

Coronavirus (COVID-19) hospital admissions by vaccination and pregnancy status, methodology: England

Methods and data sources used to estimate coronavirus (COVID-19) hospital admissions by vaccination status in pregnant and non-pregnant women in England.

Contact:
Matt Bosworth, Ryan Schofield,
Camille Harrison
health.data@ons.gov.uk
+44 1633 455046

Release date:
11 July 2022

Next release:
To be announced

Table of contents

1. [Overview of data sources and method](#)
2. [Data sources](#)
3. [Study population and design](#)
4. [Method for identifying pregnancies](#)
5. [Statistical methods](#)
6. [Future developments](#)
7. [Related links](#)

1 . Overview of data sources and method

- Data sources and methods detailed in this article were used in our [Coronavirus \(COVID-19\) hospital admissions by vaccination and pregnancy status, England: 8 December 2020 to 31 August 2021 bulletin](#).
- The main data source is the Office for National Statistics' Public Health Data Asset, which includes 2011 Census records, deaths registrations, and electronic health records for England.
- The study used a retrospective cohort design of 815,477 women in England who were first infected with COVID-19 between 8 December 2020 and 31 August 2021.
- The total number of women identified as pregnant when infected with COVID-19 in our study population was 33,549 (4.1% of the study population).
- We used Cox proportional hazards regression models to assess how the rate of COVID-19 hospital admission varies by vaccination status in women, stratified by those who were pregnant when they were first infected with COVID-19 and those who were not pregnant.

2 . Data sources

This study used data from the Office for National Statistics' (ONS) Public Health Data Asset (PHDA).

The PHDA is a unique linked dataset for England only. It was created using deterministic and probabilistic linkage methods to link the 2011 Census to the 2011 to 2013 NHS Patient Registers to retrieve NHS numbers. Then, deterministic linkage on NHS number was used to link to:

- death registrations dataset
- the [Hospital Episode Statistics \(HES\)](#) dataset
- the [General Practice Extraction Service \(GPES\) Data for Pandemic Planning and Research \(GDPPR\)](#) dataset
- the [National Immunisation Management Service](#) dataset
- the [NHS Test and Trace](#) (Pillar 1 and Pillar 2) dataset

For this analysis, we further linked the PHDA to NHS birth notifications data for 2020, 2021, and January to March 2022 using mothers' NHS number. This enabled the identification of women who were pregnant at the time of first coronavirus (COVID-19) infection and went on to have a live birth or stillbirth (see [Section 4: Method for identifying pregnancies](#) for more information).

The birth notification is a document completed by the doctor or midwife present at the birth (see [NHS Birth Notifications guidance](#)). This data source is supplied to the ONS by the NHS and was used instead of birth registrations data to provide a more timely indicator of pregnancy. There are small differences in the number of births recorded between birth notifications data and birth registrations data, but the two data sources are very similar. For more information, see our [Births in England and Wales explained: 2020 article](#).

In addition, this study used data from the 2021 Census to derive more up-to-date socio-demographic characteristics for participants in the study. The 2021 Census was deterministically linked to the NHS Personal Demographics Service to retrieve NHS numbers, with a linkage rate of 94.6%. Following clerical review of links made, the precision (proportion of true links) was estimated to be 99.4% (plus or minus 2.9%). Of these links, 1.6% involved multiple census records linked to the same NHS number, which were excluded following deduplication.

The 2021 Census was then linked to the PHDA using NHS number. We were able to update census variables for 91.6% of participants in the study population with data from the 2021 Census. The remaining 8.4% of the study population that could not be linked to the 2021 Census were included in the analysis, but their census variables were based on data from the 2011 Census.

3 . Study population and design

The study used a retrospective cohort design of 815,477 women in England who were first infected with coronavirus (COVID-19) between 8 December 2020 and 31 August 2021. The inclusion criteria for the study were:

- enumerated at the 2011 Census
- female self-reported sex according to their 2011 Census return
- aged 18 to 45 years at the start of the study period (8 December 2020)
- able to be linked to the 2011 to 2013 NHS Patient Register (to obtain an NHS number)
- able to be linked to at least one General Practice Extraction Service (GPES) Data for Pandemic Planning and Research (GDPPR) record (to identify active NHS patients at the start of the coronavirus pandemic)
- resident in England according to GDPPR and living in a private household according to the 2011 Census
- evidence of COVID-19 infection between 8 December 2020 and 31 August 2021
- no evidence of COVID-19 infection before 8 December 2020

Evidence of COVID-19 infection was determined by either a positive swab for COVID-19, using a polymerase chain reaction (PCR) test or lateral flow device (LFD) obtained through Pillar 1 or Pillar 2 testing, or a hospital inpatient episode or outpatient appointment with an International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) code for U07.1 (COVID-19, virus identified) or U07.2 (COVID-19, virus not identified) as the primary or secondary diagnosis.

We excluded anyone with evidence of COVID-19 before 8 December 2020. The index date for start of follow-up was the earliest evidence of COVID-19 infection occurring within the study period.

Data for COVID-19 vaccination were retrieved from the National Immunisation Management Service (NIMS). Vaccination status (unvaccinated, single-vaccinated or double-vaccinated) was defined as the number of doses received at least 14 days before the index date. This was to allow time for an antibody response to develop, as detailed in the [Vaccine uptake and SARS-CoV-2 antibody prevalence among 207,337 adults during May 2021 in England: REACT-2 study article](#), published on the medRxiv website.

Among women who were not pregnant when first infected with COVID-19, 68.6% were unvaccinated, 16.4% were single-vaccinated, and 15.0% were double-vaccinated. Among women who were pregnant when first infected with COVID-19, 87.2% were unvaccinated, 8.1% were single-vaccinated, and 4.7% were double-vaccinated.

4 . Method for identifying pregnancies

The total number of women identified as pregnant at index date in our study population was 33,549 (4.1% of the study population). Of these pregnancies, 30,740 were identified from birth notifications data. An additional 2,809 pregnancies were identified from Hospital Episode Statistics data.

Method for birth notifications data

Estimated conception date was calculated by subtracting gestational age from date of birth on the earliest birth notification occurring after the index date. Two weeks were then added to account for the fact that gestational age is defined from the start of the last menstrual period and conception is assumed to occur at the mid-way point (two weeks) of the four-week menstrual cycle.

Participants were classified as pregnant at index date if there was evidence of coronavirus (COVID-19) infection occurring between the estimated conception date and birth notification.

Gestational age was missing for 0.3% of the birth notifications used to estimate conception date. For these records, gestational age was imputed as 40 weeks for live births and 32 weeks for stillbirths. This is based on the approximate average length of gestation for these outcomes, as calculated from our [Birth characteristics dataset – 2020 edition](#).

Method for Hospital Episode Statistics (HES) data

We searched HES data for hospital episodes with evidence of ongoing pregnancy and end of pregnancy in the 42 weeks before the index date (see the [data tables](#) for the code lists used). Participants were classified as pregnant at index date if they met all of the following criteria:

- they had at least one episode with an International Classification of Diseases, 10th Revision (ICD-10) or Classification of Interventions and Procedures version 4 (OPCS-4) code indicating pregnancy over the 42 weeks prior to and including index date
- the most recent code over the 42 weeks prior to index date indicated ongoing pregnancy (not end of pregnancy)
- there was no end of pregnancy code in the six weeks prior to the most recent ongoing pregnancy code (some of the ongoing pregnancy codes relate to conditions that can be diagnosed in the post-partum period, up to six weeks after end of pregnancy)

We also searched the HES data for hospital episodes with evidence of ongoing pregnancy (routine obstetric scans only) and birth events occurring after the index date. Participants were classified as pregnant at index date if any of the following criteria were met:

- they had a hospital episode with an OPCS-4 code for a dating scan (R36.1) (normally performed at week 12 of pregnancy) up to 70 days after the index date
- they had a hospital episode with an OPCS-4 code for a mid-trimester scan (R36.3) (normally performed at week 20 of pregnancy) up to 126 days after the index date
- they had a hospital episode with an ICD-10 code for a live birth (Z37.0, Z37.2, Z37.5 or Z38) or mixed live and still birth (Z37.3 or Z37.6) up to 38 weeks after the index date
- they had a hospital episode with an ICD-10 code for a stillbirth (Z37.1, Z37.4 or Z37.7) up to 30 weeks after the index date

Estimating gestational age at infection

Gestational age at index date was estimated by calculating the difference between estimated conception date and index date. For pregnancies that were identified from NHS birth notifications data (91.6% of all pregnancies), the method of estimating conception date is described under the [Method for birth notifications](#) heading.

Of the pregnancies identified in HES only, we were able to estimate conception date for the 18.5% that had either a recorded hospital episode with an OPCS-4 code for a dating scan or mid-trimester scan, or an ICD-10 code for a birth outcome. Estimated conception date was calculated by:

- subtracting 70 days from the date of the earliest episode with an OPCS-4 code for a dating scan (normally performed at week 12 of pregnancy) occurring up to 28 weeks before or 10 weeks after the index date
- subtracting 126 days from the date of the earliest episode with an OPCS-4 code for a mid-trimester scan (normally performed at week 20 of pregnancy) occurring up to 20 weeks before or 18 weeks after the index date
- subtracting 266 days from the date of the earliest episode occurring up to 38 weeks after the index date with an ICD-10 code for either a live birth or a mixed live and stillbirth
- subtracting 210 days from the date of the earliest episode occurring up to 30 weeks after the index date with an ICD-10 code for a stillbirth

For pregnancies where more than one of these codes was recorded, priority was given to the mid-trimester scan (which is often timed more accurately based on the findings from the dating scan), followed by the dating scan, then live births and finally stillbirths, when calculating estimated conception date.

The estimated gestational age at index date was then calculated by adding two weeks to the number of weeks between the estimated conception date and the index date.

Stages of pregnancy at index date were then defined as trimester based on the gestational age in the following way:

- first trimester: 2 weeks to 13 weeks and 6 days (25.8%)
- second trimester: 14 weeks to 27 weeks and 6 days (35.6%)
- third trimester: 28 weeks or more (31.8%)

It was not possible to estimate gestational age for the remaining 6.8% of pregnancies.

5 . Statistical methods

Age-standardised rates

We calculated age-standardised rates of coronavirus (COVID-19) hospital admission (per 100,000 infections). This is defined as an inpatient episode with either U07.1 (COVID-19, virus identified) or U07.2 (COVID-19, virus not identified) as the primary diagnosis and occurring within 120 days of the index date.

The age-standardised rate is a weighted sum of age-specific COVID-19 hospitalisation rates. This is where the age-specific weights represent the relative age distribution in the standard population (in this case, the [2013 European Standard Population](#)).

Rates were calculated as follows:

$$\frac{\sum_i w_i r_i}{\sum_i w_i} = \times 100,000 \text{ study population}$$

where:

- i is the age group
- w_i is the number, or proportion, of individuals in the standard population in age group i
- r_i is the observed age-specific rate in the subject population in age group i

The observed age-specific rate in the subject population in age group (r_i) is calculated as:

$$r_i = d_i/n_i$$

where:

- d_i is the observed number of events (COVID-19 hospital admissions) in the subject population in age group i
- n_i is the population at risk in age-group i

The variance is the sum of the age-specific variances, and its standard error is the square root of the variance:

$$SE(ASR) = \sqrt{\frac{\sum \left(w_i^2 \frac{r_i^2}{d_i} \right)}{(\sum w_i)^2}}$$

where:

- r_i is the crude age-specific rate in the local population in age group i
- d_i is the number of COVID-19 hospital admissions in the local population in age group i

The normal approximation method for rates based on 100 or more hospital admissions were used to calculate 95% confidence intervals for the age-standardised rates. This is calculated as:

$$ASR_{LL/UL} = ASR \pm 1.96 \times SE$$

where:

- $ASR_{LL/UL}$ represents the lower and upper 95% confidence limits, respectively, for the age-standardised rate
- SE is the standard error

For rates based on fewer than 100 hospital admissions, confidence intervals were calculated using the method proposed by Dobson and others in [Confidence intervals for weighted sums of Poisson parameters](#). The full method is described in the [Association of Public Health Observatories' third technical briefing \(2008\) \(PDF, 2088KB\)](#).

In this method, confidence intervals are obtained by scaling and shifting (weighting) the exact interval for the Poisson distribution counts (number of hospital admissions). The weight used is the ratio of the standard error of the age-standardised rate to the standard error of the number of hospital admissions.

The lower and upper 95% confidence intervals are calculated as:

$$ASR_{lower} = ASR + (D_l - D) \cdot \sqrt{\frac{v(ASR)}{v(D)}}$$

$$ASR_{upper} = ASR + (D_u - D) \cdot \sqrt{\frac{v(ASR)}{v(D)}}$$

where:

- D_l and D_u are the exact lower and upper confidence limits for the number of hospital admissions, calculated using confidence limit factors from a Poisson probability distribution table
- D is the number of hospital admissions
- $v(ASR)$ is the variance of the age-standardised rate
- $v(D)$ is the variance of the number of hospital admissions

Cox proportional hazards regression

We used Cox proportional hazards regression models to assess how the rate of COVID-19 hospital admission varies by vaccination status, separately for women who were pregnant, or not pregnant, when they were first infected with COVID-19.

Hazard ratios for COVID-19 hospital admission were calculated for the single-vaccinated and double-vaccinated groups, relative to the unvaccinated group (reference group).

Study participants' follow-up time was censored at the earliest of index date plus 120 days (to avoid including outcomes relating to a subsequent infection episode, that is, reinfection), and date of death.

Analyses were adjusted for factors related to vaccination uptake and risk of COVID-19 hospital admission. We also adjusted for calendar time of infection to account for differences in COVID-19 variant and changes in hospital capacity over the study period.

Separate regression models were fitted for women who were pregnant when they were infected and women who were not pregnant when infected. In both models, we adjusted for:

- age
- calendar time of infection
- region
- Index of Multiple Deprivation decile
- Rural Urban Classification (major conurbations, minor conurbations, cities and towns, towns and fringes, villages, hamlets and other isolated dwellings)
- ethnic group (Asian, Black, Mixed, White, Other)
- English language proficiency (main language, speak English very well or well, do not speak English well or at all)
- country of birth (UK versus non-UK)
- keyworker status (yes versus no)
- highest qualification held (degree or above, two or more A-levels or equivalent, five or more GCSE passes or equivalent, one to four GCSE passes or equivalent, apprenticeship or other, no qualifications)
- disability status (not disabled, disabled and limited a little, disabled and limited a lot)
- health status (very good, good, fair, bad, very bad)

Please note, for age and calendar time of infection, restricted cubic splines with an internal knot at the fiftieth percentile and boundary knots at the tenth and ninetieth percentiles of the distributions were used.

Geographical variables were based on participants' most recent postcode in General Practice Extraction Service Data for Pandemic Planning and Research data. All other covariates were from the 2021 Census, except for 8.6% of participants who could not be linked to the 2021 Census. For these participants, 2011 Census data were used.

We tested whether the association between vaccination status and COVID-19 hospital admission varied by pregnancy trimester in the 93.2% of participants identified as being pregnant at index date and for whom gestational age could be calculated. This was achieved by assessing the [statistical significance](#) of the interaction between trimester and vaccination status.

6 . Future developments

This analysis forms part of a wider programme of work into coronavirus (COVID-19) in pregnancy. In future work, we plan to assess:

- birth outcomes following COVID-19 infection during pregnancy, by vaccination status
- COVID-19 vaccine uptake during pregnancy
- birth outcomes following COVID-19 vaccination during pregnancy

An overview of existing evidence is available in the [Systematic review and meta-analysis of the effectiveness and perinatal outcomes of COVID-19 vaccination in pregnancy article](#) in the Nature Communications online journal.

We also plan to refine the methodology for identifying pregnancy status using different data sources (for example, the [NHS Maternity Services Data Set](#)), through collaboration with academic partners with experience in [building birth cohorts from hospital records](#). More broadly, the method for identifying pregnancy from health records and births data has the potential to be used in other projects related to pregnancy.

7 . Related links

[Coronavirus \(COVID-19\) hospital admissions by vaccination and pregnancy status, England: 8 December 2020 to 31 August 2021](#)

Bulletin | Released 11 July 2022

Estimates of coronavirus (COVID-19) hospitalisation rates in England by vaccination status in pregnant and non-pregnant women.