91–9. doi:10.1007/s10654-018-0442-4
8 NHS Digital. Hospital episode statistics, 2021. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
9 Lewis JD, Bilker WB, Weinstein RB, *et al*. The relationship between time since registration and measured incidence rates in the general practice research database. *Pharmacoepidemiol Drug Saf* 2005;14:443–51.
10 National Institute for Health and Care Excellence. Hypertension in pregnancy: pre-eclampsia, 2020. Available: https://www.nice.org.uk/guidance/ng133
11 NHS. Miscarriage overview, 2021. Available: https://www.nhs.uk/conditions/miscarriage/
12 Harron K, Gilbert R, Cromwell D, *et al*. Linking data for mothers and babies in de-identified electronic health data. *PLoS One* 2016;11:e0164667–18. doi:10.1371/journal.pone.0164667
13 NHS Digital. Diagnostic imaging dataset, 2021. Available: https://did.hscic.gov.uk/
14 Care Quality Commission. 2018 survey of women's experiences of maternity care, 2020. Available: https://www.cqc.org.uk/sites/default/files/20200128_mat19_statisticalrelease.pdf

3.3 Published Supplementary Files

Appendix 1: Table S1, Key CPRD GOLD variables

Column name	*Field name*	*Description*
Last Collection Date	lcd	Date of the last collection for the practice
Up to Standard Date	uts	Date at which the practice data are deemed to be of research quality. Derived using a CPRD algorithm that primarily looks at practice death recording and gaps in the data
First Registration Date	frd	Date the patient first registered with the practice.
Current Registration Date	crd	Date the patient's current period of registration with the practice began.
Transfer Out Date	tod	Date the patient transferred out of the practice, if relevant. Empty for patients who have not transferred out
Death Date	deathdate	Patient's date of death – derived using a CPRD algorithm
Acceptable Patient Flag	accept	Flag to indicate whether the patient has met certain quality standards: 1 = acceptable, 0 = unacceptable
Event Date	eventdate	Date associated with the event, as entered by the GP
System Date	sysdate	The date on which information was entered on to the GP software system (generated automatically)

Appendix 2: Table S2, CPRD Pregnancy Register Variables

Field name	*Description*
Patid	Encrypted unique patient identifier
Pregid	Unique identifier of the pregnancy episode
Mblbabies	Number of babies the pregnancy is linked to in the MBL
babypatid[1]	Encrypted unique patient identifier (linked baby)
babymob	Baby's month of birth as recorded in the baby's medical record
babyyob	Baby's year of birth as recorded in the baby's medical record
totalpregs	Total number of identified pregnancy episodes (per woman)
pregnumber	Pregnancy episode number (per woman)
pregstart	Estimated start date of pregnancy
firstantenatal	Date of earliest antenatal record within the pregnancy
startsource	Data source used to estimate pregnancy start date: 1 = Imputed[2], 2 = EDD, 3 = LMP, 4 = Gestational age at birth, 5 = Gestational age from antenatal record, 6 = EDC
startadj	Flag to indicate whether the pregnancy start date has been adjusted: 0 = Not adjusted, 1 = Due to antenatal records in the preceding 4 weeks, 2 = Due to specific conflicts between the estimated pregnancy duration and records indicating gestational age at birth (live births and stillbirths only), 3 = Both
Secondtrim[3]	Estimated start date of second trimester
Thirdtrim[3]	Estimated start date of third trimester
pregend	Estimated end date of pregnancy. NB: For pregnancies with unknown outcome, the date of the latest antenatal record in the pregnancy episode is provided.
endsource	Data source used to estimate pregnancy end date: 1 = Delivery record, 2 = Postnatal record in the mother's medical record, 3 = Discharge date relating to a delivery, 4 = Baby's (month and) year of birth as recorded in the baby's medical record, 5 = Postnatal record

	in the baby's medical record, 6 = First consultation in the baby's medical record. Only completed for live births and stillbirths.
endadj	Flag to indicate whether the pregnancy end date has been adjusted: 0 = Not adjusted, 1 = Due to specific conflicts between the estimated pregnancy duration and records indicating gestational age, 2 = Due to prior adjustments to the start date, 3 = Both. Missing for deliveries based on late pregnancy records[4].
gestdays	Estimated duration of pregnancy episode in days (calculated as pregend minus pregstart)
matage	Mother's age at end of pregnancy (years)
outcome	Outcome of pregnancy: 1 = Live birth, 2 = Stillbirth, 3 = 1 and 2, 4 = Miscarriage, 5 = TOP, 6 = Probable TOP, 7 = Ectopic, 8 = Molar, 9 = Blighted ovum, 10 = Unspecified loss, 11 = Delivery based on a third trimester pregnancy record, 12 = Delivery based on a late pregnancy record[4], 13 = Outcome unknown
preterm_ev	Flag to indicate evidence of a premature delivery: 1=preterm, 0=no evidence of preterm, 9=not applicable (outcome not a delivery)
postterm_ev	Flag to indicate evidence of a post-term delivery: 1=post-term, 0=no evidence of post-term, 9=not applicable (outcome not a delivery)
multiple_ev	Flag to indicate evidence of a multiple pregnancy: 1=multiple, 0=no evidence of multiple. Missing for pregnancy losses.
conflict	Flag to indicate whether the pregnancy episode overlaps with another episode (within a woman): 1=conflicting, 0= non-conflicting

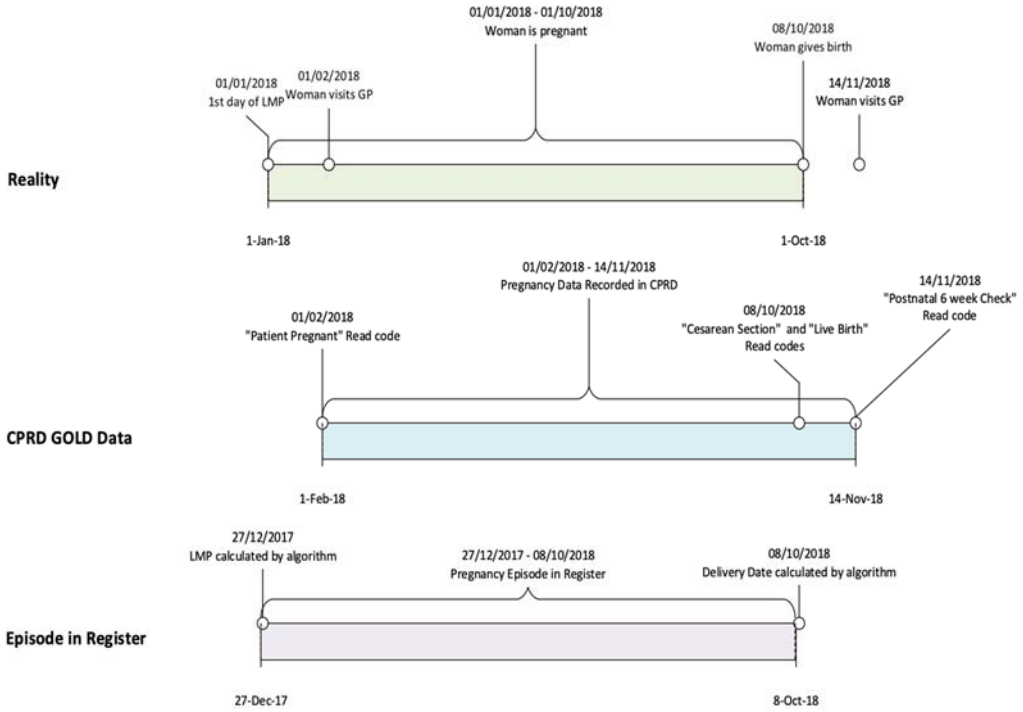
1 A single babypatid is provided. For multiple pregnancies resulting in >1 liveborn infant (when mblbabies>1), additional babypatids may be retrieved from the MBL.

2 For "Outcome unknown" pregnancies, the imputed start date is obtained by subtracting 4 weeks from the earliest antenatal record in the episode.

3 The timing of trimesters is estimated using a common convention: first trimester (first day of LMP [pregstart] to 13 completed weeks), second (weeks 14 to 26), and third (week 27 to delivery [pregend]).

4 Late pregnancy records refer to the period up to 3 weeks before delivery, e.g. "Baby overdue".

Appendix 3: Figure S1, Example of how a pregnancy may appear in the Register vs GOLD data vs reality



Appendix 4: Table S3, ICD codes indicating end of pregnancy

O00	Ectopic pregnancy	
O00.0	Abdominal pregnancy	
O00.1	Tubal pregnancy	
O00.2	Ovarian pregnancy	
O00.8	Other ectopic pregnancy	
O00.9	Ectopic pregnancy, unspecified	
O01	Hydatidiform mole	
O01.0	Classical hydatidiform mole	
O01.1	Incomplete and partial hydatidiform mole	
O01.9	Hydatidiform mole, unspecified	
O02	Other abnormal products of conception	
O02.0	Blighted ovum and nonhydatidiform mole	
O02.1	Missed abortion	
O02.8	Other specified abnormal products of conception	
O02.9	Abnormal product of conception, unspecified	
O03	Spontaneous abortion	
O03.0	Spontaneous abortion	Incomplete, complicated by genital tract and pelvic infection
O03.1	Spontaneous abortion	Incomplete, complicated by delayed or excessive haemorrhage
O03.2	Spontaneous abortion	Incomplete, complicated by embolism
O03.3	Spontaneous abortion	Incomplete, with other and unspecified complications
O03.4	Spontaneous abortion	Incomplete, without complication
O03.5	Spontaneous abortion	Complete or unspecified, complicated by genital tract and pelvic infection
O03.6	Spontaneous abortion	Complete or unspecified, complicated by delayed or excessive haemorrhage
O03.7	Spontaneous abortion	Complete or unspecified, complicated by embolism
O03.8	Spontaneous abortion	Complete or unspecified, with other and unspecified complications
O03.9	Spontaneous abortion	Complete or unspecified, without complication

O04	Medical abortion	
O04.0	Medical abortion	Incomplete, complicated by genital tract and pelvic infection
O04.1	Medical abortion	Incomplete, complicated by delayed or excessive haemorrhage
O04.2	Medical abortion	Incomplete, complicated by embolism
O04.3	Medical abortion	Incomplete, with other and unspecified complications
O04.4	Medical abortion	Incomplete, without complication
O04.5	Medical abortion	Complete or unspecified, complicated by genital tract and pelvic infection
O04.6	Medical abortion	Complete or unspecified, complicated by delayed or excessive haemorrhage
O04.7	Medical abortion	Complete or unspecified, complicated by embolism
O04.8	Medical abortion	Complete or unspecified, with other and unspecified complications
O04.9	Medical abortion	Complete or unspecified, without complication
O05	Other abortion	
O05.0	Other abortion	Incomplete, complicated by genital tract and pelvic infection
O05.1	Other abortion	Incomplete, complicated by delayed or excessive haemorrhage
O05.2	Other abortion	Incomplete, complicated by embolism
O05.3	Other abortion	Incomplete, with other and unspecified complications
O05.4	Other abortion	Incomplete, without complication
O05.5	Other abortion	Complete or unspecified, complicated by genital tract and pelvic infection
O05.6	Other abortion	Complete or unspecified, complicated by delayed or excessive haemorrhage
O05.7	Other abortion	Complete or unspecified, complicated by embolism
O05.8	Other abortion	Complete or unspecified, with other and unspecified complications
O05.9	Other abortion	Complete or unspecified, without complication
O06	Unspecified abortion	
O06.0	Unspecified abortion	Incomplete, complicated by genital tract and pelvic infection
O06.1	Unspecified abortion	Incomplete, complicated by delayed or excessive haemorrhage
O06.2	Unspecified abortion	Incomplete, complicated by embolism
O06.3	Unspecified abortion	Incomplete, with other and unspecified complications

O06.4	Unspecified abortion	Incomplete, without complication
O06.5	Unspecified abortion	Complete or unspecified, complicated by genital tract and pelvic infection
O06.6	Unspecified abortion	Complete or unspecified, complicated by delayed or excessive haemorrhage
O06.7	Unspecified abortion	Complete or unspecified, complicated by embolism
O06.8	Unspecified abortion	Complete or unspecified, with other and unspecified complications
O06.9	Unspecified abortion	Complete or unspecified, without complication
O07	Failed attempted abortion	
O07.0	Failed medical abortion, complicated by genital tract and pelvic infection	
O07.1	Failed medical abortion, complicated by delayed or excessive haemorrhage	
O07.2	Failed medical abortion, complicated by embolism	
O07.3	Failed medical abortion, with other and unspecified complications	
O07.4	Failed medical abortion, without complication	
O07.5	Other and unspecified failed attempted abortion, complicated by genital tract and pelvic infection	
O07.6	Other and unspecified failed attempted abortion, complicated by delayed or excessive haemorrhage	
O07.7	Other and unspecified failed attempted abortion, complicated by embolism	
O07.8	Other and unspecified failed attempted abortion, with other and unspecified complications	
O07.9	Other and unspecified failed attempted abortion, without complication	
O08	Complications following abortion and ectopic and molar pregnancy	
O08.0	Genital tract and pelvic infection following abortion and ectopic and molar pregnancy	
O08.1	Delayed or excessive haemorrhage following abortion and ectopic and molar pregnancy	
O08.2	Embolism following abortion and ectopic and molar pregnancy	
O08.3	Shock following abortion and ectopic and molar pregnancy	
O08.4	Renal failure following abortion and ectopic and molar pregnancy	
O08.5	Metabolic disorders following abortion and ectopic and molar pregnancy	
O08.6	Damage to pelvic organs and tissues following abortion and ectopic and molar pregnancy	
O08.7	Other venous complications following abortion and ectopic and molar pregnancy	
O08.8	Other complications following abortion and ectopic and molar pregnancy	
O08.9	Complication following abortion and ectopic and molar pregnancy, unspecified	

O60.1	Preterm spontaneous labour with preterm delivery
O60.2	Preterm spontaneous labour with term delivery
O62.3	Precipitate labour
O68	Labour and delivery complicated by fetal stress [distress]
O68.0	Labour and delivery complicated by fetal heart rate anomaly
O68.1	Labour and delivery complicated by meconium in amniotic fluid
O68.2	Labour and delivery complicated by fetal heart rate anomaly with meconium in amniotic fluid
O68.3	Labour and delivery complicated by biochemical evidence of fetal stress
O68.8	Labour and delivery complicated by other evidence of fetal stress
O68.9	Labour and delivery complicated by fetal stress, unspecified
O69	Labour and delivery complicated by umbilical cord complications
O69.0	Labour and delivery complicated by prolapse of cord
O69.1	Labour and delivery complicated by cord around neck, with compression
O69.2	Labour and delivery complicated by other cord entanglement, with compression
O69.3	Labour and delivery complicated by short cord
O69.4	Labour and delivery complicated by vasa praevia
O69.5	Labour and delivery complicated by vascular lesion of cord
O69.8	Labour and delivery complicated by other cord complications
O69.9	Labour and delivery complicated by cord complication, unspecified

O70	Perineal laceration during delivery
O70.0	First degree perineal laceration during delivery
O70.1	Second degree perineal laceration during delivery
O70.2	Third degree perineal laceration during delivery
O70.3	Fourth degree perineal laceration during delivery
O70.9	Perineal laceration during delivery, unspecified
O74	Complications of anaesthesia during labour and delivery
O74.0	Aspiration pneumonitis due to anaesthesia during labour and delivery
O74.1	Other pulmonary complications of anaesthesia during labour and delivery
O74.2	Cardiac complications of anaesthesia during labour and delivery
O74.3	Central nervous system complications of anaesthesia during labour and delivery
O74.4	Toxic reaction to local anaesthesia during labour and delivery
O74.5	Spinal and epidural anaesthesia-induced headache during labour and delivery
O74.6	Other complications of spinal and epidural anaesthesia during labour and delivery
O74.7	Failed or difficult intubation during labour and delivery
O74.8	Other complications of anaesthesia during labour and delivery
O74.9	Complication of anaesthesia during labour and delivery, unspecified
O75	Other complications of labour and delivery, not elsewhere classified
O75.0	Maternal distress during labour and delivery

O75.1	Shock during or following labour and delivery
O75.5	Delayed delivery after artificial rupture of membranes
O75.6	Delayed delivery after spontaneous or unspecified rupture of membranes
O75.7	Vaginal delivery following previous caesarean section
O75.8	Other specified complications of labour and delivery
O75.9	Complication of labour and delivery, unspecified
O80	Single spontaneous delivery
O80.0	Spontaneous vertex delivery
O80.1	Spontaneous breech delivery
O80.8	Other single spontaneous delivery
O80.9	Single spontaneous delivery, unspecified
O81	Single delivery by forceps and vacuum extractor
O81.0	Low forceps delivery
O81.1	Mid-cavity forceps delivery
O81.3	Other and unspecified forceps delivery
O81.4	Vacuum extractor delivery
O81.5	Delivery by combination of forceps and vacuum extractor
O82	Single delivery by caesarean section
O82.0	Delivery by elective caesarean section

O82.1	Delivery by emergency caesarean section
O82.2	Delivery by caesarean hysterectomy
O82.8	Other single delivery by caesarean section
O82.9	Delivery by caesarean section, unspecified
O83	Other assisted single delivery
O83.0	Breech extraction
O83.1	Other assisted breech delivery
O83.2	Other manipulation-assisted delivery
O83.4	Destructive operation for delivery
O83.8	Other specified assisted single delivery
O83.9	Assisted single delivery, unspecified
O84	Multiple delivery
O84.0	Multiple delivery, all spontaneous
O84.1	Multiple delivery, all by forceps and vacuum extractor
O84.2	Multiple delivery, all by caesarean section
O84.8	Other multiple delivery
O84.9	Multiple delivery, unspecified
P03	Fetus and newborn affected by other complications of labour and delivery
P03.0	Fetus and newborn affected by breech delivery and extraction

P03.1	Fetus and newborn affected by other malpresentation, malposition and disproportion during labour and delivery
P03.2	Fetus and newborn affected by forceps delivery
P03.3	Fetus and newborn affected by delivery by vacuum extractor [ventouse]
P03.4	Fetus and newborn affected by caesarean delivery
P03.5	Fetus and newborn affected by precipitate delivery
P03.8	Fetus and newborn affected by other specified complications of labour and delivery
P03.9	Fetus and newborn affected by complication of labour and delivery, unspecified
P04.0	Fetus and newborn affected by maternal anaesthesia and analgesia in pregnancy, labour and delivery
P20.1	Intrauterine hypoxia first noted during labour and delivery
P61.2	Anaemia of prematurity
Z37	Outcome of delivery
Z37.0	Single live birth
Z37.1	Single stillbirth
Z37.2	Twins, both liveborn
Z37.3	Twins, one liveborn and one stillborn
Z37.4	Twins, both stillborn
Z37.5	Other multiple births, all liveborn
Z37.6	Other multiple births, some liveborn
Z37.7	Other multiple births, all stillborn

Z38	Liveborn infants according to place of birth
Z38.0	Singleton, born in hospital
Z38.1	Singleton, born outside hospital
Z38.2	Singleton, unspecified as to place of birth
Z38.3	Twin, born in hospital
Z38.4	Twin, born outside hospital
Z38.5	Twin, unspecified as to place of birth
Z38.6	Other multiple, born in hospital
Z38.7	Other multiple, born outside hospital
Z38.8	Other multiple, unspecified as to place of birth
Z39.0	Care and examination immediately after delivery

Appendix 5: Table S4, OPCS codes indicating end of pregnancy

OPCS		
P141	INCISION OF INTROITUS OF VAGINA	POSTERIOR EPISIOTOMY AND DIVISION OF LEVATOR ANI MUSCLE
P142	INCISION OF INTROITUS OF VAGINA	POSTERIOR EPISIOTOMY NEC
P143	INCISION OF INTROITUS OF VAGINA	ANTERIOR EPISIOTOMY
Q101	CURETTAGE OF UTERUS	DILATION OF CERVIX UTERI AND CURETTAGE OF PRODUCTS OF CONCEP
Q102	CURETTAGE OF UTERUS	CURETTAGE OF PRODUCTS OF CONCEPTION FROM UTERUS NEC
Q111	OTHER EVACUATION OF CONTENTS OF UTERUS	VACUUM ASPIRATION OF PRODUCTS OF CONCEPTION FROM UTERUS NEC
Q112	OTHER EVACUATION OF CONTENTS OF UTERUS	DILATION OF CERVIX UTERI AND EVACUATION OF PRODUCTS OF CONCE
Q113	OTHER EVACUATION OF CONTENTS OF UTERUS	EVACUATION OF PRODUCTS OF CONCEPTION FROM UTERUS NEC
Q115	OTHER EVACUATION OF CONTENTS OF UTERUS	VACUUM ASPIRATION/PRODUCTS OF CONCEPTION/UTERUS USING RIGID
Q116	OTHER EVACUATION OF CONTENTS OF UTERUS	VACUUM ASPIRATION/PRODUCTS OF CONCEPTION/UTERUS USING FLEXI
Q141	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	INTRA-AMNIOTIC INJECTION OF PROSTAGLANDIN
Q142	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	INTRA-AMNIOTIC INJECTION OF ABORTIFACIENT NEC
Q143	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	EXTRA-AMNIOTIC INJECTION OF PROSTAGLANDIN
Q144	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	EXTRA-AMNIOTIC INJECTION OF ABORTIFACIENT NEC
Q145	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	INSERTION OF PROSTAGLANDIN PESSARY
Q146	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	INSERTION OF ABORTIFACIENT PESSARY NEC
Q148	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	OTHER SPECIFIED
Q149	INTRODUCTION OF ABORTIFACIENT INTO UTERINE CAVITY	UNSPECIFIED
R031	SELECTIVE DESTRUCTION OF FETUS	EARLY SELECTIVE FETICIDE
R032	SELECTIVE DESTRUCTION OF FETUS	LATE SELECTIVE FETICIDE
R038	SELECTIVE DESTRUCTION OF FETUS	OTHER SPECIFIED
R039	SELECTIVE DESTRUCTION OF FETUS	UNSPECIFIED
R141	SURGICAL INDUCTION OF LABOUR	FOREWATER RUPTURE OF AMNIOTIC MEMBRANE
R142	SURGICAL INDUCTION OF LABOUR	HINDWATER RUPTURE OF AMNIOTIC MEMBRANE
R148	SURGICAL INDUCTION OF LABOUR	OTHER SPECIFIED
R149	SURGICAL INDUCTION OF LABOUR	UNSPECIFIED
R151	OTHER INDUCTION OF LABOUR	MEDICAL INDUCTION OF LABOUR
R158	OTHER INDUCTION OF LABOUR	OTHER SPECIFIED

R159	OTHER INDUCTION OF LABOUR	UNSPECIFIED
R171	ELECTIVE CAESAREAN DELIVERY	ELECTIVE UPPER UTERINE SEGMENT CAESAREAN DELIVERY
R172	ELECTIVE CAESAREAN DELIVERY	ELECTIVE LOWER UTERINE SEGMENT CAESAREAN DELIVERY
R178	ELECTIVE CAESAREAN DELIVERY	OTHER SPECIFIED
R179	ELECTIVE CAESAREAN DELIVERY	UNSPECIFIED
R181	OTHER CAESAREAN DELIVERY	UPPER UTERINE SEGMENT CAESAREAN DELIVERY NEC
R182	OTHER CAESAREAN DELIVERY	LOWER UTERINE SEGMENT CAESAREAN DELIVERY NEC
R188	OTHER CAESAREAN DELIVERY	OTHER SPECIFIED
R189	OTHER CAESAREAN DELIVERY	UNSPECIFIED
R191	BREECH EXTRACTION DELIVERY	BREECH EXTRACTION DELIVERY WITH VERSION
R198	BREECH EXTRACTION DELIVERY	OTHER SPECIFIED
R199	BREECH EXTRACTION DELIVERY	UNSPECIFIED
R201	OTHER BREECH DELIVERY	SPONTANEOUS BREECH DELIVERY
R202	OTHER BREECH DELIVERY	ASSISTED BREECH DELIVERY
R208	OTHER BREECH DELIVERY	OTHER SPECIFIED
R209	OTHER BREECH DELIVERY	UNSPECIFIED
R211	FORCEPS CEPHALIC DELIVERY	HIGH FORCEPS CEPHALIC DELIVERY WITH ROTATION
R212	FORCEPS CEPHALIC DELIVERY	HIGH FORCEPS CEPHALIC DELIVERY NEC
R213	FORCEPS CEPHALIC DELIVERY	MID FORCEPS CEPHALIC DELIVERY WITH ROTATION
R214	FORCEPS CEPHALIC DELIVERY	MID FORCEPS CEPHALIC DELIVERY NEC
R215	FORCEPS CEPHALIC DELIVERY	LOW FORCEPS CEPHALIC DELIVERY
R218	FORCEPS CEPHALIC DELIVERY	OTHER SPECIFIED
R219	FORCEPS CEPHALIC DELIVERY	UNSPECIFIED
R221	VACUUM DELIVERY	HIGH VACUUM DELIVERY
R222	VACUUM DELIVERY	LOW VACUUM DELIVERY
R223	VACUUM DELIVERY	VACUUM DELIVERY BEFORE FULL DILATION OF CERVIX
R228	VACUUM DELIVERY	OTHER SPECIFIED
R229	VACUUM DELIVERY	UNSPECIFIED
R231	CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRESENTATION OF	MANIPULATIVE CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRESENT
R232	CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRESENTATION OF	NON-MANIPULATIVE CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRE
R238	CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRESENTATION OF	OTHER SPECIFIED
R239	CEPHALIC VAGINAL DELIVERY WITH ABNORMAL PRESENTATION OF	UNSPECIFIED
R249	NORMAL DELIVERY	ALL
R251	OTHER METHODS OF DELIVERY	CAESAREAN HYSTERECTOMY
R252	OTHER METHODS OF DELIVERY	DESTRUCTIVE OPERATION TO FACILITATE DELIVERY

R258	OTHER METHODS OF DELIVERY	OTHER SPECIFIED
R259	OTHER METHODS OF DELIVERY	UNSPECIFIED
R271	OTHER OPERATIONS TO FACILITATE DELIVERY	EPISIOTOMY TO FACILITATE DELIVERY
R278	OTHER OPERATIONS TO FACILITATE DELIVERY	OTHER SPECIFIED
R279	OTHER OPERATIONS TO FACILITATE DELIVERY	UNSPECIFIED
R281	INSTRUMENTAL REMOVAL/PRODUCTS/CONCEPTION FROM DEL.UTERU	CURETTAGE OF DELIVERED UTERUS
R288	INSTRUMENTAL REMOVAL/PRODUCTS/CONCEPTION FROM DEL.UTERU	OTHER SPECIFIED
R289	INSTRUMENTAL REMOVAL/PRODUCTS/CONCEPTION FROM DEL.UTERU	UNSPECIFIED
R291	MANUAL REMOVAL/PRODUCTS/CONCEPTION FROM DELIVERED UTERU	MANUAL REMOVAL OF PLACENTA FROM DELIVERED UTERUS
R298	MANUAL REMOVAL/PRODUCTS/CONCEPTION FROM DELIVERED UTERU	OTHER SPECIFIED
R299	MANUAL REMOVAL/PRODUCTS/CONCEPTION FROM DELIVERED UTERU	UNSPECIFIED
R301	OTHER OPERATIONS ON DELIVERED UTERUS	REPOSITIONING OF INVERTED DELIVERED UTERUS
R302	OTHER OPERATIONS ON DELIVERED UTERUS	EXPRESSION OF PLACENTA
R303	OTHER OPERATIONS ON DELIVERED UTERUS	INSTRUMENTAL EXPLORATION OF DELIVERED UTERUS NEC
R304	OTHER OPERATIONS ON DELIVERED UTERUS	MANUAL EXPLORATION OF DELIVERED UTERUS NEC
R308	OTHER OPERATIONS ON DELIVERED UTERUS	OTHER SPECIFIED
R309	OTHER OPERATIONS ON DELIVERED UTERUS	UNSPECIFIED
R321	REPAIR OF OBSTETRIC LACERATION	REPAIR OF OBSTETRIC LACERATION OF UTERUS OR CERVIX UTERI
R322	REPAIR OF OBSTETRIC LACERATION	REPAIR OF OBSTETRIC LACERATION OF PERINEUM AND SPHINCTER
R323	REPAIR OF OBSTETRIC LACERATION	REPAIR OF OBSTETRIC LACERATION OF VAGINA AND FLOOR OF PELVIS
R324	REPAIR OF OBSTETRIC LACERATION	REPAIR OF MINOR OBSTETRIC LACERATION
R325	REPAIR OF OBSTETRIC LACERATION	REPAIR OBSTETRIC LACERATION PERINEUM SPHINCTER MUCOSA ANUS
R328	REPAIR OF OBSTETRIC LACERATION	OTHER SPECIFIED
R329	REPAIR OF OBSTETRIC LACERATION	UNSPECIFIED

Appendix 6: Table S5, HES Maternity Values to indicate delivery

Variable	*Definition*	*Acceptable values*
numbaby	Number of babies delivered	1-4
delmeth	Method used to deliver a baby that is a registrable birth	0-9
delplac	Actual type of delivery place	0-8
delprean	Anaesthetic or analgesic administered before and during labour and delivery	1-7
delposan	Anaesthetic or analgesic administered after delivery	1-7
neodur	Baby's age in days	>=1
neocare	Neonatal level of care	0-3
postdur	Postnatal days of stay	>=1

Appendix 7: Table S6, Pregnancy Read codes identified as likely to be recorded as useful pregnancy history

medcode	read_oxmis_code	read_oxmis_term
164	635..13	Premature baby
165	L04..11	Miscarriage
255	L05..12	Termination of pregnancy
364	7F13111	Lower uterine segment caesarean section (LSCS) NEC
618	L398400	Delivery by emergency caesarean section
683	Q420.00	Haemolytic disease due to rhesus isoimmunisation
720	L398.00	Caesarean delivery
740	7F12.00	Elective caesarean delivery
863	L398200	Caesarean section - pregnancy at term
974	Q4z..15	Stillbirth NEC
1413	L264.00	Intrauterine death
1492	L36..00	Postpartum haemorrhage (PPH)
1744	L03..00	Ectopic pregnancy
2240	Q4z..12	Neonatal death
2638	L1...00	Pregnancy complications
2639	E204.11	Postnatal depression
2664	L180900	Gestational diabetes mellitus
2787	L11..11	Antepartum haemorrhage
2923	62T1.00	Puerperal depression
2924	7E06600	Hysterotomy and termination of pregnancy
3029	L166500	Infections of kidney in pregnancy
3085	7F12z00	Elective caesarean delivery NOS
3327	L13..11	Hyperemesis gravidarum
3874	L031200	Tubal abortion
4367	L362.00	Secondary and delayed postpartum haemorrhage
4530	L00..00	Hydatidiform mole
4607	L414.00	Postnatal deep vein thrombosis
4638	7F13.00	Other caesarean delivery
4786	L213200	Multiple delivery, all by caesarean section
4979	Eu53012	[X]Postpartum depression NOS
5113	L39y411	Postnatal vaginal discomfort
5464	L11y100	Other antepartum haemorrhage - delivered
7174	L43..00	Obstetric pulmonary embolism
7670	L398z00	Caesarean delivery NOS
7916	Z254500	Delivered by caesarean section - pregnancy at term
8147	L264.11	Fetal death in utero
8295	Q48D100	[X]Macerated stillbirth
8446	L180811	Gestational diabetes mellitus
8776	Q48D.00	[X] Stillbirth
8906	ZV27.12	[V]Stillbirth
9067	L125.00	Severe pre-eclampsia
9668	7F12100	Elective lower uterine segment caesarean delivery
9800	L398300	Delivery by elective caesarean section

10049	7F12111	Elective lower uterine segment caesarean section (LSCS)
10278	L180800	Diabetes mellitus arising in pregnancy
11359	L180.00	Diabetes mellitus during pregnancy/childbirth/puerperium
11947	L181500	Postpartum thyroiditis
11986	7E13300	Excision of ruptured ectopic tubal pregnancy
12090	L126.00	Eclampsia
12118	7F13300	Emergency caesarean section
12320	L09..11	Complications following abortion/ectopic/molar pregnancies
13307	Eu53011	[X]Postnatal depression NOS
13584	3885	Edinburgh postnatal depression scale
15061	L13..12	Hyperemesis of pregnancy
15514	7F13000	Upper uterine segment caesarean delivery NEC
15533	L451400	Obstetric breast abscess with postnatal complication
16250	L414.12	Phlegmasia alba dolens - obstetric
16281	L45z400	Obstetric breast infection NOS with postnatal complication
16321	L360.00	Third-stage postpartum haemorrhage
17614	Eu53111	[X]Puerperal psychosis NOS
17744	7F13100	Lower uterine segment caesarean delivery NEC
18258	L167.00	Liver disorder in pregnancy
18369	ZV27100	[V]Single stillbirth
18702	6G00.00	Postnatal depression counselling
18770	Q20yz13	Renal injury due to birth trauma
18830	L414.11	DVT - deep venous thrombosis, postnatal
20152	L090y00	Sepsis NOS following abortion/ectopic/molar pregnancy
20165	L363.00	Postpartum coagulation defects
20307	L091.00	Delayed/excessive haemorrhage following abortive pregnancy
20573	Q48D000	[X]Fresh stillbirth
22775	L11y.00	Other antepartum haemorrhage
23015	6334	Twins - 1 still + 1 live born
23588	L414200	Postnatal deep vein thrombosis with postnatal complication
23642	Eu53z00	[X]Puerperal mental disorder, unspecified
24089	L356z00	Obstetric damage to pelvic joints and ligaments NOS
24927	Eu53.00	[X]Mental and behav disorders assoc with the puerperium NEC
24951	L18C.00	Endocrine nutrition+metab dis complic pregn,childbirth+puerp
25028	L09z.00	Complication NOS following abortion/ectopic/molar pregnancy
25415	Q411.00	Perinatal intraventricular haemorrhage
28364	Q420.12	Rhesus isoimmunisation of the newborn
28861	L398500	Delivery by caesarean hysterectomy
29155	7F1A000	Caesarean hysterectomy
31203	6332	Single stillbirth
31857	Q204.00	Spine or spinal cord injury due to birth trauma

32950	L03y100	Cornual pregnancy
33477	L398100	Caesarean delivery - delivered
33724	L03z.00	Ectopic pregnancy NOS
34136	L120z00	Benign essential hypertension in preg/childb/puerp NOS
34173	L12B.00	Proteinuric hypertension of pregnancy
34299	L240.00	Congenital abnormality of uterus in preg/childbirth/puerp
34502	6335	Twins - both still born
34639	L180100	Diabetes mellitus during pregnancy - baby delivered
34868	L4...00	Complications of the puerperium
35190	7F13z00	Other caesarean delivery NOS
35309	6755	Post miscarriage counselling
36421	L167z00	Liver disorder in pregnancy NOS
37280	L36z.00	Postpartum haemorrhage NOS
39117	L126500	Eclampsia in pregnancy
40224	Eu53000	[X]Mild mental/behav disorder assoc with the puerperium NEC
40500	Eu53100	[X]Severe mental and behav disorder assoc wth puerperium NEC
40730	L125z00	Severe pre-eclampsia NOS
42088	L125100	Severe pre-eclampsia - delivered
42598	L175.00	Maternal rubella in pregnancy, childbirth and the puerperium
44494	L441z00	Caesarean wound disruption NOS
45806	L070x00	Unspecified abortion with complication NOS
46756	L184.00	Mental disorders in pregnancy, childbirth and the puerperium
47227	ZV27300	[V]Twins, one live born and one stillborn
47542	L362200	Secondary postpartum haemorrhage with postnatal problem
47546	7F12y00	Other specified elective caesarean delivery
47607	L440.11	CVA - cerebrovascular accident in the puerperium
47686	L181.00	Thyroid dysfunction in pregnancy/childbirth/puerperium
47741	L127000	Pre-eclampsia or eclampsia with hypertension unspecified
47863	Lyu5200	[X]Other single delivery by caesarean section
48500	Q49..00	Cardiovascular disorders originating in the perinatal period
49363	Q200100	Subdural haemorrhage unspecified, due to birth trauma
50093	L093000	Oliguria following abortive pregnancy
52875	L398000	Caesarean delivery unspecified
52967	Lyu0B00	[X]Complic following abortion & ectopic & molar preg, unspec
53141	L241.00	Tumour of uterine body in pregnancy/childbirth/puerperium
54652	L362z00	Secondary and delayed postpartum haemorrhage NOS
55304	L131z00	Hyperemesis gravidarum with metabolic disturbance NOS
56279	L440.12	Stroke in the puerperium
57236	L400200	Puerperal endometritis with postnatal complication
58156	L03y.00	Other ectopic pregnancy

58982	L186.00	Other cardiovascular diseases in pregnancy/childbirth/puerp
61204	L414z00	Postnatal deep vein thrombosis NOS
61578	L441000	Caesarean wound disruption unspecified
62052	L092500	Uterus damage following abortive pregnancy
62358	L167000	Liver disorder in pregnancy unspecified
62919	L125200	Severe pre-eclampsia - delivered with postnatal complication
63277	L393.00	Acute renal failure following labour and delivery
64127	L121000	Renal hypertension in pregnancy/childbirth/puerp unspecified
64384	L180z00	Diabetes mellitus in pregnancy/childbirth/puerperium NOS
66213	Q20yz12	Kidney injury due to birth trauma
66594	L186.11	Heart disease during pregnancy
67006	L096400	Pulmonary embolism following abortive pregnancy
68319	L351300	Rupture of uterus during/after labour with postnatal problem
70891	L126400	Eclampsia with postnatal complication
71314	L093.00	Renal failure following abortive pregnancy
71717	L121100	Renal hypertension in pregnancy/childbirth/puerp - delivered
72215	L241z00	Uterine body tumour in pregnancy/childbirth/puerperium NOS
72230	L241100	Tumour of uterine body - baby delivered
72458	L393000	Post-delivery acute renal failure unspecified
72513	7F13200	Extraperitoneal caesarean section
73407	L261200	Rhesus isoimmunisation with antenatal problem
73617	L261000	Rhesus isoimmunisation unspecified
73647	L188000	Abnormal GTT - unspec whether during pregnancy/puerperium
86756	Qyu3600	[X]Other chronic resp diseases originating/perinatal period
93710	Q317y00	Other specified perinatal chronic respiratory disease
94718	L121z00	Renal hypertension in pregnancy/childbirth/puerperium NOS
97367	L43z100	Obstetric pulmonary embolism NOS - delivered
99188	L173.00	Maternal tuberculosis in pregnancy/childbirth/puerperium
103465	Qyu3B00	[X]Cardiovasc disord origin in the perinat period, unspecif
103677	Eu32B00	[X]Antenatal depression
110868	L181000	Thyroid dysfunction - unspec whether in pregnancy/puerperium
111574	L114z00	Antepartum haemorrhage with trauma NOS

Appendix 8: Table S7, Antenatal Read codes identified as pregnancy advice codes

medcode	read_oxmis_code	read_oxmis_term
30351	67A6.00	Drugs in pregnancy advice
36903	67AZ.00	Pregnancy advice NOS
102359	67AF.00	Pregnancy advice for patients with epilepsy
107892	67Iu.00	Advice on risk harm to fetus from maternl medictn dur preg
110888	67It.00	Advice on risk harm to mother from maternl medictn dur preg

Appendix 9: Table S8, Read codes potentially misclassified as antenatal rather than outcomes

medcode	read_oxmis_code	read_oxmis_term
424	L281.00	Premature rupture of membranes
906	L100.00	Threatened abortion
1413	L264.00	Intrauterine death
1737	L02..00	Missed abortion
1879	L071.00	Unspecified abortion incomplete
3004	L14..11	Premature labour
6730	L051.12	Surgical abortion - incomplete
7114	L044.00	Inevitable abortion incomplete
7413	L041.00	Spontaneous abortion incomplete
8076	8H7W.00	Refer to TOP counselling
8147	L264.11	Fetal death in utero
8173	L043.00	Inevitable abortion unspecified
12241	L02..11	Missed miscarriage
12337	L051.00	Legal abortion incomplete
17625	L044.11	Inevitable miscarriage incomp
20621	ZV25313	[V]Admission for termination of pregnancy
20809	L14..00	Early or threatened labour
20933	6776	Preg. termination counselling
25883	L071y00	Unspecified incomplete abortion + no mention of complication
28605	L051z00	Incomplete legal abortion NOS
29439	L041z00	Incomplete spontaneous abortion NOS
33964	L0A4.00	Failed medical abortion, without complication
35184	L071z00	Unspecified incomplete abortion NOS
35273	L097.00	Readmission for abortive pregnancy (NHS codes)
35701	L100000	Threatened abortion unspecified
37831	L264z00	Intrauterine death NOS

39754	L051.11	Medal abortion - incomplete
41118	L08z.00	Failed attempted abortion NOS
41783	L041100	Incomp spontaneous abortion + delayed/excessive haemorrhage
47376	L0A1.00	Failed medical abortion complic by genital tract/pelvic infn
47435	L097200	Readmission for retained produc of concept, illegal abortion
50903	L0A2.00	Failed medical abortion comp by delayed/excessive haem'ge
53201	ZV25B00	[V]Admission for administration of abortifacient
59572	L0A3.00	Failed medical abortion, complicated by embolism
59789	L14z.00	Early or threatened labour NOS
65716	Q011.00	Fetus/neonate affected maternal premature rupture membrane
68683	7E0B.00	Introduction of abortifacient into uterine cavity
96418	L06z.00	Illegally induced abortion NOS
97391	L281200	Premature rupture of membranes with antenatal problem
99205	7E0Bz00	Introduction of abortifacient into uterine cavity NOS
101959	7E0B300	Extraamniotic injection of abortifacient NEC
102362	389B.00	Assessment for termination of pregnancy
102494	8Hh3.00	Self referral to termination of pregnancy service
105048	7E0By00	Introduction of abortifacient into uterine cavity OS

Appendix 10: Table S9, Outcome Groupings

Pregnancy Outcomes will be grouped together with those pregnancies which would have similar rules applied and combinations of outcome group for each pair will be coded.

Group	*Pregnancy Register codes*	*Group*
Early Pregnancy Loss	4, 5, 6, 10, 7, 8, 9	1
Delivery	1, 2, 3, 11, 12	2
Unknown Outcome	13	3

Appendix 11: Table S10, Read Codes identified as likely to only be recorded during current pregnancy

medcode	read_oxmis_term
30979	[SO]Fetus
36441	[V]Amniocentesis to screen for chromosomal anomalies
61455	[V]Amniotic fluid to screen for alphafetoprotein levels
6298	[V]Antenatal screening
49665	[V]Antenatal screening for chromosomal anomalies
35912	[V]Pregnancy confirmed
43428	[V]Screening for fetal growth retardation using ultrasonics
103341	[V]Screening for isoimmunisation
7536	[V]Screening for malformations using ultrasonics
13167	A/N 12 weeks examination
13166	A/N 16 week examination
29364	A/N 20 week examination
13169	A/N 24 week examination
26554	A/N 28 week examination
29627	A/N 30 week examination
13171	A/N 32 week examination
13170	A/N 34 week examination
29727	A/N 35 week examination
29610	A/N 36 week examination
26552	A/N 37 week examination
26553	A/N 38 week examination
26551	A/N 39 week examination
29280	A/N 40 week examination
37029	A/N 41 week examination
55605	A/N 42 week examination
3517	A/N booking examination
13984	Antenatal ultrasound confirms ectopic pregnancy
12260	A/N Rh antibody screen
68089	A/N Rh antibody screen NOS
70616	A/N sickle cell screen done
102099	A/N sickle cell screen NOS
64141	A/N syphilis screen-blood sent
14086	A/N U/S scan abnormal
27057	A/N U/S scan for ? abnormality
64537	A/N U/S scan for slow growth
37221	A/N U/S scan normal +? dates
35826	A/N U/S scan normal += dates
106588	Antenatal 22 week examination
106923	Antenatal 25 week examination
106425	Antenatal 31 week examination
13168	Antenatal examination NOS
10056	Antenatal examinations
13416	Antenatal sickle cell screen
13417	Antenatal syphilis screen
42326	Antenatal syphilis screen NOS
13968	Antenatal ultrasound confirms intra-uterine pregnancy

2029	Antenatal ultrasound scan
27056	Antenatal ultrasound scan at 17-22 weeks
39611	Antenatal ultrasound scan at 22-40 weeks
14084	Antenatal ultrasound scan at 9-16 weeks
14083	Antenatal ultrasound scan NOS
14085	Antenatal ultrasounds scan at 4-8 weeks
12890	Confirmation of pregnancy
50546	Dating scan
9462	Dating/booking US scan
100164	Detailed structural scan
103741	Doppler ultrasound scan of middle cerebral artery of fetus
102885	Doppler ultrasound scan of umbilical artery
95166	Doppler ultrasound scan of uterine artery
46126	Double test
13414	Downs screen - blood test
38358	Downs screen blood test abnormal
34508	Downs screen blood test normal
64832	Downs screening - blood sent
39173	Downs screening blood test NOS
103893	Fetal ascites scan
19720	Fetal monitoring
19590	Fetal movements felt
55493	Fetal movements seen
53420	Fetal tachycardia
9164	Fetal U-S scan
31110	Fundal height equal to dates
25875	Fundal height high for dates
37039	Fundal height low for dates
37038	Girth of pregnant abdomen
91773	Good baseline variability in fetal heart rate
105992	Height of uterine fundus
92171	Mid trimester scan
85992	Non routine obstetric scan for fetal observations
95875	Non routine obstetric scan for fetal observations NOS
38846	Normal fetal heart baseline pattern
13997	Nuchal scan
95881	O/E - fetal heart < 40
101119	O/E - fetal heart > 200
68996	O/E - fetal heart 100-120
26707	O/E - fetal heart 120-160
62903	O/E - fetal heart 160-180
62898	O/E - fetal heart 180-200
72837	O/E - fetal heart 40-80
70856	O/E - fetal heart 80-100
7681	O/E - fetal heart heard
22815	O/E - fetal movements
25153	O/E - fetal movements felt
52857	O/E - fetal movements NOS
53687	O/E - fetal movements seen
27801	O/E - fetal movemnt.diminished
26710	O/E - fetal presentation

67186	O/E - fetal presentation NOS
69819	O/E - fetal station NOS
24701	O/E - fetus very active
26708	O/E - fundal size = dates
37049	O/E - fundus = term size
26705	O/E - fundus 12-16 week size
37051	O/E - fundus 16-20 week size
26704	O/E - fundus 20-24 week size
26709	O/E - fundus 24-28 week size
30802	O/E - fundus 28-32 week size
30803	O/E - fundus 32-34 week size
26703	O/E - fundus 34-36 week size
26706	O/E - fundus 36-38 week size
13318	O/E - fundus size - obstetric
30804	O/E - gravid uterus size
62897	O/E - gravid uterus size NOS
37180	O/E - lie of fetus
29788	O/E - multiple presentation
63024	O/E -fetal presentation unsure
37050	O/E -fundus 38 weeks-term size
49519	Observation of position of pregnancy
12625	Obstetric monitoring
44173	Obstetric X-ray - fetus
56727	Obstetric X-ray - placenta
85951	Other non routine obstetric scan NOS
96343	Other specified routine obstetric scan
13165	Patient currently pregnant
127	Patient pregnant
14899	Patient pregnant NOS
38669	Placenta U-S scan
9986	Pregnancy care
4536	Pregnancy confirmed
15338	Pregnancy unplanned ? wanted
14877	Pregnant - ? planned
30817	Pregnant - blood test confirms
51298	Pregnant - on abdom. palpation
20240	Pregnant - planned
16215	Pregnant - urine test confirms
35592	Pregnant - V.E. confirms
10173	Pregnant abdomen observation
15567	Pregnant -unplanned-not wanted
107698	Pregnant uterus displaced laterally
32975	Pregnant, diaphragm failure
29692	Pregnant, IUD failure
14994	Pregnant, sheath failure
11989	Referral for termination of pregnancy
2278	Requests pregnancy termination
69815	Rh screen - 1st preg. sample
29623	Rh screen - 2nd preg. sample
109416	Rh screen - 3rd preg. sample
93946	Rhesus detailed scan
86011	Routine obstetric scan

85245	Routine obstetric scan NOS
6095	Seen in antenatal clinic
29205	Serum pregnancy test positive
70845	Sinusoidal pattern of fetal heart
27614	Triple test
39218	Ultrasonic doppler for fetal heart sounds
19800	Ultrasound in obstetric diagn.
12837	Ultrasound monitoring of early pregnancy
13965	Ultra-sound scan - obstetric
3030	Urine pregnancy test positive
2382	U-S obstetric diagn. scan NOS
29685	U-S obstetric scan abnormal
4797	U-S obstetric scan normal
45963	U-S scan - fetal abnormality
72159	U-S scan - fetal cephalometry
42093	U-S scan - fetal maturity
41919	U-S scan - fetal presentation
41937	U-S scan - multiple fetus
35558	U-S scan - obstetric, diagn.
68858	U-S scan -placental localisatn
67047	Viability scan
37147	Viability US scan
10306	Weeks pregnant

Appendix 12: Table S11, Outcome Group Combinations

Within conflicting pairs combinations of outcome groups will be coded as follows:

Outcome Group combination	*Variable Code*
1 1 (Loss- Loss)	1
1 2 (Loss- Delivery)	2
1 3 (Loss- Unknown)	3
2 2 (Delivery- Delivery)	4
2 3 (Delivery- Unknown)	5
3 3 (Unknown- Unknown)	6

Appendix 13: Table S12, Read codes for Antenatal scan

medcode	read_oxmis_code	Read term
2029	62G..00	Antenatal ultrasound scan
13965	584..13	Ultra-sound scan - obstetric
9462	584A.00	Dating/booking US scan
2382	584Z.00	U-S obstetric diagn. scan NOS
13997	584G.00	Nuchal scan
42093	5846	U-S scan - fetal maturity
37147	584B.00	Viability US scan
4797	5842	U-S obstetric scan normal
27019	5841	U-S obstetric scan requested
9164	584..11	Fetal U-S scan
14083	62GZ.00	Antenatal ultrasound scan NOS
35826	62G6.00	A/N U/S scan normal += dates
14084	62GC.00	Antenatal ultrasound scan at 9-16 weeks
35558	584..12	U-S scan - obstetric, diagn.
50546	7F26000	Dating scan
29012	7F27300	Nuchal translucency scan
27056	62GD.00	Antenatal ultrasound scan at 17-22 weeks
39611	62GE.00	Antenatal ultrasound scan at 22-40 weeks
47415	62G5.00	A/N U/S scan awaited
37220	62G2.00	A/N U/S scan offered
14085	62GB.00	Antenatal ultrasounds scan at 4-8 weeks
29685	5843	U-S obstetric scan abnormal
72159	5845	U-S scan - fetal cephalometry
45963	5847	U-S scan - fetal abnormality
27057	62G9.00	A/N U/S scan for ? abnormality
41919	5849	U-S scan - fetal presentation
30885	62G4.00	A/N U/S scan wanted
86011	7F26.00	Routine obstetric scan
68858	5844	U-S scan -placental localisatn
67047	7F26100	Viability scan
41937	5848	U-S scan - multiple fetus
14086	62G8.00	A/N U/S scan abnormal
85992	7F27.00	Non routine obstetric scan for fetal observations
37221	62G7.00	A/N U/S scan normal +? dates
38669	5844.11	Placenta U-S scan
78449	7F28.00	Other non routine obstetric scan
100164	7F27100	Detailed structural scan
92171	7F26200	Mid trimester scan
95166	7F2A111	Doppler ultrasound scan of uterine artery
64537	62GA.00	A/N U/S scan for slow growth

47116	7F28000	Placental localisation scan
85245	7F26z00	Routine obstetric scan NOS
102885	7F2A011	Doppler ultrasound scan of umbilical artery
96343	7F26y00	Other specified routine obstetric scan
95875	7F27z00	Non routine obstetric scan for fetal observations NOS
85951	7F28z00	Other non routine obstetric scan NOS
98261	7F27y00	OS non routine obstetric scan for fetal observations
95698	7F28y00	Other specified other non routine obstetric scan

Appendix 14: Table S13, DID Snomed fetal scan codes

Dating/booking ultrasound scan (procedure)	169229007
Fetal anatomy study (procedure)	271442007
Fetal biophysical profile (procedure)	21623001
Fetal echocardiography (procedure)	433235006
Magnetic resonance imaging of multiple pregnancy (procedure)	450825001
Placental localization (procedure)	164817009
Ultrasonography of multiple pregnancy for fetal anomaly (procedure)	445866007
Ultrasonography of multiple pregnancy for fetal nuchal translucency (procedure)	446810002
Ultrasound scan for amniotic fluid volume (procedure)	241494004
Ultrasound scan for fetal growth (procedure)	241493005

Appendix 15: Table S14, Number of episodes with a suitably timed outcome in linked HES data

Dataset in which evidence of a suitably timed pregnancy outcome was found.	**N pregnancy episodes where evidence of an outcome was found (% of episodes which were eligible for this linked data source)**	**N pregnancy episodes which were during current registration and UTS follow up**	**Total number of pregnancy episodes with recorded outcome missing which were eligible for HES linkage to each source**
HES Diagnosis (Part of HES APC)	24,902 (5.9%)	16,389 (65.8%)	424,375
HES Maternity (Part of HES APC)	163,483 (38.5%)	109,393 (66.9%)	424,375
HES Procedures (Part of HES APC)	201,731 (47.5%)	133,077 (66.0%)	424,375
HES Episodes (Part of HES APC)	185,436 (43.7%)	122,350 (66.0%)	424,375
HES Outpatient	735 (0.2%)	560 (76.2%)	311,982
Any HES Source	211,070 (49.7%)	139,084 (65.9%)	424,375

Appendix 16: Table S15, Numbers of pregnancy episodes with recorded outcome missing which were within practice UTS follow-up and patient's current registration period that were consistent with applied criteria for each scenario

Scenario	**Description**	**N pregnancy episodes which meet this scenario (% of total episodes with missing outcome)**	**N pregnancy episodes which only meet this scenario (% of the total episodes with missing outcome)**	**N pregnancy episodes with evidence of an outcome in linked HES (% of linkage eligible episodes)**
Denominator		475,664	475,664	265,264
Problem 1: The women was pregnant at the time of the database record, but the outcome was not captured in CPRD primary care data.				
Scenario 1a	The pregnancy outcome occurred in hospital or elsewhere and information wasn't fed back to the practice.	139,084 (29.2%)	1,825 (0.4%)	139,084 (52.4%)
Scenario 1b	The outcome of the pregnancy is recorded in the primary care data but has no event date associated with it.	475 (0.1%)	28 (0.0%)	113 (0.0%)
Scenario 1c	The pregnancy occurred before the patient was registered at the practice or before UTS	-	-	-

Problem 2: The women was pregnant at the time of the database record, but the pregnancy was still ongoing at the end of available follow up in the database.				
Scenario 2a	The patient transferred out before the putative end of pregnancy	117,571 (24.7%)	34,659 (7.3%)	52,601 (19.8%)
Scenario 2b	The last collection date of the practice was before the putative end of pregnancy	58,698 (12.3%)	20,122 (4.2%)	21,702 (8.2%)
Problem 3: The patient was not pregnant at the time of the database record.				
Scenario 3a	Episode is derived from historical pregnancy information recorded in the first few months after the patient joined the practice	3,875 (0.8%)	386 (0.1%)	1,271 (0.5%)
Scenario 3b	Patient asks for advice whilst planning a pregnancy	0 (0.0%)	0 (0.0%)	0 (0.0%)
Problem 4: The pregnancy record belongs to another pregnancy episode in the Register.				
Scenario 4a	Delay in recording the outcome of a pregnancy, algorithm calculates LMP too late and uncovers records at the beginning of pregnancy creating this pregnancy episode with no outcome recorded.	35,255 (7.4%)	8,265 (1.7%)	14,402 (5.4%)

Scenario 4b	The LMP is derived from the data and is wrong resulting in early codes being uncovered creating this episode	17,110 (3.6%)	3,715 (0.8%)	6,651 (2.5%)
Scenario 4c	The LMP has been shifted backwards uncovering records at the end of the pregnancy	0 (0.0%)	0 (0.0%)	0 (0.0%)
Scenario 4d	A code recorded relating to the patient's delivery history is incorrectly identified by the algorithm as a delivery uncovering records at the end.	219,505 (46.1%)	109,161 (22.9%)	65,883 (24.8%)
Scenario 4e	The outcome of the pregnancy episode has been misclassified as antenatal	18,222 (3.8%)	7,418 (1.6%)	3,990 (1.5%)
Pregnancy Episodes which didn't meet any scenario	These pregnancy episodes did not meet the criteria for any identified scenarios.	94,769 (19.9%)	0 (0.0%)	0 (0.0%)

Appendix 17: Table S16, Numbers of conflicting pregnancy episodes which were within practice UTS follow-up and patient's current registration period that were consistent with applied criteria for each scenario

Scenario	Description	N pregnancy pairs (% of total conflicting pregnancy pairs)	N which only fit this scenario (% of the total pairs meeting this scenario)	N of pairs with a linked baby in the MBL (% of the total pairs meeting this scenario)	N pairs with evidence of pregnancy in linked HES
Denominator		144,670	144,670	144,670	93,100
Problem 1: Both pregnancies are true but one is a current pregnancy and one is a historical pregnancy					
Scenario 1a	The GP records a past delivery or loss during a current pregnancy with the same outcome resulting in another episode being created	1,981 (1.4%)	317 (0.2%)	1,782 (1.2%)	1,875 (2.0%)
Scenario 1b	A patient has a record relating to a loss recorded during a pregnancy ending in delivery or vice-versa. Conflicting episodes are generated by the algorithm	31,526 (21.8%)	15,453 (10.7%)	8,275 (5.7%)	11,410 (12.3%)
Problem 2: Both pregnancies are historical					

Scenario 2a	A patient has information on historical pregnancies recorded with the current date rather than the actual date.	12,557 (8.7%)	0 (0.0%)	97 (0.1%)	4,309 (4.6%)
Problem 3: Both pregnancies are true and current but the gestation of the second pregnancy estimated by the algorithm is too long.					
Scenario 3a	The woman has two losses which are >8weeks and <12weeks apart.	2,284 (1.6%)	3 (0.0%)	0 (0.0%)	635 (0.7%)
Scenario 3b	The woman has two pregnancies close together and the second ends in delivery. If the LMP information is wrong for this pregnancy, then algorithm episodes may overlap.	13,464 (9.3%)	2,387 (1.6%)	1,113 (0.8%)	4,502 (4.8%)
Problem 4: : The pregnancy is true and current but is split into separate episodes by the rules of the algorithm					
Scenario 4a	The GP records further information about a pregnancy outcome >25 weeks later for deliveries or >8weeks <12 weeks later for losses.	2,347 (1.6%)	183 (0.1%)	2,155 (1.5%)	2,255 (2.4%)

Scenario 4b	The GP records further antenatal information after the end of a pregnancy. Conflicting episodes are generated by the algorithm	27,131 (18.8%)	25,097 (17.3%)	11,097 (7.7%)	11,668 (12.5%)
Scenario 4c	The patient has a follow up scan after a pregnancy loss. The scan is recorded in the data as an antenatal scan, a conflicting episode is then generated by the algorithm.	2,088 (1.4%)	0 (0.0%)	0 (0.0%)	587 (0.6%)
Scenario 4d	The GP records information about a pregnancy but no outcome with >6 weeks between records. If the second episode has gestational information the start may be assigned before the start of the first episode.	9,911 (6.9%)	9,911 (6.9%)	0 (0.0%)	5,079 (5.5%)
Scenario 4e	The pregnancy dates have been shifted backwards by the rules of the algorithm leaving uncovered records. Conflicting episodes are generated by the algorithm.	55,205 (38.2%)	53,044 (36.7%)	43,945 (30.4%)	33,057 (35.5%)
None	These pairs of pregnancies did not meet the criteria for any identified scenarios.	15,650 (10.8%)	-	8,921 (6.2%)	8,235 (8.8%)

Appendix 18: Table S17, Number of conflicting episode pairs by outcome combination

Outcome Combination	*N pairs (% of total conflicting pairs)*
two losses	65,826 (26.2%)
one loss one delivery	73,222 (29.2%)
one loss one unknown	62,776 (25.0%)
two deliveries	10,204 (4.1%)
one delivery one unknown	24,303 (9.7%)
two unknowns	14,695 (5.9%)
Total Pairs	251,026 (100%)

3.4 Chapter Summary

Investigating Uncertain Pregnancy Episodes

- Whilst the CPRD Pregnancy Register represents an advancement in our ability to use EHR data to study pregnancy there remain questions around how to utilise uncertain pregnancy data.
- This work outlines the details of 12 scenarios which may result in pregnancy episodes with no recorded outcome and 10 scenarios which may result in conflicting pregnancy episodes in the CPRD Pregnancy Register.
- Criteria were outlined and applied to the data to investigate the frequency with which evidence of these scenarios occurred.
- Linked secondary care data was investigated to see whether any further evidence could be found.
- Recommendations as to the implications of excluding or including pregnancy episodes meeting each scenario are given dependent on the purpose of the research being conducted.
- This work shows evidence that most uncertain pregnancy episodes are consistent with true and current pregnancies for which the data contain valuable information.
- As with any algorithmic approach this work is limited by the fact that underlying assumptions may not be true in all cases.
- This work represents a valuable advancement in our understanding of pregnancy records in EHR data and provides useful advice for researchers.
- This work was approved by the MHRA's Research Data Governance Committee (RDG) (formerly known as the Independent Scientific Advisory Committee, ISAC) protocol numbers 17_285R2 (Appendix 2) and 19_140 (Appendix 3). It was also approved by London School of Hygiene and Tropical Medicine Ethics Online (LEO) committee reference numbers 14660 and 17244.

Chapter 4: A Systematic Review of COVID-19 and Pregnancy Loss

4.1 Introduction

In order to put into practice the methodological recommendations developed and outlined in Chapter 3 an epidemiological question considering a pregnancy exposure was needed. Given the timing of my PhD research coincided with the COVID-19 pandemic I chose to investigate whether having COVID-19 during pregnancy was associated with an increased risk of pregnancy loss. I began by conducting a systematic review of the evidence to date on this question. The methods for this systematic review were published apriori in the *BMJ Open*, this article and accompanying supplementary information are presented below (section 4.2 and 4.3). The results of the systematic review have been submitted to the *BMJ Open* and this submitted paper and its supplementary files are presented in section 4.4 and 4.5.

The review looked for evidence as to whether COVID-19 in pregnancy increases the risk of pregnancy loss including both miscarriage and stillbirth. In order to ensure that the latest available research was gathered both published and pre-print studies were included. The review included all studies which attempted to quantitatively assess this question and included a COVID-19 free comparison group.

4.2 Published Paper: Systematic Review Protocol

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	71130	**Title**	Mrs
First Name(s)	Jennifer		
Surname/Family Name	Campbell		
Thesis Title	Investigation and Application of a Pregnancy Register Based on Electronic Primary Care Data		
Primary Supervisor	Krishnan Bhaskaran		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open (Open Access please see the following link for copright approval https://www.bmj.com/company/wp-content/uploads/2019/03/Author-Permissions-Policy.pdf)		
When was the work published?	03/10/2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This protocol was written by JC with co-authors performing critical review.

SECTION E

Student Signature	[redacted]
Date	01/03/23

Supervisor Signature	[redacted]
Date	7/3/23

BMJ Open

COVID-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review protocol

Jennifer Campbell ,[1,2] Rachael Williams,[2] Mia Harley,[1] Krishnan Bhaskaran[1]

To cite: Campbell J, Williams R, Harley M, *et al*. COVID-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review protocol. *BMJ Open* 2022;**12**:e065588. doi:10.1136/bmjopen-2022-065588

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-065588).

Received 10 June 2022
Accepted 23 September 2022

[1]NCDE, London School of Hygiene & Tropical Medicine Faculty of Epidemiology and Population Health, London, UK
[2]Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, London, UK

Correspondence to
Jennifer Campbell;
jennifer.campbell@mhra.gov.uk

ABSTRACT

Introduction The COVID-19 pandemic has led to concerns about potential adverse pregnancy outcomes associated with infection, resulting in intensive research. Numerous studies have attempted to examine whether COVID-19 is associated with an increased risk of pregnancy loss. However, studies and reviews to date have drawn differing conclusions. The aim of this systematic review is to provide a summary of all quantitative research on the relationship between pregnancy loss and COVID-19 infection and, if appropriate, to synthesise the evidence into an overall effect estimate.

Methods and analysis Three publication databases (Embase, PubMed and Cochrane) and four preprint databases (medRxiv, Lancet Preprint, Gates Open Research and Wellcome Open Research) will be searched. Boolean logic will be used to combine terms associated with pregnancy loss and COVID-19. The population of interest are pregnant women. Retrieved results will be assessed in two phases: (1) abstract screening and (2) full text evaluation. All studies which compare pregnancy loss outcomes in women who had COVID-19 versus those who did not quantitatively will be included. Narrative and non-English studies will be excluded. Two reviewers will screen independently, with results compared and discrepancies resolved by the study team. Study quality and risk of bias will be assessed using a quality appraisal tool. Results will be summarised descriptively and where possible synthesised in a meta-analysis.

Ethics and dissemination This systematic review requires no ethical approval. This review will be published in a peer-reviewed journal and provide an important update in a rapidly evolving field of research.

PROSPERO registration number CRD42022327437.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This systematic review will include both published and preprint studies in an attempt to capture the very latest data and minimise publication bias.
⇒ Study selection, data extraction and quality assessment will be performed independently by two researchers, which will ensure that all relevant studies are included without personal biases.
⇒ All included studies will be assessed for quality using the National Institute for Health and Care Excellence quality appraisal checklist for quantitative studies reporting correlations and associations.
⇒ Studies which are not published in English will not be included. This limitation may cause language bias.

BACKGROUND

SARS-CoV-2 emerged as a new coronavirus at the end of 2019 spreading rapidly to cause a global pandemic of its associated illness COVID-19. Many millions of people around the world have been infected with the virus including pregnant women. However, due to the novelty of COVID-19 little is known about its potential effect on the unborn fetus and pregnancy outcomes. Aetiological hypotheses have been proposed as to ways in which COVID-19 may adversely affect pregnancy outcomes including potential increased risk of loss mediated by placental damage.[1] COVID-19 in pregnancy has therefore been the subject of intense research and there have been numerous studies which have examined any potential adverse effect leading to reviews which have attempted to summarise the evidence.[2–4]

As both the virus itself and our knowledge of its effects are constantly evolving both studies and reviews to date have drawn differing conclusions. Some have concluded an increased risk of pregnancy loss associated with COVID-19 infection[2 5–9] while others have concluded no increased risk.[10–13] Many early reviews of this question included only case reports as no comparative studies were available.[4 14–16] The latest published systematic review on this question by Pathirathna *et al* included studies published prior to June 2021 just over 1 year into the COVID-19 pandemic and like all reviews to date on this topic they noted the need for further research.[8] Since this review, there have been numerous additional studies published and there has been a global roll-out of vaccinations for COVID-19 to pregnant women. It is therefore important that we continue to review all emerging evidence in order to provide a full and current picture of any potential adverse risk.

Table 1 Inclusion and exclusion criteria

Inclusion	Exclusion
► Epidemiological studies which attempt to quantitatively assess any association between pregnancy loss and COVID-19. (Study designs may include prospective and retrospective cohort studies, case–control studies and cross-sectional studies.)	► Non-English language publications including those where the summary is in English but not the full text. ► Narrative review articles, guidelines, editorials or comments. ► Studies without a control or comparison group, for example, case reports. ► Conference presentations.

The overall aim of this study is to identify and summarise all studies to date which have quantitatively compared pregnancy loss outcomes in women who contracted COVID-19 while pregnant versus those who did not. Where possible, quantitative estimates of associations between COVID-19 and pregnancy loss will be synthesised into an overall effect estimate.

METHODS

Study registration

This protocol is prepared in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols statement (online supplemental appendix 1).[17] This protocol is registered on the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42022327437).

Eligibility criteria

The review will include all studies which have attempted to quantitatively assess the potential association between having COVID-19 during pregnancy and pregnancy loss.

The population of interest are pregnant women at any maternal age or gestation of pregnancy. The exposure of interest will be COVID-19 during pregnancy. We will include all studies which attempt to ascertain COVID-19 exposure in pregnancy regardless of the method of diagnosis. The comparator population will be women who did not have COVID-19 during pregnancy. The outcome of interest will be pregnancy loss (miscarriage or stillbirth).

Table 1 gives the inclusion and exclusion criteria that will be applied to identified studies.

Information sources

Publication databases to be searched: Embase (Ovid), PubMed, Cochrane.

Table 2 Database search strategy

Database	Dates of search coverage	Miscarriage/stillbirth	COVID-19
PubMed	1 March 2020 to current date	'Abortion, Spontaneous' [MeSH] OR 'Fetal Death' [MeSH] OR 'Stillbirth' [MeSH] OR (miscarriage [MeSH Terms])) OR (miscarriages [MeSH Terms] OR Miscarriage* OR pregnancy loss* OR spontaneous abortion* OR fetal loss* OR foetal loss* OR foetal death* OR fetal death*	'coronavirus' [MeSH] OR 'coronavirus infections' [MeSH Terms] OR 'coronavirus' [All Fields] OR 'covid 2019' [All Fields] OR 'SARS2' [All Fields] OR 'SARS-CoV-2' [All Fields] OR 'SARS-CoV-19' [All Fields] OR 'severe acute respiratory syndrome coronavirus 2' [supplementary concept] OR 'coronavirus infection' [All Fields] OR 'severe acute respiratory pneumonia outbreak' [All Fields] OR 'novel cov' [All Fields] OR '2019ncov' [All Fields] OR 'sars cov2' [All Fields] OR 'cov22' [All Fields] OR 'ncov' [All Fields] OR 'covid19' [All Fields] OR 'covid 19' [All Fields] OR 'covid-19' [All Fields] OR 'coronaviridae' [All Fields] OR 'corona virus' [All Fields]
Embase	1 March 2020 to current date	spontaneous abortion/exp OR stillbirth/exp OR stillbirth.m.p OR pregnancy loss/exp OR pregnancy loss.mp OR foetal death.m.p OR fetus death OR fetus death/exp NOT [medline]/lim	'coronavirinae'/exp OR 'coronavirinae' OR 'coronaviridae infection'/exp OR 'coronaviridae infection' OR 'coronavirus disease 2019'/exp OR 'coronavirus'/exp OR coronavirus OR 'coronavirus infection'/de NOT [medline]/lim
Cochrane	1 March 2020 to current date	Search for 'stillbirth' OR 'miscarriage' OR 'foetal death rates' OR 'foetal death rate' OR 'fetal death' OR 'fetal death rate' OR 'pregnancy loss rate' OR 'pregnancy loss-rate' OR pregnancy 'loss-rates'	Search for 'coronavirus' in the Title Abstract Keyword fields

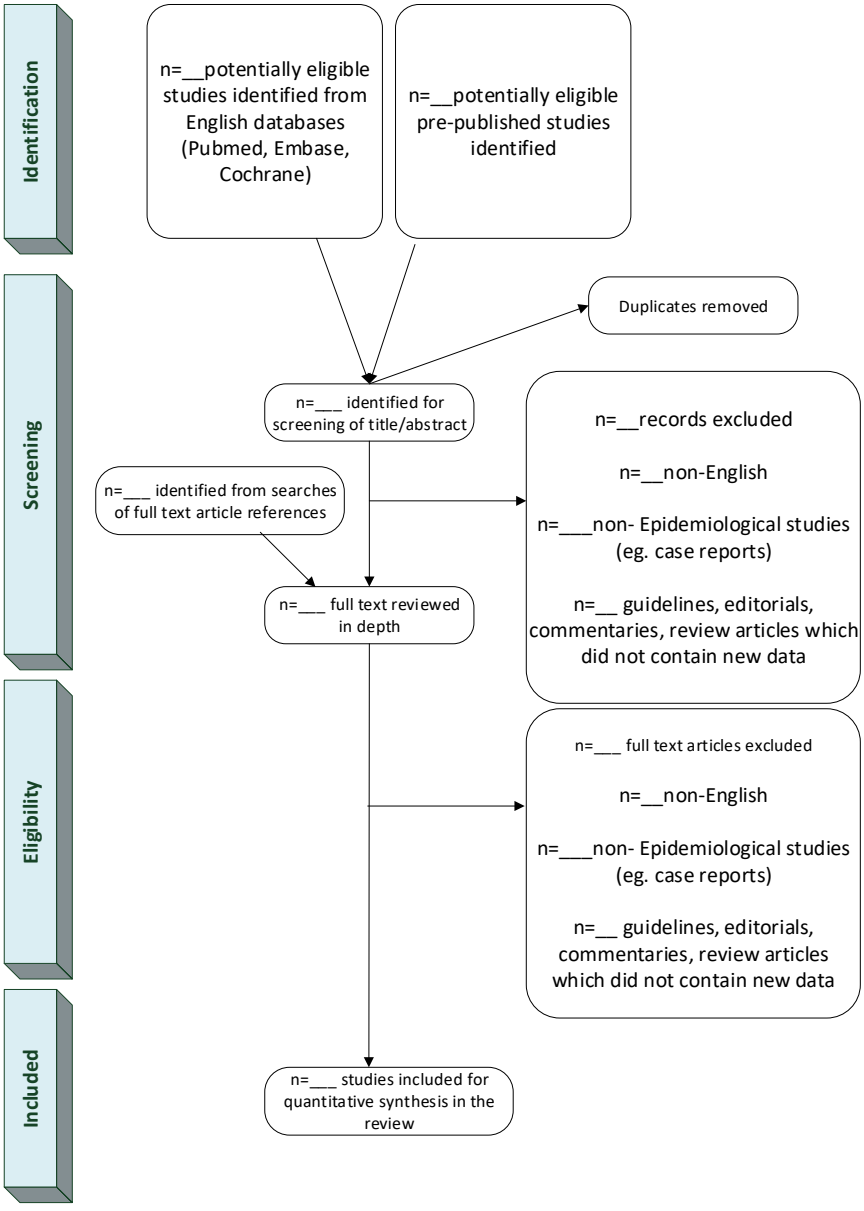


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of study selection process.

Preprint platforms to be searched: medRxiv, Lancet Preprint, Gates Open Research, Wellcome Open Research.

Search strategy

Search terms listed in table 2 will be applied in the respective databases. Terms related to pregnancy loss will be combined with terms related to COVID-19 using AND logic. Only publications after 1 March 2020 will be searched.

To further increase the sensitivity of our search, the list of references from review articles relating to COVID-19 and pregnancy loss will be screened manually to identify other potentially eligible articles.

Due to the fast-moving nature of COVID-19 research we will also search databases of preprint articles.[18] The medRxiv database will be searched via Embase using the search terms detailed above. The Lancet Preprint database will be searched for Obstetrics and Gynaecology articles which contain the term 'Covid-19'. Gates Open Research and Wellcome Open Research will also be searched for 'Covid-19' and 'Pregnancy'. Preprint databases were selected from a systematic examination of preprint platforms by Kirkham *et al*.[19] Preprint articles will be flagged as such in any presentation of results.

Data management and selection process

Searches will be performed across all databases by reviewer 1. Records of the search terms, results from the search and the date of last run will be saved. Results will be exported into Mendeley where any duplicate results will be removed.[20] Each article will be given a study ID. The remaining articles will be screened for eligibility based on titles and abstracts by two independent reviewers applying the inclusion/exclusion criteria described

Table 3 Example of data collection form

Study ID	First author, year	Study design	Location	Exposure definition	Outcome definition	Subjects (n)	Exposed (n)	Miscarriage among the exposed (n)	Stillbirth among the exposed (n)	Miscarriage among the unexposed (n)	Stillbirth among the unexposed (n)	Statistical measure and result reported in the paper	Was the study before or after vaccine roll-out?
-													

above. Discrepancies will be discussed and, where necessary, will be decided by the whole study team. Full text articles will be obtained for all articles deemed eligible for inclusion from the initial screening. Articles will be divided and assessed independently by two reviewers after which the final selection will be agreed. Any reasons for exclusion will be recorded. The study selection process is outlined in figure 1.

Data collection process

The example data capture form (table 3) will be pilot tested on a random sample of five included studies and revised if necessary. The finalised data capture form will then be completed by reviewers 1 and 2 independently for a sample of 10 studies to check concordance, after which each study will be examined by one reviewer.

Assessment of study quality

All included studies will be assessed for bias by reviewers using an adapted version of the National Institute for Health and Care Excellence (NICE) quality appraisal checklist for quantitative studies reporting correlations and associations (online supplemental appendix 2).[21] The NICE tool was chosen as it is designed for identifying rigour in observational studies that explore and generate hypotheses about causal relationships and can be used for multiple study designs. The NICE tool consists of five major items: study population and participants; selection and methods; outcomes; analysis; and summary.

Appraisal will be done using an Excel format to allow for easy compilation of responses. Decisions will be discussed and any discrepancies resolved. Each study will then be awarded an overall study quality grade for external and internal validity from one of the three categories below which are based on the checklist criteria (online supplemental appendix 1).

- ++All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
- +Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- – Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Studies deemed to be low quality (category) will be excluded from any meta-analysis.

Data synthesis

We will use Higgins and Thompson's I^2 statistic to quantify heterogeneity, and if I^2 is >50% meta-analysis will be conducted in Stata using a random-effects model.[22] Where meta-analysis is attempted funnel plots will be used to assess publication bias.[23] Where statistical pooling is not possible, findings will be presented in narrative form using tables to aid in data presentation. If possible, we will conduct subgroup analyses of studies reporting miscarriage and stillbirth separately. We will also look at any potential impact of the widespread use of COVID-19 vaccines by grouping studies into those conducted before and after vaccine roll-out if possible. We will use 1 March 2021 as the cut-off date for studies considered to be post-vaccine roll-out. For studies after this date we will examine the national vaccine roll-out programme for the country in which the study was conducted to assess the likelihood that pregnant women within the study would have been vaccinated. We will also consider a subgroup analysis of hospitalised versus non-hospitalised COVID-19 cases if there are enough studies which consider this.

Patient and public involvement

There will be no patient or public involvement in this project.

DISCUSSION

The COVID-19 pandemic has been a challenging time for pregnant women, knowledge on the potential risks of infection to them and their unborn babies is ever evolving. With COVID-19 now circulating widely in many countries and limited risk reduction measures in place it is important to try and fully understand the risks so that pregnant women can be advised appropriately. Reviews and studies to date on whether COVID-19 increases the risk of pregnancy loss have drawn mixed conclusions.[2–4 8 13] COVID-19 research is a fast-moving area; therefore, it is important that reviews are regularly updated. This systematic review aims to provide a comprehensive overview of the latest evidence.

COVID-19 research moves very quickly, and preprint literature has become a key outlet for new research with many researchers opting to make their work available as quickly as possible. Including prepublications in this review, something which previous reviews have not done, will allow us to obtain as current a picture as possible of all

of the evidence. Inclusion of preprint literature may also help mitigate any risk of publication bias.

Vaccination against COVID-19 became widely available globally in 2021.[24] In the UK, pregnant women have been routinely advised to receive COVID-19 vaccination together with the rest of the population, according to their age and underlying health conditions since 16 April 2021.[25] The widespread introduction of COVID-19 vaccination may have led to a decrease in potential risk or pregnancy loss. We hope to identify enough studies to allow us to examine separately those which were conducted before and after the vaccination roll-out in order to provide an insight into any impact the vaccine may have had.

The results of this review can be used to inform public health messaging for pregnant women around the potential risks of COVID-19 infection. This research will also help inform any future research studies planned on this question.

Contributors This protocol was written by JC with KB, RW and MH performing critical review. JC will act as the guarantor of the review.

Funding This work is funded by Wellcome (Senior Research Fellowship for KB, grant number: 220283/Z/20/Z). This work forms part of JC's PhD which is funded by the Clinical Practice Research Datalink, a division of the UK Medicines and Healthcare Products Regulatory Agency.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Jennifer Campbell http://orcid.org/0000-0002-0684-4437

REFERENCES

1 Arthurs AL, Jankovic-Karasoulos T, Roberts CT. COVID-19 in pregnancy: what we know from the first year of the pandemic. *Biochim Biophys Acta - Mol Basis Dis* 202;1867.
2 Kazemi SN, Hajikhani B, Didar H, *et al*. COVID-19 and cause of pregnancy loss during the pandemic: a systematic review. *PLoS One* 2021;16:e0255994.
3 Soheili M, Moradi G, Baradaran HR, *et al*. Clinical manifestation and maternal complications and neonatal outcomes in pregnant women with COVID-19: a comprehensive evidence synthesis and meta-analysis. *J Matern Fetal Neonatal Med* 2021;22:1–14.
4 Juan J, Gil MM, Rong Z, *et al*. Effect of coronavirus disease 2019 (COVID-19) on maternal, perinatal and neonatal outcome: systematic review. *Ultrasound Obstet Gynecol* 2020;56:15–27.
5 Hazari KS, Abdeldayem R, Paulose L, *et al*. Covid-19 infection in pregnant women in Dubai: a case-control study. *BMC Pregnancy Childbirth* 2021;21:1–10.
6 González Rodríguez L, Oreja Cuesta AB, Pardo Pumar MI, *et al*. SARS-CoV-2 infection in early first-trimester miscarriages: a prospective observational study. *Reprod Biomed Online* 2022;44:127–30.
7 Webster WS, Freeman JAD. Prescription drugs and pregnancy. *Expert Opin Pharmacother* 2003;4:949–61.
8 Pathirathna ML, Samarasekara BPP, Dasanayake TS, *et al*. Adverse perinatal outcomes in COVID-19 infected pregnant women: a systematic review and meta-analysis. *Health Care* 2022;10:203.
9 Wei SQ, Bilodeau-Bertrand M, Liu S, *et al*. The impact of COVID-19 on pregnancy outcomes: a systematic review and meta-analysis. *Can Med Assoc J* 2021;193:E540–8.
10 Sacinti KG, Kalafat E, Sukur YE, *et al*. Increased incidence of first-trimester miscarriage during the COVID -19 pandemic. *Ultrasound Obstet Gynecol* 2021;57:1013–4.
11 Francesco D'Antonio MD, MDDM CS. *Maternal and perinatal outcomes in high compared to Working group on coronavirus disease 2019*, 2020.
12 Cosma S, Carosso AR, Cusato J, *et al*. Coronavirus disease 2019 and first-trimester spontaneous abortion: a case-control study of 225 pregnant patients. *Am J Obstet Gynecol* 2021;224:391.e1–391.e7.
13 Cavalcante MB, de Melo Bezerra Cavalcante CT, Cavalcante ANM, *et al*. COVID-19 and miscarriage: from immunopathological mechanisms to actual clinical evidence. *J Reprod Immunol* 2021;148:103382.
14 Medeiros KS, Sarmento ACA, Costa APF, *et al*. Consequences and implications of the coronavirus disease (COVID-19) on pregnancy and newborns: a comprehensive systematic review and meta-analysis. *Intl J Gynecology & Obste* 2022;156:394–405.
15 Bellos I, Pandita A, Panza R. Maternal and perinatal outcomes in pregnant women infected by SARS-CoV-2: a meta-analysis. *Eur J Obstet Gynecol Reprod Biol* 2021;256:194–204.
16 Banaei M, Ghasemi V, Saei Ghare Naz M, *et al*. Obstetrics and neonatal outcomes in pregnant women with covid-19: a systematic review. *Iran J Public Health* 2020;49:38–47.
17 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647–25.
18 Clyne B, Walsh KA, O'Murchu E, *et al*. Using preprints in evidence synthesis: commentary on experience during the COVID-19 pandemic. *J Clin Epidemiol* 2021;138:203–10.
19 Kirkham JJ, Penfold NC, Murphy F, *et al*. Systematic examination of preprint platforms for use in the medical and biomedical sciences setting. *BMJ Open* 2020;10:e041849.
20 Mendeley Ltd. Mendeley reference manager, 2020. Available: https://www.mendeley.com
21 NICE. Methods for the development of NICE public health guidance (third edition) Natl Inst Heal Clin Excell; 2012: 284. https://www.nice.org.uk/process/pmg4/resources/methods-for-the-development-of-nice-public-health-guidance-third-edition-pdf-2007967445701
22 StataCorp. *Stata statistical software: release 17*. College Station, TX: StataCorp LLC, 2021.
23 Egger M, Davey Smith G, Schneider M, *et al*. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.
24 Ritchie H, Mathieu E, Rodés-Guirao L. Coronavirus pandemic (COVID-19) OurWorldInData.org; 2020. https://ourworldindata.org/coronavirus
25 UK Health Security Agency. COVID-19: The Green Book chapter 14a. In: *The green book*, 2022. https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a%0A

4.3 Published Supplementary Files

Appendix 1: Table S1, Prisma-P checklist

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item
ADMINISTRATIVE INFORMATION		
Title:		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:		
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:		
Sources	5a	Indicate sources of financial or other support for the review
Sponsor	5b	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS		

Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records:		
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

Appendix 2: Table S2, NICE Quality appraisal checklist for quantitative studies reporting correlations and associations

Checklist items are worded so that 1 of 5 responses is possible: ++ Indicates that for that particular aspect of study design, the study has been designed or conducted in such a way as to minimise the risk of bias. + Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias for that particular aspect of study design. − Should be reserved for those aspects of the study design in which significant sources of bias may persist. Not reported (NR) Should be reserved for those aspects in which the study under review fails to report how they have (or might have) been considered. Not applicable (NA) Should be reserved for those study design aspects that are not applicable given the study design under review (for example, allocation concealment would not be applicable for case–control studies).

Study identification: Include full citation details		
Study design: • Refer to the glossary of study designs (appendix D) and the algorithm for classifying experimental and observational study designs (appendix E) to best describe the paper's underpinning study design		
Assessed by:		
Section 1: Population		
1.1 Is the source population or source area well described?	++	Comments:

• Was the country (e.g. developed or non-developed, type of health care system), setting (primary schools, community centres etc), location (urban, rural), population demographics etc adequately described?	+ – NR NA	
1.2 Is the eligible population or area representative of the source population or area? • Was the recruitment of individuals, clusters or areas well defined (e.g. advertisement, birth register)? • Was the eligible population representative of the source? Were important groups underrepresented?	++ + – NR NA	Comments:
1.3 Do the selected participants or areas represent the eligible population or area? • Was the method of selection of participants from the eligible population well described? • What % of selected individuals or clusters agreed to participate? Were there any sources of bias?	++ + – NR NA	Comments:

• Were the inclusion or exclusion criteria explicit and appropriate?		
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised? • How was selection bias minimised?	++ + – NR NA	Comments:
2.2 Was the selection of explanatory variables based on a sound theoretical basis? • How sound was the theoretical basis for selecting the explanatory variables?	++ + – NR NA	Comments:
2.3 How well were likely confounding factors identified and controlled?	++	Comments:

• Were there likely to be other confounding factors not considered or appropriately adjusted for? • Was this sufficient to cause important bias?	+ – NR NA	
2.4 Is the setting applicable to the UK? • Did the setting differ significantly from the UK?	++ + – NR NA	Comments:
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable? • Were outcome measures subjective or objective (e.g. biochemically validated nicotine levels ++ vs self-reported smoking −)?	++ + –	Comments:

• How reliable were outcome measures (e.g. inter- or intra-rater reliability scores)? • Was there any indication that measures had been validated (e.g. validated against a gold standard measure or assessed for content validity)?	NR NA	
3.2 Were the outcome measurements complete? • Were all or most of the study participants who met the defined study outcome definitions likely to have been identified?	++ + – NR NA	Comments:
3.3 Were all the important outcomes assessed? • Were all the important benefits and harms assessed? • Was it possible to determine the overall balance of benefits and harms of the intervention versus comparison?	++ + − NR NA	Comments:

3.4 Was there a similar follow-up time in exposure and comparison groups? • If groups are followed for different lengths of time, then more events are likely to occur in the group followed-up for longer distorting the comparison. • Analyses can be adjusted to allow for differences in length of follow-up (e.g. using person-years).	++ + – NR NA	Comments:
3.5 Was follow-up time meaningful? • Was follow-up long enough to assess long-term benefits and harms? • Was it too long, e.g. participants lost to follow-up?	++ + – NR NA	Comments:
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)?	++ +	Comments:

• A power of 0.8 (i.e. it is likely to see an effect of a given size if one exists, 80% of the time) is the conventionally accepted standard. • Is a power calculation presented? If not, what is the expected effect size? Is the sample size adequate?	− NR NA	
4.2 Were multiple explanatory variables considered in the analyses? • Were there sufficient explanatory variables considered in the analysis?	++ + − NR NA	Comments:
4.3 Were the analytical methods appropriate? • Were important differences in follow-up time and likely confounders adjusted for?	++ + − NR NA	Comments:

4.6 Was the precision of association given or calculable? Is association meaningful? • Were confidence intervals or p values for effect estimates given or possible to calculate? • Were CIs wide or were they sufficiently precise to aid decision-making? If precision is lacking, is this because the study is under-powered?	++ + – NR NA	Comments:
Section 5: Summary		
5.1 Are the study results internally valid (i.e. unbiased)? • How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? • Were there significant flaws in the study design?	++ + –	Comments:
5.2 Are the findings generalisable to the source population (i.e. externally valid)? • Are there sufficient details given about the study to determine if the findings are generalisable to the source population?	++ + –	Comments:

• Consider: participants, interventions and comparisons, outcomes, resource and policy implications.		

4.4 Research Paper Submitted for Publication

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	71130	**Title**	Mrs
First Name(s)	Jennifer		
Surname/Family Name	Campbell		
Thesis Title	Investigation and Application of a Pregnancy Register Based on Electronic Primary Care Data		
Primary Supervisor	Krishnan Bhaskaran		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open (Open Access please see the following link for copright approval https://www.bmj.com/company/wp-content/uploads/2019/03/Author-Permissions-Policy.pdf)		
When was the work published?	03/10/2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This protocol was written by JC with co-authors performing critical review.

SECTION E

Student Signature	[redacted]
Date	01/03/23

Supervisor Signature	[redacted]
Date	7/3/23

COVID-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review and meta-analysis of the evidence to date

Jennifer CAMPBELL MSc [1 & 2], Mia HARLEY MSc [1], Rachael WILLIAMS PhD [2], Krishnan BHASKARAN PhD [1]

1. *London School of Hygiene and Tropical Medicine (LSHTM)*
2. *Clinical Practice Research Datalink (CPRD)*

Correspondence to Jennifer Campbell; jennifer.campbell@mhra.gov.uk

Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, 10 South Colonnade, Canary Wharf, London, E14 4PU, United Kingdom

Structured Abstract (350-word limit)

Objective

To assimilate the evidence to date as to whether having COVID-19 during pregnancy is associated with a higher risk of pregnancy loss (miscarriage or stillbirth).

Data sources

Pubmed, Embase (Ovid), Cochrane, medRxiv, Lancet Preprint, Gates Open Research, Wellcome Open Research from 01/03/2020 to 11/05/2022.

Study eligibility criteria

Studies which attempt to quantitatively assess the association between having COVID-19 during pregnancy and pregnancy loss. Only studies which included a COVID-19 free control group were included.

Study appraisal and synthesis methods

Two reviewers independently identified studies for inclusion. Data was extracted by one reviewer who also assessed the risk of bias using an adapted National Institute for Clinical Excellence (NICE) quality appraisal tool. Studies were grouped according to the type of pregnancy loss outcomes included and how COVID-19 exposure was defined. Where no measures of association were presented, crude risk-ratios were calculated directly from outcomes and denominator information.

Following assessment for heterogeneity, meta-analysis was carried out using a random effects model on a subgroup of studies.

Results

Thirty-one studies were included, of which 7 investigated miscarriage, 21 investigated stillbirth and 3 investigated both types of pregnancy loss. There was large variation in the study quality in relation to the study question of this review with 12/31 studies found to be at risk of bias in at least one domain. Of 5 studies (n=5) investigating risk of pregnancy loss and COVID-19 exposure throughout the entire pregnancy, 4 observed more miscarriage in the COVID-exposed group, but most confidence intervals spanned the null and the number of studies was too low for formal synthesis. Eight studies which considered COVID-19 infection at delivery and risk of stillbirth were meta-analysed, and a higher risk was seen in the COVID-19 group (pooled risk ratio 1.54, 1.21-1.88).

Conclusions

Research to date suggests that having COVID-19 at point of delivery may be associated with an increased risk of stillbirth, but the quality of studies is variable. There is insufficient good quality evidence to reliably assess whether having COVID-19 during pregnancy is associated with an increased risk of miscarriage.

Key Words:

COVID-19, Miscarriage, Pregnancy Outcomes, SARS-CoV-2, Spontaneous Abortion, Stillbirth, Systematic Review

Introduction

COVID-19, caused by infection with the SARS-CoV-2 virus, emerged as a new disease in 2019 leading to a global pandemic and approximately 6.4 million deaths to date [1]. As with other respiratory infections it has been established that pregnant women are vulnerable to developing severe COVID-19 and the complications associated with it [2,3]. Less is known however about the potential association between infection during pregnancy and the risk of pregnancy loss. It has been

hypothesised that contracting COVID-19 during pregnancy may lead to an increased risk of miscarriage or stillbirth potentially mediated by placental damage [4–8]. COVID-19 during pregnancy has therefore been the subject of intense research over the last two years.

During the early part of the pandemic rapid research was important as the world tried to understand a new disease therefore many early reviews of COVID-19 and it's affects on pregnancy outcomes were based on case studies and case series [9–12]. Unfortunately this type of evidence is limited and can often have inherrent biases [13]. There have now been a number of more indepth studies published which use comparative methods to try to estimate any increased risk of pregnancy loss. Furthermore, over time our knowledge of COVID-19 has evolved as has the virus itself and various studies and reviews have reached differing conclusions. As risk minimisation measures around the world have now largely been lifted and we learn to live with COVID-19 in circulation it is important that we fully understand the risks that it may pose to pregnant women and their unborn children.

Objective

This review attempts to bring together the evidence to date on whether having COVID-19 during pregnancy is associated with a higher risk of the pregnancy ending in loss, either miscarriage or stillbirth.

Methods

This review was registered in the International Prospective Register of Systematic Reviews (PROSPERO 2022: CRD42022327437). A priori methods were outlined in the published protocol [14].

Data Sources and Identification

Three publication databases (Embase, Pubmed, Cochrane) and four pre-print databases (MedRxiv, Lancet Preprint, Gates Open Research and Wellcome Open Research) were searched. Studies published between the 01/03/2020 and 31/05/2022 were eligible for inclusion. A search strategy combing search terms related to COVID-19 with those related to pregnancy loss was developed for each database (Appendix 1). Searches of all databases were carried out on the 11/05/2022. The bibliographies of included studies and previous reviews were manually screened for additional studies by JC.

Study eligibility

Studies were considered to be eligible if they attempted to quantitatively assess the association between having COVID-19 during pregnancy and pregnancy loss. Only studies which included a COVID-19 free control group were included. Studies where it was unclear whether the COVID-19 had occurred during or prior to pregnancy, such as those studies measuring the presence of antibodies as an indicator of exposure were excluded. Studies which reported a composite outcome of adverse pregnancy events or those which compared pregnancy loss rates in the pandemic vs pre-pandemic time period without looking at individuals' exposure were also excluded. Studies which were not published in English were excluded.

Study Selection

The eligibility of each study was assessed by two reviewers (JC and MH) independently reviewing all retrieved abstracts. Where sufficient detail was not available in the title and abstract to determine eligibility then the full paper was obtained and reviewed. Disagreements over eligibility were resolved by discussion including the rest of the study team where necessary.

Data Extraction

Data were extracted by one reviewer (JC) on study characteristics (title, study design, location), the outcome and exposure definitions, the number of exposed and unexposed individuals, the number of recorded outcomes in each group and the statistical effect measure reported by the paper. All results relating to COVID-19 exposure and pregnancy loss were collected. The timing of the study was also noted in relation to the rollout of COVID-19 vaccinations.

Risk of Bias in Individual Studies

The quality of each individual study was assessed by JC using an adapted version of the NICE Quality appraisal checklist for quantitative studies reporting correlations and associations (Appendix 2) [15]. The tool is designed to assess quality across five categories: study population and participants; selection and methods; outcomes; analysis; and summary. For each category, studies were given a summary score equivalent to the lowest score awarded for a question in that category (Figures 2 and 3). Each study was then awarded an overall study quality grade for external and internal validity from one of the three categories below which are based on the checklist criteria (Appendix 3).

- ++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.

- + Some of the checklist criteria have been fulfilled, where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- – Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Data Synthesis

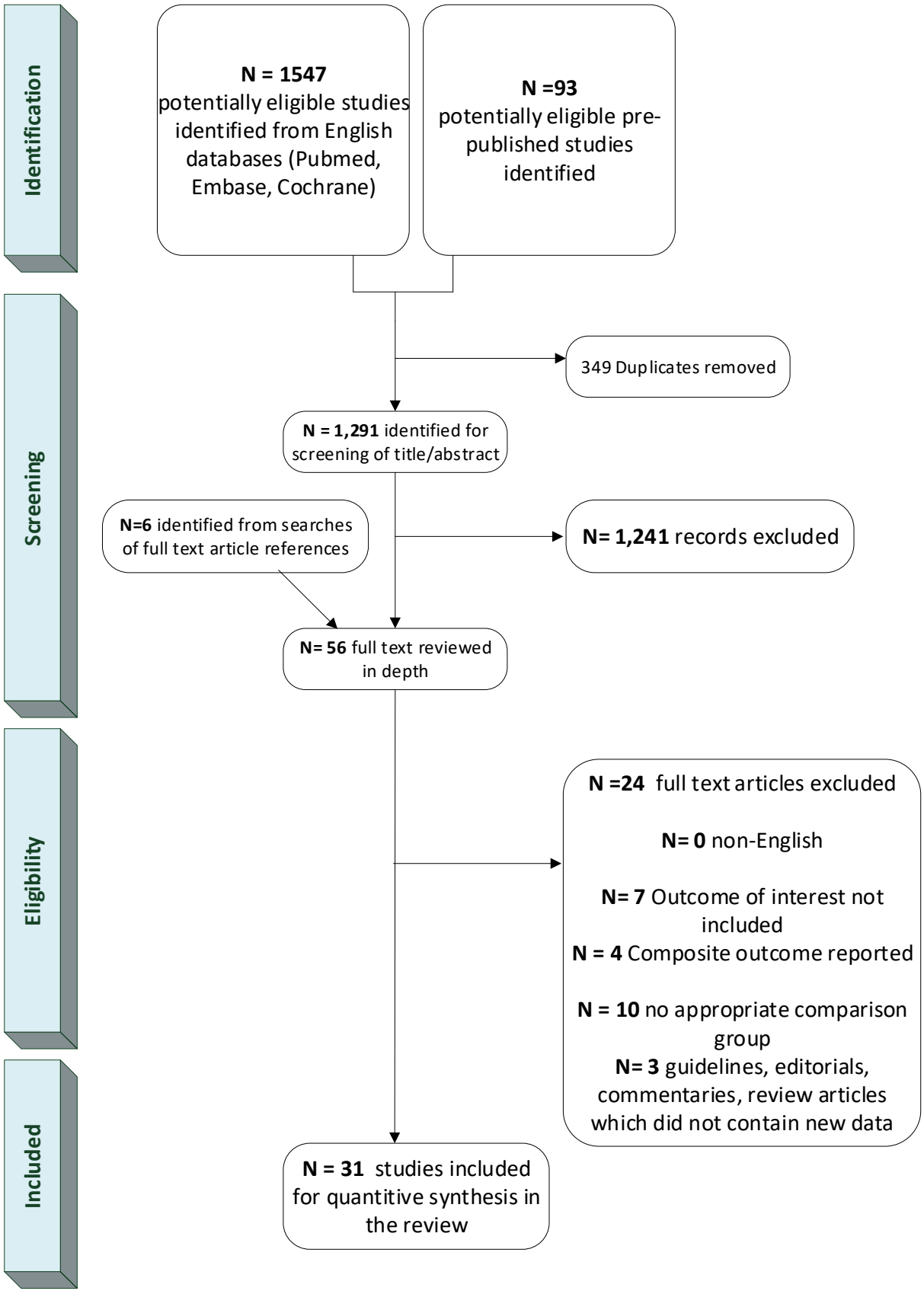
Where adjusted ratios were reported by the study these were extracted otherwise crude risk ratios were calculated from the given data (Figures 2 and 3). Studies were compared in terms of exposure and outcome definitions and quality assessment scores. When deciding whether to perform data-synthesis we looked at how studies measured COVID-19 exposure and at what time point in the pregnancy. We also compared outcome definitions, for example at what gestation a loss was regarded as stillbirth or a miscarriage. Only studies which were at low or medium risk of bias according to the quality assessment were eligible for inclusion in data synthesis. A group of studies which looked at stillbirth as the outcome and COVID-19 diagnosis at delivery were deemed to be sufficiently comparable and of high enough quality for meta-analysis to be performed. Higgins and Thompson's I^2 statistic was used to quantify heterogeneity between these studies. Stata was used to conduct the meta-analysis using a random-effects model [16]. Funnel plots were used to assess publication bias between the studies included in the meta-analysis (Appendix 4)

Results

Included Studies

Database searches returned 1,291 results of which 93 came from pre-print databases. A further 6 studies were identified by hand searching. After initial screening of 1,242 were assessed as not meeting the eligibility criteria based on the abstract. 55 papers were then selected for full text review of these 24 were excluded based on the study criteria (Figure 1)

Figure 1: Systematic Review PRISMA flow chart of included studies


Identification
Screening
Eligibility
Included
N = 1547 potentially eligible studies identified from English databases (Pubmed, Embase, Cochrane)
N =93 potentially eligible pre-published studies identified
349 Duplicates removed
N = 1,291 identified for screening of title/abstract
N=6 identified from searches of full text article references
N= 1,241 records excluded
N= 56 full text reviewed in depth
N =24 full text articles excluded
N= 0 non-English
N= 7 Outcome of interest not included
N = 4 Composite outcome reported
N = 10 no appropriate comparison group
N= 3 guidelines, editorials, commentaries, review articles which did not contain new data
N = 31 studies included for quantitive synthesis in the review

Miscarriage

Amongst the final 31 there were 7 studies which looked solely at miscarriage as the outcome of interest [17–23] and a further 3 included it along with all types of pregnancy loss [24–26]. The definition of miscarriage varied from only including pregnancy losses in the first trimester to including pregnancy loss before 20 weeks gestation (Table 1a). Three studies measured COVID-19 exposure at a single time point by a positive PCR test at the time of admission to hospital due to miscarriage and 4 studies attempted to ascertain whether women were infected at any point during their pregnancy either through self-reporting or antibody tests. There were an additional three studies which looked at miscarriage as part of an assessment of all pregnancy loss outcomes (Table 1c). Two of these studies used electronic health data and considered outcomes and exposure across the whole pregnancy [27,28]. The third study only included late miscarriage (>14 weeks) and COVID-19 status on admission to hospital.

Five of the 10 studies which included miscarriage as an outcome were found to be at a high risk of bias [29–33], none of these studies presented adjusted results or adequately attempted to control for confounding. Two studies were found to be at medium risk of bias [34,35] and 2 were found to be at low risk [36,37] (Figure 2) (Appendix 3).

The type of statistical measure reported by the studies varied with a large number just reporting the proportions of events between the exposed and the unexposed. Only 2/10 studies reported an adjusted odds or risk ratio (Figure 2); crude ratios were calculated for the remaining studies. Amongst the studies which measured COVID-19 exposure at the time of miscarriage 3/4 observed higher risks in the COVID-19 group all increases were all non-significant. For those which attempted to measure COVID-19 exposure across the whole pregnancy 4/5 observed an increased risk amongst the COVID-19 group, though again results were generally non-significant.

There were insufficient numbers of high-quality studies with comparable exposure and outcome definitions for a meta-analysis of miscarriage studies to be performed.

Stillbirth

Twenty-four of the included studies looked at stillbirth as an outcome of interest, including the three studies which looked at all pregnancy loss outcomes (Table 1b and 1c). The gestation after which the fetal death was considered to be a stillbirth varied from 19 to 24 weeks with most studies not specifying a minimum gestation (Table 1b). Methods to define exposure also varied, 13 of these studies only tested for COVID-19 at the point of admission for delivery [35,38–45], 9 studies included a

positive PCR test at any point during pregnancy [37,46–53] and 2 tested on admission for delivery but also looked for the presence of antibodies as an indicator that infection had occurred earlier in pregnancy[54,55].

For studies which looked at stillbirth there were 8 studies deemed to be high quality [37,56–62], 8 studies which were at medium risk of bias [35,53,63–68]. Five studies were found to have significant risk of bias in one of more area in relation to our question of interest based on the reported paper [69–75]. One study did not detect any pregnancy loss in their results and was therefore deemed ineligible for quality assessment [76]. The main areas of concern were around the methods of selecting the exposure and control groups.(Figure 3).

Ratios for studies which looked at stillbirth are shown in Figure 3; three of the studies did not have enough outcome events for ratios to be calculated [70,74,76]. Six of the twelve studies which looked at COVID-19 exposure at the point of delivery reported adjusted ratios as did 6/9 of those looking at exposure across the whole pregnancy. Of those measuring COVID-19 at delivery 11/12 showed an increased risk of stillbirth amongst the exposed group, 5/11 of these were significant results including the majority of studies which were deemed to be at medium or low risk of bias. Amongst the 9 studies looking at COVID-19 exposure across the whole pregnancy risk/odds ratios ranged from 0.33-9.30. Five of the 9 studies reported an increased risk amongst the exposed group although 4 of these were non-significant.

There was insufficient consistency in the exposure definitions and quality of studies for a meta-analysis looking at whether having COVID-19 at any point in pregnancy may increase the risk of stillbirth. However, there was a sub-group of 8 studies which looked at risk of stillbirth dependent on COVID-19 status at time of delivery which were deemed to be of sufficient quality for a sub-group meta-analysis [35,57,58,65–67,77,78] (3/8 low risk of bias and 5/8 medium risk of bias). Risk/odds ratios reported by these studies ranged from 0.97 to 4.70, five of the studies reported adjusted ratios all of which showed a higher risk of stillbirth amongst the COVID-19 group . Meta-analysis of these studies showed a statistically significant increase in risk of stillbirth amongst women who had COVID-19 at time of delivery with a pooled risk ratio of 1.54 (1.21-1.88) (Figure 4).

COVID-19 Vaccination status

The majority of studies (28/31) were conducted before the widespread availability of COVID-19 vaccines. Of the three studies which also included follow-up time post vaccine roll-out only one gave

results split into pre and post vaccination periods [66]. None of the included studies reported vaccination status of the women directly.

Table 1a: Characteristics of included studies which looked at miscarriage as the outcome of interest

Author, Date	Country	Reported Study Design	Exposure Definition	Outcome Definition	N Exposed	N Unexposed	Pre/post vaccination rollout
Balachandren et al, 2022[79]	UK	Cohort	Self-reported COVID-19	Pregnancy Loss <13 weeks	77	2,669	Pre-vaccine
Cosma et al, 2021[36]	Italy	Case-Control	Prescence if IgG and IgM non-neutralizing antibodies against SARS- CoV-2 or a positive PCR test	Pregnancy loss at <13 weeks gestation	23	102	Pre-vaccine
Gajbhiye et al, 2021[29]	India	Cohort	Positive PCR test on admission to hospital	Pregnancy loss at < 20 weeks or delivery of a dead fetus weighing less than 500 g.	487	11,952	Both
Gonzalez-Rodriguez et al, 2022 [30]	Spain	Cohort	Positive PCR test on admission to hospital	Pregnancy loss at <13 weeks gestation	*	*	Pre-vaccine
Jacoby et al, 2021[31]	USA	Cohort	Positive PCR test at <14 weeks gestation	Pregnancy Loss at <20 weeks	94	15	Pre-vaccine
Khoiwal et al, 2022[32]	India	Matched Cohort	Positive PCR test on admission to hospital	Pregnancy Loss at <20 weeks	60	60	Pre-vaccine
Kiremitli et al, 2022 [33]	Turkey	Cohort	Positive PCR test during the 1st trimester	Complete spontaneous abortion in the first trimester including blighted ovum and biochemical pregnancies. Ectopic pregnancies were excluded.	52	53	Pre-vaccine

* reported numbers of COVID-19 infected women among those who miscarried and those who delivered a live baby

Table 1b: Characteristics of included studies which looked at stillbirth as the outcome of interest

Author, Date	Country	Reported Study Design	Exposure Definition	Outcome Definition	N Exposed	N Unexposed	Pre/post vaccination rollout
Ackerman et al, 2022[77]	USA	Cohort	COVID-19 diagnosis recorded alongside hospitalisation for delivery.	Stillbirth	8584	465 318	Pre-vaccine
Adhikari et al, 2020[73]	USA	Cohort	Positive PCR test during pregnancy	Stillbirth	252	3122	Pre-vaccine
Ahlberg et al, 2020 [65]	Sweden	Cross-sectional	Positive PCR test on admission to hospital.	Stillbirth	155	604	Pre-vaccine
Cruz Melguzio et al,[70]	Spain	Cohort	Positive PCR test on admission to hospital.	Stillbirth	1347	1607	Pre-vaccine
Cruz-Lemini et al, 2021[80]	Spain	Cohort	Asymptomatic with Positive PCR test on admission to hospital.	Stillbirth	174	430	Pre-vaccine
DeSisto et al, 2021[66]	USA	Cohort	COVID-19 diagnosis documented at delivery hospitalisation	Stillbirth >20 weeks gestation	21,653	1,227,981	Both pre and post vaccination time periods included
Ferrara et al, 2022[59]	USA	Cohort	Positive PCR test from 30 days before the last menstrual period	Stillbirth	1332	42 554	Pre-vaccine

			to 7 days after delivery. Or a diagnoses of COVID-19 infection taken from ICD-10 codes.				
Gurol-Urganci et al, 2021[67]	England	Cohort	Positive PCR test on admission to hospital.	Stillbirth > 24 weeks gestation	3527	338,553	Pre-vaccine
Jering et al, 2021[58]	USA	Cohort	ICD 10 Code for COVID-19 recorded when admission for delivery	Stillbirth	6380	400,066	Pre-vaccine
Ko et al, 2021[78]	USA	Cohort	Positive PCR test on admission to hospital.	Stillbirth	6550	482,921	Pre-vaccine
Litman et al, 2022[60]	UK	Cohort	Positive PCR test during pregnancy	Stillbirth >19 weeks gestation	2708	39,562	Pre-vaccine
Martinez Perez et al, 2021[81]	Spain	Cohort	Positive PCR test on admission to hospital.	Stillbirth	246	763	Pre-vaccine
Piekos et al, 2022[53]	USA	Matched Cohort	Positive PCR test during pregnancy	Stillbirth	882	889	Both
Pirjani et al, 2020[71]	Iran	Cohort	Positive PCR test on admission to hospital plus clinical symptoms of COVID-19	Stillbirth	66	133	Pre-vaccine

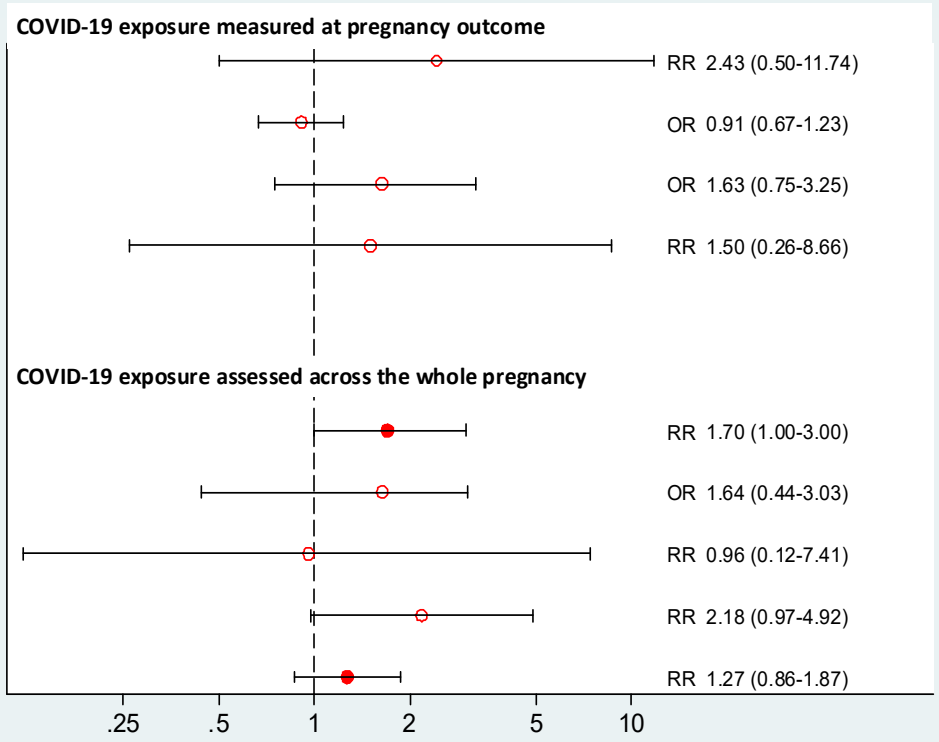
Prabhu et al, 2020[72]	USA	Cohort	Positive PCR test on admission to hospital.	Stillbirth	70	605	Pre-vaccine
Saviron-Cornudella et al, 2021[74]	Spain	Cohort	Positive PCR test on admission to hospital or presence of antibodies indicating past infection.	Stillbirth	43	1146	Pre-vaccine
Son et al, 2021[61]	USA	Cohort	Positive PCR test or Covid diagnosis recorded during pregnancy	Stillbirth >24 weeks	7432	100635	Pre-vaccine
Steffen et al, 2021[75]	USA	Cohort	Positive PCR test on admission to hospital or presence of antibodies indicating past infection.	Stillbirth	65	971	Pre-vaccine
Stephansson et al, 2022[68]	Sweden	Cohort	Positive PCR test during pregnancy	Stillbirth	794	13,871	Pre-vaccine
Vousden et al, 2021[62]	UK	Cohort	Hospitalised with COVID-19 during pregnancy.	Stillbirth	722	694	Pre-vaccine
Wilkinson et al, 2022[76]	England	Matched Cohort	Positive PCR test during pregnancy	Stillbirth >24 weeks gestation	214	214	Pre-vaccine

Table 1c: Characteristics of included studies which looked at all types of pregnancy loss

Author, Date	Country	Reported Study Design	Exposure Definition	Outcome Definition	N Exposed	N Unexposed	Pre/post vaccination rollout
Cardona-Perez et al, 2021 [35]	Mexico	Retrospective Case-control	Positive PCR test on admission to hospital.	Records of miscarriage or stillbirth recorded in EHR data	70	170	Pre-vaccine
Hcini et al, 2021[57]	West Guiana	Cohort	Positive PCR test on admission to hospital.	late miscarriages (>14 weeks) and stillbirth >20 weeks.	137	370	Pre-vaccine
Regan et al, 2022[37]	USA	Cohort	Record of COVID-19 during pregnancy either from physician diagnosis or PCR test result.	Records of miscarriage or stillbirth taken from medical claims data.	2655	75,628	Pre-vaccine

Figure 2: Summary Data Quality scores and reported Risk/Odds ratios for studies which included miscarriage as an outcome

Study	Population	Method of Selection of Exposure (or comparison) group	Outcomes	Analyses	Summary
Cardona-Perez	+	+	+	+	+
Gajbhiye	+	-	-	-	-
Gonzalez Rodriguez	+	+	-	-	-
Khoiwal	+	+	++	-	-
Balachandren	+	+	+	+	+
Cosma	+	+	++	+	+
Jacoby	+	-	+	-	-
Kiremitli	+	-	+	-	-
Regan	++	++	++	++	++



• Filled circle indicates an adjusted ratio

o Hollow circle indicates an unadjusted ratio

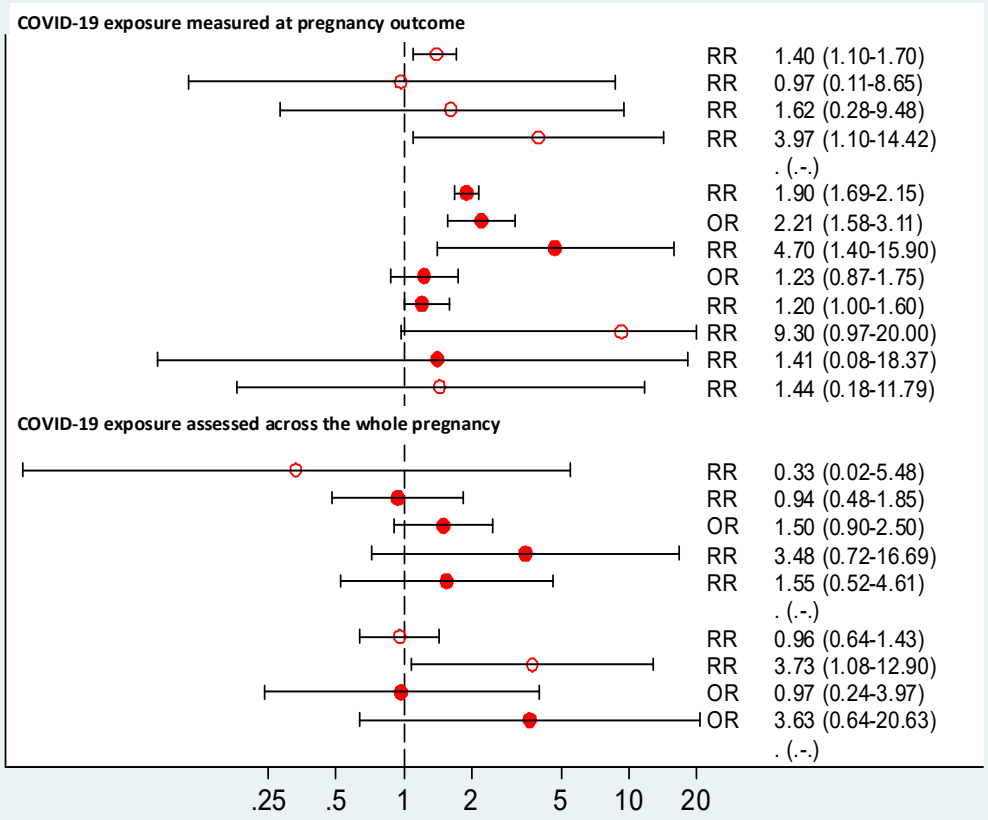
++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.

+ Some of the checklist criteria have been fulfilled where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.

- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

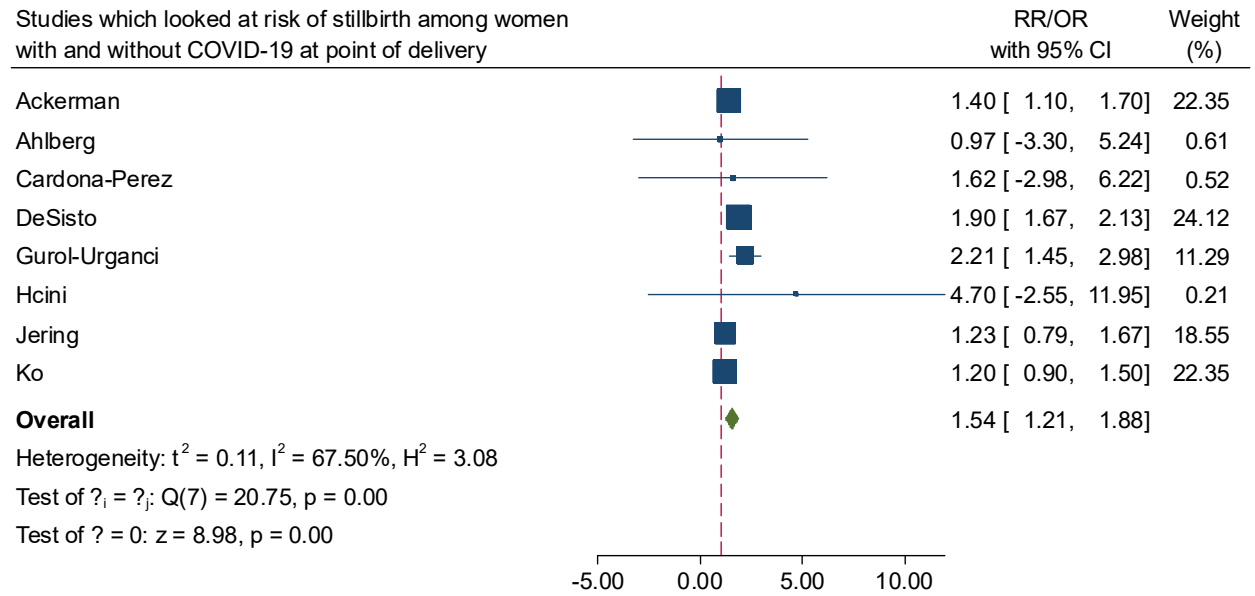
Figure 3 Summary Data Quality scores and reported Risk/Odds ratios for studies which included stillbirth as an outcome

Study	Population	Method of Selection of Exposure (or comparison) group	Outcomes	Analyses	Summary
Ackerman	++	+	++	+	+
Ahlberg	+	+	++	+	+
Cardona-Perez	+	+	+	+	+
Cruz Melguzio	+	-	+	+	-
Cruz-Lemini	+	+	++	++	+
DeSisto	+	+	++	++	+
Gurol-Urganci	++	+	+	++	+
Hcini	+	++	+	++	++
Jering	++	++	++	++	++
Ko	++	+	++	++	++
Martinez Perez	++	-	+	++	-
Pirjani	+	-	+	+	-
Prabhu	+	-	++	+	-
Adhikari	++	-	+	-	-
Ferrara	++	++	++	++	++
Litman	++	++	++	++	++
Piekos	++	+	++	+	+
Regan	++	++	++	++	++
Saviron-Cornudella	+	-	+	+	-
Son	+	++	+	++	++
Steffen	+	-	++	++	-
Stephansson	++	-	++	+	+
Vousden	++	++	+	++	++
Wilkinson	++	++	++	+	NA



• Filled circle indicates an adjusted ratio
o Hollow circle indicates an unadjusted ratio
> Arrow symbol indicates that upper confidence interval has been truncated to allow
++ All or most of the checklist criteria have been fulfilled, where they have not been fulfilled the conclusions are very unlikely to alter.
+ Some of the checklist criteria have been fulfilled where they have not been fulfilled, or not adequately described, the conclusions are unlikely to alter.
- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Figure 4: Meta Analysis of the risk of stillbirth among women with COVID-19 at point of delivery



Comment

This review attempts to assimilate the evidence as to whether having COVID-19 during pregnancy increases the risk of pregnancy loss, either through miscarriage or stillbirth. We found 7 studies which investigated miscarriage, 21 which investigated stillbirth and only 3 which looked at both types of pregnancy loss. The majority of studies looking at miscarriage observed an increased risk amongst women with COVID-19 this was the same regardless of the way in which COVID-19 exposure was assessed however, study quality was variable, and most results were not statistically significant. There was also an increased risk observed amongst the majority of studies which looked at stillbirth as an outcome. Observed increased risk was more common and more likely to be statistically significant among studies which looked at COVID-19 status at point of delivery. There was only one study by Regan et al which considered COVID-19 exposure across the whole pregnancy and all types of pregnancy loss outcomes [37]. This study was deemed to be of low risk of bias and reported an increase in risk of both types of pregnancy loss associated with having COVID-19 whilst pregnant however, these increases were not found to be statistically significant.

Miscarriage is a difficult outcome to study due to underreporting. Considering this and combined with the short timeframe since the emergence of COVID-19 it is unsurprising that there have been only a few studies which looked at the potential risk of this type of pregnancy loss. None of the studies found reported a statistically significant increase in risk of miscarriage associated with COVID-19. However, outcome and exposure definitions varied considerably meaning results could not be synthesised. A number of studies were found which considered stillbirth as an outcome. However, most of these studies were hospital based and included stillbirth as one of a range of adverse maternal and neonatal outcomes rather than the focus of the study therefore, information presented on this outcome is often limited.

In relation to the question of interest for this review the majority of studies found were deemed to be cross-sectional (despite being reported as cohort studies) only measuring COVID-19 status at the time of pregnancy outcome. Whilst these studies can give us useful information about whether having COVID-19 during delivery may increase the risk of stillbirth they do not address the overall question as to whether having the disease at any time during pregnancy may increase the risk of loss. Due to the number of studies which took this approach it was decided post-hoc to meta-analyse the sub- group of studies which looked at risk of stillbirth among women with COVID-19 at point of delivery. The results of this meta-analysis indicate that pregnant women with COVID-19

were 1.5 times more likely to deliver a stillborn baby than those without. This result provides some evidence of a potential link between COVID-19 and an increased risk of pregnancy loss.

A previous review by Pathirathna et al which looked at the impact of COVID-19 on multiple perinatal outcomes also reported a limited number of studies which considered miscarriage as an outcome[82]. They also reported a combined odds ratio which showed a statistically significant increase in the risk of stillbirth among women who had COVID-19, they included all studies which considered miscarriage regardless of COVID-19 exposure timing. Systematic reviews conducted earlier in the pandemic reached mixed conclusions regarding risk of stillbirth although these reviews included mainly case reports and noncomparative study designs [83–90].

The quality of the included studies in relation to the question posed by this review was variable. The main concern highlighted by the quality assessment was a lack of consideration of potential confounding factors either within the study design or within the analysis. For studies which reported both a crude and adjusted risk ratio adjustment for potential confounding factors reduced the risk ratio towards 1 suggesting that studies which did not adjust may be reporting an overestimation in effect (Appendix 5). Other concerns were around studies equivalence in the length of follow-up time between cases and controls and the time period over which COVID-19 exposure was assessed. Measuring COVID-19 status only at pregnancy end may mean that women who have had COVID-19 earlier in their pregnancy are misclassified as unexposed. This does not allow for testing of the hypothesis that COVID-19 may lead to placental damage resulting in pregnancy loss and would result in an underestimation of any effect.

The methods for this review were published and peer-reviewed a-priori helping to ensure a robust approach. The search strategy encompassed both published and pre-print databases ensuring that the very latest available studies were included, this is an important strength of this review as COVID-19 research is fast paced. Hopefully the inclusion of pre-print studies also goes someway to reducing publication bias however, it is not possible to assess whether there were some studies which remain unreported entirely. It is conceivable that studies carried out prospectively in a clinical setting, such as many of those we found, are more inclined to report when their results show an increased risk and therefore have clinical practice implications.

Some studies which looked at stillbirth as one of a number of adverse perinatal and maternal outcomes were only found as a result of searching reference lists, it is possible that other studies like this may have been missed. This highlights a potential limitation of our search strategy which specifically focused on pregnancy loss studies. Whilst two reviewers independently assessed the abstracts for inclusion due to resource constraints only one reviewer carried out the data extraction

and quality assessment for this review, we acknowledge that this may have resulted in assessment bias or mistakes in data extraction. Unfortunately there were insufficient studies found to allow us to assess the potential impact of global COVID-19 vaccination programme.

This review has highlighted a need for further research to address the question as to whether having COVID-19 during pregnancy is associated with an increased risk of pregnancy loss. Whilst there appears to be evidence to suggest an increased risk of stillbirth amongst women with COVID-19 during delivery we found insufficient evidence to assess whether the same risk of loss applies to women who get COVID-19 earlier in pregnancy. Nevertheless, this review provides support to the strong recommendation that pregnant women are vaccinated against COVID-19.

References

1. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard.

2. Babarinsa IA, Okunoye GO, Odukoya O Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-1) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infections in pregnancy - An overview. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2021; 263, 171-175

3. Saccone G, Sen C, Di Mascio D, et al. Maternal and perinatal outcomes of pregnant women with SARS-CoV-2 infection. *Ultrasound in Obstetrics and Gynecology*. 2021;57(2):232-241.

4. Schwartz DA, Avvad-Portari E, Babál P, et al. Placental Tissue Destruction and Insufficiency from COVID-19 Causes Stillbirth and Neonatal Death from Hypoxic-Ischemic Injury: A Study of 68 Cases with SARS-CoV-2 Placentitis from 12 Countries. *Arch Pathol Lab Med*. Published online February 2022. doi:10.5858/ARPA.2022-0029-SA

5. Bouachba A, Allias F, Nadaud B, et al. Placental lesions and SARS-Cov-2 infection: Diffuse placenta damage associated to poor fetal outcome. *Placenta*. 2021;112:97-104. doi:10.1016/j.placenta.2021.07.288

6. Aghaamoo S, Ghods K, Rahmanian M. Pregnant women with COVID-19: the placental involvement and consequences. *J Mol Histol*. 2021;52(3):427-435. doi:10.1007/s10735-021-09970-4

7. Zaigham M, Gisselsson D, Sand A, et al. Clinical-pathological features in placentas of pregnancies with SARS-CoV-2 infection and adverse outcome: case series with and without congenital transmission. *BJOG*. 2022;129(8):1361-1374. doi:10.1111/1471-0528.17132

8. Ville Y. The placenta in COVID-19 infection in pregnancy. *BJOG*. 2022;129(8):1375. doi:10.1111/1471-0528.17162

9. Juan J, Gil MM, Rong Z, Zhang Y, Yang H, Poon LC. Effect of coronavirus disease 2019 (COVID-19) on maternal, perinatal and neonatal outcome: systematic review. *Ultrasound in Obstetrics and Gynecology*. 2020;56(1):15-27. doi:10.1002/uog.22088

10. de Medeiros KS, Sarmento ACA, Costa APF, et al. Consequences and implications of the coronavirus disease (COVID-19) on pregnancy and newborns: A comprehensive systematic review and meta-analysis. *International Journal of Gynecology and Obstetrics*. 2022;156(3):394-405.

11. Bellos I, Pandita A, Panza R. Maternal and perinatal outcomes in pregnant women infected by SARS-CoV-2: A meta-analysis. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2021;256:194-204.

12. Banaei M, Ghasemi V, Saei Ghare Naz M, et al. Obstetrics and neonatal outcomes in pregnant women with covid-19: A systematic review. *Iran J Public Health*. 2020;49:38-47. doi:10.18502/IJPH.V49IS1.3668

13. Ang XL, Chonkar SP, Chua MSQ, Sulaiman S, Lee JCS. Problems with Early Systematic Reviews: The Case of Coronavirus Disease 2019 (COVID-19) in Pregnancy. *Matern Child Health J*. 2021;25(1):38-41. doi:10.1007/s10995-020-03046-7

14. Campbell J, Williams R, Harley M, Bhaskaran K. Covid-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review protocol. *BMJ Open*. 2022;12. doi:10.1136/bmjopen-2022-065588

15. NICE. Methods for the development of NICE public health guidance (third edition). *National Institute of Health and Clinical Excellence*. 2012;(September):284.

16. StataCorp. Stata Statistical Software: Release 17. Published online 2021.

17. Gajbhiye RK, Tilve A, Kesarwani S, et al. Increased rate of miscarriage during second wave of COVID-19 pandemic in India. *Ultrasound in Obstetrics and Gynecology*. 2021;58(6):946-949. doi:10.1002/uog.24784

18. Kiremitli S, Kiremitli T, ULUG P, et al. Does being infected with SARS-CoV-2 in the first-trimester increase the risk of miscarriage? 2022;94(2):e20211283. doi:10.1590/0001-3765202220211283

19. Jacoby VL, Murtha A, Afshar Y, et al. Risk of pregnancy loss before 20 weeks' gestation in study participants with COVID-19. *Am J Obstet Gynecol*. 2021;225(4):456-457. doi:10.1016/j.ajog.2021.06.080

20. Khoiwal K, Agarwal A, Gaurav A, et al. Obstetric and perinatal outcomes in pregnant women with COVID-19: an interim analysis. *Women Health*. 2022;62(1):12-20. doi:10.1080/03630242.2021.2007199

21. González Rodríguez L, Oreja Cuesta AB, Pardo Pumar MI, et al. SARS-CoV-2 infection in early first-trimester miscarriages: a prospective observational study. *Reprod Biomed Online*. 2022;44(1):127-130. doi:10.1016/j.rbmo.2021.09.010

22. Cosma S, Carosso AR, Cusato J, et al. Coronavirus disease 2019 and first-trimester spontaneous abortion: a case-control study of 225 pregnant patients. *Am J Obstet Gynecol*. 2021;224(4):391.e1-391.e7. doi:10.1016/j.ajog.2020.10.005

23. Piekos SN, Roper RT, Hwang YM, et al. The effect of maternal SARS-CoV-2 infection timing on birth outcomes: a retrospective multicentre cohort study. *Lancet Digit Health*. 2022;4(2):e95-e104. doi:10.1016/S2589-7500(21)00250-8

24. Cardona-Pérez JA, Villegas-Mota I, Helguera-Repetto AC, et al. Prevalence, clinical features, and outcomes of SARS-CoV-2 infection in pregnant women with or without mild/moderate symptoms: Results from universal screening in a tertiary care center in Mexico City, Mexico. *PLoS One*. 2021;16(4 April):1-19. doi:10.1371/journal.pone.0249584

25. Hcini N, Maamri F, Picone O, et al. Maternal, fetal and neonatal outcomes of large series of SARS-CoV-2 positive pregnancies in peripartum period: A single-center prospective comparative study. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2021;257:11-18. doi:10.1016/j.ejogrb.2020.11.068

26. Regan AK, Arah OA, Fell DB, Sullivan SG. SARS-CoV-2 Infection during Pregnancy and Associated Perinatal Health Outcomes: A National US Cohort Study. *Journal of Infectious Diseases*. 2022;225(5):759-767. doi:10.1093/infdis/jiab626

27. Cardona-Pérez JA, Villegas-Mota I, Helguera-Repetto AC, et al. Prevalence, clinical features, and outcomes of SARS-CoV-2 infection in pregnant women with or without mild/moderate symptoms: Results from universal screening in a tertiary care center in Mexico City, Mexico. *PLoS One*. 2021;16(4 April):1-19. doi:10.1371/journal.pone.0249584

28. Regan AK, Arah OA, Fell DB, Sullivan SG. SARS-CoV-2 Infection during Pregnancy and Associated Perinatal Health Outcomes: A National US Cohort Study. *Journal of Infectious Diseases*. 2022;225(5):759-767. doi:10.1093/infdis/jiab626

29. Gajbhiye RK, Tilve A, Kesarwani S, et al. Increased rate of miscarriage during second wave of COVID-19 pandemic in India. *Ultrasound in Obstetrics and Gynecology*. 2021;58(6):946-949. doi:10.1002/uog.24784

30. González Rodríguez L, Oreja Cuesta AB, Pardo Pumar MI, et al. SARS-CoV-2 infection in early first-trimester miscarriages: a prospective observational study. *Reprod Biomed Online*. 2022;44(1):127-130. doi:10.1016/j.rbmo.2021.09.010

31. Jacoby VL, Murtha A, Afshar Y, et al. Risk of pregnancy loss before 20 weeks' gestation in study participants with COVID-19. *Am J Obstet Gynecol*. 2021;225(4):456-457. doi:10.1016/j.ajog.2021.06.080

32. Khoiwal K, Agarwal A, Gaurav A, et al. Obstetric and perinatal outcomes in pregnant women with COVID-19: an interim analysis. *Women Health*. 2022;62(1):12-20. doi:10.1080/03630242.2021.2007199

33. Kiremitli S, Kiremitli T, ULUG P, et al. Does being infected with SARS-CoV-2 in the first-trimester increase the risk of miscarriage? 2022;94(2):e20211283. doi:10.1590/0001-3765202220211283

34. Balachandren N, Davies M, Hall J, D M. First trimester pregnancy outcomes after confirmed SARS-CoV-2 infection in the community; a nationwide prospective longitudinal study of 10,000 pregnant women from the COVID-19 pandemic. *Human Reproduction*. 2021;36 (Suppl:i311.

35. Cardona-Pérez JA, Villegas-Mota I, Helguera-Repetto AC, et al. Prevalence, clinical features, and outcomes of SARS-CoV-2 infection in pregnant women with or without mild/moderate symptoms: Results from universal screening in a tertiary care center in Mexico City, Mexico. *PLoS One*. 2021;16(4 April):1-19. doi:10.1371/journal.pone.0249584

36. Cosma S, Carosso AR, Cusato J, et al. Coronavirus disease 2019 and first-trimester spontaneous abortion: a case-control study of 225 pregnant patients. *Am J Obstet Gynecol*. 2021;224(4):391.e1-391.e7. doi:10.1016/j.ajog.2020.10.005

37. Regan AK, Arah OA, Fell DB, Sullivan SG. SARS-CoV-2 Infection during Pregnancy and Associated Perinatal Health Outcomes: A National US Cohort Study. *Journal of Infectious Diseases*. 2022;225(5):759-767. doi:10.1093/infdis/jiab626

38. Ahlberg M, Neovius M, Saltvedt S, et al. Association of SARS-CoV-2 Test Status and Pregnancy Outcomes. *JAMA*. 2020;324(17):1782-1785. doi:10.1001/JAMA.2020.19124

39. Cruz Melguizo S, de la Cruz Conty ML, Carmona Payán P, et al. Pregnancy outcomes and SARS-CoV-2 infection: The Spanish obstetric emergency group study. *Viruses*. 2021;13(5):3-13. doi:10.3390/v13050853

40. DeSisto CL, Wallace B, Simeone RM, et al. Risk for Stillbirth Among Women With and Without COVID-19 at Delivery Hospitalization — United States, March 2020–September 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(47):1640-1645.

41. Gurol-Urganci I, Jardine JE, Carroll F, et al. Maternal and perinatal outcomes of pregnant women with SARS-CoV-2 infection at the time of birth in England: national cohort study. *Am J Obstet Gynecol*. 2021;225(5):522.e1-522.e11. doi:10.1016/j.ajog.2021.05.016

42. Hcini N, Maamri F, Picone O, et al. Maternal, fetal and neonatal outcomes of large series of SARS-CoV-2 positive pregnancies in peripartum period: A single-center prospective comparative study. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2021;257:11-18.

43. Jering KS, Claggett BL, Cunningham JW, et al. Clinical Characteristics and Outcomes of Hospitalized Women Giving Birth With and Without COVID-19. *JAMA Intern Med*. 2021;181(5):714-717. doi:10.1001/JAMAINTERNMED.2020.9241

44. Pirjani R, Hosseini R, Soori T, et al. Maternal and neonatal outcomes in COVID-19 infected pregnancies: A prospective cohort study. *J Travel Med*. 2020;27(7):1-7. doi:10.1093/JTM/TAAA158

45. Prabhu M, Cagino K, Matthews KC, et al. Pregnancy and postpartum outcomes in a universally tested population for SARS-CoV-2 in New York City: a prospective cohort study. *BJOG*. 2020;127(12):1548-1556. doi:10.1111/1471-0528.16403

46. Adhikari EH, Moreno W, Zofkie AC, et al. Pregnancy Outcomes among Women with and without Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *JAMA Netw Open*. 2020;3(11):1-11. doi:10.1001/jamanetworkopen.2020.29256

47. Ferrara A, Hedderson MM, Zhu Y, et al. Perinatal Complications in Individuals in California with or Without SARS-CoV-2 Infection during Pregnancy. *JAMA Intern Med*. 2022;182(5):503-512. doi:10.1001/jamainternmed.2022.0330

48. Litman EA, Yin Y, Nelson SJ, Capbarat E, Kerchner D, Ahmadzia HK. Adverse perinatal outcomes in a large United States birth cohort during the COVID-19 pandemic. *Am J Obstet Gynecol MFM*. 2022;4(3):100577. doi:10.1016/j.ajogmf.2022.100577

49. Son M, Gallagher K, Lo JY, et al. Coronavirus Disease 2019 (COVID-19) Pandemic and Pregnancy Outcomes in a U.S. Population. *Obstetrics and gynecology*. 2021;138(4):542-551. doi:10.1097/AOG.0000000000004547

50. Stephansson O, Pasternak B, Ahlberg M, et al. SARS-CoV-2 and pregnancy outcomes under universal and non-universal testing in Sweden: register-based nationwide cohort study. doi:10.1111/1471-0528.16990

51. Vousden N, Bunch K, Morris E, et al. The incidence, characteristics and outcomes of pregnant women hospitalized with symptomatic and asymptomatic SARS-CoV-2 infection in the UK from

March to September 2020: A national cohort study using the UK Obstetric Surveillance System (UKOSS). *PLoS One*. 2021;16(5 May):1-19. doi:10.1371/journal.pone.0251123

52. Wilkinson M, Johnstone ED, Simcox LE, Myers JE. The impact of COVID-19 on pregnancy outcomes in a diverse cohort in England. *Sci Rep*. 2022;12(1):1-10. doi:10.1038/s41598-022-04898-5

53. Piekos SN, Roper RT, Hwang YM, et al. The effect of maternal SARS-CoV-2 infection timing on birth outcomes: a retrospective multicentre cohort study. *Lancet Digit Health*. 2022;4(2):e95-e104. doi:10.1016/S2589-7500(21)00250-8

54. Saviron-Cornudella R, Villalba A, Esteban LM, et al. Screening of severe acute respiratory syndrome coronavirus-2 infection during labor and delivery using polymerase chain reaction and immunoglobulin testing. *Life Sci*. 2020;14(4)(January):293. doi:https://doi.org/10.1016/j.lfs.2021.119200

55. Steffen HA, Swartz SR, Jackson JB, et al. SARS-CoV-2 Infection during Pregnancy in a Rural Midwest All-delivery Cohort and Associated Maternal and Neonatal Outcomes. *Am J Perinatol*. 2021;38(6):614-621. doi:10.1055/s-0041-1723938

56. Ko JY, DeSisto CL, Simeone RM, et al. Adverse Pregnancy Outcomes, Maternal Complications, and Severe Illness Among US Delivery Hospitalizations With and Without a Coronavirus Disease 2019 (COVID-19) Diagnosis. *Clin Infect Dis*. 2021;73(Suppl 1):S24-S31. doi:10.1093/cid/ciab344

57. Hcini N, Maamri F, Picone O, et al. Maternal, fetal and neonatal outcomes of large series of SARS-CoV-2 positive pregnancies in peripartum period: A single-center prospective comparative study. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2021;257:11-18.

58. Jering KS, Claggett BL, Cunningham JW, et al. Clinical Characteristics and Outcomes of Hospitalized Women Giving Birth With and Without COVID-19. *JAMA Intern Med*. 2021;181(5):714-717. doi:10.1001/JAMAINTERNMED.2020.9241

59. Ferrara A, Hedderson MM, Zhu Y, et al. Perinatal Complications in Individuals in California with or Without SARS-CoV-2 Infection during Pregnancy. *JAMA Intern Med*. 2022;182(5):503-512. doi:10.1001/jamainternmed.2022.0330

60. Litman EA, Yin Y, Nelson SJ, Capbarat E, Kerchner D, Ahmadzia HK. Adverse perinatal outcomes in a large United States birth cohort during the COVID-19 pandemic. *Am J Obstet Gynecol MFM*. 2022;4(3):100577. doi:10.1016/j.ajogmf.2022.100577

61. Son M, Gallagher K, Lo JY, et al. Coronavirus Disease 2019 (COVID-19) Pandemic and Pregnancy Outcomes in a U.S. Population. *Obstetrics and gynecology*. 2021;138(4):542-551. doi:10.1097/AOG.0000000000004547

62. Vousden N, Bunch K, Morris E, et al. The incidence, characteristics and outcomes of pregnant women hospitalized with symptomatic and asymptomatic SARS-CoV-2 infection in the UK from March to September 2020: A national cohort study using the UK Obstetric Surveillance System (UKOSS). *PLoS One*. 2021;16(5 May):1-19. doi:10.1371/journal.pone.0251123

63. Ackerman CM, Nguyen JL, Ambati S, et al. Clinical and Pregnancy Outcomes of Coronavirus Disease 2019 among Hospitalized Pregnant Women in the United States. *Open Forum Infect Dis*. 2022;9(2):1-7. doi:10.1093/ofid/ofab429

64. Cruz-Lemini M, Perez EF, De La Cruz Conty ML, et al. Obstetric outcomes of sars-cov-2 infection in asymptomatic pregnant women. *Viruses*. 2021;13(1):1-12. doi:10.3390/v13010112

65. Ahlberg M, Neovius M, Saltvedt S, et al. Association of SARS-CoV-2 Test Status and Pregnancy Outcomes. *JAMA*. 2020;324(17):1782-1785. doi:10.1001/JAMA.2020.19124

66. DeSisto CL, Wallace B, Simeone RM, et al. Risk for Stillbirth Among Women With and Without COVID-19 at Delivery Hospitalization — United States, March 2020–September 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(47):1640-1645.

67. Gurol-Urganci I, Jardine JE, Carroll F, et al. Maternal and perinatal outcomes of pregnant women with SARS-CoV-2 infection at the time of birth in England: national cohort study. *Am J Obstet Gynecol*. 2021;225(5):522.e1-522.e11. doi:10.1016/j.ajog.2021.05.016

68. Stephansson O, Pasternak B, Ahlberg M, et al. SARS-CoV-2 and pregnancy outcomes under universal and non-universal testing in Sweden: register-based nationwide cohort study. doi:10.1111/1471-0528.16990

69. Martinez-Perez O, Prats Rodriguez P, Muner Hernandez M, et al. The association between SARS-CoV-2 infection and preterm delivery: a prospective study with a multivariable analysis. *BMC Pregnancy Childbirth*. 2021;21(1):1-11. doi:10.1186/s12884-021-03742-4

70. Cruz Melguizo S, de la Cruz Conty ML, Carmona Payán P, et al. Pregnancy outcomes and SARS-CoV-2 infection: The Spanish obstetric emergency group study. *Viruses*. 2021;13(5):3-13. doi:10.3390/v13050853

71. Pirjani R, Hosseini R, Soori T, et al. Maternal and neonatal outcomes in COVID-19 infected pregnancies: A prospective cohort study. *J Travel Med*. 2020;27(7):1-7. doi:10.1093/JTM/TAAA158

72. Prabhu M, Cagino K, Matthews KC, et al. Pregnancy and postpartum outcomes in a universally tested population for SARS-CoV-2 in New York City: a prospective cohort study. *BJOG*. 2020;127(12):1548-1556. doi:10.1111/1471-0528.16403

73. Adhikari EH, Moreno W, Zofkie AC, et al. Pregnancy Outcomes among Women with and without Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *JAMA Netw Open*. 2020;3(11):1-11. doi:10.1001/jamanetworkopen.2020.29256

74. Saviron-Cornudella R, Villalba A, Esteban LM, et al. Screening of severe acute respiratory syndrome coronavirus-2 infection during labor and delivery using polymerase chain reaction and immunoglobulin testing. *Life Sci*. 2020;14(4)(January):293. doi:https://doi.org/10.1016/j.lfs.2021.119200

75. Steffen HA, Swartz SR, Jackson JB, et al. SARS-CoV-2 Infection during Pregnancy in a Rural Midwest All-delivery Cohort and Associated Maternal and Neonatal Outcomes. *Am J Perinatol*. 2021;38(6):614-621. doi:10.1055/s-0041-1723938

76. Wilkinson M, Johnstone ED, Simcox LE, Myers JE. The impact of COVID-19 on pregnancy outcomes in a diverse cohort in England. *Sci Rep*. 2022;12(1):1-10. doi:10.1038/s41598-022-04898-5

77. Ackerman CM, Nguyen JL, Ambati S, et al. Clinical and Pregnancy Outcomes of Coronavirus Disease 2019 among Hospitalized Pregnant Women in the United States. *Open Forum Infect Dis*. 2022;9(2):1-7. doi:10.1093/ofid/ofab429

78. Ko JY, DeSisto CL, Simeone RM, et al. Adverse Pregnancy Outcomes, Maternal Complications, and Severe Illness Among US Delivery Hospitalizations With and Without a Coronavirus Disease 2019 (COVID-19) Diagnosis. *Clin Infect Dis*. 2021;73(Suppl 1):S24-S31. doi:10.1093/cid/ciab344

79. Balachandren N, Davies MC, Hall JA, et al. SARS-CoV-2 infection in the first trimester and the risk of early miscarriage: A UK population-based prospective cohort study of 3041 pregnancies conceived during the pandemic. *Human Reproduction*. 2022;37(6):1126-1133. doi:10.1093/humrep/deac062

80. Cruz-Lemini M, Perez EF, De La Cruz Conty ML, et al. Obstetric outcomes of sars-cov-2 infection in asymptomatic pregnant women. *Viruses*. 2021;13(1):1-12. doi:10.3390/v13010112

81. Martinez-Perez O, Prats Rodriguez P, Muner Hernandez M, et al. The association between SARS-CoV-2 infection and preterm delivery: a prospective study with a multivariable analysis. *BMC Pregnancy Childbirth*. 2021;21(1):1-11. doi:10.1186/s12884-021-03742-4

82. Pathirathna ML, Samarasekara BPP, Dasanayake TS, Saravanakumar P, Weerasekara I. Adverse Perinatal Outcomes in COVID-19 Infected Pregnant Women: A Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2022;10(2). doi:10.3390/healthcare10020203

83. Pérez-López FR, Savirón-Cornudella R, Chedraui P, et al. Obstetric and perinatal outcomes of pregnancies with COVID 19: a systematic review and meta-analysis. *Journal of Maternal-Fetal and Neonatal Medicine*. 2022;0(0):1-17. doi:10.1080/14767058.2022.2051008

84. Cavalcante MB, de Melo Bezerra Cavalcante CT, Cavalcante ANM, Sarno M, Barini R, Kwak-Kim J. COVID-19 and miscarriage: From immunopathological mechanisms to actual clinical evidence. *J Reprod Immunol*. 2021;148(September):103382. doi:10.1016/j.jri.2021.103382

85. Jha D, Adhikari M, Gautam JS, et al. Effect of COVID-19 on maternal and neonatal services. *Lancet Glob Health*. 2021;9(2).

86. Agolli A, Agolli O, Velazco DFS, et al. Fetal Complications in COVID-19 Infected Pregnant Woman: A Systematic Review and Meta-Analysis. *Avicenna J Med*. 2021;11(4):200-209. doi:10.1055/s-0041-1736540

87. Kasraeian M, Zare M, Vafaei H, et al. COVID-19 pneumonia and pregnancy; a systematic review and meta-analysis. *Journal of Maternal-Fetal and Neonatal Medicine*. 2022;35(9):1652-1659. doi:10.1080/14767058.2020.1763952

88. Di Toro F, Gjoka M, Di Lorenzo G, et al. Impact of COVID-19 on maternal and neonatal outcomes: a systematic review and meta-analysis. *Clinical Microbiology and Infection*. 2021;27(1):36-46.

89. Lanzoni G, Linetsky E, Correa D, et al. Umbilical cord mesenchymal stem cells for COVID-19 acute respiratory distress syndrome: A double-blind, phase 1/2a, randomized controlled trial. *Stem Cells Transl Med*. 2021;10(5):660-673. doi:10.1002/sctm.20-0472

90. Banaei M, Ghasemi V, Saei Ghare Naz M, et al. Obstetrics and neonatal outcomes in pregnant women with covid-19: A systematic review. *Iran J Public Health*. 2020;49:38-47. doi:10.18502/IJPH.V49IS1.3668

4.5 Research Paper Supplementary Files

Appendix 1: Table S1, Search Strategy

Database	Dates of search coverage	Miscarriage/Stillbirth	Covid-19
Pubmed	01/03/2020-Current date	“Abortion, Spontaneous” [MeSH] OR “Fetal Death” [MeSH] OR “Stillbirth” [MeSH] OR (miscarriage[MeSH Terms])) OR (miscarriages[MeSH Terms] OR Miscarriage* OR pregnancy loss* OR spontaneous abortion* OR fetal loss* OR fetal loss* OR fetal death* OR fetal death*	"coronavirus"[MeSH] OR "coronavirus infections"[MeSH Terms] OR "coronavirus"[All Fields] OR "covid 2019"[All Fields] OR "SARS2"[All Fields] OR "SARS-CoV-2"[All Fields] OR "SARS-CoV-19"[All Fields] OR "severe acute respiratory syndrome coronavirus 2" [supplementary concept] OR “coronavirus infection”[All Fields] OR "severe acute respiratory pneumonia outbreak”[All Fields] OR "novel cov"[All Fields] OR “2019ncov”[All Fields] OR “sars cov2”[All Fields] OR “cov22”[All Fields] OR “ncov”[All Fields] OR “covid19”[All Fields] OR “covid 19”[All Fields] OR “covid-19”[All Fields] OR “coronaviridae”[All Fields] OR "corona virus"[All Fields]

Embase	01/03/2020-Current date	spontaneous abortion/exp OR stillbirth/exp OR stillbirth.m.p OR pregnancy loss/exp OR pregnancy loss.mp OR fetal death.m.p OR fetus death OR fetus death/exp **NOT [medline]/lim**	**'coronavirinae'/exp OR 'coronavirinae' OR 'coronaviridae infection'/exp OR 'coronaviridae infection' OR 'coronavirus disease 2019'/exp OR 'coronavirus'/exp OR coronavirus OR 'coronavirus infection'/de NOT [medline]/lim**
Cochrane	01/03/2020-Current date	Search for “stillbirth” OR “miscarriage” OR “fetal death rates” OR ““fetal death rate” OR “fetal death” OR “fetal death rate” OR “pregnancy loss rate” OR “pregnancy loss-rate” OR pregnancy “loss-rates”	Search for “coronavirus” in the Title Abstract Keyword fields

Appendix 2: Table S2, NICE Quality appraisal checklist for quantitative studies reporting correlations and associations

(Already presented in Section 4.3 Appendix 2 of this thesis)

Appendix 3: Table S3, Quality Assessment Results in Full

	Population			Method of Selection of Exposure (or comparison) group			Outcomes					Analyses				Summay	
Study	Is the source population well described?	Is the eligible population or area representative of the source population or area?	Do the selected participants or areas represent the eligible population or area?	Selection of exposure (and comparison) group. How was selection bias minimised?	Was the selection of explanatory variables based on sound theoretical basis ?	How well were likely confounding factors identified and controlled?	Were the outcome measures and procedures reliable?	Were the outcome measurements complete?	Were all the important outcomes assessed?	Was there a similar follow-up time in exposure and comparison groups?	Was follow-up time meaningful?	Was the study sufficiently powered to detect an intervention effect (if one exists)?	Were multiple explanatory variables considered in the analysis	Were the analytical methods appropriate	Was the precision of association given or calculable? Is association meaningful ?	Are the study results internally valid (i.e. unbiased ?)	Are the findings generalisable to the source population?
Ackerman	++	++	++	++	++	+	++	++	++	++	++	++	+	+	+	+	+
Adhikari	++	++	++	++	++	-	++	++	++	+	++	-	++	+	++	-	++
Ahlberg	++	+	++	+	+	+	++	++	++	++	+	++	+	++	++	+	+
Balachandren	++	++	+	++	+	+	+	+	++	++	++	++	+	++	++	++	++
Cardona-Perez	+	+	+	+	++	+	++	+	++	++	++	+	++	+	++	+	++
Cosma	++	++	+	++	++	+	++	++	++	++	++	+	++	+	++	+	+
Cruz Melguzio	++	++	+	+	++	-	++	++	++	+	+	++	+	++	+	-	+
Cruz-Lemini	++	+	+	+	++	++	++	++	++	++	+	++	++	++	++	+	+
DeSisto	+	+	+	+	++	++	++	++	++	++	++	++	++	++	++	+	+
Ferrara	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Gajbhiye	++	+	+	-	-	-	++	++	++	-	-	++	-	++	++	-	-
Gonzalez	++	++	+	+	+	+	+	++	++	-	-	+	+	-	-	-	-
Gurol-Urganci	++	++	++	+	++	++	++	++	++	+	+	++	++	++	++	+	++
Hcini	++	+	+	++	++	++	++	++	++	+	++	++	++	++	++	++	++
Jacoby	++	+	+	+	++	-	++	+	++	++	++	+	-	-	-	-	+
Jering	++	++	++	+	++	++	++	++	++	++	+	++	++	++	++	++	++
Khoiwal	++	++	+	++	++	+	++	++	++	++	++	+	++	-	-	-	-
Kiremitli	+	+	+	+	++	-	++	++	++	+	-	-	-	-	+	-	++
Ko	++	++	++	+	++	++	++	++	++	++	++	++	++	++	++	++	++
Litman	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Martinez Perez	++	++	++	-	++	++	++	++	++	+	+	++	++	++	++	-	++
Piekos	++	++	++	+	++	++	++	++	++	++	++	++	++	++	+	+	+
Pirjani	+	++	+	-	++	++	++	+	++	+	++	+	++	++	++	-	+
Prabhu	++	+	+	+	++	-	++	++	++	++	++	+	+	+	++	-	+
Regan	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Saviron-Cornudella	+	+	++	++	++	-	++	++	++	+	++	++	+	++	++	-	+
Son	++	++	+	++	++	++	++	++	++	+	+	++	++	++	++	++	++
Steffen	++	+	++	++	-	-	++	++	++	++	++	++	++	++	++	-	+
Stephansson	++	++	++	++	+	-	++	++	++	++	++	+	+	+	+	+	+
Vousden	++	++	++	++	++	++	++	++	++	+	++	++	++	++	++	++	++
Wilkinson	++	++	++	++	++	++	++	++	++	++	++	++	++	++	+	NR	NA

Figure S1: Funnel plot of studies included in Meta Analysis of the risk of stillbirth among women with COVID-19 at point of delivery

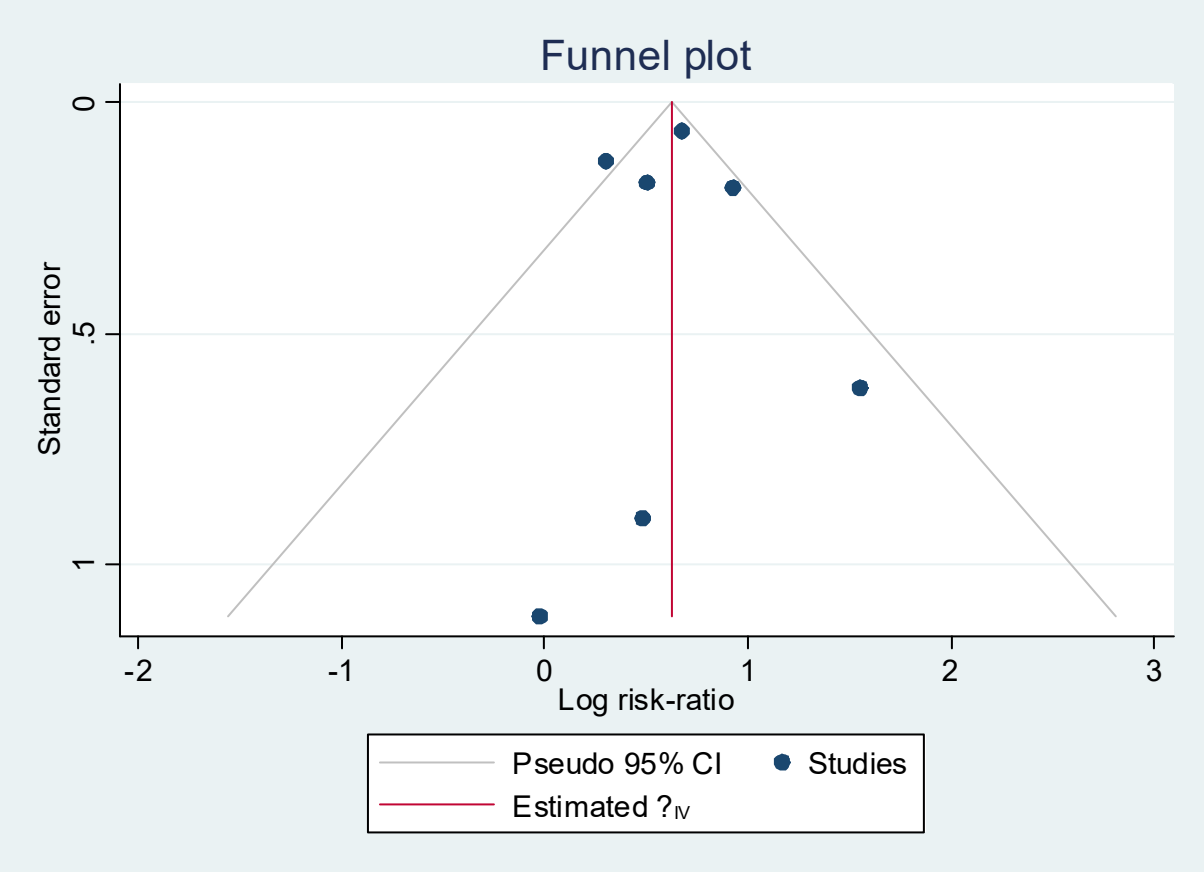
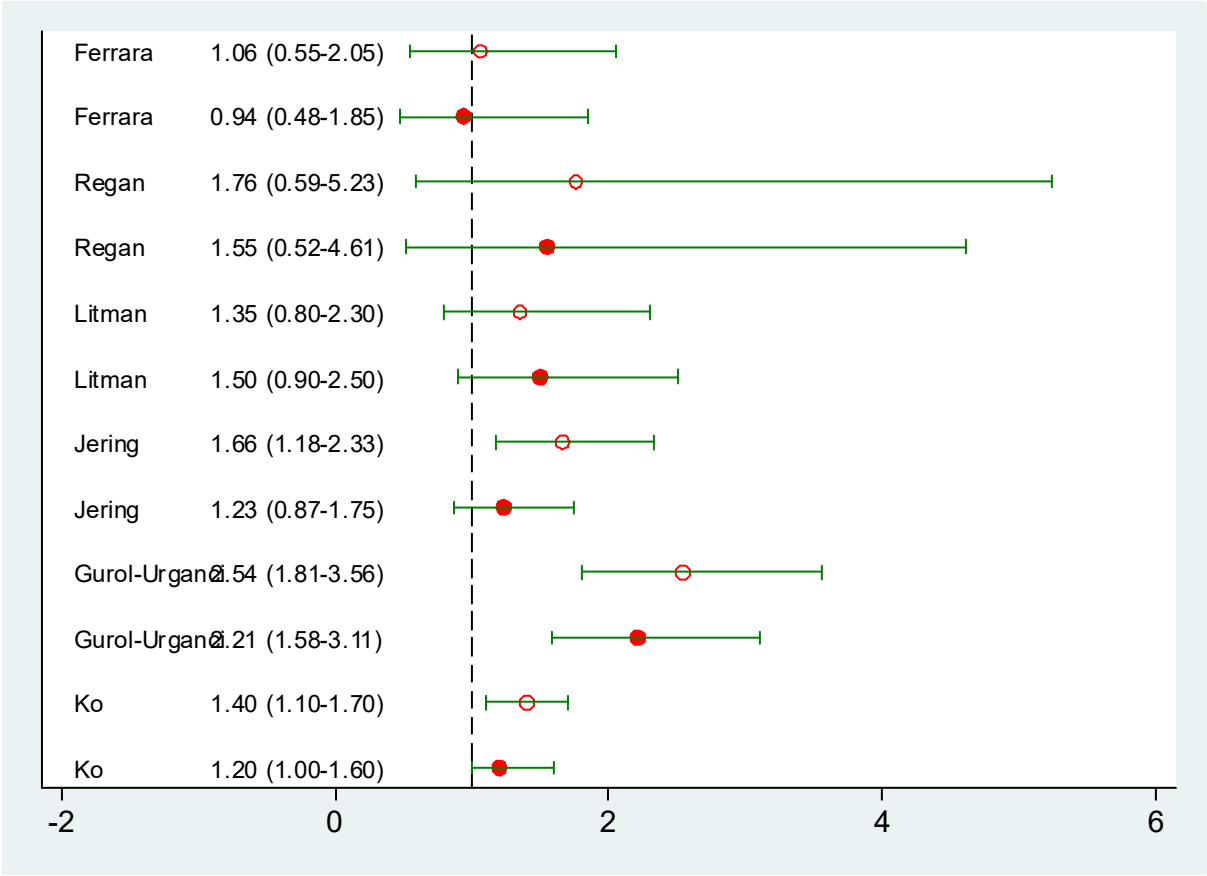


Figure S2: Comparison of adjusted vs crude ratios for studies which presented both



- Indicates an adjusted RR/OR
- Indicates a crude RR/OR

4.6 Chapter Summary

COVID-19 and risk of pregnancy loss evidence to date

- A review of all published and pre-print studies attempting to quantitatively assess the relationship between COVID-19 and pregnancy loss was carried out.
- Thirty-one studies were included in the review, of which 7 investigated miscarriage, 21 investigated stillbirth and 3 investigated all pregnancy loss.
- Quality assessment of the included studies showed variation in the study quality with 12 of the 31 included studies found to be at risk of bias.
- Variations in study quality, the definition of COVID-19 exposure and the definition of miscarriage meant that it was not appropriate to carry out a meta-analysis of all included studies.
- A subgroup meta-analysis found that there is evidence to suggest that having COVID-19 at point of delivery may be associated with an increased risk of stillbirth.
- This systematic review highlighted a need for further research looking at the relationship between COVID-19 in pregnancy and risk of pregnancy loss.

Chapter 5: COVID-19 and Risk of Pregnancy Loss: An Applied Example

5.1 Introduction

In this chapter I present an applied study which builds on my learnings about the CPRD Pregnancy Register to address a pressing pregnancy-related question of public health importance: Is having COVID-19 in pregnancy associated with an increased risk of pregnancy loss?

The study utilised a matched cohort design and the initial study included a cohort of women who had a record of pregnancy which began in the first year of the COVID-19 pandemic in the UK (01/03/2020-01/03/2021). The study compared the outcomes of pregnancies which were exposed to COVID-19 to both contemporary and historical control pregnancies by using a Cox regression model to calculate adjusted hazard ratios. This work is presented as a paper which will be submitted for publication (section 5.2) along with supplementary information (section 5.3).

Uncertain pregnancy episodes were excluded from the primary analysis for this work. I therefore conducted further investigation into the potential impact of this decision based on recommendations that I have previously outlined in the methodological paper presented in section 3.2. The findings of these further investigations are discussed in section 5.4.

5.2 Research Paper Intended for Publication

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	71130	**Title**	Mrs
First Name(s)	Jennifer		
Surname/Family Name	Campbell		
Thesis Title	Investigation and Application of a Pregnancy Register Based on Electronic Primary Care Data		
Primary Supervisor	Krishnan Bhaskaran/ Caroline Minassian		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	The Lancet
Please list the paper's authors in the intended authorship order:	Jennifer Campbell, Rachael Williams, Caroline Minassian, Michael Algeo and Krishnan Bhaskaran
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	JC contributed to the initiation, planning and design of the study. JC performed all of the analysis . JC wrote the manuscript with critical revision from supervisors.

SECTION E

Student Signature	[redacted]
Date	01/03/23

Supervisor Signature	[redacted]
Date	7/3/23

Covid-19 during pregnancy and risk of miscarriage or stillbirth: A matched cohort study using English electronic health data

Jennifer Campbell [1, 2], Rachael Williams [2], Caroline Minassian [1], Michael Algeo[3] and Krishnan Bhaskaran [1]

1. *Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine (LSHTM)*
2. *Clinical Practice Research Datalink (CPRD), Medicines and Healthcare Products Regulatory Agency*
3. *North Bristol NHS Trust*

Corresponding Author:

Jennifer Campbell; jennifer.campbell@mhra.gov.uk

Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency, 10 South Colonnade, Canary Wharf, London, E14 4PU, United Kingdom

Key Words:

COVID-19, Miscarriage, Pregnancy Outcomes, SARS-CoV-2, Spontaneous Abortion, Stillbirth,

Abstract (max 300 words)

Background

COVID-19 is a new disease for which a detrimental impact on pregnant women has been established however, studies assessing its impact on pregnancy outcomes have drawn mixed conclusions (1).

Methods

A matched cohort study design using English primary care electronic health records from the Clinical Practice Research Datalink (CPRD) Aurum database, linked Hospital Episodes Statistics (HES) data and COVID-19 test records. We compared 7,847 pregnancies recorded 01/03/2020 -01/03/2021 with a record of COVID-19 between 28 days gestation and the end of pregnancy. These exposed pregnancies were matched on maternal and gestational age to 23,502 contemporary and 25,488

historical controls (whose pregnancy began between 01/03/2018 and 01/03/2019). A Cox regression model was used to calculate hazard ratios which were adjusted for potential confounders.

Findings

Amongst the exposed pregnancies 301 (3.8%) ended in pregnancy loss compared to 764 (3.3%) among the contemporary controls and 792 (3.1%) among the historical controls. The adjusted hazard ratio for all pregnancy loss showed some evidence of an increased risk amongst the exposed group when compared to the contemporary control group (1.18 95% CI 1.01-1,37). An increased risk was also seen when looking at miscarriage separately (Adj HR 1.17 95% CI 1.00-1.36). We observed no increased risk of stillbirth when compared to the contemporary controls. However, we found an increased risk of all types of pregnancy loss when comparing to historical pre-pandemic controls (Adj HR's: all pregnancy loss 1.39 (1.20-1.60); miscarriage 1.35 (1.16-1.58); stillbirth 2.04 (1.10-3.77)).

Interpretation

We found evidence of that there may be an increased risk of both miscarriage and stillbirth, associated with having COVID-19 during pregnancy, strongly supporting current recommendations that pregnant women are vaccinated against COVID-19 to minimise associated risks.

Funding

This work is funded by Wellcome and CPRD, a division of the UK Medicines and Healthcare Products Regulatory Agency.

Introduction

SARS-CoV-2 emerged as a new coronavirus at the end of 2019 spreading rapidly to cause a global pandemic of its associated illness COVID-19. Many millions of people around the world have been infected with the virus including pregnant women (2). In the UK the first wave of Covid-19 cases began in February 2020 (3).

Since the start of the pandemic there have been a number of studies which looked at the impact of COVID-19 in pregnancy. It is now established that, as with other respiratory illnesses, COVID-19 has a detrimental impact on pregnant women. There has been shown to be an increased risk of severe maternal morbidity and mortality especially among the unvaccinated (4–6). However, studies specifically examining the relationship between COVID-19 and risk of pregnancy loss have found

mixed results (1,7). Many previous studies have been small and often only assessed COVID-19 status at the end of pregnancy, potentially missing any impacts of COVID-19 earlier in pregnancy. Our recently conducted systematic review showed evidence of a potential increased risk of stillbirth among women who had COVID-19 at point of delivery (pooled risk ratio 1.54, 1.21-1.88). (1,8). However we found there were insufficient high-quality studies to draw conclusions about the relationship between COVID-19 and the risk of miscarriage.

A recently developed pregnancy register in a database of UK primary care electronic health records (EHR) Clinical Practice Research Datalink (CPRD) Aurum (9) represented an opportunity to examine the relationship between COVID-19 and pregnancy loss in a large retrospective cohort study. The objective of this study was therefore to evaluate the association between having Covid-19 during pregnancy and the risk of pregnancy loss (miscarriage or stillbirth) using a matched cohort study design.

Methods

Data Sources

CPRD Aurum is a UK database which contains primary care electronic health data for ~70 million patients (10). These data include all information recorded by a General Practitioner (GP) in relation to their patient's care including diagnoses, referrals, tests, and medications prescribed (11). The CPRD Aurum Pregnancy Register is a list of all of the pregnancy episodes within the CPRD Aurum database developed using an algorithmic approach in order to facilitate pregnancy research using CPRD data (9,12). This study utilised the May 2022 version of CPRD Aurum and corresponding Pregnancy Register (13). Investigators had full access to the database population used to create the study population.

In addition to the primary care data approximately 93% of patients in CPRD Aurum are linked at a person-level to English secondary care data sources (14). These data sources include: the Hospital Episode Statistics (HES) Admitted Patient Care which contains details of admissions to hospitals in England (15); Second Generation Surveillance System (SGSS) which contains routine laboratory data on infectious diseases in England including tests for COVID-19; and socioeconomic status data including the 2019 English Index of Multiple Deprivation (IMD)(16) In this study HES APC and SGSS data were used to obtain additional records of COVID-19 and linked IMD data was used as an indicator or socioeconomic status.

Study Population, Exposure and Outcomes

Women with a record of a pregnancy beginning between 01/03/2020 and 01/03/2021 (the first year of the COVID-19 pandemic in the UK) in the CPRD Aurum Pregnancy Register who were linked to secondary care data and whose records met basic data quality criteria were eligible for inclusion in the base population from which exposed, and control patients were selected. Women were required to have at least 280 days of CPRD Aurum data follow-up after their pregnancy start. Women who had a record of COVID-19 or SARS-COv-2 infection in either CPRD Aurum, HES APC or SGSS data prior to their pregnancy were excluded. This was to prevent any bias from the potential effects of long COVID-19. For women who had further pregnancies during the follow-up period only the first pregnancy was eligible for inclusion in the study. Pregnancies which were flagged as conflicting with another pregnancy episode for the same woman in the CPRD Aurum Pregnancy Register were excluded, as were pregnancies which had no outcome recorded in the data, because the timings and outcome of these pregnancies could not be reliably determined. We also included women with a pregnancy beginning between 01/03/2018 and 01/03/2019 as a source of additional historical controls.

Women were classified as being exposed if they had a record of SARS-CoV-2 infection or COVID-19 in either CPRD Aurum, HES APC or SGSS data which was dated between the start and end of pregnancy (code lists provided in Appendix 1). Henceforth exposure will be referred to as COVID-19 for simplicity. Women who did not have a record of COVID-19 before day 28 of pregnancy were eligible to be controls. Exposed patients were randomly matched to 3 controls on both maternal and gestational age at index date. Exposure status was time varying with exposed women contributing time as controls to the analysis until their COVID-19 diagnosis. In addition to contemporary controls a second analysis was conducted comparing the exposed cohort to historical controls who had no risk of COVID-19 and therefore no risk of exposure misclassification. Historical controls pregnancies were matched to the exposed pregnancies 3:1. A flow-diagram detailing the number of pregnancies excluded at each stage is given in Appendix 2.

Outcomes of interest were defined as a miscarriage any time after four weeks of pregnancy or a stillbirth outcome in the CPRD Aurum Pregnancy Register (9).

Covariates

Maternal age was defined as the year of pregnancy start minus the mother's year of birth. Smoking status was taken as the last record related to smoking prior to the pregnancy start and classified as either non-smoker, current-smoker, or ex-smoker. Body-Mass Index (BMI) was taken as the last recorded measurement prior to the pregnancy start. Where BMI was not recorded specifically it was calculated based on the woman's most recent height measurement and her last weight measurement before pregnancy start. BMI measurements were categorised as Not Obese (<30), Obese class 1 (30-34.9), Obese class 2 (35-39.9), Obese Class 3 (>40). Maternal age, smoking status and BMI were all defined using records in CPRD Aurum data. Ethnicity was classified using an ethnicity algorithm created by Shiekh et al which takes data from both CPRD Aurum and linked data to find the most likely ethnicity for each patient (17). Linked IMD quintiles were used as a measure of socioeconomic status.

Chronic Health conditions which were considered to be potential confounders included Diabetes (Type 1 and 2), Gestational diabetes, long term kidney disease, HIV infection, Rheumatoid arthritis and Immunosuppressive drug use regardless of indication. These were defined as any record of the condition in CPRD Aurum prior to pregnancy start with the exception of gestational diabetes which was defined as any record during pregnancy. Code lists used to define covariates are provided in Appendix 3. Only women who had complete data for all variables were included in the analysis

Statistical Analysis

Women contributed follow-up time from day 28 of pregnancy, the earliest date possible for the outcome to occur. Follow-up ended at the earliest of pregnancy end, death or their end of follow-up in CPRD Aurum. The index date was the date of their earliest COVID-19 record. A Cox regression model was used to ascertain the hazard ratio of the risk of pregnancy loss amongst women who had COVID-19 during their pregnancy compared with those who did not. The model was used to adjusted for all of the covariates outlined above.

Hazard ratios were calculated for all pregnancy loss, miscarriage only and stillbirth only (Table 3). Log-log survival curves were used to check the proportional hazards assumption before partitioning the time axis and fitting an interaction by trimester at index date. Trimester specific hazard ratios were calculated for all types of pregnancy loss combined (Table 4). Statistical analysis was carried out using STATA 17(18).

The number of GP surgery consultations women had recorded in CPRD Aurum between the start and end of pregnancy was examined as an indication of level of care given. The number and percentage of women who had at least one of each type of contact was calculated along with the mean number of contacts per pregnancy by type (Table 5) using records in the CPRD Aurum Consultation Table (19).

Results

From the CPRD Aurum Pregnancy Register there were 205,884 pregnancies in the main cohort pool and 280,979 in the historical pool which met the cohort inclusion criteria. Of these there were 7,847 pregnancies where the woman had her first ever record of COVID-19 between 28 days gestation and the end of pregnancy. These exposed pregnancies were matched to 23,502 contemporary controls from the main pool and 25,488 historical controls.

Tables 1a and 1b show the mean follow-up time and covariate distribution for the exposed pregnancies compared to the contemporary and historical control pregnancies. Mean follow-up time was comparable between the exposed pregnancies and both groups of controls (~34 weeks). The vast majority of the pregnant women were between the ages of 21 and 40 (94.2%). The distribution of socioeconomic status, smoking status, ethnicity, body-mass index, and chronic co-morbidities was similar between the exposed group and the controls (Table 1a and 1b). Table 2 shows the distribution of pregnancy outcomes. Amongst the exposed pregnancies there were 301 (3.8%) which ended in pregnancy loss compared to 764 (3.3%) among the contemporary controls and 792 (3.1%) among the historical controls.

After adjusting for potential confounders (outlined in Table 1) the hazard ratio for all pregnancy loss across the whole pregnancy showed some evidence of an increased risk amongst the exposed group when compared to the contemporary control group (1.18 95% CI 1.01-1,37). An increased risk was also seen when looking at miscarriage separately (Adj HR 1.17 95% CI 1.00-1.36). We observed no increased risk of stillbirth when compared to the contemporary controls. However, for all three outcomes, the estimated hazard ratios were substantially larger when we used historical pre-pandemic controls. There was no evidence that the HR varied according to stage of pregnancy (p=0.48 and 0.47) for the contemporary and historical cohorts respectively.

The proportion of women who had at least one GP consultation recorded between the start and end of their pregnancy was higher amongst the 2018 cohort as was the mean number of contacts per woman (Table 5).

Table 1a: Distribution of Outcomes and Covariates among the exposed women and the matched contemporary controls.

	Exposed N= 7,847	**Contemporary Controls N= 23,502**
Follow-up		
Mean Follow-up (days)	239.09	241.50
Age Group		
11-20	221 (2.8%)	655 (2.8%)
21-30	3,592 (45.8%)	10,764 (45.8%)
31-40	3,797 (48.4%)	11,391 (48.5%)
41-49	237 (3.0%)	692 (2.9%)
IMD Quintile (Patient Level)		
1 (most deprived)	1,090 (13.9%)	4,054 (17.2%)
2	1,371 (17.5%)	4,357 (18.5%)
3	1,540 (19.6%)	4,532 (19.3%)
4	1,806 (23.0%)	5,078 (21.6%)
5 (least deprived)	2,040 (26.0%)	5,481 (23.3%)
Smoking Status		
Non smoker	5,154 (65.7%)	15,236 (64.8%)
Current Smoker	1,590 (20.3%)	5,078 (21.6%)
Ex-Smoker	1,103 (14.1%)	3,188 (13.6%)
Ethnicity		
White	5,933 (75.6%)	18,563 (79.0%)
Mixed	176 (2.2%)	591 (2.5%)
Asian	1,259 (16.0%)	2,851 (12.1%)
Black	423 (5.4%)	1,308 (5.6%)
Other	56 (0.7%)	189 (0.8%)
BMI		
Not Obese	5,765 (73.5%)	17,623 (75.0%)
Obese class 1	1,076 (13.7%)	3,076 (13.1%)
Obese class 2	494 (6.3%)	1,291 (5.5%)
Obese class 3	512 (6.5%)	1,512 (6.4%)

History of Chronic Conditions		
Chronic Kidney Disease	14 (0.2%)	49 (0.2%)
HIV	0	14 (0.1%)
Rheumatoid Arthritis	26 (0.3%)	65 (0.3%)
Gestational Diabetes	576 (7.3%)	1,637 (7.0%)
Diabetes	90 (1.1%)	249 (1.1%)
Immunosuppressive Drug Use	25 (0.3%)	75 (0.3%)

Table 1b: Distribution of Outcomes and Covariates among the exposed women and the matched historical controls.

	Exposed N= 7847	Historical Controls N= 25488
Follow-up		
Mean Follow-up (days)	239.10	240.20
Age Group		
11-20	222 (2.8%)	728 (2.9%)
21-30	3,593 (45.8%)	11,631 (45.6%)
31-40	3,797 (48.4%)	12,362 (48.5%)
41-49	236 (3.0%)	767 (3.0%)
IMD Quintile (Patient Level)		
1 (most deprived)	1,092 (13.9%)	4,419 (17.3%)
2	1,370 (17.5%)	4,628 (18.2%)
3	1,540 (19.6%)	4,823 (18.9%)
4	1,804 (23.0%)	5,718 (22.4%)
5 (least deprived)	2,042 (26.0%)	5,900 (23.1%)
Smoking Status		
Non smoker	5,157 (65.7%)	16,254 (63.8%)
Current Smoker	1,589 (20.2%)	5,483 (21.5%)
Ex-Smoker	1,102 (14.0%)	3,751 (14.7%)
Ethnicity		
White	5,934 (75.6%)	20,030 (78.6%)
Mixed	176 (2.2%)	612 (2.4%)
Asian	1,259 (16.0%)	3,179 (12.5%)

Black	423 (5.4%)	1,516 (5.9%)
Other	56 (0.7%)	151 (0.6%)
BMI		
Not Obese	5,766 (73.5%)	19,195 (75.3%)
Obese class 1	1,075 (13.7%)	3,356 (13.2%)
Obese class 2	495 (6.3%)	1,440 (5.6%)
Obese class 3	512 (6.5%)	1,497 (5.9%)
History of Chronic Conditions		
Chronic Kidney Disease	14 (0.2%)	44 (0.2%)
HIV	0	22 (0.1%)
Rheumatoid Arthritis	27 (0.3%)	85 (0.3%)
Gestational Diabetes	576 (7.3%)	1,443 (5.7%)
Diabetes	90 (1.1%)	256 (1.0%)
Immunosuppressive Drug Use	25 (0.3%)	96 (0.4%)

Table 2: Distribution of pregnancy outcomes

	Exposed N= 7,847	**Contemporary Controls N= 23,502**	**Historical Controls N= 25488**
Pregnancy Outcomes			
Miscarriage	266 (3.4%)	681 (2.9%)	718 (2.8%)
Stillbirth	35 (0.4%)	83 (0.4%)	74 (0.3%)
Livebirth	7,284 (92.8%)	22,001 (93.6%)	23,632 (92.7%)
Other	262 (3.3%)	737 (3.1%)	1,064 (4.2%)

Table 3: Hazard ratios by type of pregnancy loss for exposed vs contemporary and historical controls

	Contemporary Controls		**Historical Controls**	
	Crude HR (95%CI)	**Adjusted HR (95%CI)**	**Crude HR (95%CI)**	**Adjusted HR (95%CI)**
All Pregnancy Loss	1.15 (0.99 - 1.33)	1.18 (1.01 - 1.37)	1.38 (1.20-1.60)	1.39 (1.20-1.60)
Miscarriage	1.14 (0.98 -1.32)	1.17 (1.00-1.36)	1.34 (1.16-1.56)	1.35 (1.16-1.58)
Stillbirth	1.30 (0.70-2.40)	1.35 (0.67-2.76)	1.91 (1.13-3.23)	2.04 (1.10-3.77)

Table 4: Hazard ratios for all pregnancy loss by pregnancy trimester for exposed vs contemporary and historical controls

	Contemporary Controls		Historical Controls	
	Crude HR (95%CI)	**Adjusted(95%CI) HR**	**Crude HR (95%CI)**	**Adjusted(95%CI) HR**
1st trimester	1.10 (0.95- 1.29)	1.11 (0.95-1.30)	1.25 (1.08-1.46)	1.26 (1.08-1.47)
2nd trimester	1.57 (0.76- 3.26)	1.58 (0.74- 3.35)	1.87 (0.94-3.72)	1.80 (0.89- 3.62)
3rd trimester	1.26 (0.63- 2.56)	1.33 (0.66- 2.79)	1.50 (0.80- 2.82)	1.50 (0.79-2.86)

Table 5: Descriptive statistics of the number GP surgery consultations during pregnancy for women in each cohort.

	2020/21 Cohort		2018/19 Cohort	
Type of Contact	**Number & percentage of women who had at least one of each contact type N= 30,291**	**Mean number of contacts per pregnancy by contact type***	**Number & percentage of women who had at least one of each contact type N= 25,486**	**Mean number of contacts per pregnancy by contact type***
Specialist Clinic	92 (0.3%)	1.58	334 (1.3%)	2.23
GP Appointment	26,196 (86.5%)	5.19	24,413 (95.8%)	7.90
GP Home Visit	5 (0.0%)	1.00	104 (0.4%)	1.55
GP Telephone	24,687 (81.5%)	4.48	12,627 (49.5%)	2.97
Nurse	39 (0.1%)	1.59	14 (0.1%)	1.43

*Among women who had at least one contact record

Discussion

We found evidence that women who had COVID-19 during pregnancy had an 18% higher risk of pregnancy loss compared with women who were pregnant during the same calendar period but with no record of COVID-19, and a 39% higher risk of pregnancy loss when compared with women who were pregnant pre-pandemic. The increased risks for stillbirths were particularly pronounced, with women exposed to COVID-19 having around double the risk for this outcome compared with pre-pandemic controls.

The work was conducted in a large cohort of women who are generalisable to the UK population. Only a few studies have previously looked at COVID-19 exposure and risk of miscarriage (1) . Our findings build on previous work by Balachandran et al who reported an increased risk of miscarriage among women who were infected with COVID-19 during their first trimester (20). Another large study by Regan et al assessed a range of adverse pregnancy outcomes including loss in US claims data and found an increased risk of miscarriage and fetal harm (6). Calvert et al examined risk of miscarriage following COVID-19 compared to both contemporary and historical controls in Scottish EHR data, however, they found no evidence of an increased risk in either comparison their work was conducted using data post vaccine roll-out (21). Large studies which have looked at COVID-19 exposure across the whole pregnancy and risk of stillbirth include Ferrara et al and Litman et al which both found no increased risk (22,23). However, several smaller studies have observed an increased risk. We previously conducted a meta-analysis of eight studies which looked at risk of stillbirth among women who were infected with COVID-19 at point of delivery and found a higher risk in the exposed women after pooling the results (1).

Previous work by Mansfield et al found that there was a reduction in the number of healthcare contacts during the early stages of the COVID-19 pandemic (24). We therefore calculated the proportion of women in our pandemic and historical cohorts who contacted their GP at least once during pregnancy, by type of contact, in order to investigate whether there was any change in the pattern of care between the two time periods which could have contributed to an increased risk of pregnancy loss in the pandemic cohort. Whilst our results showed a reduction in the mean number of times the women who were pregnant during the pandemic saw their GP in person there was an increase in the mean number of telephone consultations and 86.5% of the women had a record of seeing a GP in person at least once. It is difficult to comment without further investigation as to whether change in care may have had a direct impact on risk of pregnancy loss. Nevertheless given

that we still found an increased risk of pregnancy loss when comparing to unexposed women who were pregnant during the pandemic, change in care would not offer a full explanation.

Our study period covers the first year of the COVID-19 pandemic in the UK when the original Wuhan strain of the SARS-CoV2 variant was in circulation. During this period, testing for COVID-19 was initially not available to the UK public with mass testing first becoming available in May 2020 for symptomatic cases only (25). COVID-19 is not always symptomatic, and many cases may have gone unreported. It is therefore possible that some of the women in our contemporary control group are actually misclassified exposed patients which would reduce the hazard ratio seen if there was a true increased risk of pregnancy loss associated with COVID-19. This may explain the higher hazard ratio found when comparing to women who were pregnant in 2018 who definitely did not have COVID-19.

EHR data are a very useful resource for large scale observational studies such as this one. UK primary care health data such as CPRD Aurum are particularly useful due to the gate-keeper healthcare system meaning a large proportion of health events are recorded in primary care allowing for longitudinal follow-up and good capture of events. However, data are not collected primarily for research purposes and can therefore be incomplete. Our decision to exclude pregnancy episodes without an outcome recorded from the study cohort may have resulted in some pregnancy loss outcomes being undetected (26) . However, unless COVID-19 patients were at higher risk of pregnancy loss and had a different probability of having their outcome being unrecorded than the unexposed controls this is unlikely to have resulted in a difference between the two groups. We plan to conduct further work to look at whether obtaining additional pregnancy outcomes from linked secondary care data results in any changes to our findings.

One potential limitation of our study is that we were unable to find a set of matches for all of the cases in the cohort due to the requirement to have three matches per case. This resulted in some cases being dropped before the analysis. We intend to carry investigations as to whether reducing to two or one matched control per case has any impact on the observed association.

A further limitation of our study is that we were not able to assess COVID-19 severity and whether it has any impact on the observed relationship. However, given that many of our exposed patients had a COVID-19 record before routine testing was introduced it is likely that they were at least symptomatic. Given that the data used for this study was from the early stages of the pandemic we are unable to comment on how our findings generalise to currently circulating variants. It is also important to note that the study period is before COVID-19 vaccination was available to pregnant women in the UK (27) and therefore all women in the cohort are unvaccinated. Further research is

warranted to investigate whether vaccination status has an impact on the observed risk of pregnancy loss. However, our findings lend weight to the importance of pregnant women being vaccinated to protect against COVID-19 during pregnancy.

Conclusion

The results of this study suggest that there may be an increased risk of pregnancy loss, both miscarriage and stillbirth, associated with having COVID-19 during pregnancy, strongly supporting current recommendations that pregnant women are vaccinated against COVID-19 to minimise associated risks.

References

1. Jennifer Campbell, Mia Harley, Rachael Williams, Krishnan Bhaskaran. COVID-19 and risk of pregnancy loss (miscarriage or stillbirth): a systematic review and meta-analysis. Under Review- BMJ Open. 2023;
2. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2022 Aug 15]. Available from: https://covid19.who.int/
3. UK Government. Coronavirus (COVID-19) in the UK [Internet]. [cited 2022 Jan 24]. Available from: https://coronavirus.data.gov.uk/
4. Villar J, Soto Conti CP, Gunier RB, Ariff S, Craik R, Cavoretto PI, et al. Pregnancy outcomes and vaccine effectiveness during the period of omicron as the variant of concern, INTERCOVID-2022: a multinational, observational study. The Lancet. 2023 Feb;
5. Allotey J, Chatterjee S, Kew T, Gaetano A, Stallings E, Fernández-García S, et al. SARS-CoV-2 positivity in offspring and timing of mother-to-child transmission: Living systematic review and meta-analysis. The BMJ. 2022;
6. Regan AK, Arah OA, Fell DB, Sullivan SG. SARS-CoV-2 Infection during Pregnancy and Associated Perinatal Health Outcomes: A National US Cohort Study. Journal of Infectious Diseases. 2022 Mar 1;225(5):759–67.

7. Pathirathna ML, Samarasekara BPP, Dasanayake TS, Saravanakumar P, Weerasekara I. Adverse Perinatal Outcomes in COVID-19 Infected Pregnant Women: A Systematic Review and Meta-Analysis. Healthcare (Switzerland). 2022;10(2).

8. Campbell J, Williams R, Harley M, Bhaskaran K. Covid-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review protocol. BMJ Open [Internet]. 2022;12. Available from: https://bmjopen.bmj.com/content/12/10/e065588.full

9. Campbell J, Shepherd H, Welburn S, Barnett R, Oyinlola J, Oues N, et al. Methods to refine and extend a Pregnancy Register in the <scp>UK</scp> Clinical Practice Research Datalink primary care databases. Pharmacoepidemiol Drug Saf [Internet]. 2022 Dec 15; Available from: https://onlinelibrary.wiley.com/doi/10.1002/pds.5584

10. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile : Clinical Practice Research Datalink (CPRD) Aurum. 2019;1–8.

11. CPRD. CPRD: Clinical Practice Research Datalink [Internet]. 2021 [cited 2018 Sep 17]. Available from: www.cprd.com

12. Minassian C, Williams R, Meeraus WH, Smeeth L, Campbell OMR, Thomas SL. Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. Pharmacoepidemiol Drug Saf [Internet]. 2019;(April):1–11. Available from: http://doi.wiley.com/10.1002/pds.4811

13. CPRD. Release Notes: CPRD Aurum May 2022. https://cprd.com/cprd-aurum-may-2022-dataset. 2022.

14. Gallagher AM, Puri S van ST. Linkage of the General Practice Research Database (GPRD) with Other Data Sources. In: Linkage of the General Practice Research Database (GPRD) with Other Data Sources. International Conference of Pharmacoepidemiology; 2011.

15. NHS Digital. Hospital Episode Statistics [Internet]. 2021 [cited 2018 Sep 17]. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

16. Department for Communities and Local Government. English Indices of Deprivation 2015- Statistical Release. 2015.

17. Shiekh SI, Harley M, Ghosh RE, Ashworth M, Myles P, Booth HP, et al. Completeness, agreement, and representativeness of ethnicity recording in the United Kingdom's Clinical Practice Research Datalink (CPRD) and linked Hospital Episode Statistics (HES). Popul Health Metr [Internet]. 2023 Mar

14;21(1):3. Available from: https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-023-00302-0

18. StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.; 2021.

19. CPRD. CPRD Aurum Data Specification [Internet]. 2022 [cited 2023 Feb 15]. Available from: https://cprd.com/primary-care-data-public-health-research

20. Balachandren N, Davies MC, Hall JA, Stephenson JM, David AL, Barrett G, et al. SARS-CoV-2 infection in the first trimester and the risk of early miscarriage: A UK population-based prospective cohort study of 3041 pregnancies conceived during the pandemic. Human Reproduction. 2022 Jun 1;37(6):1126–33.

21. Calvert C, Carruthers J, Denny C, Donaghy J, Hillman S, Hopcroft LEM, et al. A population-based matched cohort study of early pregnancy outcomes following COVID-19 vaccination and SARS-CoV-2 infection. Nat Commun. 2022 Dec 1;13(1).

22. Litman EA, Yin Y, Nelson SJ, Capbarat E, Kerchner D, Ahmadzia HK. Adverse perinatal outcomes in a large United States birth cohort during the COVID-19 pandemic. Am J Obstet Gynecol MFM. 2022 May 1;4(3).

23. Ferrara A, Hedderson MM, Zhu Y, Avalos LA, Kuzniewicz MW, Myers LC, et al. Perinatal Complications in Individuals in California with or Without SARS-CoV-2 Infection during Pregnancy. JAMA Intern Med. 2022 May 1;182(5):503–12.

24. Mansfield KE, Mathur R, Tazare J, Henderson AD, Mulick AR, Carreira H, et al. Indirect acute effects of the COVID-19 pandemic on physical and mental health in the UK: a population-based study. Lancet Digit Health. 2021;3(4):e217–30.

25. Briggs A, Jenkins D, Fraser C. NHS Test and Trace: the journey so far [Internet]. 2020 [cited 2023 Feb 15]. Available from: https://www.health.org.uk/publications/long-reads/nhs-test-and-trace-the-journey-so-far

26. Campbell J, Bhaskaran K, Thomas S, Williams R, Mcdonald HI, Minassian C. Investigating the optimal handling of uncertain pregnancy episodes in the CPRD GOLD Pregnancy Register: a methodological study using UK primary care data. BMJ Open [Internet]. 2022;12:55773. Available from: http://dx.doi.org/10.1136/bmjopen-2021-055773

27. UK Health Security Agency. COVID-19: The Green Book chapter 14a. In: The Green Book [Internet]. 2022. Available from: https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a%0A

5.3 Research Paper Supplementary Files

Appendix 1: COVID-19 code lists

Table S1, ICD 10 codes used in HES APC

ICD-10	TERM
U07.1	Covid-19, Virus Identified
U10	Multisystem inflammatory syndrome associated with COVID-19
U10.9	Multisystem inflammatory syndrome associated with COVID-19, unspecified

Table S2, Codes used in CPRD Aurum

medcodeid	term
13045121000006100.00	2019-nCoV (novel coronavirus) IgA detected
13032641000006100.00	2019-nCoV (novel coronavirus) IgG detected
13032661000006100.00	2019-nCoV (novel coronavirus) IgM detected
13053191000006100.00	2019-nCoV (novel coronavirus) antibody detection result positive
13053051000006100.00	2019-nCoV (novel coronavirus) antigen detection result positive
12990741000006100.00	2019-nCoV (novel coronavirus) detected
13483601000006100.00	2019-nCoV (novel coronavirus) detection result positive at the limit of detection
13483161000006100.00	2019-nCoV (novel coronavirus) ribonucleic acid detected
14168601000006100.00	Acute COVID-19
13486461000006100.00	Acute COVID-19 infection
13486511000006100.00	Acute bronchitis caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
14168611000006100.00	Acute disease caused by Severe acute respiratory syndrome coronavirus 2
13486471000006100.00	Acute disease caused by severe acute respiratory syndrome coronavirus 2 infection

13485251000006100.00	Acute hypoxemic respiratory failure due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13485211000006100.00	Acute kidney injury due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13485231000006100.00	Acute kidney injury due to disease caused by Severe acute respiratory syndrome coronavirus 2
13486661000006100.00	Acute respiratory distress syndrome due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486691000006100.00	Acute respiratory distress syndrome due to disease caused by Severe acute respiratory syndrome coronavirus 2
13486311000006100.00	Assessment using C19-YRS (COVID-19 Yorkshire Rehabilitation Screening) tool
13002271000006100.00	Assessment using COVID-19 severity scale
13486281000006100.00	Assessment using Newcastle post-COVID syndrome Follow-up Screening Questionnaire
13486351000006100.00	Assessment using PCFS (Post-COVID-19 Functional Status) Scale patient self-report
13486411000006100.00	Assessment using PCFS (Post-COVID-19 Functional Status) Scale structured interview
13486421000006100.00	Assessment using Post-COVID-19 Functional Status Scale structured interview
13486571000006100.00	Asymptomatic COVID-19
13486561000006100.00	Asymptomatic SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infection
13486291000006100.00	C19-YRS (COVID-19 Yorkshire Rehabilitation Screening) tool
13012431000006100.00	COVID-19
13483031000006100.00	COVID-19
13486301000006100.00	COVID-19 Yorkshire Rehabilitation Screening tool
13486521000006100.00	COVID-19 acute bronchitis
13486671000006100.00	COVID-19 acute respiratory distress syndrome
13012441000006100.00	COVID-19 caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)

13002281000006100.00	COVID-19 confirmed by laboratory test
13002311000006100.00	COVID-19 confirmed clinically
13002291000006100.00	COVID-19 confirmed using clinical diagnostic criteria
13483951000006100.00	COVID-19 detected
13486761000006100.00	COVID-19 lower respiratory infection
13484091000006100.00	COVID-19 pneumonia
13002261000006100.00	COVID-19 severity scale
13002251000006100.00	COVID-19 severity score
13484031000006100.00	Cardiomyopathy due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
14168631000006100.00	Chronic post-COVID-19 syndrome
12802201000006100.00	Confirmed 2019-nCoV (novel coronavirus) infection
13485901000006100.00	Conjunctivitis due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13485401000006100.00	Detection of 2019 novel coronavirus antigen
13485791000006100.00	Detection of 2019-nCoV (novel coronavirus)
12990531000006100.00	Detection of 2019-nCoV (novel coronavirus) using polymerase chain reaction technique
13485771000006100.00	Detection of COVID-19
13485461000006100.00	Detection of RNA (ribonucleic acid) of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) in nasopharyngeal swab
13485521000006100.00	Detection of RNA (ribonucleic acid) of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) in oropharyngeal swab
13485581000006100.00	Detection of RNA (ribonucleic acid) of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) in sputum
13485701000006100.00	Detection of RNA (ribonucleic acid) of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) using polymerase chain reaction
13485751000006100.00	Detection of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13485341000006100.00	Detection of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) antibody

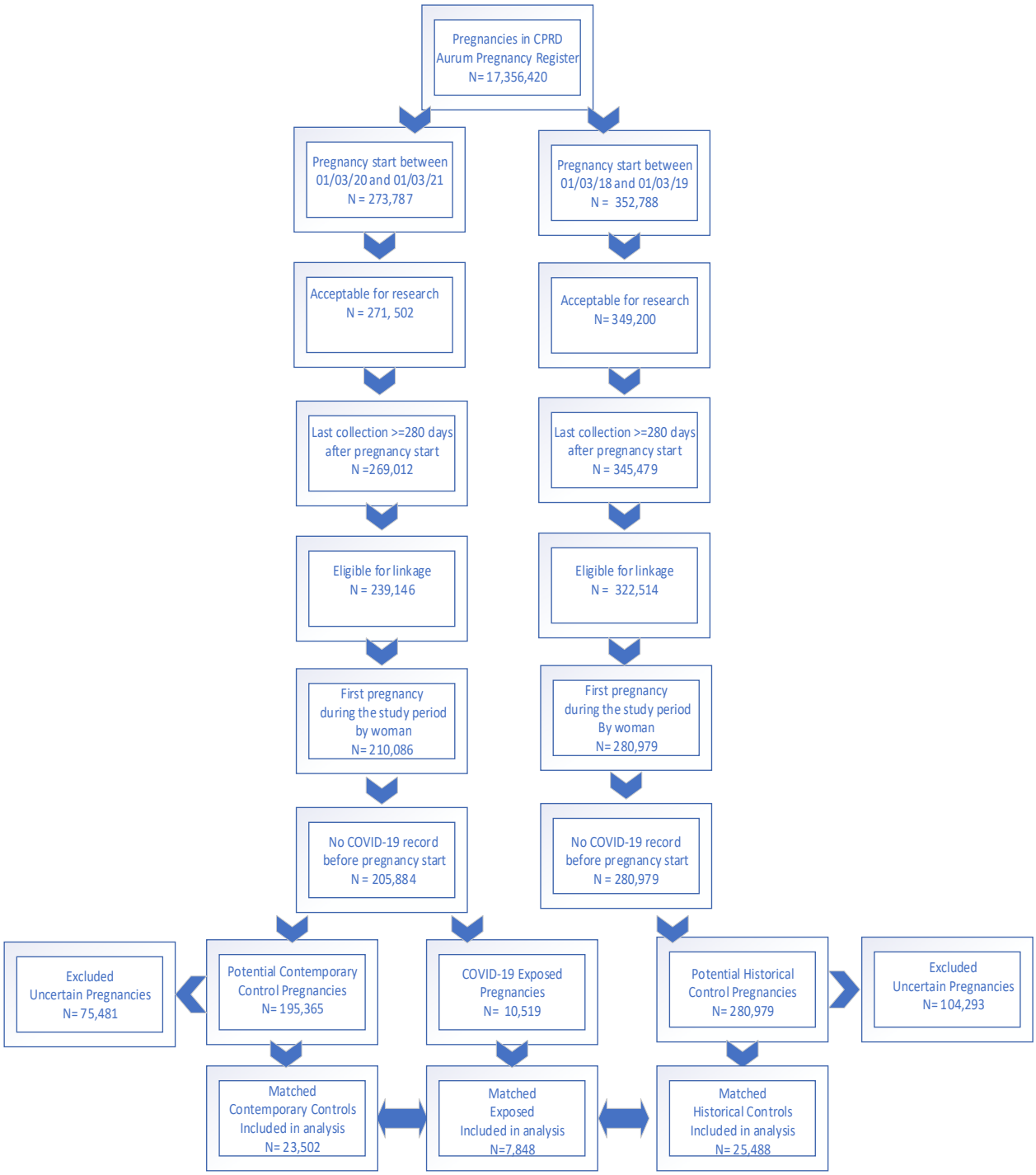
13485381000006100.00	Detection of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) antigen
13012231000006100.00	Detection of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) using polymerase chain reaction technique
13485471000006100.00	Detection of ribonucleic acid of 2019 novel coronavirus in nasopharyngeal swab
13485711000006100.00	Detection of ribonucleic acid of COVID-19 using polymerase chain reaction
13485421000006100.00	Detection of ribonucleic acid of Severe acute respiratory syndrome coronavirus 2
13483041000006100.00	Disease caused by 2019 novel coronavirus
13483051000006100.00	Disease caused by 2019-nCoV
12991131000006100.00	Disease caused by 2019-nCoV (novel coronavirus)
13483061000006100.00	Disease caused by Severe acute respiratory syndrome coronavirus 2
12991141000006100.00	Disease caused by Wuhan 2019-nCoV (novel coronavirus)
13485991000006100.00	Dyspnoea caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13012281000006100.00	Encephalopathy caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13483911000006100.00	Encephalopathy due to COVID-19
13483891000006100.00	Encephalopathy due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13483921000006100.00	Encephalopathy due to disease caused by Severe acute respiratory syndrome coronavirus 2
13485961000006100.00	Fever caused by 2019 novel coronavirus
13485951000006100.00	Fever caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
12990711000006100.00	Gastroenteritis caused by 2019-nCoV (novel coronavirus)
13012291000006100.00	Gastroenteritis caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486631000006100.00	History of COVID-19
13486611000006100.00	History of disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)

13483871000006100.00	Infection of upper respiratory tract caused by 2019 novel coronavirus
13483881000006100.00	Infection of upper respiratory tract caused by Severe acute respiratory syndrome coronavirus 2
13486741000006100.00	Lower respiratory infection caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486771000006100.00	Lower respiratory infection caused by Severe acute respiratory syndrome coronavirus 2
13484801000006100.00	Lymphocytopenia due to SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13484821000006100.00	Lymphopenia due to COVID-19
12990591000006100.00	Myocarditis caused by 2019-nCoV (novel coronavirus)
13012251000006100.00	Myocarditis caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13483841000006100.00	Myocarditis due to COVID-19
13483821000006100.00	Myocarditis due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486271000006100.00	Newcastle post-COVID syndrome Follow-up Screening Questionnaire
13486481000006100.00	Ongoing symptomatic COVID-19
13486491000006100.00	Ongoing symptomatic disease caused by severe acute respiratory syndrome coronavirus 2
13012241000006100.00	Otitis media caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13483791000006100.00	Otitis media due to COVID-19
13483771000006100.00	Otitis media due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486331000006100.00	PCFS (Post-COVID-19 Functional Status) Scale patient self-report
13486371000006100.00	PCFS (Post-COVID-19 Functional Status) Scale patient self-report final scale grade
13486431000006100.00	PCFS (Post-COVID-19 Functional Status) Scale structured interview

13486391000006100.00	PCFS (Post-COVID-19 Functional Status) Scale structured interview final scale grade
13484101000006100.00	Pneumonia caused by 2019 novel coronavirus
13484121000006100.00	Pneumonia caused by 2019-nCoV (novel coronavirus)
12990651000006100.00	Pneumonia caused by 2019-nCoV (novel coronavirus)
13484081000006100.00	Pneumonia caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13012271000006100.00	Pneumonia caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
13486451000006100.00	Post-COVID-19 syndrome
14176451000006100.00	Post-COVID-19 syndrome resolved
14168621000006100.00	Post-acute COVID-19
13486261000006100.00	Referral to Your COVID Recovery rehabilitation platform
14259201000006100.00	Referral to long term effects of COVID-19 assessment clinic
13486251000006100.00	Referral to post-COVID assessment clinic
13483611000006100.00	SARS (severe acute respiratory syndrome) coronavirus 2 RNA
6874141000006110.00	SARS-CoV
6604671000006110.00	SARS-CoV infection
13045111000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) IgA detection result positive
13032631000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) IgG detection result positive
13032501000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) IgG qualitative existence in specimen
13032651000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) IgM detection result positive
13483141000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) RNA (ribonucleic acid) detection result positive
13052351000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) RNA (ribonucleic acid) detection result positive
13483591000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) RNA (ribonucleic acid) detection result positive at the limit of detection

13052381000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) RNA (ribonucleic acid) qualitative existence in specimen
14333901000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) anti-spike IgG detection result positive
13053181000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) antibody detection result positive
14146661000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) antibody detection result positive
13053041000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) antigen detection result positive
13012301000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) detected
13039281000006100.00	SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) detection result positive
13484481000006100.00	SARS-CoV-2 - severe acute respiratory syndrome coronavirus 2
14168991000006100.00	SARS-CoV-2 mRNA
14164811000006100.00	SARS-CoV-2 viraemia
13485171000006100.00	Sepsis due to disease caused by COVID-19
13485161000006100.00	Sepsis due to disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
14146671000006100.00	Severe acute respiratory syndrome coronavirus 2 antibody test positive
13483941000006100.00	Severe acute respiratory syndrome coronavirus 2 detected
13483151000006100.00	Severe acute respiratory syndrome coronavirus 2 ribonucleic acid detected
13484871000006100.00	Thrombocytopenia due to SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
12990621000006100.00	Upper respiratory tract infection caused by 2019-nCoV (novel coronavirus)
13012261000006100.00	Upper respiratory tract infection caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)
12990751000006100.00	Wuhan 2019-nCoV (novel coronavirus) detected

Appendix 2: Figure S1, Cohort Selection



Appendix 3: Code Lists used to define covariates

All covariate code lists are available at the following link: https://doi.org/10.17037/DATA.00003323.

5.4 Investigating the Potential Impact of Excluding Uncertain Pregnancy Episodes

In the research paper presented in section 5.2, uncertain pregnancy episodes (pregnancies with no recorded outcome and those which conflict with another pregnancy episode for the same woman) were excluded at the cohort selection stage. The assumption was made that the distribution of these pregnancies and whether or not they ended in loss would not differ between the women exposed to COVID-19 and the control groups. This section of the thesis looks at these uncertain pregnancy episodes in more detail as per the recommendations in the paper presented in section 3.2 in which I outlined the following four considerations for studies investigating an exposure in pregnancy and its impact on outcomes:

1. Consider ensuring pregnancy start is at least 9 months before the last data collection date to allow for attainment of outcomes.
2. Exclude uncertain pregnancy episodes which are likely to be derived from historical data based on our described scenarios.
3. Consider utilising linked data to obtain additional outcomes restricting the study population to those patients eligible for linkage
4. Consider merging conflicting episodes which are likely to represent a true and current pregnancy that has been split into two separate conflicting episodes by the rules of the algorithm (problem 4 (1)). Based on the scenarios described in section 3.2 identify the likely true outcome and adjust the pregnancy start and end dates accordingly.

The first of these recommendations was implemented when selecting the original cohort. Only pregnancies with at least 280 days (9 months) of CPRD follow-up were included in order to allow for ascertainment of pregnancy outcomes. The second recommendation was redundant, as all uncertain pregnancy episodes were excluded as part of the cohort definition. Below I present some further analyses that incorporate the remaining recommendations (numbers 3 and 4, above)

5.4.1 Obtaining Additional Pregnancy Outcomes from HES linked data

I investigated the potential implications of the third recommendation, to look for additional outcomes in the linked HES data.

Methods

The latest available linked HES APC data has a coverage period of 01/04/1997- 31/03/2021. In order to investigate whether there were additional pregnancy outcomes available in HES I restricted to pregnancy episodes with no outcome where the start date in the register was between 01/03/2020 and the 10/06/2020. This was to allow for enough HES APC follow-up to ascertain pregnancy outcomes. I correspondingly reduced the recruitment period for the historical controls to 01/03/2018 – 10/06/2018.

I used ICD-10 and OPCS code lists to look for evidence of outcomes for these pregnancy episodes in the HES APC Episodes, Diagnosis and Maternity tables using the methodology detailed in the paper in section 3.2. Only outcome codes which were able to be clearly categorised were included, for example OPCS codes for which it was impossible to differentiate between a miscarriage and a termination were excluded (e.g. Q10 Curettage of uterus).

Where classifiable additional outcomes were found the length of the pregnancy episode was adjusted so that the pregnancy end date was that of the pregnancy outcome found in HES. I looked for COVID-19 records between the start and the new end date of pregnancy and reclassified pregnancy episodes as exposed accordingly. The remaining pregnancy episodes with no classifiable outcome in HES were excluded from the cohort. I then re-ran the Cox regression model described in section 5.2 first using the new smaller cohort and then again including the additional HES outcomes plus timing and exposure adjustments. Conflicting episodes were still excluded from this analysis.

Results

There were 19,544 contemporary pregnancy episodes (exposed and controls) and 29,445 historical pregnancy episodes within the shortened time periods which had no outcome recorded in the register and which otherwise met the inclusion criteria for the study. Of the 19,544 pregnancies with no recorded outcome in the new shortened contemporary cohort 8,167 (42%) had a potential outcome recorded in HES APC. There were also 12,360/29,455 (42%) pregnancies with no recorded outcome in the historical control cohort for which an outcome was found in HES. This is slightly lower than the results seen in chapter 3 where outcomes were found in HES for 50% of pregnancies with no recorded outcome. The adjustment of pregnancy dates resulted in an additional 278 pregnancies being classified as exposed to COVID-19 (Table 1).

Table 1 The distribution of additional pregnancy outcomes found in HES before and after the adjustment of exposure timing (before matching).

	Shortened Cohort			*Shortened Cohort adjusted for HES outcome exposure timing*		
Pregnancy Outcomes	*Exposed N= 2,058*	*Contemporary Controls N= 6161*	*Historical Controls N= 6,691*	*Exposed N= 2,309*	*Contemporary Controls N= 6,908*	*Historical Controls N= 6691*
Miscarriage	*12 (0.6%)*	*27 (0.4%)*	*41 (0.6%)*	*14 (0.6%)*	*33 (0.5%)*	*41 (0.6%)*
Stillbirth	*11 (0.5%)*	*26 (0.4%)*	*15 (0.2%)*	*13 (0.6%)*	*24 (0.3%)*	*15 (0.2%)*
Livebirth	*2,008 (97.6%)*	*6,039 (98.0%)*	*6,482 (96.9%)*	*2,247 (97.3%)*	*6,754 (97.8%)*	*6,482 (96.9%)*
Other	*27 (1.3%)*	*69 (1.1%)*	*153 (2.3%)*	*35 (1.5%)*	*97 (1.4%)*	*153 (2.3%)*

Table 2 The distribution of pregnancy outcomes after matching for all pregnancies in the shortened cohort.

	Shortened Cohort			**Shortened Cohort adjusted for HES outcomes**		
Pregnancy Outcomes	**Exposed N= 2,552**	**Contemporary Controls N= 60,881**	**Historical Controls N= 87,220**	**Exposed N= 2,830**	**Contemporary Controls N= 60,603**	**Historical Controls N= 87,220**
Miscarriage	21 (0.8%)	6,468 (10.6%)	8,658 (9.9%)	24 (0.8%)	6,867 (11.3%)	9,339 (10.7%)
Stillbirth	15 (0.6%)	158 (0.3%)	189 (0.2%)	17 (0.6%)	176 (0.3%)	230 (0.3%)
Livebirth	2,428 (95.1%)	29,944 (49.2%)	40,398 (46.3%)	2,716 (96.0%)	36,924 (60.9%)	51,123 (58.6%)
Outcome unknown	40 (1.6%)	19,504 (32.0%)	29,445 (33.8%)	15 (0.5%)	11,362 (18.7%)	17,085 (19.6%)
Other	48 (1.9%)	4,807 (7.9%)	8,530 (9.8%)	58 (2.0%)	5,274 (8.7%)	9,443 (10.8%)

Table 3 Crude and adjusted hazard ratios for pregnancies in the shortened cohort before HES outcomes were included

	Contemporary Controls		Historical Controls	
	Crude HR (95%CI)	**Adjusted(95%CI) HR**	**Crude HR (95%CI)**	**Adjusted(95%CI) HR**
All Pregnancy Loss	1.33 (0.77-2.14)	1.29 (0.62-2.70)	1.77 (1.00-3.14)	1.82 (0.83-2.97)
Miscarriage	1.41 (0.69-2.84)	1.18 (0.43-3.26)	1.43 (0.71-2.86)	1.58 (0.59-2.94)
Stillbirth	1.24 (0.52-2.97)	1.10 (0.29-4.09)	3.04 (1.03-8.98)	4.30 (0.44-6.98)

Table 4 Crude and adjusted hazard ratios for pregnancies in the shortened cohort after additional HES outcomes were included and exposure status and follow-up time were adjusted

	Contemporary Controls		Historical Controls	
	Crude HR (95%CI)	**Adjusted(95%CI) HR**	**Crude HR (95%CI)**	**Adjusted(95%CI) HR**
All Pregnancy Loss	1.48 (0.83-2.63)	1.79 (0.85-3.79)	2.03 (1.14-3.62)	2.04 (1.04-3.99)
Miscarriage	1.77 (0.85-3.68)	1.39 (0.54-3.55)	1.12 (0.56-2.25**)**	1.56 (0.57-4.24)
Stillbirth	2.88 (1.07-7.70)	2.31 (0.86-6.19)	3.10 (0.81-5.96)	4.50 (0.98-9.73)

Implications

The end dates for pregnancy episodes with missing outcomes in the CPRD Pregnancy Registers are taken as the last antenatal record found (49), these dates are effectively a "best guess" and may be incorrect. Including all pregnancies with unknown outcome may lead to misclassification of exposure status due to uncertainty around the end dates of those pregnancies. As demonstrated here it is necessary to ascertain the correct pregnancy end date and then update the exposure status based on new pregnancy timings.

Utilising the HES data to look for additional pregnancy outcomes did not substantially change the distribution of outcomes. The main difference was a slight increase in the proportion of live births. This is likely to be due to the fact that live births are easier to ascertain in HES than miscarriage outcomes which can be difficult to differentiate from termination records. The proportion of miscarriages and stillbirths remained constant among the exposed patients but increased slightly in the controls. Given the relatively small numbers of exposed patients experiencing these events in the shortened cohort it is difficult to comment on the implications this may have in a larger group.

Compared to the whole register analysis (Chapter 3) there were fewer outcomes found in HES. This is likely to be due to the fact that this study was looking specifically at miscarriage and therefore only included codes which allowed these to be differentiated from terminations. The previous analysis grouped pregnancy outcomes into losses or deliveries and utilised all codes.

As expected, hazard ratios for the shortened cohort showed a similar pattern to those of the one-year cohort but confidence intervals were much wider due to the reduced number of pregnancies. Across all pregnancy loss categories the addition of further outcomes from HES did not markedly change the adjusted hazard ratios.

5.4.2 Pregnancy Episodes which have been split by the algorithm into conflicting episodes.

The final recommendation refers to four scenarios which may result in a single pregnancy episode being incorrectly split by the rules of the algorithm. These scenarios were outlined in the methodological work in chapter 3 and are re-capped (along with the criteria by which they are defined) in table 5.

Methods

I arranged the conflicting episodes which were excluded from the COVID-19 and pregnancy loss cohort study into their conflicting pairs and applied the criteria outlined in table 5 to ascertain how many had evidence of each scenario. I included overlapping pairs even if only one of the episodes met the cohort inclusion criteria. Where pregnancy episodes conflicted with multiple other episodes, they were included multiple times.

Results

The proportions of conflicting episodes which met the criteria for each scenario are described in Table 5 along with a re-cap of the proportions found in the whole register analysis in Chapter 3. Scenarios are not mutually exclusive so pairs may be consistent with multiple scenarios. For each of the scenarios outlined there were some conflicting pregnancies within the cohort which met the criteria. The most common scenario was 4d, that the GP had recorded information about a pregnancy but no outcome resulting in the pregnancy being split into multiple episodes which then overlap based on algorithm generated timings. Whilst the proportions of pregnancy pairs meeting each scenario are similar between the contemporary and historical cohorts, they are very different to those seen in the whole register analysis.

Implications

Work presented in this chapter utilised CPRD Aurum and the original work on uncertain pregnancy episodes utilised CPRD GOLD, therefore the differing distribution of scenarios may be due to the way in which records are entered into the two databases. For example, the criteria for 4a, 4b and 4c all involve utilising scan records, the proportion of 4a and 4C which rely on a scan record being present has decreased and conversely 4b which relies on a scan record not being present has increased. It may be that scan records in CPRD Aurum were less likely to be in the coded data due to the way in which data are captured in the EMIS software.

A high proportion of the conflicting episodes were consistent with scenario 4d, which involves a GP recording information about a pregnancy >6 weeks apart but ultimately no outcome for that pregnancy. In this scenario both conflicting episodes will be outcome unknown pregnancies. It is therefore difficult to ascertain useful information from these pregnancy episodes. If utilising the HES data to obtain additional outcomes it would be important to merge these pregnancies first in order to avoid double counting them. However, it would be significant work to merge the conflicting episodes and would rely on multiple assumptions being applied which may result in further error.

5.4.3 Conclusions

Including additional outcomes from linked HES data in order to utilise the pregnancies with no recorded outcome appeared to have little impact on the observed relationship between COVID-19 and pregnancy loss. However, investigations into the impact of uncertain episodes were limited by having to shorten the cohort and therefore reduce the number of pregnancies within the analysis. Changes in exposure status found after adjusting pregnancy ends based on HES demonstrated the risk of exposure misclassification which may be associated with including uncertain episodes without consideration. The relationship between conflicting pregnancy episodes and those which have no outcome recorded is intertwined and can be difficult to untangle given the often-limited information about them.

Table 5 Conflicting pregnancy episode pairs which have evidence of scenarios indicating a pregnancy has been split into two episodes.

Scenario	Description	Criteria applied to identify evidence of each scenario	Exposed N=493	Contemporary Controls N= 14687	Historical Controls N= 22981	Proportion found in whole CPRD GOLD Pregnancy Register
	Problem 4: The pregnancy is true and current but is split into separate episodes by the rules of the algorithm					
4a	The GP records further information about a pregnancy outcome > 25 weeks after the delivery date for pregnancies ending in delivery OR >8 weeks but <12 weeks for pregnancies ending in loss. The algorithm assumes this further information is a different pregnancy and generates a new episode, which may overlap with the "true" episode.	- The outcome combination of the two episodes must be delivery/delivery or loss/loss (Section 3.3, Appendix 12) - The first episode had an antenatal code from a list deemed likely to only be recorded if the patient was currently pregnant (Section 3.3, Appendix 11) OR a scan record in the HES DID data	77 (15.6%)	2,066 (14.1%)	3,169 (13.8%)	1.2%

		between firstantenatal* and pregend*.				
4b	The GP records further antenatal information about a pregnancy after delivery or pregnancy loss. This will then be used to generate a new pregnancy without outcome episode by the algorithm. If the code is within 4 weeks of the end of the true pregnancy episode the two will overlap	- The first episode must have outcome= 1-10 in the register (Section 3.3, Appendix 2) and must have endadj* =0 - The second episode must have no recorded outcome (outcome= 13) - The second episode must have a gestdays* =28 (likely to consist of one code) and there must NOT be a scan code (Section 3.3, Appendix 13) with an eventdate* = pregend* of the second episode.	0	344 (2.3%)	799 (3.5%)	17.4%
4c	The patient has a follow up scan after a pregnancy loss. This is recorded in the data by the GP as an antenatal scan. The algorithm then	- The outcome combination of the two episodes must be loss/missing.	2 (0.4%)	122 (0.8%)	209 (0.9%)	1.1%

	creates a second pregnancy episode based on the antenatal scan code which becomes a pregnancy without outcome in the register.	- The second episode must have a gestdays* =28 (likely to consist of one code) and there must be a scan code (Section 3.3, Appendix 13) with an eventdate* = pregend* of the second episode. - The outcome combination of the two episodes must be missing/missing. - The pregend* of the first episode is > 42 days before the firstantenatal* date of the second episode.				
4d	The GP records information about a pregnancy but no information about the outcome. If records relating to this pregnancy are more than 6 weeks apart, they will be turned into multiple episodes. Once estimated start dates are generated for these	- The outcome combination of the two episodes must be missing/missing. - The pregend* of the first episode is > 42 days before the firstantenatal* date of the second episode.	261 (52.9%)	6,912 (47.1%)	11,401 (49.6%)	5.9%

	episodes based on the data recorded episodes may overlap. For example, if there is gestational information included in the second episode the start of this episode will be assigned before the start of the previous episode resulting in a nested pregnancy episode.					
4e	The first pregnancy episode ended in delivery and has been shifted backwards by the rules of the algorithm leaving unassigned late pregnancy or third trimester records. These records will then be identified by the algorithm as end of pregnancies and new conflicting episodes will be created.	- The first episode must have a delivery outcome code and endadj* variable not = to 0 - The second episode must have outcome= to 11, 12 or 13.	84 (17.0%)	1,735 (11.8%)	2,712 (11.8%)	29.2%

5.5 Chapter Summary

Using the CPRD Aurum Pregnancy Register to investigate the relationship between COVID-19 and pregnancy loss

- I utilised the CPRD Aurum Pregnancy register to conduct a matched cohort study investigating the relationship between COVID-19 and risk of pregnancy loss. Women who had a record of COVID-19 during their pregnancy were compared to both contemporary and pre-pandemic controls using a Cox regression model.

- Adjusted hazard ratios showed some evidence of an increased risk of pregnancy loss when compared to the contemporary controls. The hazard ratios for all types of pregnancy loss were higher when compared to historical controls.

- I assessed the potential impact of having excluded all uncertain pregnancy episodes from the cohort study.

- There was evidence that the inclusion of pregnancies with no recorded outcome could result in exposure misclassification due to incorrect pregnancy timings.

- Additional pregnancy outcomes in the linked HES data enabled me to utilise some of the pregnancy episodes which had no outcome recorded in the CPRD Aurum register in a secondary analysis. However, this did not appear to change the observed relationship between COVID-19 and risk of pregnancy loss.

- This work was approved by the MHRA's RDG committee, protocol number 22_001695 (Appendix 4) and LEO reference number 27021.

Chapter 6: Discussion

6.1 Introduction

The first aim of this thesis was to conduct a detailed investigation into uncertain pregnancy episodes in the CPRD Pregnancy Register and generate appropriate recommendations for handling these data in research. The second linked aim was to utilise the CPRD Pregnancy Register to answer an important and topical clinical question about whether COVID-19 infection during pregnancy is associated with pregnancy loss, and as part of this study, to implement some of my recommendations on the handling of uncertain pregnancy episodes. Detailed discussion of each of the pieces of work I conducted to address these aims are included in the relevant chapters of this thesis (chapters 3-5) This chapter brings together the key findings and discussion points from each piece of work. I also outline the strengths and potential limitations of the work I have conducted. Finally, I discuss the implications for policy and future research including recommendations for researchers, data providers and clinicians before outlining the conclusions of this thesis.

6.2 Overview of Key Findings

6.2.1 Uncertain Pregnancy Episodes in the CPRD Pregnancy Registers

- I identified 12 scenarios which may result in pregnancy episodes with no recorded outcome and 10 scenarios which may result in conflicting pregnancy episodes (those which seemingly overlap with another pregnancy episode for the same woman) in the CPRD Pregnancy Register.

- I established criteria as to how each of the scenarios might appear in the data and systematically applied these to identify which uncertain pregnancy episodes had evidence of each scenario. Linked secondary care data was used to look for pregnancy events not captured in the primary care data.

- Most pregnancy episodes with missing outcomes in the CPRD GOLD Pregnancy Register had evidence that they were true and contemporaneous pregnancies which would be missed if all uncertain episodes are excluded from an analysis. This is important to note for studies where pregnancies are the denominator such as vaccine uptake in pregnancy.

- A sizeable proportion of uncertain episodes generated by the algorithm appear to be due to historical outcomes being recorded by the GP during an ongoing pregnancy. This can lead to the algorithm generating an additional episode with outcome missing or two separate episodes with outcomes. In both cases these episodes may conflict with one another.

- Whilst some conflicting episodes may be caused by poor quality data there are many conflicting episodes for which it may be possible to clarify which time period is likely to be the true pregnancy. I found that episode conflicts were more likely to occur for pregnancies ending in loss; this is of little surprise given the wider variation around the true gestation of such pregnancies.

- Uncertain episodes may appear more frequently for women with a history of complicated pregnancy outcomes. The exclusion of overlapping pregnancies might therefore systematically exclude those with a history of pregnancy complications, introducing bias.

- I found some evidence that pregnancies with serious complications are more likely to have an uncertain episode in the Register. For example, women with pre-eclampsia are more likely to have consultant led antenatal care carried out in hospital increasing the chances that their primary care record is incomplete and has no recorded outcome.

- The implications of including or excluding uncertain pregnancy episodes will vary depending on the purpose and design of the study being conducted. I generated a series of recommendations for a tailored approach to including or excluding uncertain pregnancy episodes. However, in order to implement these researchers will need access to the original primary care pregnancy records and linked data.

6.2.2 A Systematic Review of COVID-19 and Pregnancy Loss

- Epidemiological studies to date have drawn differing conclusions about the relationship between COVID-19 and risk of pregnancy loss. Systematic reviews conducted early in the pandemic reached mixed conclusions regarding risk of stillbirth although these reviews included mainly case reports and noncomparative study designs(34,35,56–61).

- A systematic review by Pathirathna et al published in 2021 looked at multiple adverse pregnancy outcomes and reported a limited number of studies which considered miscarriage as an outcome (62). They also reported a combined odds ratio which showed a statistically significant increase in the risk of stillbirth among women who had COVID-19.

- I conducted a systematic review of published and pre-print literature for studies which attempt to quantitatively assess the association between having COVID-19 during pregnancy and pregnancy loss. Only studies which included a COVID-19 free control group were included.

- My systematic review identified 21 studies which looked at risk of stillbirth, 7 which looked at risk of miscarriage and 3 which looked at risk of all pregnancy loss among women who had COVID-19 during pregnancy compared to those who did not.

- There were discrepancies in the way in which both COVID-19 as the exposure and pregnancy loss as an outcome were defined from study to study. I also found reasons for concern about the risk of bias for over half of the studies based on the published accounts. The most common concerns were a lack of consideration for potential confounding factors and a lack of clarity as to whether the study population was representative of the source population.

- I found 5 studies which looked at COVID-19 exposure and risk of pregnancy loss across the whole pregnancy. Of these studies 4 observed more miscarriage in the COVID-19 exposed group however, most confidence intervals spanned the null and the number of studies retrieved was too low for a formal synthesis.

- The majority of identified studies which considered COVID-19 exposure and risk of stillbirth were cross-sectional and measured COVID-19 exposure at point of delivery.

- I conducted a meta-analysis of a subgroup of 8 of these stillbirth studies which were deemed to be high quality. The results of this pooled analysis suggested that women who had COVID-19 at point of delivery were 1.5 times more likely to have a stillbirth than women who did not have COVID-19.

- I found insufficient numbers of high-quality studies to assess whether there is an increased risk of pregnancy loss associated with COVID-19 earlier in pregnancy highlighting a need for further research.

6.2.3 COVID-19 and risk of Pregnancy Loss: An Applied Example

- I conducted a matched cohort study, using the CPRD Aurum Pregnancy Register and linked secondary care data. Pregnancies with a record of COVID-19 were matched to both contemporary pregnancies without a COVID-19 record and historical control pregnancies. A Cox regression model was used to calculate hazard ratios adjusted for potential confounders.

- I found evidence that women who had COVID-19 during pregnancy had an 18% higher risk of pregnancy loss compared with women who were pregnant during the same calendar period but with no record of COVID-19.

- When women with a record of COVID-19 in pregnancy were compared to women who were pregnant pre-pandemic a 39% higher risk of pregnancy loss was observed.

- The increased risks for stillbirths were particularly pronounced, with women exposed to COVID-19 having around double the risk for this outcome compared with pre-pandemic controls. This contradicts the findings of some studies which have looked at COVID-19 exposure and risk of stillbirth notably Ferrara et al and Litman et al which both found no increased risk (22,23). However, several studies ((36,37,63,64)) have observed an increased risk and the pooled results of my systematic review meta-analysis also found a higher risk of stillbirth in the COVID-19 exposed group.

- Sensitivity analyses to investigate the impact of including pregnancies where the outcome was missing from CPRD but could be obtained in linked HES data did not markedly change the observed relationship between COVID-19 and pregnancy loss. However, these were conducted on a shortened cohort with relatively small numbers of exposed patients. It is therefore difficult to comment on the implications this may have in a larger group.

- Changes to exposure status after adjusting pregnancy timings based on data found in HES highlighted the potential risk of exposure misclassification associated with including all uncertain episodes without consideration.

6.3 Summary of the strengths of this work

My work investigating uncertain pregnancy episodes in CPRD data (presented in chapter 3) offers valuable insight for researchers enabling them to make informed decisions about whether to include incomplete and uncertain pregnancy data when designing studies in electronic health records. I have provided detailed descriptions of scenarios which may result in uncertain pregnancy episodes. These scenarios were developed utilising clinician advice and clinical guidelines as well as in-depth investigations into the patterns of pregnancy data recorded in CPRD. I also developed a series of criteria which researchers can apply to ascertain which uncertain pregnancy episodes may fit each scenario allowing them to tailor their study population based on the design and purpose of their research. This work provides a level of information previously unavailable. Not only does this further enhance the usefulness of the CPRD Pregnancy Registers but many outlined scenarios will also be applicable to other EHR pregnancy data sources.

My systematic review of epidemiological studies which investigated the relationship between COVID-19 and pregnancy loss (presented in chapter 4) included both published and pre-print literature ensuring that the latest information was captured. This is particularly important given that COVID-19 research is fast paced and evolving rapidly. The inclusion of pre-print studies will also have helped to reduce the risk of publication bias. Another major strength of this review was that the methods were published and peer-reviewed a-priori helping to ensure a robust approach. I was able to provide a valuable narrative of epidemiological studies to date which have used a comparative approach to investigate the relationship between having COVID-19 in pregnancy and risk of pregnancy loss.

The CPRD Pregnancy Registers are an extremely useful tool for research, they offer an opportunity to study large cohorts of pregnant women relatively easily and give details on the pregnancy timings and outcomes. Utilising the CPRD Aurum Pregnancy Register alongside the corresponding CPRD Aurum data allowed me to conduct a large-scale cohort study examining the risk of pregnancy loss associated with having COVID-19 during pregnancy. The longitudinal nature of the EHR data in CPRD Aurum meant that unlike many previous studies I was able to examine COVID-19 exposure across the whole pregnancy and not just at the point of pregnancy outcome. Capture of exposure data was further strengthened by utilising linked secondary care data and COVID-19 lab test results. Furthermore, comparing to both a pre-pandemic as well as a contemporary control group meant that I was able to assess the potential impact of COVID-19 without the concern of exposure misclassification amongst the controls. The richness of the data available in CPRD meant I was able to control for a range of potential confounders. I also conducted a sensitivity analysis to examine the impact of my decision to exclude pregnancies with no outcome recorded in the CPRD data and those which conflict with another pregnancy, thus providing further insight for researchers wishing to utilise EHR data for this kind of research in the future.

6.4 Summary of the potential limitations of this work

The main limitation of my work investigating uncertain pregnancy episodes is that it relies heavily on the assumption that real life scenarios will consistently result in the same data patterns. Electronic health data are not collected for the purposes of research and can be messy for a variety of reasons, some of which may not have been captured in this work. As with any algorithmic approach there will always be cases which fall outside of the criteria which I have outlined. However, it is not possible to examine each case individually in data of this scale.

The two main limitations of my systematic review were the result of time and resource constraints. Firstly, the review only included studies which were published in English which may have resulted in some studies being missed. The second limitation was that whilst two reviewers were available to independently assess the abstracts for inclusion, the data extraction and quality assessment was carried out by me alone. This may have resulted in assessment bias or mistakes in data extraction.

The study period for my investigation of COVID-19 and pregnancy loss only covers the first year of the COVID-19 pandemic in the UK. Since then, SARS-CoV-2 has evolved into new strains which may not carry the same risk of pregnancy loss as the original strains. The early study period also meant that I was unable to examine whether vaccination had any impact on the observed risk.

Furthermore, I was unable to assess whether COVID-19 severity has any impact; there may well be a difference between the risk of pregnancy loss between women with asymptomatic COVID-19 and those who end up in hospital with severe disease.

Finally, whilst the sensitivity analyses I conducted provided useful insight, the restrictions on available HES follow-up meant that they were limited to a three-month follow-up period significantly reducing the size of the cohort. It was also not possible to ascertain whether some loss records recorded in secondary care pertained to a spontaneous or induced abortion and therefore these records had to be excluded from the analysis. Identifying outcomes in HES allowed me to adjust the dates of the pregnancy episodes. However, whilst HES data is useful as a complementary source of information it is also an EHR database derived from data that were not collected for research purposes and there may be gaps in recording. It is, however, less likely that pregnancy outcome events which happen in hospital will be recorded retrospectively and therefore dates of recorded outcomes may be considered more reliable.

6.5 Implications for policy and future research

6.5. 1 Implications for EHR researchers and data providers

It is important that researchers conducting pregnancy research in EHR data carefully consider how they handle uncertain pregnancy episodes when designing studies. There are implications to both including and excluding all uncertain episodes which vary depending on the type and purpose of the research being conducted. Ideally a tailored approach should be developed based on the recommendations outlined in this work.

This work highlights the value of having linked data alongside a patient's primary care records. Some of the scenarios described in my assessment of uncertain pregnancy episodes can only be identified through the use of secondary care data to look for evidence of pregnancy outcomes or fetal scan records. Linked secondary care data provides the opportunity to ascertain additional pregnancy outcomes which are not recorded in the primary care records which in turn allows for the adjustment of pregnancy timings. Researchers wishing to carry out pregnancy research utilising EHR data should consider obtaining linked data when possible. It is also necessary for researchers to have access to the code lists and pregnancy data used to generate the Pregnancy registers in order to establish evidence of some of the scenarios I have described. Data providers such as CPRD could

potentially facilitate this by creating summary variables which indicate evidence of the scenarios described in this work and by incorporating information from secondary care such as pregnancy outcomes or fetal scan records into the Pregnancy Registers. These additions to the CPRD Pregnancy Registers could have a highly beneficial public health impact given the importance of EHR data in facilitating pregnancy research.

6.5.2 Implications for clinicians and public health

This work suggests that there is a potential increased risk of pregnancy loss associated with having COVID-19 in pregnancy. In particular I found evidence from both my systematic review and cohort study that having COVID-19 may increase the risk of stillbirth. Women who have COVID-19 towards the end of pregnancy and during delivery should be closely monitored, especially those who have not been vaccinated. My findings support the importance of COVID-19 vaccination campaigns for pregnant women to reduce the risk of them contracting the disease.

6.5.3 Recommendations for future research

As outlined in this thesis there are several types of pregnancy research which can be conducted using EHR data. My applied example tested the recommendations for handling uncertain episodes in a study considering an exposure during pregnancy and the risk of specific outcomes. It would be useful for further example studies to be conducted testing the recommendations made for different applications of pregnancy EHR data e.g. vaccine uptake studies or excluding pregnant women from cohorts.

There is a need for further research into the potential risks of pregnancy loss associated with having COVID-19. Follow-up studies are required which look specifically at how an individual's vaccination status may affect the relationship between COVID-19 exposure and pregnancy loss. It is also important that further work is carried out to investigate whether more recent strains of the SARS-CoV-2 virus carry the same risk and whether disease severity also plays a part.

6.6 Conclusions

The first aim of my thesis was to conduct a detailed investigation into uncertain pregnancy episodes in the CPRD Pregnancy Register and generate appropriate recommendations for handling these data in research. This work goes beyond what anyone has previously attempted to do and offers useful insight and advice for researchers wishing to utilise EHR data to study pregnancy. I have shown that whilst there are many reasons that uncertain pregnancy episodes occur in EHR data they are often true and contemporaneous offering evidence that a woman was really pregnant at that time point. Blanket decisions to include or exclude uncertain episodes may lead to under ascertainment of pregnancies, biased study populations and errors in analysis such as exposure misclassification. Researchers utilising EHR data such as CPRD to study pregnancy should consider a tailored approach dependent on the design and purpose of their study. I have outlined advice on how researchers may wish to develop a tailored study population. However, untangling uncertain pregnancy episodes can be challenging. Data providers could help to facilitate more efficient higher quality pregnancy research by providing summary variables based on the scenarios outlined in this work.

The second aim was to utilise the CPRD Pregnancy Register to answer an important and topical clinical question about whether COVID-19 infection during pregnancy is associated with pregnancy loss. The application of the CPRD Pregnancy Register for this study demonstrates the importance of EHR data for pregnancy research. I found evidence to suggest an increased risk of pregnancy loss associated with having COVID-19 during pregnancy. However, questions remain as to the impact of newer variants of COVID-19 and also whether risk of pregnancy loss is lower in vaccinated women. Nevertheless, this work offers useful insight into the risks of having COVID-19 in pregnancy and supports the importance of protecting pregnant women from COVID-19 by vaccination.

Bibliography

1. Campbell J, Bhaskaran K, Thomas S, Williams R, Mcdonald HI, Minassian C. Investigating the optimal handling of uncertain pregnancy episodes in the CPRD GOLD Pregnancy Register: a methodological study using UK primary care data. BMJ Open [Internet]. 2022;12:55773. Available from: http://dx.doi.org/10.1136/bmjopen-2021-055773

2. Campbell J, Williams R, Harley M, Bhaskaran K. Covid-19 during pregnancy and risk of pregnancy loss (miscarriage or stillbirth): a systematic review protocol. BMJ Open [Internet]. 2022;12. Available from: https://bmjopen.bmj.com/content/12/10/e065588.full

3. NHS England. Patients Registered at a GP Practice, February 2023. https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/february-2023#. 2023.

4. National Institute for Health and Care Excellence. Antenatal care for uncomplicated pregnancies. Clinical Guideline [Internet]. 2008;(March):1–55. Available from: https://www.nice.org.uk/guidance/cg62/resources/antenatal-care-for-uncomplicated-pregnancies-975564597445

5. Meador K, Reynolds MW, Crean S, Fahrbach K, Probst C. Pregnancy outcomes in women with epilepsy: A systematic review and meta-analysis of published pregnancy registries and cohorts. Vol. 81, Epilepsy Research. 2008. p. 1–13.

6. Medicines and Healthcare Products Regulatory Agency. Antiepileptic drugs: review of safety of use during pregnancy. https://www.gov.uk/government/publications/public-assesment-report-of-antiepileptic-drugs-review-of-safety-of-use-during-pregnancy/antiepileptic-drugs-review-of-safety-of-use-during-pregnancy. 2021.

7. Toussi M, Shlaen M, Coste F, de Voogd H, Dimos V, Kaplan S. Effectiveness of risk minimisation measures for valproate: A drug utilisation study in Europe. Pharmacoepidemiol Drug Saf. 2021 Mar 1;30(3):292–303.

8. Public Health England. Pertussis Vaccination Programme for Pregnant Women: vaccine coverage estimates in England, April to August 2014 - GOV.UK. 2014;11(34). Available from: https://www.gov.uk/government/publications/pertussis-immunisation-in-pregnancy-vaccine-coverage-estimates-in-england-october-2013-to-march-2014/pertussis-vaccination-programme-for-pregnant-women-vaccine-coverage-estimates-in-england-april-to-august-2014

9. Walker JL, Rentsch CT, McDonald HI, Bak J, Minassian C, Amirthalingam G, et al. Social determinants of pertussis and influenza vaccine uptake in pregnancy: a national cohort study in England using electronic health records. BMJ Open. 2021;11(6):e046545.

10. Minassian C, Thomas SL, Williams DJ, Campbell O, Smeeth L. Acute Maternal Infection and Risk of Pre-Eclampsia: A Population-Based Case-Control Study. PLoS One. 2013 Sep 3;8(9).

11. Margulis A v., Palmsten K, Andrade SE, Charlton RA, Hardy JR, Cooper WO, et al. Beginning and duration of pregnancy in automated health care databases: Review of estimation methods and validation results. Vol. 24, Pharmacoepidemiology and Drug Safety. John Wiley and Sons Ltd; 2015. p. 335–42.

12. Devine S, West S, Andrews E, Tennis P, Hammad TA, Eaton S, et al. The identification of pregnancies within the general practice research database. Pharmacoepidemiol Drug Saf. 2010 Jan;19(1):45–50.

13. Webster WS, Freeman JAD. Prescription drugs and pregnancy. Expert Opin Pharmacother. 2003 Jun;4(6):949–61.

14. Hardy JR, Holford TR, Hall GC, Bracken MB. Strategies for identifying pregnancies in the automated medial records of the General Practice Research Database. Pharmacoepidemiol Drug Saf. 2004 Nov;13(11):749–59.

15. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. PLoS One. 2018;13(2):1–20.

16. Charlton R, Snowball J, Sammon C, de Vries C. The Clinical Practice Research Datalink for drug safety in pregnancy research: an overview. Therapie. 2014;69(1):83–9.

17. World Health Organization (WHO). Coronavirus disease (COVID-19) pandemic. 2023.

18. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2022 Aug 15]. Available from: https://covid19.who.int/

19. UK Government. Coronavirus (COVID-19) in the UK [Internet]. [cited 2022 Jan 24]. Available from: https://coronavirus.data.gov.uk/

20. Briggs A, Jenkins D, Fraser C. NHS Test and Trace: the journey so far [Internet]. 2020 [cited 2023 Feb 15]. Available from: https://www.health.org.uk/publications/long-reads/nhs-test-and-trace-the-journey-so-far

21. UK Government. Regulatory approval of Pfizer/BioNTech vaccine for COVID-19. https://www.gov.uk/government/publications/regulatory-approval-of-pfizer-biontech-vaccine-for-covid-19. 2022.

22. UK Health Security Agency. COVID-19: The Green Book chapter 14a. In: The Green Book [Internet]. 2022. Available from: https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a%0A

23. UK Health Security Agency. Vaccine uptake among pregnant women increasing but inequalities persist. https://www.gov.uk/government/news/vaccine-uptake-among-pregnant-women-increasing-but-inequalities-persist. 2022.

24. Mak TK, Mangtani P, Leese J, Watson JM, Pfeifer D. Infl uenza vaccination in pregnancy: current evidence and selected national policies [Internet]. Vol. 8. 2008. Available from: http://infection.thelancet.com

25. Mosby LG, Rasmussen SA, Jamieson DJ. 2009 pandemic influenza A (H1N1) in pregnancy: A systematic review of the literature. Vol. 205, American Journal of Obstetrics and Gynecology. 2011. p. 10–8.

26. de Souza Silva GA, da Silva SP, da Costa MAS, da Silva AR, de Vasconcelos Alves RR, Ângelo Mendes Tenório F das C, et al. SARS-CoV, MERS-CoV and SARS-CoV-2 infections in pregnancy and fetal development. J Gynecol Obstet Hum Reprod. 2020 Dec 1;49(10).

27. Arthurs AL, Jankovic-Karasoulos T, Roberts CT. COVID-19 in pregnancy: What we know from the first year of the pandemic. Biochim Biophys Acta Mol Basis Dis. 2021 Dec 1;1867(12).

28. Pathirathna ML, Samarasekara BPP, Dasanayake TS, Saravanakumar P, Weerasekara I. Adverse Perinatal Outcomes in COVID-19 Infected Pregnant Women: A Systematic Review and Meta-Analysis. Healthcare (Switzerland). 2022;10(2).

29. Knight M, Bunch K, Vousden N, Morris E, Simpson N, Gale C, et al. Characteristics and outcomes of pregnant women admitted to hospital with confirmed SARS-CoV-2 infection in UK: National population based cohort study. The BMJ. 2020 Jun 8;369.

30. Saccone G, Sen C, di Mascio D, Galindo A, Grünebaum A, Yoshimatsu J, et al. Maternal and perinatal outcomes of pregnant women with SARS-CoV-2 infection. Ultrasound in Obstetrics and Gynecology. 2021 Feb 1;57(2):232–41.

31. Jamieson DJ, Honein MA, Rasmussen SA, Williams JL, Swerdlow DL, Biggerstaff MS, et al. H1N1 2009 infl uenza virus infection during pregnancy in the USA. www.thelancet.com [Internet]. Available from: http://www.cdc.gov/h1n1fl

32. Schwartz DA, Dhaliwal A. Coronavirus Diseases in Pregnant Women, the Placenta, Fetus, and Neonate. Adv Exp Med Biol. 2021;1318:223–41.

33. Kazemi SN, Hajikhani B, Didar H, Hosseini SS, Haddadi S, Khalili F, et al. COVID-19 and cause of pregnancy loss during the pandemic: A systematic review. PLoS One. 2021 Aug 1;16(8 August).

34. Pérez-López FR, Savirón-Cornudella R, Chedraui P, López-Baena MT, Pérez-Roncero G, Sanz-Arenal A, et al. Obstetric and perinatal outcomes of pregnancies with COVID 19: a systematic review and meta-analysis. Journal of Maternal-Fetal and Neonatal Medicine [Internet]. 2022;0(0):1–17. Available from: https://doi.org/10.1080/14767058.2022.2051008

35. Agolli A, Agolli O, Velazco DFS, Ahammed MR, Patel M, Cardona-Guzman J, et al. Fetal Complications in COVID-19 Infected Pregnant Woman: A Systematic Review and Meta-Analysis. Avicenna J Med [Internet]. 2021 Oct [cited 2022 May 6];11(4):200–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34881203

36. Gurol-Urganci I, Jardine JE, Carroll F, Draycott T, Dunn G, Fremeaux A, et al. Maternal and perinatal outcomes of pregnant women with SARS-CoV-2 infection at the time of birth in England: national cohort study. Am J Obstet Gynecol. 2021 Nov 1;225(5):522.e1-522.e11.

37. DeSisto CL, Wallace B, Simeone RM, Polen K, Ko JY, Meaney-Delman D, et al. Risk for Stillbirth Among Women With and Without COVID-19 at Delivery Hospitalization — United States, March 2020–September 2021. MMWR Morb Mortal Wkly Rep. 2021 Nov 26;70(47):1640–5.

38. Stephansson O, Pasternak B, Ahlberg M, Hervius Askling H, Aronsson B, Appelqvist E, et al. SARS-CoV-2 and pregnancy outcomes under universal and non-universal testing in Sweden: register-based nationwide cohort study. Available from: www.bjog.org

39. Cavalcante MB, de Melo Bezerra Cavalcante CT, Cavalcante ANM, Sarno M, Barini R, Kwak-Kim J. COVID-19 and miscarriage: From immunopathological mechanisms to actual clinical evidence. Vol. 148, Journal of Reproductive Immunology. Elsevier Ireland Ltd; 2021.

40. Cardona-Pérez JA, Villegas-Mota I, Helguera-Repetto AC, Acevedo-Gallegos S, Rodríguez-Bosch M, Aguinaga-Ríos M, et al. Prevalence, clinical features, and outcomes of SARS-CoV-2 infection in pregnant women with or without mild/moderate symptoms: Results from universal screening in a tertiary care center in Mexico City, Mexico. PLoS One. 2021 Apr 1;16(4 April).

41. Balachandren N, Davies MC, Hall JA, Stephenson JM, David AL, Barrett G, et al. SARS-CoV-2 infection in the first trimester and the risk of early miscarriage: A UK population-based prospective cohort study of 3041 pregnancies conceived during the pandemic. Human Reproduction. 2022 Jun 1;37(6):1126–33.

42. Calvert C, Carruthers J, Denny C, Donaghy J, Hillman S, Hopcroft LEM, et al. A population-based matched cohort study of early pregnancy outcomes following COVID-19 vaccination and SARS-CoV-2 infection. Nat Commun. 2022 Dec 1;13(1).

43. CPRD. CPRD: Clinical Practice Research Datalink [Internet]. 2021 [cited 2018 Sep 17]. Available from: www.cprd.com

44. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T Van, et al. Data Resource Profile : Clinical Practice Research Datalink (CPRD). 2015;(June):827–36.

45. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile : Clinical Practice Research Datalink (CPRD) Aurum. 2019;1–8.

46. CPRD. CPRD Aurum Data Specification [Internet]. 2022 [cited 2023 Feb 15]. Available from: https://cprd.com/primary-care-data-public-health-research

47. Padmanabhan S. CPRD GOLD Data Specification [Internet]. 2017 [cited 2020 Jul 8]. Available from: https://cprd.com/sites/default/files/CPRD GOLD Full Data Specification v2.0_0.pdf

48. Boggon R, Gallagher A, Williams T van ST. Creating a Mother Baby Link and Pregnancy Register for a UK Population. Pharmacoepidemiol Drug Saf. 2011;20(S44).

49. Minassian C, Williams R, Meeraus WH, Smeeth L, Campbell OMR, Thomas SL. Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. Pharmacoepidemiol Drug Saf [Internet]. 2019;(April):1–11. Available from: http://doi.wiley.com/10.1002/pds.4811

50. Campbell J, Shepherd H, Welburn S, Barnett R, Oyinlola J, Oues N, et al. Methods to refine and extend a Pregnancy Register in the <scp>UK</scp> Clinical Practice Research Datalink primary care databases. Pharmacoepidemiol Drug Saf [Internet]. 2022 Dec 15; Available from: https://onlinelibrary.wiley.com/doi/10.1002/pds.5584

51. NHS Digital. Hospital Episode Statistics [Internet]. 2021 [cited 2018 Sep 17]. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

52. Gallagher AM, Puri S van ST. Linkage of the General Practice Research Database (GPRD) with Other Data Sources. In: Linkage of the General Practice Research Database (GPRD) with Other Data Sources. International Conference of Pharmacoepidemiology; 2011.

53. Murray J, Saxena S, Modi N, Majeed A, Aylin P, Bottle A. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. Journal of Public Health (United Kingdom). 2013;

54. Hannah Knight, Meulen DDCPJ van der, Johnson DIGUDDRDTMFDDRFPATMADMS. Patterns of Maternity Care in English NHS Hospitals 2011/12. 2013;1–62.

55. Department for Communities and Local Government. English Indices of Deprivation 2015- Statistical Release. 2015.

56. Banaei M, Ghasemi V, Saei Ghare Naz M, Kiani Z, Rashidi-Fakari F, Banaei S, et al. Obstetrics and neonatal outcomes in pregnant women with covid-19: A systematic review. Iran J Public Health. 2020;49:38–47.

57. Cavalcante MB, de Melo Bezerra Cavalcante CT, Cavalcante ANM, Sarno M, Barini R, Kwak-Kim J. COVID-19 and miscarriage: From immunopathological mechanisms to actual clinical evidence. J Reprod Immunol [Internet]. 2021;148(September):103382. Available from: https://doi.org/10.1016/j.jri.2021.103382

58. di Toro F, Gjoka M, di Lorenzo G, de Santo D, de Seta F, Maso G, et al. Impact of COVID-19 on maternal and neonatal outcomes: a systematic review and meta-analysis. Clinical Microbiology and Infection. 2021 Jan 1;27(1):36–46.

59. Jha D, Adhikari M, Gautam JS, Tinkari BS, Mishra SR, Khatri RB. Effect of COVID-19 on maternal and neonatal services. Lancet Glob Health. 2021 Feb 1;9(2):e114–5.

60. Kasraeian M, Zare M, Vafaei H, Asadi N, Faraji A, Bazrafshan K, et al. COVID-19 pneumonia and pregnancy; a systematic review and meta-analysis. Journal of Maternal-Fetal and Neonatal Medicine. 2022;35(9):1652–9.

61. Lanzoni G, Linetsky E, Correa D, Messinger Cayetano S, Alvarez RA, Kouroupis D, et al. Umbilical cord mesenchymal stem cells for COVID-19 acute respiratory distress syndrome: A double-blind, phase 1/2a, randomized controlled trial. Stem Cells Transl Med. 2021 May 1;10(5):660–73.

62. Pathirathna ML, Samarasekara BPP, Dasanayake TS, Saravanakumar P, Weerasekara I. Adverse Perinatal Outcomes in COVID-19 Infected Pregnant Women: A Systematic Review and Meta-Analysis. Healthcare (Basel) [Internet]. 2022 Jan 20 [cited 2022 May 6];10(2). Available from: http://www.ncbi.nlm.nih.gov/pubmed/35206820

63. Hcini N, Maamri F, Picone O, Carod JF, Lambert V, Mathieu M, et al. Maternal, fetal and neonatal outcomes of large series of SARS-CoV-2 positive pregnancies in peripartum period: A single-center prospective comparative study. European Journal of Obstetrics and Gynecology and Reproductive Biology. 2021 Feb 1;257:11–8.

64. Pereira A, Cruz-Melguizo S, Adrien M, Fuentes L, Marin E, Perez-Medina T. Clinical course of coronavirus disease-2019 in pregnancy. Acta Obstet Gynecol Scand. 2020 Jul 1;99(7):839–47.