

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)



Swansea University  
Prifysgol Abertawe

## Data resource profile: Scottish Linked Pregnancy and Baby Dataset (SLiPBD)

Laura Lindsay<sup>1,†</sup>, Kate Mark<sup>1,†</sup>, Emily Moore<sup>1,†</sup>, Jade Carruthers<sup>1</sup>, Leanne Hopkins<sup>1</sup>, Denise Jennings<sup>1</sup>, and Rachael Wood<sup>1,2,\*</sup>

### Submission History

Submitted:	29/02/2024
Accepted:	31/05/2024
Published:	18/07/2024

<sup>1</sup>Public Health Scotland, Edinburgh, EH12 9EB  
<sup>2</sup>Usher Institute, University of Edinburgh, Edinburgh, EH16 4UX  
<sup>†</sup> Joint first author

### Abstract

#### Introduction

Here we present the Scottish Linked Pregnancy and Baby Dataset (SLiPBD), a new national data resource held by Public Health Scotland (PHS).

#### Methods

SLiPBD comprises a population-based e-cohort of all fetuses and births (babies) from pregnancies to women in Scotland from 2000 onwards. It is updated monthly by linking and reconciling the following national datasets: antenatal booking records; general and maternity hospital discharge records; termination of pregnancy notifications; and statutory live and stillbirth registrations.

#### Results

Key information included on all babies in SLiPBD includes estimated date of conception, end of pregnancy date, gestation, multiple pregnancy status, pregnancy outcome, and maternal sociodemographic characteristics. For live births, additional information on the birth, the baby's sociodemographic characteristics, and subsequent infant deaths is included.

Following the cohort refresh in January 2024, SLiPBD contained 1,770,226 babies from 1,750,830 pregnancies to 898,161 women. Of the 1,770,226 babies, 1,284,461 (73%) were live births, 5,731 (0.3%) stillbirths, and 316,897 (18%) and 114,840 (6%) came from a pregnancy ending a termination or early spontaneous loss respectively. 22,414 (1%) had an unknown pregnancy outcome, and for 25,883 (1%) the pregnancy was still ongoing. Data completeness for key sociodemographic characteristics except for ethnicity was very high, and variables showed expected patterns. Ethnicity data completeness is poor on historical records but improving over time. Completeness of unique patient identifiers was very high. External validation to source datasets was reassuring.

#### Conclusion

SLiPBD can be analysed standalone or linked to other national vital event and health datasets held by PHS. It supports longitudinal and intergenerational analyses, enabling epidemiological and health service surveillance and research on maternal and child health. Researchers interested in accessing pseudonymised extracts of SLiPBD through the Scottish NHS safe haven facility should contact Research Data Scotland. PHS will continue to refine SLiPBD as source datasets improve.

#### Keywords

Pregnancy; births; e-cohort; birth cohort; administrative healthcare data; administrative data linkage

#### Key features

1. The Scottish Linked Pregnancy and Baby Dataset (SLiPBD) is a new national data resource created and maintained by Public Health Scotland to facilitate epidemiological and health service analyses focused on maternal and child health.
2. SLiPBD comprises a population-based e-cohort of all fetuses and births (babies) from pregnancies to women in Scotland from 2000 onwards. At least 68,000 babies (of which at least 46,000 are live births) are included annually.
3. SLiPBD is updated on a monthly basis by linking and reconciling records relating to ongoing and completed pregnancies from the following existing national datasets: antenatal booking records; general and maternity hospital discharge records; termination of pregnancy notifications; and statutory live and stillbirth registrations.
4. Key information included on all babies in SLiPBD includes estimated date of conception, end of pregnancy date, gestation, multiple pregnancy status, pregnancy outcome, and maternal sociodemographic characteristics. For live births, additional information on the birth, the baby's sociodemographic characteristics, and any subsequent infant deaths is included.
5. Inclusion of unique personal identifiers for the mother and (where applicable) baby used within the health service and on statutory birth registration records ensures SLiPBD provides a core intergenerational spine record, allowing linkage between mothers and babies, and to other national datasets.
6. Subject to governance approvals, researchers can access pseudonymised extracts of SLiPBD (linked to other national datasets as required) through the Scottish NHS safe haven facility, which is supported by Public Health Scotland. Interested researchers should submit an initial enquiry form to Research Data Scotland (<https://www.researchdata.scot/accessing-data/>).

\*Corresponding Author:

Email Address: [rachael.wood@phs.scot](mailto:rachael.wood@phs.scot) (Rachael Wood)

## Background

Public Health Scotland (PHS) is responsible for developing and holding Scotland's national health datasets and using them to inform population health improvement [1]. In general, Scotland's national health datasets are excellent, reflecting key healthcare events such as hospital admissions with high completeness and accuracy, however records relating to pregnancies and births are contained in multiple disparate datasets, reflecting different stages of pregnancy (ongoing or completed) and different pregnancy outcomes. In addition, some datasets relate primarily to the mother, whilst others relate primarily to the baby. To date, this has meant that any analysis requiring a complete cohort of pregnancies or births and/or intergenerational linkage of data relating to mothers and live births, has first had to undertake a bespoke linkage of the relevant national datasets. This approach has been inefficient and has risked inconsistency and error.

During the COVID-19 pandemic, PHS was responsible for surveillance and research on infection and vaccination in pregnant women. This was undertaken through the COVID-19 in Pregnancy in Scotland (COPS) study [2]. As above, the first step in COPS was to develop a robust and timely e-cohort of all pregnancies and births in Scotland which could then be linked to infection, vaccination, and outcome data [3]. Whilst the COPS study has now ended, it demonstrated the benefits that would be associated with PHS maintaining a population-based pregnancy and births e-cohort on an ongoing basis.

PHS has therefore developed the Scottish Linked Pregnancy and Baby Dataset (SLiPBD), a population-based e-cohort of all fetuses and births (babies) from pregnancies to women in Scotland from 2000 onwards. Its development has been informed by the COPS study and by reviewing the literature to identify examples of similar pregnancy and/or birth e-cohorts maintained in other countries and their associated methodologies.

Examples of existing UK cohorts derived from routine administrative data that contain all pregnancies regardless of outcome include the UK Clinical Practice Research Datalink (CPRD) [4] and the Royal College of General Practitioners Research and Surveillance Centre (RSC) [5] pregnancy registers. These are based on algorithms identifying coded pregnancy-related events recorded in primary care records from selected general practices across the UK (CPRD) or in England and Wales (RSC). Examples from North America include pregnancy registers derived from insurance claim [6], secondary care [7], and integrated health delivery system [8] records. Like the CPRD and RSC, these all cover selected subsets of the population. Other European examples include the recently established Polish pregnancy register, however this has raised ethical and privacy concerns due to the very restricted access to termination of pregnancy in that nation [9].

Examples of complete population-based cohorts containing live births only include the Nordic nations' birth registers which are based on mandatory central data collection from patient records [10, 11]. A birth cohort for England and Wales based on linkage of statutory live birth registrations and health service records is also available [12, 13].

SLiPBD is based on linkage and reconciliation of records relating to ongoing and completed pregnancies from the following existing national datasets that are held by PHS:

antenatal booking records; general and maternity hospital discharge records; termination of pregnancy notifications; and statutory live and stillbirth registrations. It is updated monthly to add information on recent conceptions and end of pregnancy events. Key information included on all babies in SLiPBD includes estimated date of conception, end of pregnancy date, gestation, multiple pregnancy status, pregnancy outcome, and maternal sociodemographic characteristics. For live births, additional information on the birth, the baby's sociodemographic characteristics, and any subsequent infant deaths is included. Inclusion of unique personal identifiers for the mother and (where applicable) baby used within the health service (the Community Health Index [CHI] number) and on statutory birth registration records (National Records of Scotland [NRS] triplicate identifiers) ensures SLiPBD provides a core intergenerational spine record, allowing linkage between mothers and babies, and to other national datasets.

SLiPBD represents an important new national data resource for Scotland that will facilitate accurate and efficient epidemiological and health service analyses focused on maternal and child health, without the need for initial bespoke linkage to generate analysis cohorts. Within PHS it will underpin ongoing surveillance of maternal and child health. It will also be available for research, both within PHS and to external academic teams through the NHS Scotland Safe Haven facility. In this paper, we describe the key features of SLiPBD, the methodology used to generate it, and its importance to improving maternal and child health at population level.

## Methods

### Population

The SLiPBD cohort is created at baby level, with one 'row' per fetus (from pregnancies ending in a spontaneous early loss or termination) or birth (pregnancies ending in a live or stillbirth). A unique SLiPBD pregnancy identifier number is included, allowing babies from a multiple pregnancy to be identified and ensuring that the cohort can be collapsed to pregnancy level as required for analysis. Inclusion of the maternal CHI number also ensures that sequential pregnancies to the same woman can be identified.

SLiPBD includes babies from pregnancies that were ongoing on 1 January 2000 and all subsequent pregnancies conceived from 1 January 2000 onwards. Babies from all pregnancies that have a record in a relevant national dataset (see below) are included, regardless of the pregnancy duration or outcome. SLiPBD was launched in September 2023 and is updated on a monthly basis.

### Data sources used

Table 1 summarises the existing national datasets held by PHS that are used to identify babies from ongoing and completed pregnancies for inclusion in SLiPBD. Further information on how the source records that feed into these datasets are generated and returned to PHS, and the variables and codes used to identify specific end of pregnancy events within the records are shown in Supplementary Appendix 1 and

Table 1: Existing Scottish national datasets used to identify babies from ongoing and completed pregnancies for inclusion in SLiPBD

Dataset	Description	Used to identify
NRS statutory live birth registrations	Statutory record of all babies born at any gestation showing signs of life	Live births
NRS statutory stillbirth registrations	Statutory record of all babies born at 24 weeks gestation or over showing no signs of life	Spontaneous stillbirths Stillbirths resulting from termination of pregnancy
Termination of pregnancy notifications (Abortion Act Scotland [AAS] statutory records to April 2022, Termination of Pregnancy Submissions Scotland [ToPSS] non-statutory records from May 2022)	Health service notifications of all terminations of pregnancy carried out under the Abortion Act 1967	Women having a termination of pregnancy (and hence babies from these pregnancies)
Scottish Morbidity Record (SMR) 01 general hospital discharge records	Health service record of all patients discharged following NHS day case or inpatient general care. Covers all specialities except neonatal, maternity, and mental health care	ICD10 diagnostic codes used to identify women having a spontaneous early pregnancy loss (and hence babies from these pregnancies)
Scottish Morbidity Record (SMR) 02 maternity hospital discharge records	Health service record of all patients discharged following NHS day case or inpatient maternity care	Hard coded variables used to identify women having a spontaneous early pregnancy loss, termination of pregnancy, or birth (and hence babies from these pregnancies) and to provide additional clinical information on live and stillbirths
Antenatal Booking Collection (ABC) records	Health service record of all women booking for NHS antenatal care from April 2019 onwards	Women booking for antenatal care (and hence babies from ongoing pregnancies)

Spontaneous early pregnancy loss includes miscarriage (spontaneous loss at up to 23 weeks and 6 days [23+6] gestation), ectopic pregnancy, and molar pregnancy. See <https://publichealthscotland.scot/services/national-data-catalogue/> for information on the various source national datasets held by Public Health Scotland.

Supplementary Appendix 2. Table 2 summarises the national datasets that are used to provide supplementary information on the babies included in SLiPBD.

## Creating SLiPBD

The SLiPBD cohort is created as follows. First, all live and stillbirths are identified using NRS statutory birth registration records, and these are linked to their corresponding NHS (SMR02) delivery record using a combination of maternal CHI, baby CHI (live births only), and date of birth/delivery to create an initial SLiPBD birth cohort. Births from multiple pregnancies are appropriately flagged. The linked SMR02 delivery records enrich the birth registration records by providing additional clinical information including gestation at delivery in completed weeks, birthweight, and mode of delivery.

As the statutory birth registration records are not health service records, these do not contain CHI numbers when submitted by NRS to PHS. PHS uses an established process (separate to SLiPBD) to 'seed' maternal and (for live births) baby CHI number onto the NRS birth registration records based on the other personal identifiers they contain. For a small number of birth registration records, no maternal and/or baby CHI can be seeded. In these cases, the birth registration

records are retained within SLiPBD to ensure a complete birth cohort, but 'dummy' CHIs are allocated, and no corresponding SMR02 delivery record will be linked. The unlinked 'orphan' SMR02 delivery records are dropped from SLiPBD to avoid creating duplicates.

Next, fetuses from pregnancies ending in a spontaneous early pregnancy loss (miscarriage at up to 23 weeks and 6 days [23+6] gestation, ectopic pregnancy, or molar pregnancy) or termination of pregnancy are identified from general (SMR01) and maternity (SMR02) discharge records and termination of pregnancy (AAS and ToPSS) notifications, and those from ongoing pregnancies are identified from antenatal booking (ABC) records. Fetuses from 'selective reduction' terminations of pregnancy (where some fetuses within a high order multiple pregnancy are terminated to allow the pregnancy to safely continue) are flagged as from a multiple pregnancy. No other information on number of fetuses is available on these source records, hence all other fetuses are assumed to be singletons. The SLiPBD birth cohort is linked to these other records using maternal CHI, end of pregnancy date, and estimated conception date to create an interim SLiPBD fetuses and births ('baby') cohort. Records that appear to relate to the same baby are grouped using date-based rules. Broadly, records are grouped if they have the same maternal CHI number and

Table 2: Existing Scottish national datasets used to provide supplementary information on babies included in SLiPBD

Dataset	Description	Used to identify
NRS statutory death registrations	Statutory record of all deaths occurring in Scotland. Deaths must be registered within 7 days of occurrence	Maternal deaths occurring during pregnancy or within 6 weeks of the end of pregnancy
NRS statutory infant death registrations	Statutory death records for children dying at <2 years of age. NRS appends the corresponding statutory live birth registration unique identifiers onto this subset of death records to support record linkage	Infant deaths at <1 year of age
Child Health Systems Programme: Pre-school (CHSP-PS)	National child health information system used to support delivery of universally offered child health reviews and immunisations for pre-school children	Baby ethnicity for live births
Community Health Index (CHI) database	NHS Scotland master patient index system. The CHI number is the unique patient identifier used on all health records in Scotland	Maternal emigration during pregnancy; maternal date of birth; maternal postcode at booking if missing from other sources; reconciliation to latest CHI numbers as required

See <https://publichealthscotland.scot/services/national-data-catalogue/> for information on the various source national datasets held by Public Health Scotland.

either the end of pregnancy dates, or the estimated dates of conception, are within 83 days of each other. Some source records include gestation in completed weeks at the date of event (antenatal booking or end of pregnancy), but others do not. Notably, gestation is not available on NRS statutory live birth registrations or SMR01 records. Where gestation information is missing, clinically informed rules are used to impute a likely value (Table 3). This allows imputation of an estimated date of conception, supporting identification of records relating to the same baby within the interim SLiPBD baby cohort.

Within the interim SLiPBD baby cohort, the majority of sets of records assigned to an individual baby indicate a clinically feasible series of events leading to a clear pregnancy outcome for that baby, e.g., an antenatal booking record

followed by a live birth record for the same woman, with gestation information on both records indicating a similar estimated conception date. In some cases, however, there are conflicting records within a set. For example, records indicating a miscarriage and, separately, ectopic pregnancy at a similar time for the same woman; or a miscarriage record followed by a live birth record for the same woman, with gestation information on both records indicating a similar estimated conception date. These conflicts can arise for various reasons including inaccurate clinical coding on source records (e.g., coding an early pregnancy bleed as a completed miscarriage), record linkage error, or very rapid sequential pregnancies.

The next step therefore resolves any conflicting pregnancy outcomes for babies within the interim SLiPBD baby cohort to create a resolved SLiPBD baby cohort. Each baby is

Table 3: Imputed values for gestation at specified event where this is missing from source records

Event (end of pregnancy or antenatal booking)	Imputed gestation
Live Birth (singleton)	40 weeks and 0 days [40+0]
Live Birth (multiple)	37+0
Stillbirth	32+0
Termination (AAS/ToPSS)	8+0
Termination (SMR02)	16+0
Ectopic	8+0
Molar pregnancy	10+0
Miscarriage	10+0
Unknown	40+0
Unknown – assumed early loss	10+0
Antenatal booking	10+0

assigned a final pregnancy outcome using hierarchical rules, and unfeasible records are dropped. Broadly, the hierarchy of outcomes applied is live birth > stillbirth > termination > ectopic > molar > miscarriage > ongoing/unknown. Usually, just a primary pregnancy outcome is assigned for an individual baby, but in some specific situations, a primary and secondary pregnancy outcome is retained. Specifically, a baby can be assigned termination of pregnancy and any other known pregnancy outcome (live birth, stillbirth, ectopic, molar, or miscarriage) if source records indicate both outcome types. Termination of pregnancy at later gestations will result in a still (or, less commonly, a live) birth, hence both outcomes are valid. It is less clear why (separate) source records may classify the same event as a termination and an early spontaneous loss. It may be that a pre-termination scan identifies a spontaneous loss. Currently in SLiPBD, for babies with termination plus another outcome, the termination is always recorded as the primary outcome.

In the first conflict scenario outlined above, the baby would be assigned an ectopic pregnancy outcome, and the concurrent miscarriage record would be dropped. Similarly, in the second scenario the baby would be assigned a live birth outcome, and the prior miscarriage record would be dropped.

Some babies in the interim baby cohort have only an antenatal booking record, with no linked end of pregnancy record available. The last step deals with these to generate the final SLiPBD baby cohort. Firstly, a check is run against the CHI database to identify any maternal emigrations during pregnancy and assign the relevant censored outcome for those babies. A check is also run against NRS statutory death records to identify maternal deaths during pregnancy. If the baby has another known outcome (live birth, stillbirth, termination, ectopic, molar, or miscarriage), maternal death is recorded as the secondary outcome. If no other outcome information is available for the baby, maternal death is recorded as the primary outcome.

For the remaining babies, they are assigned a censored 'unknown' pregnancy outcome at 40+0 weeks gestation, or an interim 'ongoing' outcome if the pregnancy has not reached 40+0 at the time of the SLiPBD update. The exception is when this would lead to this pregnancy overlapping with a subsequent pregnancy to the same women: in this case the baby is assigned a censored 'assumed early loss' outcome at the date of booking or at 10+0 weeks gestation, whichever is later. The sequential steps involved in creating SLiPBD are summarised in Figure 1 below.

The final SLiPBD baby cohort is then enriched by linking in information from other national datasets held by PHS. For example, the CHI database is used to append maternal date of birth. For live births, the CHSP-PS system is used to append baby ethnicity, and NRS infant death records are used to append information on deaths occurring at <1 year of age.

By design, a relatively restricted number of variables is retained for each baby in the final SLiPBD cohort. Retained variables are those that support identification of specific (sub) cohorts of types of pregnancies or babies, longitudinal data linkage for individual mothers or live births, and intergenerational data linkage (i.e., between mothers and live births), along with core sociodemographic characteristics. Additional detail is contained within the different source datasets held by PHS (e.g., clinical details of delivery care are

contained in the SMR02 maternity discharge record dataset) but this is not duplicated in SLiPBD. Given the availability of unique personal identifiers in both SLiPBD and source dataset records, this additional detail can be linked to SLiPBD extracts as required for specific analyses.

## Updating SLiPBD

SLiPBD was launched in September 2023 and is updated by PHS on a monthly basis, ensuring that the cohort remains as up to date as possible. The source datasets used to generate SLiPBD are returned to PHS according to different schedules (see Supplementary Appendix 1), hence the composite lag inherent in SLiPBD is nuanced. Overall, the rule of thumb is that, following any month's update, SLiPBD will be reasonably complete for conceptions and end of pregnancy events occurring up to 3 months previously.

In any month's SLiPBD update, all data for the previous 3 years is refreshed. SLiPBD is therefore a dynamic dataset, and the information included for any individual baby will be updated within that timeframe as records accrue. This updating will include changing an interim 'ongoing' outcome to a definitive pregnancy outcome and refining that as required, as well as incorporating any infant death records.

Following each month's update, a quality assurance check is run to ensure the number of added babies is feasible for the Scottish population, to assess levels of missing information for specific variables, and to identify babies with unfeasible gestations for their assigned pregnancy outcome. The number of source records that have not been linked into SLiPBD, e.g., SMR02 delivery records and NRS infant death records, are also checked to flag any problems with record linkage.

All the R code used to generate SLiPBD (and the analytical code used to generate results presented here) is available on PHS's public GitHub site at [https://github.com/Public-Health-Scotland/SLiPBD\\_public](https://github.com/Public-Health-Scotland/SLiPBD_public).

## Results

Here we report results from SLiPBD based on analysis of the cohort as updated in early January 2024. Following this update, the cohort contained 1,770,226 babies (fetuses and births) from 1,750,830 completed or ongoing pregnancies to 898,161 women.

Estimated dates of conception ranged from 29 March 1999 to 20 December 2023 (babies conceived prior to 1 January 2000 are included if the pregnancy was still ongoing on 1 January 2000), and end of pregnancy dates for the subset of babies from completed pregnancies ranged from 1 January 2000 to 9 January 2024.

## Pregnancy outcomes

Table 4 shows the distribution of pregnancy outcomes for the 1,770,226 babies in the cohort and Figure 2 shows the number of babies conceived per year, by pregnancy outcome group.

SLiPBD includes at least 68,000 fetuses and births from pregnancies conceived in each of the years with complete data (2000–2022). The total number of babies conceived increased from 2000 to 2008 then subsequently declined, reflecting



Figure 1: Schematic diagram showing the sequential steps in creating SLiPBD

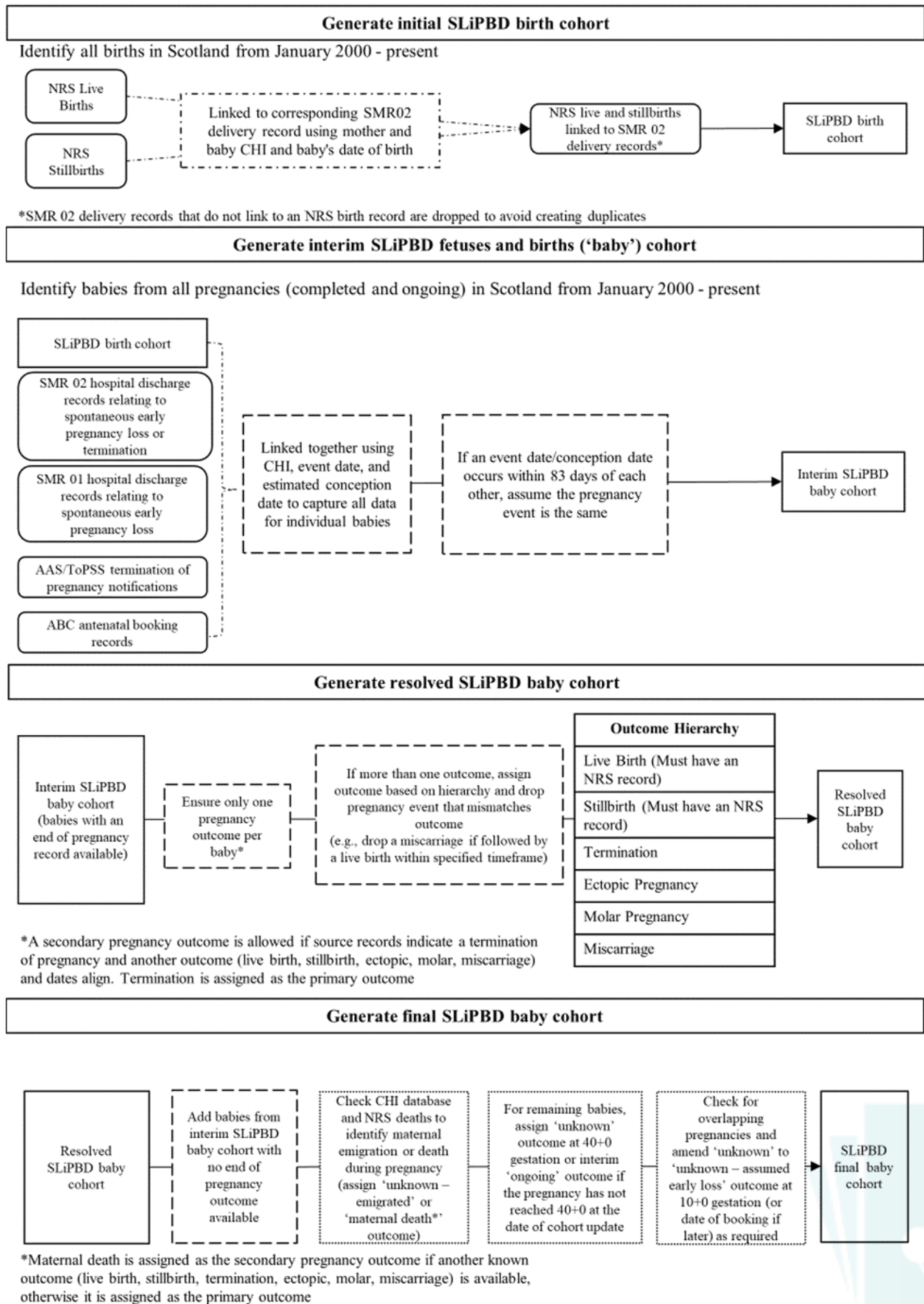
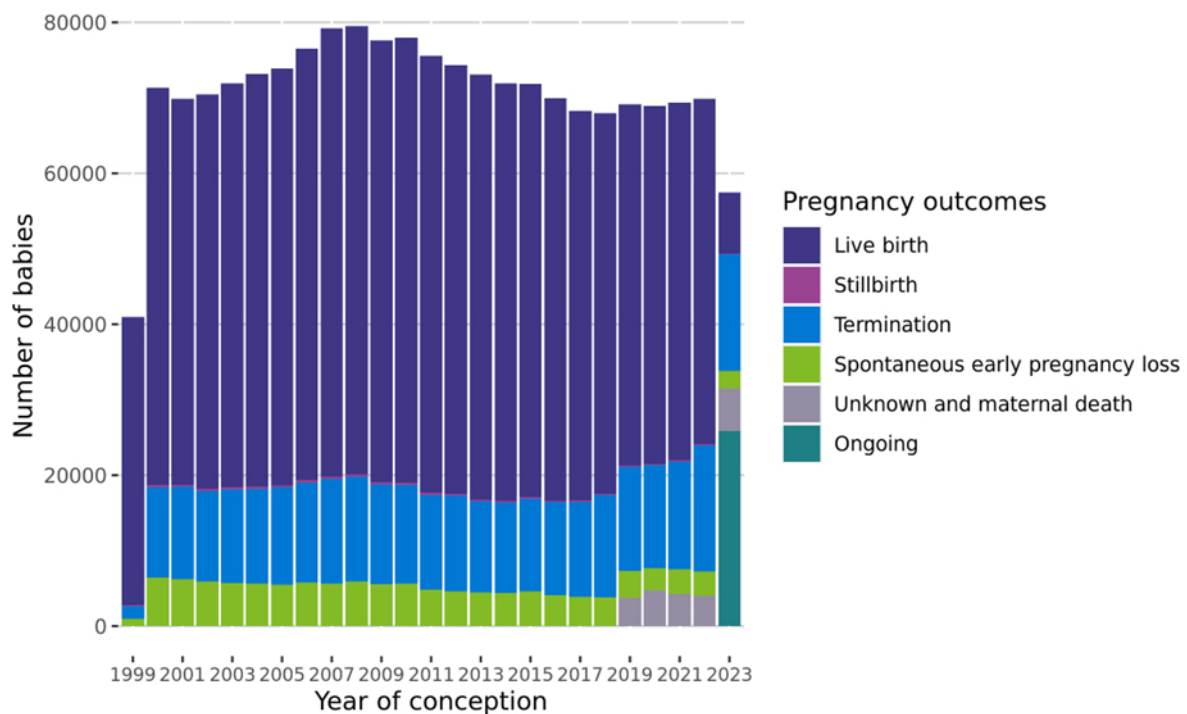


Table 4: Pregnancy outcome for babies in the SLiPBD cohort

Pregnancy outcome	Number (%) of babies
Live birth	1,284,461 (72.6%)
Stillbirth	5,731 (0.3%)
Termination	316,897 (17.9%)
Ectopic pregnancy	15,449 (0.9%)
Molar pregnancy	1,393 (0.1%)
Miscarriage	97,998 (5.5%)
Maternal death	13 (<0.1%)
Unknown	20,611 (1.2%)
Unknown – emigrated	774 (<0.1%)
Unknown - assumed early loss	1,016 (0.1%)
Ongoing pregnancy	25,883 (1.5%)
<i>Total</i>	<i>1,770,226</i>

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. SLiPBD allows a baby to be assigned a primary and secondary pregnancy outcome in certain circumstances: here only the primary outcome has been counted for each baby.

Figure 2: Babies in the SLiPBD cohort by pregnancy outcome group and year of conception



Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. Spontaneous early pregnancy loss includes miscarriage (spontaneous loss at up to 23 weeks and 6 days [23+6] gestation), ectopic pregnancy, and molar pregnancy.

known patterns of population size and fertility in the Scottish population [14, 15]. Overall, 72.6% of all babies in the cohort were live births, 0.3% were stillbirths, and 17.9% and 6.5% respectively were from pregnancies ending in a termination or early spontaneous loss. The number and proportion of babies with termination of pregnancy as their pregnancy outcome has increased over time [16] whilst the number and proportion with early spontaneous loss has decreased. The decrease in early spontaneous losses may reflect changing clinical practice, with a reduction in the proportion of women with these outcomes receiving in-patient care (and hence declining ascertainment of

these events) and/or termination acting as a competing risk for early spontaneous loss.

A high proportion of babies conceived in 1999 that are included in SLiPBD were live births. As pregnancies had to be ongoing on 1 January 2000 to be included in the cohort, this will have selected for conceptions in 1999 with longer pregnancy duration.

Babies from pregnancies with unknown outcome are seen from 2019 onwards, as ABC antenatal booking records are available from April 2019. Prior to that point, pregnancies were ascertained from records indicating end of pregnancy

events only (and information on antenatal booking was only available from subsequent SMR02 delivery records for the subset of women who booked then went on to deliver a live or stillbirth).

Among the babies conceived in 2023, approximately half had 'ongoing' pregnancy status at the time the cohort was refreshed for this analysis in January 2024, indicating that an antenatal booking record was available, but no end of pregnancy record was yet available, and the pregnancy had not yet reached 40+0 gestation. Among babies conceived in 2023 with pregnancy outcome available, a high proportion had termination or early spontaneous loss assigned. This reflects the fact that outcome information will have been differentially available for pregnancies with shorter duration at the time of analysis.

Figure 3 shows the gestation at end of pregnancy for the subset of 1,721,929 babies in the cohort with a known pregnancy outcome. Table 5 shows the source of information on gestation at end of pregnancy for this group. Reflecting the source records used to ascertain different pregnancy outcomes, and the availability (or lack) of information on gestation on different source record types, gestation at end of pregnancy is known for the substantial majority of babies where the pregnancy ends in a live or stillbirth or termination. By contrast it is only known for around half of babies where the pregnancy ends in a miscarriage, and for less than 15% of babies from an ectopic pregnancy (and consequently it is imputed for the remaining babies following the rules shown in Table 3). This accounts for the sharp 'spikes' of babies from ectopic pregnancies and miscarriages ending at 8 and 10 completed weeks respectively.

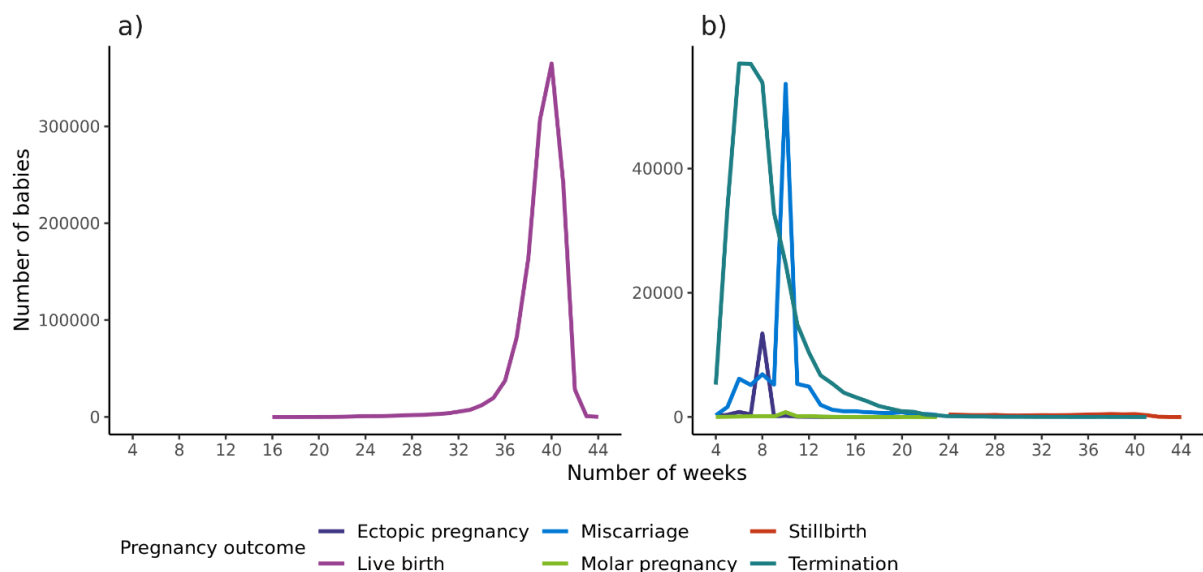
## Sociodemographic characteristics

Table 6 shows maternal sociodemographic characteristics for the 1,770,226 babies and 1,750,830 pregnancies in the SLiPBD

cohort. Information on maternal age and deprivation level is highly complete and follows expected patterns [17]. There is a high level of missing information on maternal ethnicity. The missingness of maternal ethnicity varies over time and by pregnancy outcome type, reflecting data availability and quality on source records. Of the source record types shown in Table 1, maternal ethnicity is available on SMR01, SMR02, and antenatal booking records and its completeness on these sources (in particular on SMR records) has improved over time. Maternal ethnicity is also available on ToPSS (but not AAS) termination notifications, but this variable is not yet feeding through to SLiPBD. Among babies in SLiPBD conceived from 1 January 2022 onwards, 95% of live births had maternal ethnicity available, as did 83% of stillbirths, 88% of babies from a spontaneous early loss, and 94% of babies from an ongoing pregnancy. Only 5% of these babies from a termination of pregnancy had maternal ethnicity available, but this is expected to improve as ethnicity data on ToPSS records feeds through to SLiPBD.

Table 7 shows infant sociodemographic characteristics for the 1,284,461 live births in the SLiPBD cohort. Information on infant sex is highly complete. Information on gestation is fully complete as would be imputed if missing (see Table 5). The completeness of information on birthweight reflects the high proportion of live birth records within SLiPBD with a linked SMR02 delivery record available (see section on Quality assurance of personal identifiers and record linkage following Table 7) and the high completeness of birthweight on SMR02 records. The distribution of gestation and birthweight shows expected patterns for babies from singleton and multiple pregnancies [17]. There is a high level of missing information on baby ethnicity for live births. Baby ethnicity is sourced from the results of universal child health reviews provided by Health Visitors and recorded on the CHSP-PS system [18]. Ethnicity has been recorded at reviews provided at 27-30 months from April 2013, at 10 days and 6-8 weeks from

Figure 3: Gestation at end of pregnancy for babies in the SLiPBD cohort with known pregnancy outcome: a) Live births, b) other outcomes



Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000.



Table 5: Source of information on gestation at end of pregnancy for babies in the SLiPBD cohort with known pregnancy outcome

Pregnancy outcome	Total number of babies	Number (%) with information on gestation at end of pregnancy		
		Available on end of pregnancy record	Calculated from gestation on antenatal booking record	Imputed based on pregnancy outcome type
Live birth	1,284,461	1,243,420 (96.8%)	5,968 (0.5%)	35,073 (2.7%)
Stillbirth	5731	5,729 (>99.9%)	0 (0.0%)	2 (<0.1%)
Termination	316,897	316,691 (99.9%)	2 (<0.1%)	204 (0.1%)
Ectopic pregnancy	15,449	2,016 (13.1%)	157 (1.0%)	13,276 (85.9%)
Molar pregnancy	1,393	653 (46.9%)	64 (4.6%)	676 (48.5%)
Miscarriage	97,998	46,443 (47.4%)	3,686 (3.8%)	47,869 (48.9%)
Total	1,721,929	1,614,952 (93.8%)	9,877 (0.6%)	97,100 (5.6%)

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. Information on gestation at end of pregnancy is available on NRS stillbirth, AAS/ToPSS termination, and SMR02 maternity discharge records. It is not available on NRS live birth records but is infilled from linked SMR02 maternity discharge records for live births. It is not available on SMR01 general discharge records. SMR01 is the sole source of information on a high proportion of pregnancies ending in miscarriage and on molar and ectopic pregnancies. This accounts for the high proportion of babies with these pregnancy outcomes that have imputed gestation at end of pregnancy.

Table 6: Key maternal sociodemographic characteristics for babies and pregnancies in the SLiPBD cohort

Characteristic	Number (%) of babies	Number (%) of pregnancies
<20	170,104 (9.6%)	169,503 (9.7%)
20-24	342,781 (19.4%)	340,591 (19.5%)
25-29	466,821 (26.4%)	462,013 (26.4%)
30-34	477,214 (27.0%)	470,397 (26.9%)
35-39	251,535 (14.2%)	247,399 (14.1%)
≥40	56,397 (3.2%)	55,556 (3.2%)
Unknown/missing	5,374 (0.3%)	5,371 (0.3%)
<i>Median maternal age (min-max)</i>	<i>29 years (11-55)</i>	<i>29 years (11-55)</i>
<b>Maternal deprivation level (SIMD quintile)</b>		
1 – most deprived	450,798 (25.5%)	446,586 (25.5%)
2	372,544 (21.0%)	368,751 (21.1%)
3	326,930 (18.5%)	323,347 (18.5%)
4	323,252 (18.3%)	319,353 (18.2%)
5 – least deprived	291,945 (16.5%)	288,071 (16.5%)
Unknown	4,757 (0.2%)	4,722 (0.2%)
<b>Maternal ethnicity</b>		
White	625,345 (35.3%)	617,129 (35.3%)
South Asian	20,785 (1.2%)	20,564 (1.2%)
Black/Caribbean/African	11,160 (0.6%)	10,974 (0.6%)
Other or mixed ethnicity	21,207 (1.2%)	20,997 (1.2%)
Unknown/missing	1,091,729 (61.7%)	1,081,166 (61.8%)
<i>Total</i>	<i>1,770,226</i>	<i>1,750,830</i>

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. Maternal age at conception of <11 or >55 years is recoded as unknown within SLiPBD. Postcode at end of pregnancy or, if missing, at antenatal booking has been used to assign the Scottish Index of Multiple Deprivation (SIMD) quintile [19].

February 2016, and at 13-15 months and 4-5 years from April 2017. Within SLiPBD, baby ethnicity is therefore first available for babies born in 2011, and since then has become more complete over time. 95% of live births in SLiPBD conceived from 1 January 2022 onwards have baby ethnicity available.

### Quality assurance of personal identifiers and record linkage

Table 8 shows the completeness of valid maternal CHI number for all babies in the SLiPBD cohort. Babies without a valid maternal CHI number available are assigned a 'dummy'

Table 7: Key infant sociodemographic characteristics for live births in the SLiPBD cohort

Characteristic	Number (%) of live births		
<b>Baby sex</b>			
Male	659,035 (51.3%)		
Female	625,409 (48.7%)		
Unknown	17 (<0.1%)		
<b>Baby Ethnicity</b>			
White	554,377 (43.2%)		
South Asian	21,138 (1.7%)		
Black/Caribbean/African	10,240 (0.8%)		
Other or mixed ethnicity	27,269 (2.1%)		
Unknown/missing	671,437 (52.3%)		
<b>Gestation at birth (completed weeks)</b>			
	<b>N (%) of LBs from singleton pregnancy</b>	<b>N (%) of LBs from multiple pregnancy</b>	<b>N (%) of all LBs</b>
Very preterm (<32)	11,192 (0.9%)	3,751 (9.8%)	14,943 (1.2%)
Moderately preterm (32-36)	63,464 (5.1%)	18,198 (47.5%)	81,662 (6.4%)
Term (37-41)	1,142,660 (91.7%)	16,382 (42.7%)	1,159,042 (90.2%)
Post-term (≥42)	28,807 (2.3%)	7 (<0.1)	28,814 (2.2%)
Unknown/missing	0 (0.0%)	0 (0.0%)	0 (0.0%)
<i>Median gestation (min-max)</i>	<i>40 weeks (16-44)</i>	<i>36 weeks (17-43)</i>	<i>39 weeks (16-44)</i>
<b>Birthweight (grams)</b>			
Very low (<1500)	9,524 (0.8%)	3,473 (9.1%)	12,997 (1.0%)
Low (1500-2499)	55,958 (4.5%)	17,168 (44.8%)	73,126 (5.7%)
Typical (2500-3999)	979,070 (78.6%)	16,142 (42.1%)	995,212 (77.5%)
High (≥4000)	161,665 (13.0%)	27 (0.1%)	161,692 (12.6%)
Unknown/missing	39,906 (3.2%)	1,528 (4.0%)	41,434 (3.2%)
<i>Median birthweight (min-max)</i>	<i>3440 grams (100-7900)</i>	<i>2420 grams (100-5500)</i>	<i>3420 grams (100-7900)</i>
<i>Total</i>	<i>1,246,123</i>	<i>38,338</i>	<i>1,284,461</i>

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. Gestation at live birth of <16 or >44 weeks on source records is recorded as missing (and would then be imputed if not available from any other source record). For this analysis, birthweight <100 or >7999 grams has been recorded as missing.

maternal CHI, and these records will inevitably not link to other national datasets. Availability of maternal CHI is generally very high (available for 98.2% [1,738,582 of 1,770,226] babies overall). Babies from a termination of pregnancy and stillbirths have the lowest completeness of maternal CHI at 92% and 95% respectively. This reflects known issues with seeding of maternal CHI on historic AAS records for terminations occurring before 2020 and on NRS stillbirth registration records.

Of the 1,284,461 live births in the SLiPBD cohort, 1,281,435 (99.8%) had a valid baby CHI number available.

The number of babies included in SLiPBD with specified known pregnancy outcomes was compared to the number of corresponding end of pregnancy events recorded in relevant source datasets to explore the number and patterns of source records being dropped from the SLiPBD cohort and as an external validation check.

Table 9 shows the number of babies in SLiPBD with the various pregnancy outcomes recorded as their primary outcome, and the total number that had the specified outcome recorded as their primary or secondary outcome.

As termination is currently always retained as the primary pregnancy outcome in SLiPBD, the number of babies with this as their primary outcome, and the total number with this outcome, is the same. 2,224 (0.1%) of the 1,721,929 of babies with a known pregnancy outcome in SLiPBD had termination and another outcome type recorded, hence 314,673 babies had termination as their only outcome.

The total number of live births in SLiPBD is very close to the number of statutory live birth registrations over the same period (with a very small number of records being dropped as possibly relating to the same baby), and the total number of stillbirths exactly matches the number of statutory stillbirth registrations. The total number of babies in SLiPBD assigned an early spontaneous loss pregnancy outcome is slightly lower than the total number of relevant events (e.g., miscarriages) recorded on source records, indicating that only a small number of source records are being dropped to resolve conflicts within the SLiPBD cohort (see Figure 1). The total number of babies in SLiPBD assigned a termination pregnancy outcome is slightly higher than the number of notified terminations. This is because SLiPBD can ascertain terminations from other

Table 8: Completeness of valid maternal CHI number for babies in the SLiPBD cohort by pregnancy outcome

Pregnancy outcome	Total number of babies	N (%) with valid maternal CHI
Live birth	1,284,461	1,279,359 (99.6%)
Stillbirth	5,731	5,436 (94.9%)
Termination	316,897	292,192 (92.2%)
Ectopic pregnancy	15,449	15,038 (97.3%)
Molar pregnancy	1,393	1,385 (99.4%)
Miscarriage	97,998	97,312 (99.3%)
Maternal death	13	13 (100.0%)
Unknown	20,611	20,278 (98.4%)
Unknown – emigrated	774	774 (100.0%)
Unknown - assumed early loss	1,016	1,016 (100.0%)
Ongoing pregnancy	25,883	25,779 (99.6%)
<i>Total</i>	<i>1,770,226</i>	<i>1,738,582 (98.2%)</i>

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000.

Table 9: Number of babies in SLiPBD with known pregnancy outcome compared to number of end of pregnancy events recorded in relevant source datasets

Pregnancy outcome	Number of babies with this pregnancy outcome in SLiPBD cohort Primary outcome (Primary or secondary outcome)	Number of specified end of pregnancy events recorded as occurring from 1 January 2000 to the date of cohort refresh in relevant source datasets	
		Source record	N of events
Live birth	1,284,461 (1,284,563)	NRS statutory live birth registrations	1,284,580
Stillbirth	5,731 (6,120)	NRS statutory still birth registrations	6,120
Termination	316,897 (316,897)	AAS/ToPSS	313,777
Ectopic pregnancy	15,449 (15,550)	SMR01/SMR02	15,960
Molar pregnancy	1,393 (1,417)	SMR01/SMR02	1,428
Miscarriage	97,998 (99,606)	SMR01/SMR02	101,597
<i>Total</i>	<i>1,721,929</i>	<i>n/a</i>	<i>n/a</i>

Data from SLiPBD cohort updated in early January 2024: includes babies from pregnancies in Scotland ongoing on, or conceived after, 1 January 2000. No total number of specified end of pregnancy events recorded across all source datasets is provided due to double counting within and across datasets. Double counting may be appropriate (e.g., a termination of pregnancy performed at 24 weeks gestation recorded on both on AAS/ToPSS and NRS stillbirths) or due to recording error (e.g., an early pregnancy bleed recorded on SMR01 as a completed miscarriage then an NRS livebirth record for the same baby).

sources, specifically SMR02 or diagnostic coding on NRS stillbirth records.

Of the 1,284,563 and 6,120 total live and stillbirths included in the SLiPBD cohort, 1,244,460 (96.9%) and 5,077 (83.0%) respectively had a linked SMR02 delivery record available. At the time SLiPBD was refreshed in early January 2024, a total of 1,264,671 SMR02 delivery records indicating a live or stillbirth in 2000 onwards were available, indicating that 15,134 (1.2%) of all available SMR02 delivery records had been dropped from SLiPBD as 'orphan' records that could not be linked to a corresponding NRS statutory birth registration record. Following any cohort refresh, it is likely that some of the unlinked SMR02 'orphans' are records relating to very recent births for which the NRS statutory birth registration has not yet been received. These will be incorporated in subsequent cohort refreshes.

247,715 babies from 245,001 pregnancies within the SLiPBD cohort had an ABC antenatal booking record

available. At the time SLiPBD was refreshed in early January 2024, a total of 245,068 (pregnancy-level) ABC records were available for women booking from April 2019 onwards (when this national dataset was established), indicating that 67 (<0.1%) of all available ABC records had been dropped from SLiPBD. The ABC dataset is partially de-duplicated on receipt, however sometimes more than one booking record for an individual pregnancy is retained. These duplicate records will be dropped from SLiPBD.

Finally, of the 5,285 NRS statutory infant death records for babies born from 2000 onwards and aged <1 year at death that were available at the date the cohort was refreshed, 63 (1.2%) were not linked to a corresponding live birth within SLiPBD. Of the unlinked death records, 56 (88.9%) did not have the corresponding NRS unique live birth identifiers available on the death record, suggesting that the baby may have been born outwith Scotland and therefore not included in the SLiPBD cohort.

## Discussion

The Scottish Linked Pregnancy and Baby Dataset (SLiPBD) is an important new national data resource created and maintained by Public Health Scotland (PHS). SLiPBD comprises a population-based e-cohort of all fetuses and births (babies) from pregnancies (regardless of duration or outcome) to women in Scotland from 2000 onwards. It is updated monthly by linking and reconciling relevant national vital event and health datasets held by PHS that provide information on ongoing and completed pregnancies. The resulting dataset is subject to robust quality assurance. Following the cohort refresh in January 2024, SLiPBD contained 1,770,226 babies from 1,750,830 pregnancies to 898,161 women.

Examples of existing pregnancy and birth cohorts available in other nations are provided in the Background. The methodologies used to derive the other algorithm-driven cohorts containing all pregnancies (which are most comparable to SLiPBD) vary, reflecting the different source data available in different jurisdictions. However, attention to the quality of source data and data linkage and reconciliation is universally important [20, 21]. In general, SLiPBD compares well to these other cohorts in terms of key metrics including whole nation coverage; time period covered; timeliness; ascertainment of all pregnancies; accuracy of pregnancy dates, singleton/multiple status, outcome, and gestation; availability of unique identifiers for mother and (where applicable) the baby to support intergenerational linkage; transparency of methodology; quality assurance; and availability for research. Another key advantage of SLiPBD is that it can be readily linked to the other national vital event and health datasets held by PHS to identify relevant exposures and long-term (intergenerational) outcomes.

SLiPBD has a number of limitations. Lack of access to data from primary care or early pregnancy clinics currently means that pregnancies ending in an early spontaneous loss that are managed in these settings will not be included. National data development projects are underway to address these gaps. More complete ascertainment of early pregnancy loss events will reduce the number of babies in SLiPBD assigned an 'unknown' pregnancy outcome. Early pregnancy losses that are unrecognised by the woman, or for which no healthcare is sought, will inevitably not be ascertained.

Data completeness for variables within the SLiPBD dataset is high, except for maternal and baby ethnicity. Completeness of ethnicity data is improving over time, and recent developments including addition of ethnicity to ToPSS termination of pregnancy notifications will improve this further, allowing SLiPBD to inform action on ethnic inequalities in maternal health [22]. All source datasets used for SLiPBD that contain information on gestation at relevant event (antenatal booking or end of pregnancy) currently record gestation in completed weeks rather than weeks and days, leading to some inaccuracy in calculation of estimated dates of conception. Antenatal booking records currently lack information on number of fetuses, meaning that singleton/multiple status is unknown for most babies from pregnancies that end in an early loss. Again, national data development projects are underway to address these gaps.

PHS is currently using SLiPBD for analyses relating to uptake of vaccines offered in pregnancy; prescription of potentially teratogenic medicines in pregnancy [23]; exposure to infections and environmental hazards in pregnancy; and to support monitoring of pregnancy and newborn screening programmes [24].

Future plans include ongoing monthly updating of the cohort and incorporation of the planned improvements to source datasets noted above as they become available. SLiPBD will increasingly be used within PHS to support production of official statistics on maternal and child health and bespoke analyses. We will also encourage its use by external research teams. Essentially, SLiPBD can be used to support a wide range of epidemiological or health service research on maternal and child health. It is particularly useful for analyses requiring all pregnancies (not just the subset ending in a live birth) and/or longitudinal and intergenerational data. Example research questions that could be addressed using SLiPBD linked to other PHS-held data include: What (trends in) inequalities in pregnancy-related outcomes are evident by different demographic factors and measures of socioeconomic status? How is a specific exposure in pregnancy (such as a novel infection or clinical intervention) associated with maternal, pregnancy, or baby outcomes? How do changes to the fundamental determinants of health (such as an economic downturn or a change to tax and benefit policy) influence women's reproductive choices?

## Data access

Subject to governance approvals, researchers can access pseudonymised extracts of SLiPBD (linked to other national datasets as required) through the Scottish NHS Safe Haven facility, which is supported by PHS. Interested researchers should submit an initial enquiry form to Research Data Scotland (<https://www.researchdata.scot/accessing-data/>).

## Conclusions

SLiPBD comprises a population-based e-cohort of all fetuses and births (babies) from pregnancies (regardless of duration or outcome) to women in Scotland from 2000 onwards. It can be analysed standalone or linked to other national vital event and health datasets held by PHS. It supports longitudinal and intergenerational analyses, enabling epidemiological and health service surveillance and research on maternal and child health. PHS is currently using SLiPBD for surveillance of different exposures in pregnancy (vaccines, medicines, infections) and related outcomes, and to monitor pregnancy and newborn screening programmes. PHS will continue to refine SLiPBD as source datasets improve.

## Ethics statement

A data protection impact assessment for SLiPBD was approved by Public Health Scotland's Data Protection Officer (DP2223017). No additional ethical approval was required.

## Conflict of interests statement

No author has any conflict of interest to declare with regard to this paper.

## Publication consent

All authors confirm they have approved the manuscript for submission.

## Funding statement

Development of SLiPBD was supported by core Public Health Scotland funding and additional funding provided to Public Health Scotland (i) as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (grant ref MC\_PC\_20058) and (ii) by the Scottish Government to develop capacity for surveillance of medicine use in pregnancy. No external funders had any role in the development of SLiPBD or in the production of this manuscript or the decision to submit.

## Data availability statement

SLiPBD contains identifiable patient data hence cannot be made publicly available. Subject to approval from the Public Benefit and Privacy Panel for Health and Social Care (<https://www.informationgovernance.scot.nhs.uk/pbphsc/>), researchers can access pseudonymised extracts of SLiPBD (linked to other national datasets as required) through the Scottish NHS safe haven facility, which is supported by Public Health Scotland. Interested researchers should submit an initial enquiry form to Research Data Scotland (<https://www.researchdata.scot/accessing-data/>).

## References

- Public Health Scotland's privacy notice. 2022; Available at: <https://publichealthscotland.scot/our-privacy-notice>. Accessed 29 January, 2024.
- COVID-19 in Pregnancy in Scotland. 2023; Available at: <https://www.ed.ac.uk/usher/eave-ii/covid-19-in-pregnancy-in-scotland/about>. Accessed 29 January, 2024.
- Stock SJ, Carruthers J, Denny C, Donaghy J, Goulding A, Hopcroft LEM, et al. Cohort Profile: The COVID-19 in Pregnancy in Scotland (COPS) dynamic cohort of pregnant women to assess effects of viral and vaccine exposures on pregnancy. *International journal of epidemiology* 2022;51(5):e245-e255. <https://doi.org/10.1093/ije/dyab243>
- 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* 2019;28(7):923-933. <https://doi.org/10.1002/pds.4811>
- Liyanage H, Williams J, Byford R, de Lusignan S. Ontology to identify pregnant women in electronic health records: primary care sentinel network database study. *BMJ Health Care Inform* 2019;26(1):e100013. <https://doi.org/10.1136/bmjhci-2019-100013>
- 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):e0192033. <https://doi.org/10.1371/journal.pone.0192033>
- Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernández-Díaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf* 2013;22(1):16-24. <https://doi.org/10.1002/pds.3284>
- Hornbrook MC, Whitlock EP, Berg CJ, Callaghan WM, Bachman DJ, Gold R, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health Serv Res* 2007;42(2):908-927. <https://doi.org/10.1111/j.1475-6773.2006.00635.x>
- Holt E. Poland to introduce controversial pregnancy register. *The Lancet*; 399(10343): 2256-2256. [https://doi.org/10.1016/S0140-6736\(22\)01097-2](https://doi.org/10.1016/S0140-6736(22)01097-2)
- Olausson Petra. Pakkanen M. The Swedish Medical Birth Register - A summary of content and quality. Research Report from The Swedish Centre for Epidemiology 2003;112(3). [https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2003-112-3\\_20031123.pdf](https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2003-112-3_20031123.pdf)
- Langhoff-Roos J, Krebs L, Klungsøyr K, Bjarnadottir RI, Källén K, Tapper A, et al. The Nordic medical birth registers—a potential goldmine for clinical research. *Acta Obstet Gynecol Scand* 2014;93(2):132-137. <https://doi.org/10.1111/aogs.12302>
- Harper G. Linkage of Maternity Hospital Episode Statistics data to birth registration and notification records for births in England 2005–2014: Quality assurance of linkage of routine data for singleton and multiple births. *BMJ Open* 2018;8(3):e017898. <https://doi.org/10.1136/bmjopen-2017-017898>
- Coathup V, Macfarlane A, Quigley M. Linkage of maternity hospital episode statistics birth records to birth registration and notification records for births in England 2005–2006: quality assurance of linkage. *BMJ Open* 2020;10(10):e037885. <https://doi.org/10.1136/bmjopen-2020-037885>



14. Population Estimates Time Series Data. Available at: <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data>. Accessed 26 February, 2024.
15. Births Time Series Data. Available at: <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/births/births-time-series-data>. Accessed 26 February, 2024.
16. Public Health Scotland. Termination of pregnancy statistics: Year ending December 2022. 2023. <https://publichealthscotland.scot/publications/termination-of-pregnancy-statistics/termination-of-pregnancy-statistics-year-ending-december-2022/>
17. Public Health Scotland. Births in Scotland: Year ending 31 March 2023. 2023. <https://publichealthscotland.scot/publications/births-in-scotland/births-in-scotland-year-ending-31-march-2023/>
18. Public Health Scotland. Child health pre-school review coverage 2022 to 2023. 2024. <https://publichealthscotland.scot/publications/child-health-pre-school-review-coverage/child-health-pre-school-review-coverage-2022-to-2023/>
19. Scottish Index of Multiple Deprivation 2020: introduction. 2020. <https://www.gov.scot/publications/scottish-index-multiple-deprivation-2020/>
20. Tran, D.T., Havard, A. & Jorm, L.R. Data cleaning and management protocols for linked perinatal research data: a good practice example from the Smoking MUMS (Maternal Use of Medications and Safety) Study BMC Med Res Methodol 2017;17(97). <https://doi.org/10.1186/s12874-017-0385-6>
21. Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. Paediatr Perinat Epidemiol 2006;20(4):329-337. <https://doi.org/10.1111/j.1365-3016.2006.00715.x>
22. Draper ES, Gallimore ID, Kurinczuk JJ, Kenyon SL (Eds). MBRRACE-UK Perinatal Confidential Enquiry, A comparison of the care of Black and White women who have experienced a stillbirth or neonatal death: State of the Nation Report. Leicester: The Infant Mortality and Morbidity Studies, Department of Population Health Sciences, University of Leicester. 2023. <https://timms.le.ac.uk/mbrrace-uk-perinatal-mortality/confidential-enquiries/>
23. Public Health Scotland. Anti-Seizure Medicines in Pregnancy: April 2018 - September 2023. <https://publichealthscotland.scot/publications/anti-seizure-medicines-in-pregnancy/anti-seizure-medicines-in-pregnancy/>
24. Public Health Scotland. Pregnancy Screening for Down's Syndrome, Edwards' Syndrome, and Patau's Syndrome in Scotland: 1 April 2019 to 31 March 2022. <https://publichealthscotland.scot/publications/pregnancy-screening-for-down-s-syndrome-edwards-syndrome-and-patau-s-syndrome-in-scotland/>

## Abbreviations

- AAS: Abortion Act Scotland records (provides information on terminations of pregnancy carried out under the Abortion Act 1967 up to April 2022)
- ABC: Antenatal booking collection records (provides information on women booking for NHS antenatal care from April 2019 onwards)
- CHI: Community Health Index (the CHI database is the NHS Scotland master patient index, and the CHI number is the unique patient identifier used on all Scottish health records)
- COPS: COVID-19 in Pregnancy in Scotland study
- NHS: National Health Service
- NRS: National Records of Scotland (responsible for statutory vital event registration in Scotland)
- PHS: Public Health Scotland
- SLiPBD: Scottish Linked Pregnancy and Baby Dataset
- SMR01: Scottish Morbidity Record 01 (provides information on all patients discharged following NHS day case or inpatient care in a general [not neonatal, maternity, or mental health] specialty
- SMR02: Scottish Morbidity Record 02 (provides information on all patients discharged following NHS day case or inpatient care in a maternity specialty
- ToPSS: Termination of Pregnancy Submissions Scotland records (provides information on terminations of pregnancy carried out under the Abortion Act 1967 from May 2022 onwards)



## Supplementary Appendices

Supplementary Appendix 1: Information on how the source records that feed into the national datasets used to compile SLiPBD are generated and returned to PHS

Dataset	Generation of source records	Timeliness
NRS statutory live and stillbirth registrations	Parents are legally required to register live and stillbirths through their local registration office	Live and stillbirths must be registered within 21 days of birth. NRS submits completed registration records to PHS on a weekly basis
Termination of pregnancy notifications (Abortion Act Scotland [AAS] statutory records to April 2022, Termination of Pregnancy Submissions Scotland [ToPSS] non-statutory records from May 2022)	Doctors providing a termination are legally required to notify the event to the Chief Medical Officer. Previously, these notifications were then submitted by the CMO to PHS for entry into the AAS database. More recently, the NHS service providing a termination is also responsible for entering a record of the event directly into the ToPSS database which is managed by PHS	Terminations must be notified to the CMO within 7 days of occurrence. Previously, notifications were then submitted to PHS and entered into the AAS database by data management staff within 6 weeks. More recently, clinical or administrative staff within local services record terminations in ToPSS within 30 days of occurrence
Scottish Morbidity Record (SMR) 01 general, and SMR02 maternity, hospital discharge records	The NHS service providing a patient's hospital care is responsible for submitting an SMR record of the event to PHS. SMR records are partially extracted from local electronic patient record systems and completed with input from administrative and clinical coding staff	SMR records are returned to PHS within 6 weeks of the end of the month in which the patient was discharged
Antenatal Booking Collection (ABC) records	The NHS service providing a patient's booking care is responsible for submitting an ABC record of the event to PHS. ABC records are extracted from local maternity electronic patient record systems	ABC records are returned to PHS on a weekly basis

See <https://publichealthscotland.scot/services/national-data-catalogue/> for information on the various source national datasets held by Public Health Scotland.

In general, PHS runs extracted or submitted source records through automated validation and CHI seeding processes on a monthly basis, and uploads validated records into the relevant national datasets for analysis. The monthly updates of SLiPBD are timed to closely follow the monthly updates of source national datasets to maximise timeliness.



## Supplementary Appendix 2: Information on the variables and codes used to identify specific end of pregnancy events in relevant national datasets

Dataset	End of pregnancy event	Notes
NRS statutory live birth registrations	Live birth	All records indicate specified event
NRS statutory stillbirth registrations	Stillbirth Termination of pregnancy	All records indicate a stillbirth The subset with ICD-10 code P96.4 recorded also indicate termination of pregnancy
Termination of pregnancy notifications (AAS and ToPSS)	Termination of pregnancy	All records indicate specified event
Scottish Morbidity Record (SMR) 01 general hospital discharge records	Spontaneous early pregnancy loss	ICD-10 codes: <b>Ectopic Pregnancy:</b> O00 <b>Molar Pregnancy:</b> O01 <b>Miscarriage:</b> O02.0–O02.9; O03.5–O03.9; O05.5–O05.9; O065–O06.9
Scottish Morbidity Record (SMR) 02 maternity hospital discharge records	Spontaneous early pregnancy loss Termination of pregnancy	SMR02 hard coded variables <b>Ectopic Pregnancy:</b> Condition on Discharge = 2, Type of abortion = 6 <b>Molar Pregnancy:</b> Condition on Discharge = 2, Type of abortion = 3 <b>Miscarriage:</b> Condition on Discharge = 2, Type of abortion = 1, 2, 8, or 9 <b>Termination:</b> Condition on Discharge = 2, Type of abortion = 4
Scottish Morbidity Record (SMR) 02 maternity hospital discharge records	Live or stillbirth (to provide additional clinical information to supplement NRS statutory registration records)	SMR02 hard coded variables <b>Live Birth:</b> Condition on Discharge = 3, Outcome of pregnancy = 1, 3, 4, or 5 <b>Stillbirth:</b> Condition on Discharge = 3, Outcome of pregnancy = 2

See <https://publichealthscotland.scot/services/national-data-catalogue/> for information on the various source national datasets held by Public Health Scotland.

