Version 1.1

Study Protocol: Long-term trends and influence of specific risk factors in prevalence of singleton pregnancies with a diagnosis of Down's syndrome in Scotland between 2000 and 2021

Version history

Version	Date	Summary of changes
V1.0	30/05/2024	Original study protocol and statistical analysis plan (SAP)
V1.1	24/02/2025	1. Update to the use of some risk factors/socio-demographic variables in analysis: Variable 'type of trisomy screening' was noted to differ regionally until NIPT implementation in 2020, therefore we have removed this as a risk factor in the prevalence and proportion analyses. Maternal ethnicity variable in SLiCCD has a very high proportion of missing data throughout 2000-2021 and has therefore been excluded from analysis of socio-demographic associations with proportion or prevalence analyses. Infant sex is often unknown for babies with DS where the pregnancy ended in termination or late fetal loss, therefore infant sex has only been used as a socio-demographic factor for analyses regarding live births. 2. Addition of SLiCCD VS enhanced SLiCCD validation analysis.

The publication of this document has been authorised by:

Signature:  ……

Print name: Rachael Wood……

Date: 15/05/2025……

Signature:  …….

Print name: …Rute Vieira……

Date: 28/05/2025……..

Contents

Plain English summary

Background

Down's Syndrome (DS, Trisomy 21) is a congenital condition where babies have a third complete or partial copy of chromosome 21, rather than two. Biochemical screening tests (combined or quadruple screening), which look at certain markers in the blood to identify pregnancies of a baby likely to have DS, have been routinely offered to pregnant women in Scotland since 2001. Those who had a higher chance result, indicating a pregnancy of a baby with DS, were offered an invasive diagnostic test (chorionic villus sampling or amniocentesis) to confirm the diagnosis. This invasive diagnostic test carries a small risk of miscarriage. In part to reduce the number of women unnecessarily undergoing this potentially risky invasive test, a more accurate second line screening test, non-invasive prenatal testing (NIPT), was introduced to Scotland in September 2020. NIPT was introduced for those pregnancies that have been identified as having a higher chance of DS from biochemical screening. As it is not a diagnostic test, women would still have to undergo invasive prenatal testing to confirm a result from NIPT. Therefore, currently women with a 'higher' chance result after the first line (biochemical) screening are offered the choice of no further testing, second line (NIPT) screening or an invasive diagnostic testing. The introduction of NIPT in Scotland offers an important opportunity to examine the impact of this new test on the choices of pregnant women and ultimately, understand how this may affect the outcomes of babies with DS in this country. To achieve this, the pre-implementation secular trends in total birth prevalence and live birth prevalence of pregnancies diagnosed with DS must first be understood, as well as identifying maternal and infant socio-demographic factors that may be associated with prevalence.

Changes to the screening tests offered may affect the timing of and decisions regarding undergoing screening or further diagnostic testing for DS, and the likelihood of identifying babies with DS. This may impact the total or live birth prevalence of DS in Scotland, if more babies were diagnosed prior to a miscarriage or termination of pregnancy. By using the Scottish Linked Congenital Conditions Dataset (SLiCCD), this study aimed to investigate trends over time in the total and live birth prevalence of DS in Scotland, and the associations with maternal and infant socio-demographic factors between 2000-2021. This will provide important context for understanding the impact of the recent implementation of NIPT in the screening programme, obtaining accurate and up-to-date estimates for DS prevalence and exploring associated factors in the population allows for appropriate and targeted plans, resource allocation, and service delivery for improved health and social care.

What does the study involve?

This project will use SLiCCD, a dataset of linked routine healthcare and national statutory records held by Public Health Scotland for analysis of the trends in prevalence over time of pregnancies with a diagnosis of DS in Scotland. It will also examine the association between total and live birth prevalence of DS and maternal and infant socio-demographic factors.

Abbreviations

DS – Down's syndrome

NIPT – non-invasive prenatal testing

PHS – Public Health Scotland

CARDRISS - Congenital Condition and Rare Diseases Registration and Information Service for Scotland

SLiCCD - Scottish Linked Congenital Condition Dataset

SLiPBD - Scottish Linked Pregnancy and Baby Dataset

Who is the main contact?

The study is being undertaken as part of a PhD studentship funded by Medical Research Scotland and Public Health Scotland (PHD-50200-2020) being carried out at the University of Aberdeen.

PhD candidate: Ms Elinor Sebire, Institute of Applied Health Sciences, University of Aberdeen, e.sebire.21@abdn.ac.uk

Principal supervisor: Dr Rute Vieira, Senior Lecturer in Medical Statistics, Institute of Applied Health Sciences, University of Aberdeen, rute.vieira@abdn.ac.uk

External supervisor: Dr Rachael Wood, Consultant in Public Health Medicine, Public Health Scotland, rachael.wood@phs.scot

1. Introduction

Non-invasive prenatal testing (NIPT) was implemented as a second-line screening test into the existing Scottish antenatal screening pathway for Down's syndrome (DS) in September 2020. This test provides an additional, and more accurate screening test for those women with a pregnancy identified as higher chance for DS by first-line screening (combined or quadruple screening). It is important to understand how the implementation of NIPT may have changed the decisions made about undergoing screening or diagnostic tests, pregnancy decisions (including continuation or termination of pregnancy), and on the outcomes of the population of individuals with DS in Scotland. To understand the potential impact of NIPT in Scotland on pregnancies with DS since its implementation we must first examine the baseline trends in prevalence of this population. Exploring the prevalence of total and live births of babies with DS over time, along with investigating associated maternal and infant socio-demographic factors provides important context for understanding the impact of changes to the screening programme.

There is a lack of comprehensive and comparable analysis of long-term trends in the total and live birth prevalence of DS among babies of singleton pregnancies in Scotland, and of the maternal and infant factors that may influence these prevalences.

2. Aims

To investigate the long-term trends in total and live birth prevalence of singleton pregnancies with a diagnosis of DS in the period 2000 to 2021, in Scotland. To explore the association between specific maternal and infant socio-demographic factors and the total and live birth prevalence of DS.

3. Objectives

1. To summarise the time trends for total and live birth prevalence of DS among babies from pregnancies ending in Scotland between 2000 and 2021.
2. To understand how specific maternal and infant socio-demographic risk factors (maternal age, Scottish index of multiple deprivation (SIMD), infant sex and NHS health board of residence) are associated with the total and live birth prevalences of babies with DS in Scotland.
3. For babies from pregnancies ending between 1 April 2019 and 31 December 2020, to validate the completeness of recording of babies with DS within the Scottish Linked Congenital Condition Dataset (SLiCCD) used for the primary analysis in this study, against the enhanced SLiCCD dataset sourced from a combination of SLiCCD and genetic testing data.

4. Methods

Data sources

An anonymised extract of individual-level pregnancies or babies registered with DS in the SLiCCD cohort will be placed into a PHS confidential server. All analyses will be done through the PHS secure server, using R . All researchers on this project have honorary access and permissions for the anonymised PHS datasets.

Scottish Linked Congenital Condition Dataset (SLiCCD)

This is an existing data asset held by PHS. It is a linked file derived from population level national datasets relating to pregnancies with a congenital condition between 2000-2021. All pregnancies with a congenital condition meeting EUROCAT inclusion criteria are included in SLiCCD, based on diagnostic (ICD10) codes present in the source records. Pregnancies are included if the associated pregnancy ends in a termination of pregnancy at any gestation, a spontaneous loss at ≥20 weeks gestation, or a live birth at any gestation (with the baby diagnosed at any point prior to their first birthday).

SLiCCD is derived from the following existing national health datasets and statutory records held by Public Health Scotland (PHS Technical Report, 2023):

- National Records of Scotland (NRS) statutory live births, stillbirths, and deaths for babies up to 1 year.
- Hospital maternity care delivery, miscarriage, and termination of pregnancy discharge (SMR02)
- Statutory termination of pregnancy notifications (AAS)
- Hospital neonatal care discharge:
    - SMR11 records from 2000 – April 2003
    - Scottish Birth Record (SBR) from April 2003
- General hospital discharge records for babies aged up to 1 year (SMR01)
- Perinatal death enhanced surveillance:
    - Scottish Stillbirth and Infant Death (SSBID) records for 2000 – 2012
    - Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK (MBRRACE-UK) records for 2013 – 2020

NRS live and stillbirths

Aggregate data on the number of live and, separately, stillbirths statutorily registered by NRS have been used as denominators to allow calculation of DS total birth and live birth prevalence. Aggregate data has been obtained for each year from 2000 to 2021, and broken down by maternal age group (<20, 20-24, 25-29, 30-34, 35-39, ≥40, unknown); maternal deprivation level (Scottish Index of Multiple Deprivation [SIMD] quintile 1, 2, 3, 4, 5, unknown), maternal NHS health board of residence (NHS Ayrshire and Arran; NHS Borders; NHS Dumfries and Galloway; NHS Fife; NHS Forth Valley; NHS Grampian; NHS Greater Glasgow and Clyde; NHS Highland; NHS Lanarkshire; NHS Lothian; NHS Orkney; NHS Shetland; NHS Tayside; NHS Western Isles; unknown), and infant sex (male, female, unknown). In Scotland, live births at any gestation and stillbirths where the baby is born at ≥24+0 gestation or over are registerable by law.

Enhanced SLiCCD

Enhanced SLiCCD is a dataset which includes pregnancies ending 1 April 2019 to 31 December 2020 with DS, Edwards' syndrome or Patau's syndrome diagnoses sourced from both SLiCCD and genetic testing data. Genetic testing data is data returned from invasive prenatal diagnostic tests (chorionic villus sampling or amniocentesis) or post-natal karyotype testing of pregnancies or babies from the relevant laboratories. The genetic testing data was used to confirm SLiCCD records**,** update specific diagnoses within SLiCCD records, or to add records in the case where they are not present in SLiCCD. Pregnancies within enhanced SLiCCD will therefore have a 'confirmed' diagnosis if the initial SLiCCD record matches a genetic testing record or with genetics results only record, or a 'probable' diagnosis where a SLiCCD record is not matched to a genetic testing record.

Statistical methods

We will provide the R code used in analyses on a publicly accessible GitHub repository alongside dissemination of results.

Only singleton pregnancies with confirmed DS on the SLiCCD dataset from 2000 – 2021 will be included in the analysis. Pregnancy outcomes of interest:

- Late fetal loss (defined as up to 20-24 weeks gestation)

- Stillbirth (defined as ≥ 24 weeks gestation)

- Termination of pregnancy (any gestation)

- Live birth

The research cohort within SLICCD will be summarised using appropriate descriptive analysis, to understand the probabilistic distribution of the variables within the cohort, and identify missing data as described below. Descriptive statistics will include frequency/percentages, and mean (SD), median (IQR), depending on the data characteristics. The distribution of each variable within the dataset will be explored using boxplot graphs (continuous variables) or other appropriate visualisations. Tables and bar plots will describe the categorical variables within the dataset.

Missing data:

The pregnancy and post-natal outcomes held in the SLiCCD extract should not contain any missing data, as is consistent with the method for which the SLiCCD cohort is produced. However, the maternal and sociodemographic variables (maternal age, NHS health board of care, etc.) could have missing data. The number and percentage of records with missing data will be reported for each variable of interest from SLiCCD. Descriptive statistics will be produced for the missing data for each relevant variable. Analyses will be performed on complete data only, unless it is found that there is systematic missing data. If systematic, then appropriate statistical methods will be used to impute missing data and sensitivity analyses performed to understand the significance of using this imputed data (Jakobsen et al., 2017). If more than 40% of the data for a variable is missing, the variable will not be included in further analyses.

Total and live birth prevalence of DS

Using data from SLiCCD, we will report the total number of singleton pregnancies with DS from ending in 2000 to 2021. We will report the number of babies by pregnancy outcome (live birth, late fetal loss, stillbirth, and termination of pregnancy). We will report the DS total birth prevalence per 10,000 total births and the live birth prevalence per 10,000 live births, with 95% confidence intervals (CIs).

In line with PHS and EUROCAT, we will use NRS stillbirths and live births as the denominator for the total birth prevalence, with the limitation that babies with DS and pregnancy outcome of late fetal loss or termination of pregnancy before 24+0 weeks included in the numerator will not be included in the denominator. Total and live birth prevalence will also be calculated by maternal and infant socio-demographic factors of interest.

Time trends and association with maternal and infant socio-demographic factors

Overall secular trends of total and live birth prevalence for babies with DS, as well as association with maternal and infant socio-demographic factors between 2000 and 2021 in Scotland, will be explored as follows:

<u>Objective 1:</u>

1. Using appropriate regression models for rate data (prevalence/proportion) (i.e., Poisson regression or equivalent if assumptions for Poisson regression are violated), we will model the outcomes' time trends between 2000-2021 assuming: 1) a linear trend by including the variable measuring time (Year) as a covariate in the model, and 2) a non-linear time trend using restricted cubic splines to transform and model the variable Year. DS total birth prevalence and live birth prevalence will be calculated using the total NRS denominator cohort.
2. We will use the Akaike Information Criteria (Akaike, 1975) (AIC), which estimates the relative information lost by a given model, as the criteria for deciding which model represents the data better (if AIC is at least 2 units lower than the other model). If the AIC is similar (within 2 units), the simpler (linear) model will be chosen, or the model with fewest knots.
   Time trends plots:

3. Using the model chosen in 2. we will estimate the model's predicted prevalence/proportion for each year between 2000-2022 and 95% confidence intervals of the predicted values and will plot them together with the crude prevalences/proportions, which will allow us to visualise the (linear or non-linear) trend over time and its confidence interval.

   Coefficients' tables:

4. If there is evidence of a linear time trend, using the model chosen in 2 we will report the crude prevalence risk ratio (PRR), its 95% confidence interval and p-value for the annual change in prevalence between 2000 and 2021 and assess whether this change is statistically significant.

5. On the other hand, if the time trend is non-linear, to quantify the annual change in prevalence, piecewise regression will be used where we will:

   a. identify different sequential time periods between 2000 and 2021 with a linear time trend, based on the plot produced in 4.
   b. estimate crude PRRs, their 95% confidence intervals and p-values for each time period and assess whether there is evidence of a statistically significant linear trend during each time period in the prevalence.

Objective 2:

6. Three models will be used for each sociodemographic variable of interest (maternal age, SIMD, infant sex and health board), to assess whether there is evidence for an association between the sociodemographic variable and the prevalence/proportion:

   a) First, we will estimate the crude PRRs, their 95% confidence intervals and p-values to describe the association between each of the sociodemographic factors and the prevalence/proportion using appropriate regression models for rate data (i.e., Poisson regression or equivalent if assumptions for Poisson regression are violated). Separate regression models will be built for each sociodemographic variable.

   b) To estimate PRRs for each sociodemographic factor, adjusted for the time trend , we will extend the model chosen in 2. By including each sociodemographic factor as an additional variable in the model. Separate regression models will be built for each sociodemographic variable.

   c) We will then further extend the model chosen in 2. to adjust for the other sociodemographic factors obtaining our final adjusted model. If data are sparse, we will consider not including variables which are found not to be associated with the prevalence in crude analyses in the fully adjusted model; if there are still issues with data sparsity, we will collapse categories within the sociodemographic factors (specifically for maternal age and maternal deprivation), with any grouping informed by descriptive analyses.

R packages glmmTMB and VGAM will be used for the regression analyses.

Objective 3:

7. Validation of SLiCCD with enhanced SLiCCD

We will use data from the overlapping period from 1st April 2019 to 31st December 2020 to calculate the positive predictive value (PPV) and sensitivity of recording of babies with DS in SLiCCD, compared to enhanced SLiCCD. This will demonstrate the effect of adding genetic testing data to the dataset. We will only include singleton pregnancies ending at over 20 weeks gestation for the comparison. The two datasets will be linked by common variable 'mother index number'.

PPV is the percentage of babies with DS included in SLiCCD that were also recorded as having DS in enhanced SLiCCD. Sensitivity was calculated as the percentage of babies with DS included in enhanced SLiCCD that were also recorded as having DS in SLiCCD. Analyses were repeated for the PPV and sensitivity of recording babies with DS by pregnancy outcome: termination of pregnancy, stillbirth, and live birth outcomes.

5. Governance

Ethics

The decision tool provided by the NHS Health Research Authority indicates that NHS research ethics committee approval is not required for this study.

Information governance

This analysis forms part of the work towards a PhD funded by Medical Research Scotland, with contribution from Public Health Scotland as the external partner organisation. The host organisation for the PhD is University of Aberdeen. Honorary PHS contracts are in place for the PhD student and principal University of Aberdeen supervisor to facilitate this work. The following joint governance for the NIPT evaluation and PhD has been undertaken and is in place: A PHS Data Protection Impact Assessment has been completed for this programme of work and approved by the PHS Data Protection Officer and relevant Information Assess Owner (DP21220627). Additional approval from the Public Benefit and Privacy Panel for Health and Social Care was required and secured (2122-0261).

Data will be anonymised for the purpose of this project by data analysts within PHS before extracts are transferred for access by researchers in the PHS secure server space for analysis. Researchers at the University of Aberdeen will only have access to data for singleton pregnancies and screening or outcomes relating to babies with DS. All data will be analysed on PHS secure servers following PHS information and statistical governance policies.

6. Outputs

The analyses from this protocol will be submitted for publication in an appropriate journal. Research findings will also be shared at relevant conferences.

7. References

Akaike H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 1974;19: 716 – 723.

Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. BMC Medical Research Methodology 2017;17.

Public Health Scotland. Congenital conditions in Scotland 2000 to 2020: Technical Report 2023. https://publichealthscotland.scot/media/21100/2023-08-01_congenitalconditionsinscotland2020_technicalreport.pdf (accessed 02/05/2025).