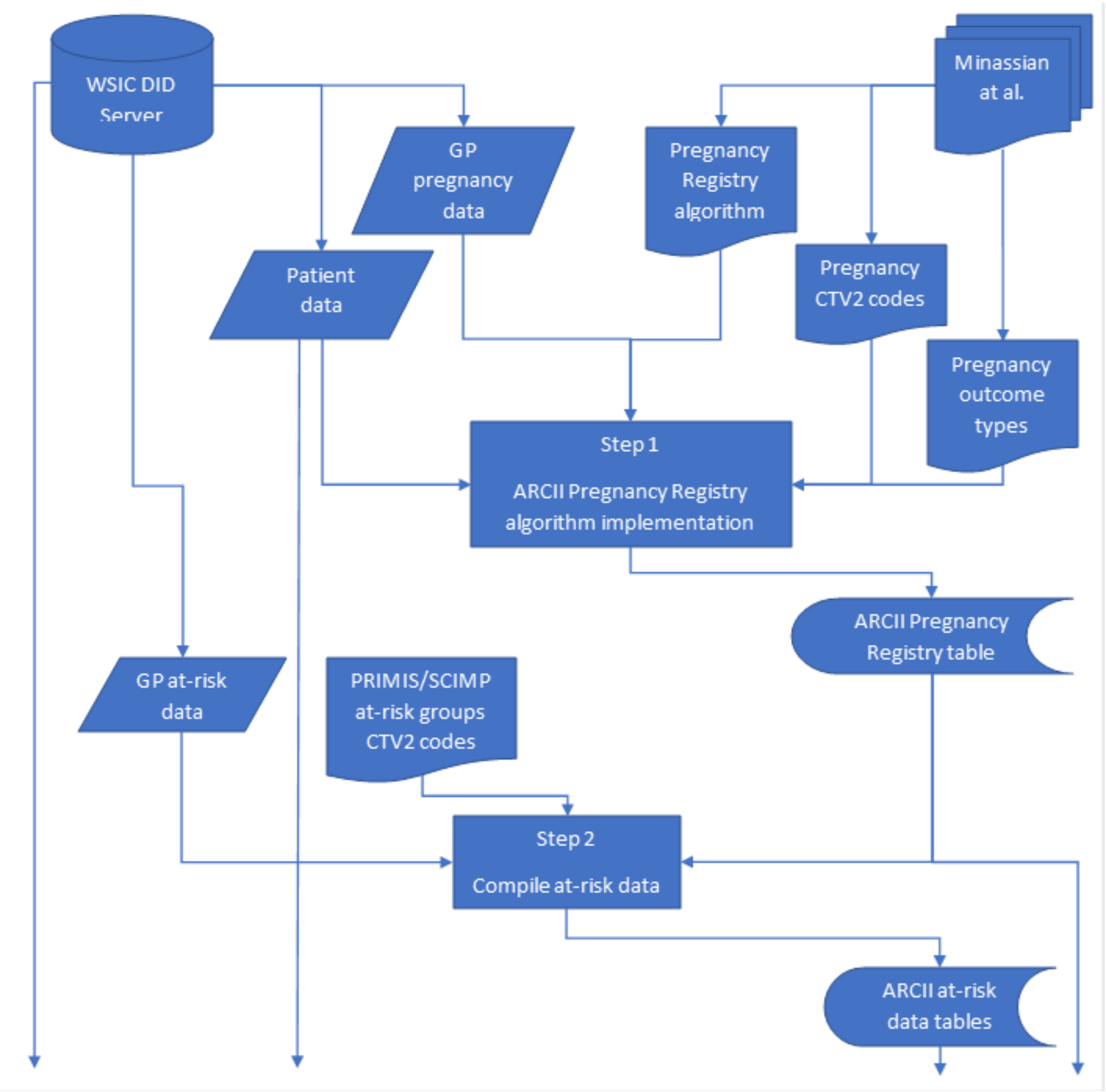
Health and socio-demographic characteristics associated with uptake of seasonal influenza vaccination amongst pregnant women: Retrospective cohort study

Supplementary Appendix S1: Generating the pregnancy register and vaccination status dataset


WSIC DID Server
Minassian at al.
GP pregnancy data
Pregnancy Registry algorithm
Patient data
Pregnancy CTV2 codes
Pregnancy outcome types
Step 1
ARCII Pregnancy Registry algorithm implementation
ARCII Pregnancy Registry table
GP at-risk data
PRIMIS/SCIMP at-risk groups CTV2 codes
Step 2
Compile at-risk data
ARCII at-risk data tables

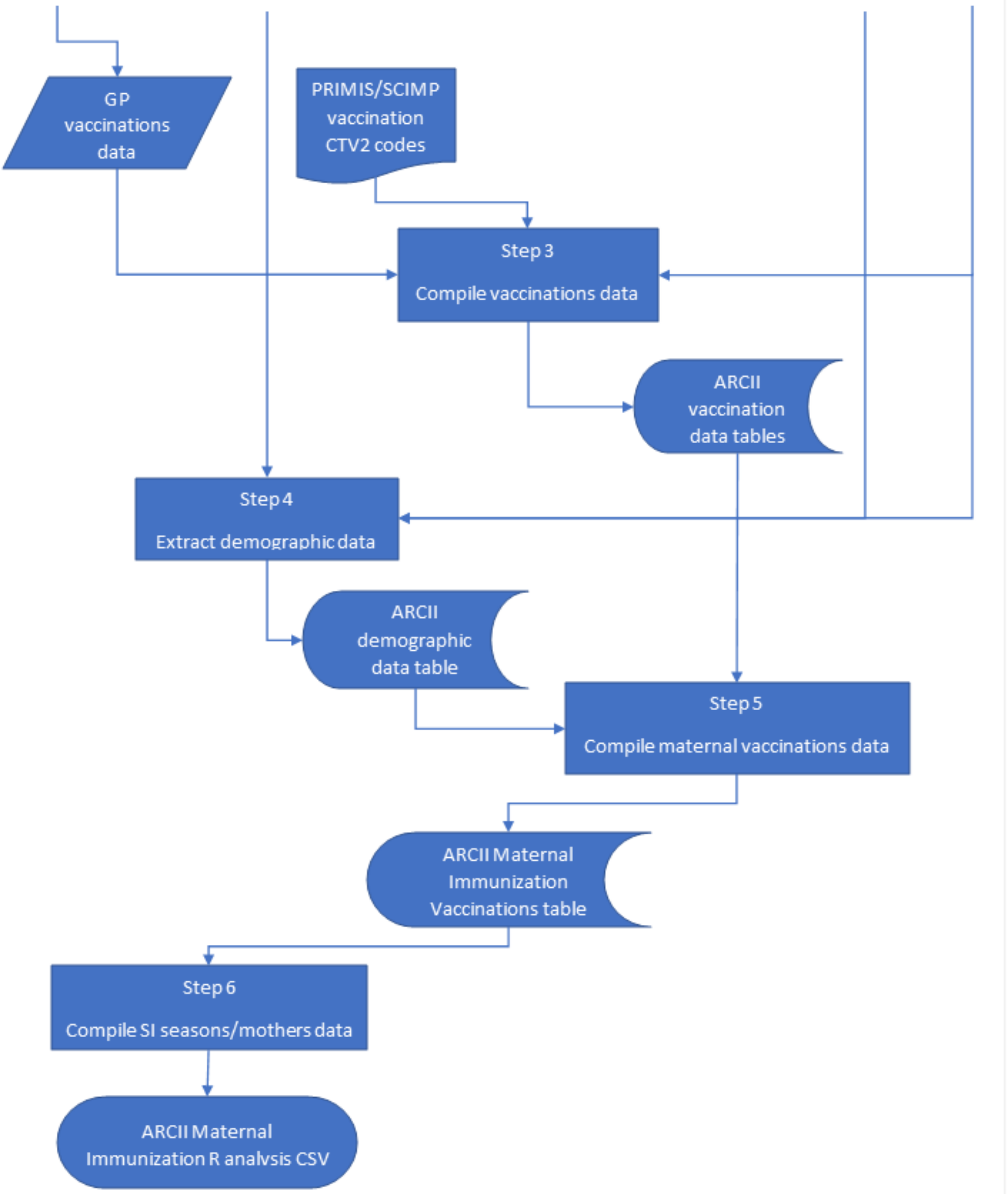

GP vaccinations data
PRIMIS/SCIMP vaccination CTV2 codes
Step 3
Compile vaccinations data
ARCII vaccination data tables
Step 4
Extract demographic data
ARCII demographic data table
Step 5
Compile maternal vaccinations data
ARCII Maternal Immunization Vaccinations table
Step 6
Compile SI seasons/mothers data
ARCII Maternal Immunization R analysis CSV

Figure S1: Algorithm for establishing maternal immunisation data set from primary care records

The generation of the maternal immunisation data set was completed in stages. We began by identifying all pregnancy episodes and thus all women registered in NWL who became pregnant during our period of observation, from September 2010 to February 2020. For this cohort of pregnant women, we then extracted their GP records indicating diagnoses of at-risk conditions, Seasonal Influenza vaccinations, and demographic data. By combining the extracted records, we compiled a maternal immunisation table to match pregnancies overlapping influenza seasons with relevant vaccinations. We then sorted the table by seasons and grouped any instances of multiple pregnancies per woman per season into a single record per woman per season. An overview of the algorithm described below is given in figure S1.

Step 1. ARCII Pregnancy Register algorithm implementation

For our analysis we needed to establish the number of pregnant women within the period of interest. We considered two methods.
Method one:
The annual estimation of the Seasonal Influenza (SI) vaccination uptake conducted by Public Health England (PHE) involves the use of a set of pregnancy-related Clinical Terms Version 2 (CTV2) read codes. A record with one of these codes in a woman's GP attendance history is considered to indicate that the woman is pregnant for the purpose of estimating the number of pregnant women during a given influenza season.
Method two:
A research team at the School of Hygiene and Tropical Medicine in London [Minassian at al.] has created an algorithm for the identification of distinctive pregnancy episodes from a set of GP attendance records. This algorithm too relies on a set of CTV2 codes, as well as on an additional set of medical entity codes. The algorithm's logic takes into consideration the type of records – delivery, perinatal, antenatal, etc., - and any additional information encoded such as gestational lengths. The final product is a register of pregnancies, each with their estimated start, end, and trimesters dates, type of outcome, mother's identifier, and age, etc.

We decided to adopt the Minassian et al. algorithm as opposed to following PHE's practice. The additional information provided by the pregnancy register allowed for a more precise match between influenza seasons, relevant pregnancy episodes and SI vaccinations. While we attempted to adhere to the method of Minassian et al. for generating the register as published, some adaptations were necessary due to the difference between the data available to us in the NWL Discover database and that available in the UK CPRD Datalink GP database.
Some of the more significant differences:

- The original work used a 7-byte format of the CTV2 codes, while the NWL Discover database only stored their 5-byte format. This prevented the utilisation of the full list of 4200 pregnancy-related codes. Only 3611 were used in our version of the algorithm.
- Data concerning any mother-baby links present in the CPRD Datalink GP database was not available in the NWL Discover database.
- Equivalents to entity codes (structured data areas in the practice software system Vision where additional data may be entered by the GP or practice staff) used in Minassian et al. pregnancy algorithm were not identified in EMIS and SystmOne or uploaded into the NWL Discover database.
- The age range for our target cohort was 15-49, instead of the 11-49 used by Minassian et al.

Step 2. Compile at-risk data

The result of Step 1 is a list of women who became pregnant at least once during our period of observation. We next identified, for each woman in the list, all GP records that implied a diagnosis of any one of several at-risk conditions. The at-risk conditions were the same used by PHE [PRIMIS]: asthma, chronic respiratory disease, chronic heart disease, chronic kidney disease, asplenia, liver disease, chronic neurological disease, diabetes, immunosuppression, and morbid obesity. For each at-risk condition, PRIMIS supplied a list of related CTV2 read codes. We determined whether a pregnant woman belonged to each one of the at-risk groups in the same manner that PHE did in their annual SI vaccination uptake estimation: presence of a GP record with any one of the at-risk-related CTV2 codes was accepted to indicate a diagnosis of the given condition.

There were some differences between our and the PHE's method for processing records related to asthma, and diabetes:

- For some GP records to be accepted as proxies for asthma diagnoses, their specific CTV2 codes required a combination with GP records with specific CTV2 codes for asthma prescriptions. Due to the much longer period of observation used in our case study compared to that used in the PHE's, we did not look for such combinations and accepted the presence of GP records with such asthma CTV2 codes as sufficient to indicate asthma diagnoses.
- In the PHE's method, GP records with diabetes CTV2 codes were not deemed to identify diabetes diagnoses if GP records with specific "diabetes resolved" CTV2 codes were also present and were within the same influenza season. Again, due to the much longer period of observation, we did not look for such "diabetes resolved" CTV2 and accepted the presence of GP records with diabetes CTV2 codes to flag diabetes diagnoses.

Step 3. Compile vaccinations data

For each woman in the list of pregnant women compiled in Step 1, we extracted all GP records indicating Seasonal Influenza vaccinations. For each vaccination type, PRIMIS supplied a list of related CTV2 read codes. We determined whether a pregnant woman was vaccinated in the same manner that PHE did: a presence of a GP record with any one of the SI-vaccination-related CTV2 codes was accepted to indicate a vaccination for the given influenza season.

Step 4. Compile demographic data

For each woman in the list of pregnant women compiled in Step 1, we extracted some demographic data such as age, GP practice code, postcode sector code, ethnicity, Local Authority district name, LSOA code, CCG code, IMD decile. Only the age, ethnicity, and IMD Decile were used in the final analysis.

In this step, for each woman, we also determined the earliest date when an at-risk condition was diagnosed, if at all. This date was used to flag the presence of the at-risk condition during subsequent pregnancy episodes.

An adjustment was made to the register records of pregnancies classified as "outcome unknown". These pregnancy episodes did not have end-of-pregnancy dates assigned by our identification method (in accordance with the Minassian et al. algorithm). For the needs of our analysis, such dates were estimated by adding one week to the date of the oldest GP antenatal record of each "outcome unknown" episode. This method was consistent with the end-of-pregnancy estimations for other types of pregnancy episodes identified by the Minassian et al. instructions.

Step 5. Compile maternal immunisation data

In this step, we determined which of the pregnancy episodes overlapped with which of the influenza seasons under consideration. This was done as follows:

Each pregnancy episode length was checked against each influenza season's start and end dates. For pregnancies overlapping with more than one influenza season, we assigned the pregnancy to one season only. Which season was determined as follows: pregnancy episodes beginning in the month of January were marked as overlapping with the previous year's season. All other pregnancy episodes were marked as overlapping with their year’s season if their end date was after or equal to the season’s start date, and their start date was before or equal to the season’s end date. The purpose of this was to avoid double counting when calculating per-season vaccination uptake statistics.

The final output of this step was a maternal immunisation vaccinations table. Each row of which constitutes a record of a pregnancy episode within an influenza season. Each record contains fields uniquely identifying the mother, the pregnancy’s start and end dates, demographic data such as the mother’s age, ethnicity, and GP practice code, LA District name, LSOA code, CCG code, and IMD Decile. It also contains fields indicating any at-risk conditions diagnoses at the start of the pregnancy, as well as flags marking SI vaccinations during or prior to the pregnancy episode, if any.

Step 6. Compile SI seasons/mothers' data

In this last step, the product of Step 5, the maternal immunization matrix is sorted by influenza seasons and some further adjustments were made. For some women, the pregnancy register algorithm identified multiple pregnancy episodes which occurred within the same influenza season. Since this could lead to erroneous double counting of vaccinations for the season, we grouped all such instances of multiple episodes per women per season into a single pregnancy.

The resulting table represents the input data set to the regression analysis code. Each row of the set represents a mother's record for a given season (if she was found to have been pregnant in that SI season). Each record contains the mother's demographics (age band, ethnicity, and IMD decile), the mother's diagnoses for at-risk conditions (binary values), and the mother's SI vaccination flag (binary value).

Supplementary Appendix S2: Multiple Imputation Analysis

The only variables in the analysis dataset that have missing values were ethnicity (5,836 of 451,954, 1.3%) and Index of Multiple Deprivation (IMD) quintile (28,118 of 451,954, 6.2%). At least one of these two fields was missing in 33,539 pregnancies (7.4%). The main mechanism by which this data could be missing not at random is through these women being less able or inclined to access the health system, which would lead to their records being incomplete and also potentially to them being less likely to get the seasonal influenza vaccination. If this were a major driver of missingness of these demographics, one would expect to frequently see both demographic variables missing together. However, this was only the case in 415 (1.2%) of the 33,539 pregnancies with at least one missing variable, or 7.1% of the 5,836 pregnancies with missing ethnicity (the least of the two). This indicates that it is plausible that reduced access is not the primary reason behind the missing data, and therefore that it is reasonable to assume that the data are missing at random given the other covariates, for the purposes of an imputation analysis.

To understand better the potential impact of this missing data on our findings, we undertook a multiple imputation analysis.

We used the Amelia II R package to produce five imputations of the data.[1] This package uses the expectation-maximisation with bootstrapping approach to multiple imputation. We used all covariates included in the main analysis in the imputation model, except for the GP practice the woman is registered with. The full mixed-effects multivariable model of the main analysis was then fitted on each of the five imputed datasets, and the results combined using Rubin's rules.[2]

Results of the combined analysis following multiple imputation are shown in table S1, alongside the results of the main analysis, as odds ratios for seasonal influenza vaccination with 95% confidence intervals. Any differences are small, and the findings of the main analysis are the same as those following multiple imputation.

Table S1: *Results of multiple imputation analysis compared with main analysis. Odds of seasonal influenza vaccination among pregnant women registered with a GP in North West London from September 2010 to February 2020.*

Characteristic	**Main analysis**			**Multiple Imputation**			**Missing Information**
	OR	**(95% CI)**		**OR**	**(95% CI)**		
Age **(15 – 19 rc)**							
20 - 24	1.32	1.25	1.40	1.33	1.26	1.40	0%
25 - 29	1.58	1.50	1.67	1.59	1.51	1.68	0%
30 - 34	1.68	1.59	1.76	1.69	1.60	1.77	0%
35 - 39	1.56	1.48	1.64	1.57	1.49	1.65	0%
40+	1.17	1.10	1.24	1.18	1.11	1.25	0%
Ethnicity **(Asian or Asian British rc)**							
Black or Black British	0.55	0.53	0.57	0.55	0.54	0.57	0%
Mixed	0.63	0.60	0.66	0.63	0.60	0.66	1%
White	0.66	0.65	0.68	0.66	0.65	0.68	2%
Other ethnic groups	0.72	0.70	0.74	0.72	0.70	0.74	1%
Unknown	0.42	0.39	0.46	-	-	-	-
IMD Quintile **(Q1 rc) (most deprived)**							
Q2	1.03	1.00	1.06	1.04	1.01	1.07	6%
Q3	1.06	1.03	1.09	1.07	1.04	1.11	7%
Q4	1.07	1.04	1.11	1.09	1.05	1.13	15%
Q5	1.16	1.11	1.21	1.17	1.13	1.22	7%
UNKNOWN	1.00	0.96	1.04	-	-	-	-
At-risk Group							
Asthma	1.50	1.46	1.54	1.50	1.46	1.54	0%
Respiratory	1.46	1.19	1.79	1.46	1.19	1.79	0%
Heart	1.43	1.30	1.57	1.43	1.30	1.57	0%
Kidney	1.18	0.96	1.44	1.18	0.96	1.45	0%
Liver	1.29	1.11	1.51	1.30	1.11	1.52	0%
Asplenia	1.59	1.43	1.76	1.59	1.44	1.76	0%
Neurological	1.27	1.12	1.44	1.27	1.12	1.45	0%
Diabetes	2.87	2.68	3.07	2.87	2.68	3.07	0%
Immunosupression	1.84	1.62	2.10	1.84	1.62	2.10	0%
Morbid_obesity	1.13	1.07	1.19	1.13	1.07	1.20	0%
Influenza Season (years)	1.14	1.14	1.15	1.14	1.14	1.15	0%

Odds ratios calculated using logistic regression. Reference category denoted by rc for each categorical variable. Missing information refers to Rubin's fraction of missing information.[2]

References

1. Honaker, James K Gary. Amelia II: A Program for Missing Data. *J. Stat. Softw.* 2011;45(7): 1–47.

2. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley; 1987.