International Journal of Population Data Science (2024) 9:2:12

International Journal of Population Data Science

Journal Website: www.ijpds.org

A maternal and child health administrative cohort in Scotland: the utility of linked administrative data for understanding early years' outcomes and inequalities

Paul Henery[1], Ruth Dundas[1], Srinivasa Vittal Katikireddi[1], Alastair H. Leyland[1], Lynda Fenton[2], Sonya Scott[2], Claire Cameron[2], and Anna Pearce[1,*]

Submission History
Submitted: 30/04/2024
Accepted: 10/09/2024
Published: xx/xx/20xx

[1]MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, G12 8TB
[2]Public Health Scotland, Glasgow, G2 6QE

Abstract

Introduction
The early years are considered one of the most impactful points in the life course to intervene to improve population health and reduce health inequalities because, for example, both ill health and social disadvantage can track into adulthood. Scotland's outstanding systems for data linkage offer untapped potential to further our understanding of when and why inequalities in child health, development and wellbeing emerge. This understanding is vital for the consideration of policy options for their reduction.

Methods
Birth registrations, hospital episodes, dispensed community prescriptions, child health reviews and immunisation records were linked for 198,483 mother-child pairs for babies born in Scotland from October 2009 to the end of March 2013, followed up until April 2018 (average age 6 years).

Results
Outcomes include birthweight and newborn health, dispensed prescriptions for mental health medications, tobacco smoke exposure, infant feeding, immunisations, hospitalisation for unintentional injuries, socio-emotional, cognitive and motor development, and overweight and obesity. Several measures are repeated throughout childhood allowing examination of timing, change and persistence. Socio-economic circumstances (SECs) include neighbourhood deprivation, relationship status of the parents, and occupational status. Descriptive analyses highlight large inequalities across all outcomes. Inequalities are greater when measured by family-level as opposed to area-level, aspects of socio-economic circumstances and for persistent or more severe outcomes. For example, 41.4% of the most disadvantaged children (living with a lone, economically inactive mother in the most deprived fifth of areas) were exposed to tobacco smoke in utero and in infancy/toddlerhood compared to <1% in the least disadvantaged children (living with a married, managerial/professional mother in the least deprived quintile of areas).

Conclusion
This novel linkage provides a longitudinal picture of health throughout the early years and how this varies according to family- and area-level measures of SECs. Future linkages could include other family members (e.g. siblings, grandmothers) and other sectors (e.g. education, social care). The creation of additional cohorts would allow for long-term and efficient evaluation of policies as natural experiments.

Keywords
child health; maternal health; inequalities; cohort; admin data

*Corresponding Author:
Email Address: anna.pearce@glasgow.ac.uk (Anna Pearce)

https://doi.org/10.23889/ijpds.v9i2.2402

Key features

1. This is Scotland's first administrative child cohort of its kind; following all children (and their mothers) born in Scotland Oct 2009 — March 2013 up until an average age of 6 years ($n \sim 200{,}000$).

2. This bespoke cohort was created to deepen our understanding of health inequalities in Scotland and to demonstrate how more comprehensive linkages of datasets across the early years can be used to consider multiple aspects of social disadvantage as well as how health changes and accumulates across the early years.

3. The dataset comprises birth registrations, hospital birth records, community prescribing dispensation data, immunisation records and child health checks.

4. Data include socio-economic characteristics, birth outcomes, parental health behaviours, physical, cognitive and socio-emotional development, and aspects of mental and physical health for the mother and child.

5. Descriptive analyses of these data confirm that there are large child health inequalities in Scotland and add to our understanding of these by showing that routine reports of inequalities, according to current area-level deprivation and snapshots of health, underestimate the detrimental impacts of social disadvantage on child and parental health.

6. These data are held by Public Health Scotland's Electronic Data Research and Innovation Service where they can be accessed by application (for a fee). The construction and analysis of this cohort demonstrates the potential of linked administrative data for furthering our understanding of health with possibilities to extend linkages to examine health across the entire life course and to evaluate policies using natural experiment study designs.

Background

The early years – broadly defined as the period from conception to the start of school – are a critically important period when key physical, cognitive, socioemotional and behavioural changes occur [1]. These changes have a powerful influence on health and wellbeing across the life course and can affect educational and economic outcomes [2–5]. Because of this, investment in the early years is considered one of the most efficient and effective ways to improve population health [2, 6]. Major investments in cohort surveys have supported a better understanding of children's health and development and their longer-term consequences. They have shown that socio-economic inequalities in health emerge from the moment of birth and persist across childhood and into adulthood [3, 7–10]. They have also supported the identification of barriers (e.g. poverty [11, 12], poor neighbourhood environments [13, 14]) and supports (e.g. social support [15, 16], early years' education [17, 18]) to children's health and development and shown how complex and inter-related they may be.

Administrative datasets hold much unharnessed potential for understanding health and health inequalities from a life course perspective [19]. For example, they are more likely to capture and retain unhealthy and disadvantaged groups than surveys, thus providing a more accurate picture of inequalities. They are less susceptible to recall or response bias because records are often triggered by objectively measured events (such as hospital admissions) or recorded using screening tools administered by health professionals [20]. Since they are collected over time, they can provide a more complete longitudinal picture of people's health and socio-economic experiences, allowing examination of timing, change and persistence. Historical data can be used to examine long-term trends and evaluate natural policy experiments [21]. Finally, they enable the examination of rare and severe outcomes and other 'data hungry' analyses, such as the investigation of intersectionality or the complex pathways through which inequalities can arise.

Public health surveillance reports and academic papers have frequently used administrative data to monitor maternal and child health outcomes and inequalities [22–25]. However, in the majority of cases, the datasets have been analysed in their unlinked forms. This means that they are limited to area-level measures of socio-economic circumstances [22] (as opposed to family-level markers available from birth records, for example) and are unable to look at the timing or accumulation of exposures, health trajectories, or the relationships between maternal and child health. Scotland - a country where health outcomes are poor and health inequalities especially high relative to the rest of western Europe [23] - has outstanding systems of data linkage. This offers currently untapped opportunity to deepen our understanding of the emergence and persistence of child health and development. The Scottish Government has some devolved powers, including health and social care services, presenting opportunities for holistic and longer-term evaluations of natural policy experiments [21].

Our aim was to create a cohort, from administrative data, of children born in Scotland during October 2009-end of March 2013. The research objectives are to: 1) examine individual and area-level inequalities in child physical and mental health, as well as maternal health and health-related behaviours; 2) consider the use of administrative data for informing eligibility criteria for early years' interventions; and 3) examine mediating pathways between socio-economic circumstances (SECs) and child health and emulate the impacts of hypothetical policy and intervention options on these pathways. By addressing these objectives, we intend to demonstrate the underutilised potential of administrative data (in the UK and beyond) for understanding health inequalities and for carrying out life course research to ultimately inform policy. The purpose of this paper is to describe the creation and characteristics of the cohort, provide a detailed report of child health inequalities and highlight the potential for future research and further linkage.

Methods

Setting and eligibility

All children born in Scotland between October 2009 and end of March 2013 were potentially eligible for inclusion, - amounting to approximately 204,000 infants (estimated using monthly birth registrations and annual births data [26]). This date range was selected to maximise historical data quality and completeness (which varied across the different datasets) whilst providing sufficient follow-up across the early years (up until April 2018 at present). Registered births with a Community Health Index (CHI), a unique ID number assigned at birth, were eligible for inclusion in the study (202,757 born to 180,230 mothers). The CHI was used to link to the administrative datasets summarised below. About 97% of births were singleton and 21% children had another sibling included in the cohort.

Figure 1 visualises the datasets included in the linkage, their key foci, and the ages covered.

Datasets

National Records of Scotland (NRS) Births: The NRS holds information on all births registered in Scotland since 1975. These records contain date and location of the birth and details of the registered parent(s), including their postcode of residence, marital/relationship status and occupational status.

Maternity records:

- *Scottish Morbidity Record 02 (SMR02):* SMR02 records all maternity inpatient and day case episodes in Scotland. Around 50% of episodes relate to births and it was these records that were requested for the purposes of building the cohort. These include demographic characteristics and information relating to the birth and clinical management.

- *The Scottish Birth Record (SBR):* The SBR is a web-based system used to record infant neonatal care in Scotland and, historically, to register a baby with a CHI number shortly after birth. SMR02 records were supplemented with SBR to improve coverage (for example in the case of missing SMR02 records for home births - although SBR in itself is unlikely to capture all homebirths). Less than <0.5% of records across all of the analytic samples used SBR as opposed to SMR02.

Scottish Morbidity Record 01 (SMR01): SMR01 is an episode-based inpatient and day case record of all non-obstetric and non-psychiatric admissions to general hospitals in Scotland. Admissions related to unintentional injuries were requested as a priority childhood outcome which is highly socially patterned [27].

Prescribing Information System (PIS): This national database records all medicines that are prescribed and dispensed in the community in Scotland. It is a reimbursement-based system for pharmacies. The cohort contains records of dispensed drugs usually used to treat depression/anxiety in mothers (as an under-researched and important determinant of children's health and development) and childhood attention deficit hyperactivity disorder in children (ADHD, a condition which is potentially on the rise, has far reaching social consequences, and requires further research [28, 29]).

Scottish Immunisation and Recall System (SIRS): SIRS has been used by all Scottish health boards since 2002. It calls children for routine immunisations according to the UK childhood immunisation schedule and records details and dates of administered immunisations including those up until the age of six.

Child Health Systems Programme (CHSP): Since the fiscal year 2011/12, this system has been used to schedule and record reviews according to the Child Health Programme. For our cohort (born October 2009 to end of March 2013), universal health visitor and child health reviews were carried out within the first fortnight of birth, at age 6-8 weeks, and once starting primary school (Primary 1 review, age 4-5 years). An additional universal review at age 27-30 months was introduced in Scotland in April 2013. Therefore, all cohort children born from January 2011 onwards were eligible for inclusion in this check with those born in October-December 2010 potentially eligible (aged 28-30 months at the point of introduction). Unscheduled visits are carried out as required. We elected to use these where a record was missing from one of the main reviews and an unscheduled visit had taken place at a similar age (and had collected the relevant information). Information collected varies by age but includes objective length and weight, feeding in infancy, assessments of physical, cognitive, and socio-emotional development at age 27-30 months, and heights, weights, and body mass index at age 27-30 months and 4-5 years.

Key variables are described in the results tables, with summaries of all variables, according to each dataset, provided in Supplementary Appendix 1. Complete lists of variables available in these datasets are provided elsewhere [30, 31].

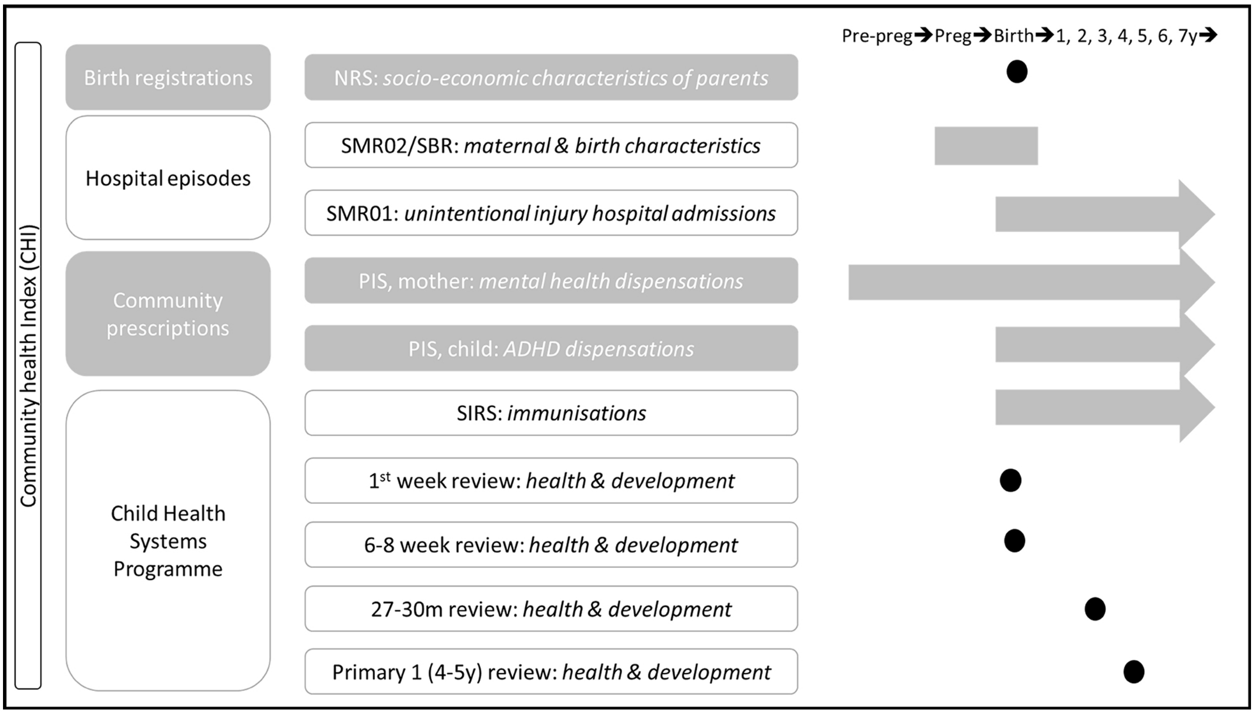
Linkage procedures, data governance and ethics

The eData Research and Innovation Service (eDRIS) within Public Health Scotland (PHS) provided access to the data, after ethical and data control criteria had been satisfied, in the form of a Public Benefit and Privacy Panel for Health and Social Care (PBPP) application (ref: 1617-0152). The eDRIS prepared and linked each individual dataset and provided them to the researchers within the National Safe Haven, a remotely accessed and secure environment where input and output is screened by assigned data controllers. As the cohort comprises secondary data sources, which have been de-identified, further ethical approval was not required. Cohort member consent was not required for the data linkage, since the data were being used for purposes that are compatible with their original collection – to support individual and public health.

Patient and public involvement

The content of the cohort and the results reported here have been guided by conversations with parents, children and young people. These conversations have taken place via a public engagement game, Data Detective, which we created to show

Figure 1: Overview of datasets, topics and study period



CHI: Community Health Index; NRS: National Records of Scotland (NRS) Births; SBR: The Scottish Birth Record; SMR: Scottish mortality record; PIS: Prescribing Information System; SIRS: Scottish Immunisation Recall System;. Preg: Pregnancy;. Y: Year;. M: Month;. ADHD: Attention deficit hyperactivity disorder.

how and why data are linked and analysed by researchers to answer public health questions [32]. We have spoken at length with a class of primary school children from one of Scotland's most disadvantaged areas born at the same time as the cohort members (and therefore included in, but not identifiable from, these data). They were generally supportive of the use of the data and highlighted neighbourhoods, families, relationships, socio-emotional wellbeing and sensory development as important areas – all of which we have striven to cover in this paper.

Socio-economic measures

There is no singular measure which can capture the complexity of family SECs in childhood [33]. We had access to three measures of SECs – each capturing different aspects of social disadvantage:

The Scottish Index of Multiple Deprivation (SIMD, 2012) was used as a relative measure of area deprivation captured at the data zone level [34]. The SIMD describes characteristics of areas, including crime levels, housing, income and education levels and employment rates. It is widely used in health inequalities research in Scotland [23, 35, 36].

Birth registration details were used to capture the relationship status of parents. This measure distinguishes between the following family structures: married, cohabiting (joint registration, unmarried), separated (joint registration, living apart) and sole registration (only one parent registered the birth). Family structure and marital status are key drivers of poverty in the UK and Scotland [37–39]. They are also strongly related to infant and child health [38], including in Scotland and using the measure described here [40–42].

The National Statistics Socio-economic Classification (NS-SEC) of the mother was categorised as: economically inactive (never worked, unemployed, student, other), routine and manual occupations, intermediate, and higher managerial and professional. It is a commonly used proxy for family socio-economic circumstances [22, 36], as it captures employment relations and work conditions, and can be related to other characteristics such as income. The father's NS-SEC is available but not reported in this paper since it is not recorded for sole birth registrations. Sensitivity analyses, taking the highest of the mother's or father's (where relevant) NS-SEC, has shown that the social patterning of health outcomes such as unintentional injuries [40] and ADHD [41] are extremely similar.

To capture a fuller picture of health inequalities, a socio-economic score was created combining relationship status, occupational status and area deprivation. Each of these original variables was collapsed into three categories and then summed. The resulting score ranged from 0 (no socioeconomic 'adversities') to 6 (all three socioeconomic 'adversities'). Figure A2 in Supplementary Appendix 2 shows the breakdown of the score according to the individual socio-economic measures, including showing how the same score can be arrived at through different combinations of the three variables.

Derived health and demographic variables

Due to small numbers, some variables were collapsed into fewer categories, either before or after the linkage stage. For example, we have combined mothers' country of birth into Scotland, rest of UK, and elsewhere, due to low numbers when broken down by other characteristics. The operational categories and prevalences for all variables are shown in the results tables, with more detailed breakdowns available from the data holders.

Mother's age at the birth of the cohort child is routinely reported. We wanted to estimate mother's age at *first* live birth. We estimated it using SMR02 records from previous births to cohort mothers, since 1980 (thereby assuming birth spacing of no more than ten years and excluding births taking place outside Scotland).

Due to small numbers in some groups, the ethnicity variable used in SMR02 was categorised as follows: White (White: Scottish, English, Welsh, Northern Irish, British, Gypsy/Traveller, Polish, Other), Mixed or multiple ethnic groups, Asian (Pakistani, Pakistani Scottish or Pakistani British, Indian, Indian Scottish or Indian British, Bangladeshi, Bangladeshi Scottish or Bangladeshi British, Chinese, Chinese Scottish or Chinese British, Other Asian), Black African (Black African, African Scottish or African British), Black Caribbean (Black Caribbean, Caribbean Scottish, Caribbean British) and Other ethnic groups (Arab, Arab Scottish or Arab British and Other disaggregated).

Low birthweight captured those with birthweights of under 2.5kg. Children were classified as experiencing underweight, overweight or obesity using age and sex adjusted International Obesity Task-Force Body Mass Index (BMI) cut-offs for 2–18 year olds [43]. It is worth noting that routine publications from PHS use the UK 1990 reference thresholds for overweight and obesity risk [44], therefore producing similar but not directly comparable results. BMI is an imperfect tool for identifying overweight and obesity in children, with the challenges and strengths of the different thresholds discussed elsewhere [43, 44]. Nevertheless, it is considered appropriate for population-level monitoring of prevalences and inequalities. The nine developmental domains covered in the 27–30-month check were combined into four: motor development (concerns on fine or motor skills); socio-emotional wellbeing (attention, behaviour, emotional, personal and social development concerns); sensory (hearing or vision concerns); and speech, language and communication concerns. A child development score was also created, summing the number of new or existing concerns across the nine development domains. This was created to look at inequalities in children who were experiencing multiple concerns but assumes each domain to be equally important.

For health outcomes measured at multiple points, longitudinal measures were created. The categorisation of these outcomes was guided by the distribution of the data, as well as the life course framework [45]. Those which are not self-explanatory from the relevant tables and figures are now described.

For immunisations, a variable was created capturing proportions fully, partially or unimmunised for all relevant immunisations (primary immunisations, the first and second doses of the Measles, Mumps and Rubella vaccine (referred to as MMR1 and MMR2 respectively) and the preschool booster immunisation) by the end of the study period.

Dispensations for anxiety and depression medications to the mother were considered during the following windows: prior to pregnancy, during pregnancy, infancy (0-6 months, 6 months-1 year), toddlerhood (1-2, 2-3 years) and early childhood (3-4, 4-5 years). A variable captured the most common patterns of dispensations for anxiety / depression medications: 'never', 'early' (pre-pregnancy-toddlerhood only), 'late' (childhood, 3 years+), 'relapse' (pre-pregnancy and again in toddlerhood-childhood), 'infancy/toddlerhood onwards' and 'across most or whole period'.

A longitudinal measure of infant feeding was derived from records at first feed, hospital discharge, first health visit and 6-8 week health visit, classified as: 'never breastfed', 'initiated breastfeeding in hospital but did not continue', 'still breastfeeding at first health visit', 'still breastfeeding at 6-8 weeks'. This variable captures exclusive breastfeeding and breastfeeding in combination with formula milk ('mixed feeding').

A longitudinal measure of smoking was created from records of maternal smoking in pregnancy and whether the child was exposed to smoke at the first, 6-8 week and 27–30 month health checks, comprising: 'never exposed to smoke', 'exposed in pregnacy only', 'exposed post-birth only', and 'exposed during pregnancy and post-birth', where post birth represented any exposure to smoking at the first, 6–8 week and 27–30 month visits.

Statistical analyses

A series of descriptive analyses were conducted; first describing the characteristics of the cohort using prevalences and means as appropriate. For binary health outcomes, we examine inequalities using the relative and slope indices of inequality (RII and SII). These compare the notionally most- and least-deprived people in the population, in relative and absolute terms [46], allowing for comparisons across different socio-economic measures (occupational status, relationship status, neighbourhood deprivation and the socio-economic score). For longitudinal outcomes, which had multiple categories, we estimate relative risk ratios (RRRs), using multinomial logistic regression, which also compare the notionally most and least deprived. All regression models were unadjusted.

Sensitivity analyses

Multiple imputation was used to fill in item missingness for the longitudinal outcomes which were most susceptible to missingness: smoking and infant feeding throughout the early years and trajectories of overweight between toddlerhood and start of school. Imputation was carried out only for children who were present within the relevant datasets. Multiple chained augmented imputation was run, under a missing at random assumption, separately for each sample and consisted of 20 iterations. Supplementary Appendix 3 provides further detail on the imputation process.

Results

Study period and loss-to-follow-up

Children have currently been followed up from birth (October 2009-March 2013 inclusive) until April 2018 (mean age 82.6

months (range: 62-103 months), or 6.9 years (5.2–8.6 years). Figure 1 shows how the coverage of the different datasets varies. For example, dispensed prescriptions in mothers were included from January 2008 (the time from which linkage to CHI was deemed to be of sufficient completeness and quality), providing information for a minimum of 12 months prior to (as well as during) pregnancy. Longitudinal records on unintentional injuries, ADHD and immunisations for the child span the entire period. Several datasets hold information at fixed time points – those relating to pregnancy and birth (SMR02) and the child health reviews (with the most recent taking place at the start of primary school (4–5 years)).

Of those with an identifier (CHI) ($n = 202{,}757$), the vast majority ($n = 198{,}483$ (98%)) could be linked to information relating to their birth, through SMR02 or SBR records.

Records of unintentional injuries, mental health prescribing, ADHD and immunisations are initiated only when a hospital episode, immunisation or drug dispensation occurs. These are referred to as 'events-based' outcomes. The absence of a record in these datasets was assumed to mean that the outcome had not occurred, although they may have been lost-to-follow-up. A total of 10,473 children had a mother who was recorded as having left the CHI dataset after their birth. This typically occurs when a patient is recorded as leaving a GP practice and not registering with another in Scotland. This was taken as indication of migration outside Scotland and these children are excluded for outcomes which were measured in SMR01 (hospitalisations), SIRS (immunisations) and PIS (prescriptions dispensed). Loss-to-follow-up from death could not be identified with the current linkage but is relatively rare. For example, the infant mortality rate over the relevant period was 3.5 per 1000 births and we estimate that ∼1000 of the total eligible cohort would have died across the study period. Children who were excluded because their mother had exited the CHI database were distributed across all socio-economic groups, although this was slightly more common in the most advantaged groups (data not shown).

The greatest degree of identifiable loss-to-follow-up was observed in the child health reviews. A total of 182,666 (92%) had a 6–8-week child health review record. Routine child health review reports imply comparable coverage. Coverage declined for the later checks. The 27–30-month review was introduced in Scotland in April 2013, so only 61% of the cohort children were eligible to receive a review. Of these, 103,596 (85%) had a record. This is slightly lower than the national average of ∼88% across the period 2013-2016 [47]. All cohort children were potentially eligible for the primary 1 child health review and 164,640 (83%) had a relevant record. This was lower than expected based on national reports over the same period, although these fluctuated from over 90% between 2012-2016, falling to 85% and 83% in 2016/17 and 2017/18 respectively [44]. It is worth noting that national reports cannot be directly compared to the figures here as they use a different base population (which, for example, includes children not born in Scotland). Of the 17% of all cohort children missing a child (4–5y) record, 27% had a mother who had exited CHI, making them theoretically ineligible due to emigration. Other reasons for the remaining missingness could include unidentified emigration and missed linkages. The characteristics of those included in each sample are considered shortly.

Analytic samples

Given the variability in eligibility for different datasets and levels of missing data, and in order to maximise representativeness and power for individual analyses, we created five analytic samples for use for different questions:

1. 'Birth Sample' ($n = 198{,}483$): children who had a CHI, NRS and SMR02/SBR records. This sample can be used to examine birth outcomes. Nested within the Birth Sample are:

2. The 'Event Sample' ($n = 188{,}010$), which excludes children whose mothers were recorded as having exited CHI over the study period and can be used to examine event- based outcomes (dispensed prescriptions, immunisations and unintentional injuries), which assume the absence of a record implies absence of the outcome;

3. The 'Infant Sample', which includes those who had a 6-8 week record, $n = 182{,}666$;

4. 'Toddler Sample': - those with a 27-30 month review, $n = 103{,}596$;

5. 'Child Sample': - those with a Primary 1 review, $n = 164{,}640$.

The samples are represented in Figure 2, along with examples of the outcomes that they could be used to examine.

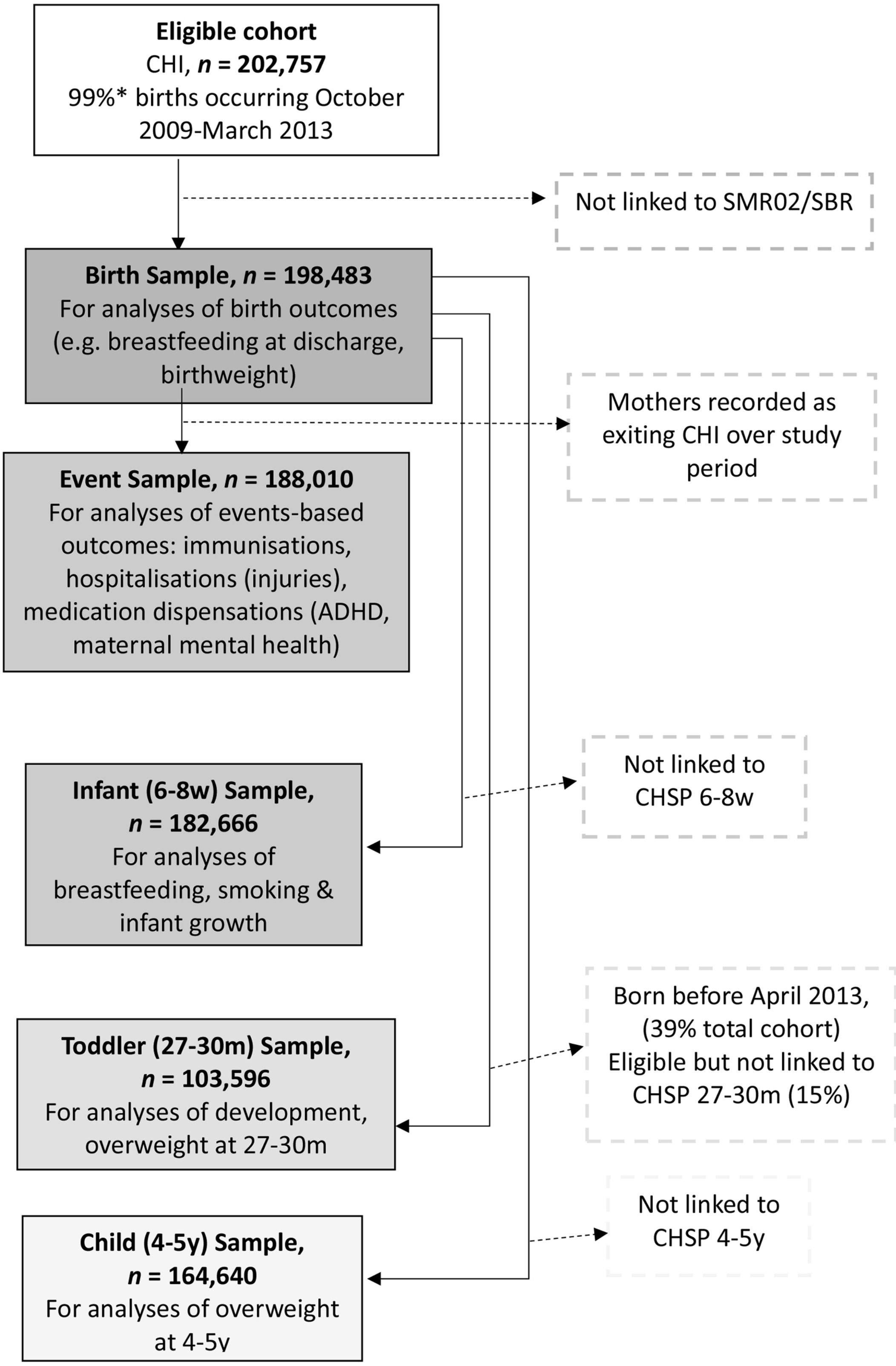
Socio-demographic, pregnancy, and birth characteristics

Table 1 presents the prevalence of socio-demographic, pregnancy and birth characteristics. Here we focus on the birth sample with comparisons to the other samples described in later sections. There were more children living in the most-deprived fifth of areas of Scotland (26%) than the least (16%) indicating that families with young children are more likely to live in more-deprived areas than average. Almost half of children had mothers who were economically inactive (22%) or in routine and manual occupations (27%). Few (5.3%) births were solely registered and 12% were registered to separated parents.

There was a large amount of missing information on ethnicity (40% of the sample were missing this information). Of the remaining 60%, 93% in the birth sample were from White ethnic backgrounds, with Asian (4.3%) and Black African (1.5%) being the most common non-White ethnic minority groups. Data on mother's country of birth, which had virtually no missingness, showed that 76% mothers were born in Scotland, 10% in the rest of the UK or Ireland, and a further 14% from elsewhere. After Scotland, the most common countries were: England (8.5%), Poland (3%), Pakistan (1.2%) and Ireland (0.9%) (data not shown).

Just over one in twenty babies (6.4%) were born low birthweight (<2.5kg), 12% were born pre-term (<37 weeks),

Figure 2: Flow diagram to demonstrate how the four analytic samples were created from the original Eligible Cohort



CHI: Community Health Index;. NRS: National Records of Scotland Births (NRS); SBR: The Scottish Birth Record; PIS: Prescribing Information System; SIRS: Scottish Immunisation Recall System;. CHSP: Child Health Surveillance Programme;. Y: Years;. M: Months;. W: Weeks.
*Estimated to be 204,000, based on annual birth data and monthly birth registrations.

16% via emergency c-section. Very few (1.4%) babies had moderately or severely depressed Apgar scores at birth or had been resuscitated. Need for additional support (based on the Health Plan Indicator [HPI]), declined across health checks, starting out common (72%) at the first check and becoming relatively rare by the primary 1 check (14%). However, levels of missingness for this variable was high, with just over one in five were missing this information. Around 17% of mothers were estimated to have given birth to their *first* child as a teenager, while 6.0% of *cohort children* were born to teenage mothers.

Parental heath and health-related behaviours

Table 2 presents the prevalence of various parental health and health-related behaviours. Here we mainly focus on the birth sample, with comparisons to the other samples described in later sections. Starting with the proportion of mothers who had been prescribed and collected medications for anxiety and depression; prior to pregnancy, this stood at around 14%. A smaller proportion (7.5% of mothers) had been prescribed and collected medications during pregnancy. In infancy this increased again, to 12% and 15% for 1–6 months and 6–12 months respectively. More than one in five mothers had been dispensed medications during each of the one-year intervals between 2–5 years, with signs of small increases with age.

During the antenatal booking, 28% and 21% mothers were recorded as experiencing overweight and obesity respectively. Experiences of underweight were relatively rare at 2.6%. This variable had relatively high levels of missingness, at 16%.

Two in three (64%) women were exclusively breastfeeding at first feed, albeit with high amounts of missing data (for one third of children). Exclusive breastfeeding at subsequent measurement points fell to just under one-in-two at the point of hospital discharge, one-in-three at the first health visit and just one-in-four by 6-8 weeks. This was accompanied by increases in formula feeding from 35% to 64% and mixed feeding from 1.0% to 10% between first feed and the 6–8 week check.

The vast majority of children were fully immunised, with the highest uptake seen for the primary immunisations (98-99% immunised by 1 year) and the lowest for the second dose of the MMR (94-95%), depending on the sample.

Almost one-in-five (19%) mothers were recorded as smoking during pregnancy, with similar proportions of primary caregivers and children who were exposed to smoke in the household (19%) (Infant Sample). The prevalence of primary caregiver smoking persisted to age 27-30 months (21%, Toddler Sample) but, by this point, the likelihood of the child being exposed in the household had fallen to 13%.

Child health and development outcomes

Table 3 presents a range of health and development outcomes for the child. At 6-8 week visit, the average weight of children was 5kg and the average length 57cm. During the 27-30 month check, the most commonly reported developmental concern (by health professionals) in the toddler sample was for speech, at 14% children affected, followed by socio-emotional development (9%). Concerns around motor skills and vision/hearing were relatively rare at 3–4%.

One in five children were experiencing overweight *including* obesity at 27–30 months, with obesity at this age relatively rare (4.0%). By the start of primary school, the prevalence of overweight including obesity was similar (19%), although with possible signs that proportions experiencing obesity was increasing (to 5.1%). It is worth noting that these prevalences were lower than those reported by PHS over a similar period, using different definitions; when overweight (including obesity) was around 22% and obesity around 10% [44].

One in twenty children had been hospitalised for an unintentional injury before age of five, and just 0.3% had been hospitalised for severe injuries (defined as requiring a transfer to another ward or ending in death). Proportions for whom ADHD medications were prescribed and collected by age seven was also very rare at 0.4% in the Birth Sample.

Comparability of the samples

The analytic samples were generally comparable indicating minimal attrition bias (in terms of what was measured). Here we note some exceptions:

Asian ethnicity and mothers born outside UK/Ireland were under-represented in the older samples when compared to the birth sample; especially so for the child sample (Table 1). Children from ethnic minority groups were far more likely to have a mother who was recorded as having left CHI records (4% in White, the next lowest rate was for Asian children at 16%, with the highest rates of loss in Black Caribbean at 31%, data not shown). The proportions of mothers born outside Scotland in the birth sample (24%) was comparable to the 2011 Census, where 26% of 30-34 year olds (the modal age range among the cohort mothers) of people were born elsewhere [48].

The toddler sample was younger than the other samples (75 months compared to 82-83 months, Table 1), because the 27-30 month health check was not introduced until April 2013 meaning that only younger children were captured. For this reason, the toddler sample differed from the other samples in several ways. ADHD dispensations were less frequent most likely because prescribing increases with age (Table 3). Dispensed prescriptions pre-pregnancy were more common in the toddler sample because women had a longer follow-up period, on average, before pregnancy (records were available from 2008) (Table 2). Births to younger mothers or that were low birthweight were less common in the Toddler Sample (Table 1), possibly explained by declines in both outcomes over time. Proportions not fully immunised were lower in both the toddler and child samples. This may signal barriers to health services experienced by young families, a role of child health checks in supporting immunisation decisions, or that these families were more likely to emigrate.

Socio-economic inequalities in snapshots of maternal and child health

Figure 3 shows relative and absolute inequalities in selected outcomes: low birth weight, moderate to depressed Apgar scores, prescribing dispensations for anxiety and depression in mothers between birth and 6 months, and two or more documented developmental concerns, across the three socio-economic measures.

Children and mothers from more, compared to less, disadvantaged backgrounds were consistently more likely

Table 1: Socio-demographic and birth characteristics of the five analytic samples, %(n)

	Birth sample, n = 198,483	Event sample, n = 188,010	Infant sample, n = 182,666	Toddler sample, n = 103,596	Child sample, n = 164,640
SIMD, quintile at birth					
1st (most deprived)	26.2 (51,876)	26.5 (49,787)	26.3 (47,941)	26.3 (27,260)	26.4 (43,387)
2nd	21.1 (41,855)	21.3 (39,978)	21.2 (38,701)	21.0 (21,762)	21.4 (35,236)
3rd	19.2 (38,078)	19.2 (35,967)	19.2 (35,116)	19.3 (19,948)	19.5 (32,000)
4th	17.9 (35,447)	17.8 (33,409)	17.7 (32,382)	17.8 (18,411)	18.0 (29,525)
5th (least deprived)	15.6 (30,961)	15.3 (28,674)	15.6 (28,395)	15.6 (16,156)	14.8 (24,374)
Missing	(266)	(195)	(131)	(59)	(118)
Mother's Occupational status at birth					
Economically inactive	21.6 (42,806)	21.1 (39,608)	21.3 (38,855)	20.7 (21,424)	20.6 (33,921)
Manual	26.8 (53,193)	27.2 (51,115)	26.7 (48,834)	26.9 (27,849)	27.5 (45,293)
Intermediate	20.5 (40,705)	20.8 (39,180)	20.6 (37,595)	20.6 (21,360)	21.2 (34,904)
Professional	31.1 (61,779)	30.9 (58,107)	31.4 (57,382)	31.8 (32,693)	30.7 (50,522)
Relationship status of parents at birth					
Sole registered	5.3 (10,451)	5.3 (10,004)	5.2 (9,456)	5.2 (5,382)	5.2 (8,622)
Separated	11.9 (23,559)	12.3 (23,076)	11.9 (21,674)	12.1 (12,562)	12.4 (20,350)
Cohabiting	34.0 (67,391)	34.6 (64,996)	34.0 (62,114)	34.4 (35,682)	35.0 (57,615)
Married	48.9 (97,082)	47.8 (89,934)	49.0 (89,422)	48.3 (50,006)	47.4 (78,053)
Sex of child					
Male	51.2 (101,539)	51.2 (96,210)	51.1 (93,421)	51.4 (53,234)	50.6 (83,342)
Female	48.8 (96,944)	48.8 (91,800)	48.9 (89,245)	48.6 (50,362)	49.4 (81,298)
Ethnicity of child					
White	93.0 (111,602)	94.1 (106,941)	93.2 (102,177)	94.1 (65,491)	94.5 (93,787)
Mixed/multiple	0.4 (421)	0.3 (349)	0.4 (388)	0.3 (211)	0.3 (294)
Asian	4.3 (5,142)	3.8 (4,306)	4.3 (4,675)	3.9 (2,689)	3.6 (3,525)
Black African	1.5 (1,775)	1.1 (1,289)	1.4 (1,485)	1.1 (756)	1.1 (1,045)
Black Caribbean	0.6 (716)	0.4 (491)	0.6 (633)	0.4 (257)	0.4 (404)
Other	0.3 (352)	0.2 (267)	0.3 (319)	0.3 (209)	0.2 (216)
Missing	(78,475)	(74,367)	(72,989)	(33,983)	65,369
Mother's country of birth					
Scotland	76.2 (151,186)	79.0 (148,553)	76.5 (139,802)	78.2 (81,039)	79.9 (131,480)
Rest of UK/Ireland	10.2 (20,202)	9.3 (17,391)	10.1 (18,526)	9.5 (9,870)	9.1 (15,028)
Rest of World	13.7 (27,093)	11.7 (22,064)	13.3 (24,336)	12.2 (12,685)	11.0 (18130)
Missing	(<10)	(<10)	(<10)	(<10)	(<10)
Number of births in pregnancy					
1	97.0 (192,611)	97.0 (182,444)	97.0 (177,262)	97.1 (100,572)	97.1 (159,896)
2	2.9 (5,776)	2.9 (5,482)	2.9 (5,320)	2.9 (2,980)	2.8 (4,678)
3+	0.05 (96)	0.04 (84)	0.05 (84)	0.04 (44)	0.04 (66)
Number of older siblings					
0	44.7 (87,838)	44.4 (82,676)	44.6 (80,767)	44.5 (45,665)	44.6 (72,663)
1	34.5 (67,878)	34.6 (64,530)	34.7 (62,737)	34.8 (35,632)	34.7 (56,626)
2	13.6 (26,763)	13.7 (25,594)	13.6 (24,624)	13.6 (13,954)	13.6 (22,219)
3+	7.2 (14,134)	7.3 (13,518)	7.1 (12,866)	7.1 (7,290)	7.1 (11,601)
Missing	(1,870)	(1,692)	(1,672)	(1,055)	(1,531)
Mother's age at first live birth (estimated)					
<20	16.8 (33,121)	17.3 (32,323)	16.7 (30,237)	16.3 (16,771)	17.4 (28,430)
20-24	25.0 (49,390)	25.5 (47,578)	25.0 (45,347)	24.9 (25,678)	25.7 (42,035)
25-29	26.6 (52,401)	26.3 (49,090)	26.6 (48,269)	26.7 (27,522)	26.4 (43,172)
30-34	21.8 (43,089)	21.4 (39,958)	22.0 (39,859)	22.2 (22,877)	21.2 (34,631)
35+	9.8 (19,302)	9.6 (17,959)	9.9 (17,879)	9.9 (10,239)	9.4 (15,406)
Missing	(1,180)	(1,102)	(1,075)	(509)	(966)
Region at birth					
East	38.5 (76,369)	38.4 (72,051)	40.6 (74,113)	38.4 (39,761)	38.3 (63,005)
South West	44.0 (87,238)	44.6 (83,781)	44.3 (80,813)	44.8 (46,428)	44.4 (73,004)

Continued.

Table 1: Continued

	Birth sample, n = 198,483	Event sample, n = 188,010	Infant sample, n = 182,666	Toddler sample, n = 103,596	Child sample, n = 164,640
North East	9.3 (18,407)	8.9 (16,719)	7.5 (13,713)	9.3 (9,646)	9.1 (15,002)
Highlands & Islands	8.2 (16,203)	8.1 (15,264)	7.6 (13,896)	7.4 (7,702)	8.2 (13,511)
Missing	(266)	(195)	(131)	(59)	(118)
Mother's age at cohort birth					
<20	6.0 (11,891)	6.1 (11,478)	6.0 (10,872)	5.5 (5,718)	6.2 (10,155)
20-24	18.2 (36,009)	18.4 (34,519)	18.1 (32,993)	17.8 (18,422)	18.7 (30,685)
25-29	27.4 (54,062)	27.2 (50,916)	27.3 (49,595)	27.2 (28,058)	27.3 (44,799)
30-34	28.6 (56,567)	28.4 (53,149)	28.7 (52,211)	29.4 (30,387)	28.2 (46,237)
35+	19.8 (39,114)	19.9 (37,193)	19.9 (36,252)	20.0 (20,662)	19.6 (32,125)
Missing	(840)	(755)	(743)	(349)	(639)
Low birth weight					
No	93.6 (185,593)	93.6 (175,757)	93.9 (171,314)	93.7 (96,947)	93.8 (154,328)
Yes	6.4 (12,712)	6.4 (12,088)	6.1 (11,194)	6.4 (6,537)	6.2 (10,190)
Missing	(178)	(165)	(158)	(76)	(122)
Gestational age					
Pre-term	12.4 (24,559)	12.4 (23,304)	12.1 (22,018)	12.5 (12,902)	12.1 (19,938)
Term	85.1 (168,835)	85.1 (159,930)	85.4 (155,977)	85.1 (88,149)	85.4 (140,479)
Post-term	2.5 (4,980)	2.5 (4,671)	2.5 (4,571)	2.4 (2,478)	2.5 (4,138)
Missing	(109)	(105)	(100)	(67)	(85)
Mode of delivery					
Vaginal	70.2 (138,773)	70.2 (131,451)	70.1 (127,630)	69.7 (72,019)	70.3 (115,417)
Caesarean (planned)	12.3 (24,285)	12.3 (23,104)	12.3 (22,387)	12.6 (12,982)	12.2 (19,981)
Caesarean (emergency)	15.8 (31,301)	15.8 (29,605)	15.8 (28,782)	16.0 (16,522)	15.7 (25,830)
Other/unspecified	1.7 (3,419)	1.7 (3,212)	1.8 (3,231)	1.7 (1,765)	1.8 (2,864)
Missing	(705)	(638)	(636)	(308)	(548)
Apgar score					
≥7:. Normal	98.6 (193,972)	98.6 (183,737)	98.7 (178,788)	98.7 (101,346)	98.7 (161,170)
<7:. Moderate-severely depressed/resuscitated	1.4 (2,800)	1.4 (2,659)	1.3 (2,352)	1.3 (1,350)	1.3 (2,139)
Missing	(1,711)	(1,614)	(1,526)	(900)	(1,331)
HPI, first visit					
Additional support	72.5 (111,062)	72.1 (104,634)	73.3 (105,856)	70.9 (57,149)	72.5 (93,132)
Missing	(45,364)	(42,846)	(38,206)	(22,935)	(36,192)
HPI, 6-8w, check					
Additional support		56.6 (90,131)	57.1 (95,621)	55.5 (49,777)	57.2 (80,153)
Missing		(28,854)	(15,281)	(13,891)	(24,394)
HPI, 27-30m check					
Additional support				16.4 (12,719)	19.4 (14,709)
Missing				(25,878)	(88,814)
HPI, 4-5y check					
Additional support					13.7 (22,094)
Missing					(2,792)
Age (full years) at end of March 2018 (end of follow-up period)					
5 years	23.6 (46,767)	23.8 (44,675)	24.1 (43,961)	38.3 (39,721)	20.8 (34,296)
6 years	28.8 (57,089)	28.8 (54,197)	29.4 (53,677)	47.1 (48,751)	29.3 (48,284)
7 years	28.7 (56,993)	28.6 (53,747)	29.2 (53,315)	14.6 (15,124)*	29.8 (49,064)
8 years	19.0 (37,634)	18.8 (35,391)	17.4 (31,713)		20.0 (32,996)
Mean	82.6 months	82.5 months	82.2 months	74.5 months	83.4 months

SIMD: Scottish Index of Multiple Deprivation; HPI: Health plan indicator; W: Weeks; M: Months; Y: Years. Missing data reported only where present.
*Combined due to small numbers.

Table 2: Parental health and health-related behaviours, in the different analytic samples (where relevant), %(n)

	Birth sample n = 198,483/ Event sample 188,010*	**Infant sample, n = 182,666/ 173,665***	**Toddler sample, n = 103,596/ 101,406***	**Child sample, n = 164,640**
Dispensed prescriptions for mental health pre-pregnancy^				
Yes	13.6 (26,925)	13.7 (25,010)	16.2 (16,813)	13.8 (22,669)
Dispensed prescriptions for mental health (during pregnancy)				
Yes	7.5 (14,908)	7.5 (13,699)	7.9 (8,158)	7.6 (12,524)
Dispensed prescriptions for mental health (post-pregnancy [0-6 m])*				
Yes	12.3 (23,158)	12.3 (21,266)	12.3 (12,456)	12.2 (19,996)
Dispensed prescriptions for mental health (post-pregnancy [6-12 m])*				
Yes	15.4 (29,030)	15.4 (26,763)	15.6 (15,833)	15.4 (25,248)
Dispensed prescriptions for mental health (1-2 y)*				
Yes	21.1 (39,772)	21.1 (36,636)	21.3 (21,626)	21.2 (34,613)
Dispensed prescriptions for mental health (2-3y)*				
Yes	22.4 (42,101)	22.4 (38,821)	22.7 (22,986)	22.5 (36,731)
Dispensed prescriptions for mental health (3-4y)*				
Yes	23.8 (44,753)	23.8 (41,320)	24.4 (24,750)	23.9 (39,085)
Dispensed prescriptions for mental health (4-5y)*				
Yes	25.5 (47,911)	25.5 (44,279)	26.1 (26,510)	25.6 (41,900)
Maternal BMI (at antenatal booking, derived from heights and weights)**				
Underweight	2.6 (4,380)	2.6 (4,001)	2.7 (2,365)	2.6 (3,608)
Healthy	48.8 (81,368)	48.8 (74,905)	48.4 (43,164)	48.4 (67,048)
Overweight	27.7 (46,239)	27.7 (42,628)	27.8 (24,728)	27.8 (38,572)
Obesity	20.9 (34,776)	20.9 (32,123)	21.2 (18,848)	21.2 (29,432)
Missing	(31,720)	(29,009)	(14,491)	(25,980)
First feed				
Breast only	63.8 (84,753)	63.6 (77,998)	63.0 (43,930)	62.2 (68,239)
Formula only	35.1 (46,606)	35.3 (43,222)	35.9 (25,025)	36.8 (40,338)
Mixed	1.0 (1,295)	0.9 (1,145)	1.0 (675)	0.9 (1,030)
Other	0.2 (210)	0.2 (187)	0.2 (104)	0.2 (164)
Missing	(65,619)	(60,114)	(33,862)	(54,869)
Feeding on discharge				
Breast only	48.8 (90,844)	48.8 (83,749)	47.8 (47,274)	47.5 (73,366)
Formula only	44.3 (82,536)	44.4 (76,215)	44.7 (44,244)	46.2 (71,371)
Mixed	6.3 (11,470)	6.1 (10,431)	6.6 (6,553)	5.7 (8,827)
Other	0.8 (1,440)	0.8 (1,311)	0.9 (922)	0.7 (1,061)
Missing	(12,256)	(10,960)	(4,603)	(10,015)
Breastfeeding (first visit)				
Breast only	35.6 (66,594)	35.8 (63,276)	34.9 (35,101)	34.1 (53,007)
Formula only	53.5 (100,106)	53.3 (94,135)	54.0 (54,250)	55.7 (86,502)
Mixed	10.9 (20,284)	10.9 (19,155)	11.1 %(11,202)	10.3 (15,941)
Missing	(11,499)	(6,100)	(3,043)	(9,190)
Breastfeeding (6-8 w visit)				
Breast only		26.0 (46,611)	25.5 (24,773)	24.6 (36,845)
Formula only		63.8 (114,494)	64.4 (62,451)	65.8 (98,593)
Mixed		10.3 (18,450)	10.1 (9,764)	9.7 (14,519)
Missing		(3,111)	(6,608)	(14,683)
Primary immunisation status at 12 m*				
Immunised	97.9 (184,145)	98.4 (170,867)	98.7 (100,050)	98.5 (161,047)
Partially immunised	1.3 (2,488)	1.2 (2,036)	0.9 (933)	1.1 (1,867)
Unimmunised	0.7 (1,377)	0.4 (762)	0.4 (423)	0.4 (568)

Continued.

Table 2: Continued

	Birth sample n = 198,483/ Event sample 188,010*	Infant sample, n = 182,666/ 173,665*	Toddler sample, n = 103,596/ 101,406*	Child sample, n = 164,640
MMR1 status at 24 m*				
Immunised	96.4 (181,139)	96.8 (168,058)	97.6 (98,954)	97.2 (158,869)
Unimmunised	3.7 (6,871)	3.2 (5,607)	2.4 (2,452)	2.8 (4,613)
MMR2 at 60 m*				
Immunised	94.2 (177,086)	94.6 (164,240)	95.4 (96,814)	95.7 (156,393)
Unimmunised	5.8 (10,924)	5.4 (9,425)	4.5 (4,592)	4.3 (7,089)
Preschool booster, at 60 m*				
Immunised	94.9 (178,321)	95.2 (165,370)	96.1 (97,419)	96.3 (157,407)
Unimmunised	5.2 (9,689)	4.8 (8,295)	3.9 (3,987)	3.7 (6,075)
Maternal smoking in pregnancy				
Yes	19.1 (35,529)	18.9 (32,999)	19.0 (18,482)	19.6 (30,206)
Missing	(12,557)	(11,626)	(6,339)	(10,185)
Primary caregiver smoking at first visit				
Yes	17.6 (32,270)	17.4 (30,103)	17.2 (16,892)	18.1 (27,554)
Missing	(15,281)	(9,659)	(5,181)	(12,279)
Child exposed to second hand smoke at first visit				
Yes	18.5 (26,591)	18.3 (24,931)	20.3 (17,959)	18.9 (22,650)
Missing	(54,483)	(46,529)	(14,896)	(44,683)
Exposed to second-hand smoke (6–8 w)				
Yes		16.6 (22,729)	16.6 (11,860)	17.0 (19,453)
Missing		(45,754)	(32,148)	(50,164)
Primary caregiver smoking at 27–30 m				
Yes			20.8 (20,883)	21.0 (17,726)
Missing			(3,390)	(80,358)
Exposed to second-hand smoke (27–30 m)				
Yes			12.5 (12,432)	12.7 (10,669)
Missing			(4,044)	(80,907)

BMI: Body Mass Index; n: number; MMR: Measles, mumps and rubella vaccine; M: Months;, Y: Years.
*Excluding mothers who were recorded as leaving Scotland before the end of the study period.
^ $\geq$ 12 month before estimated conception.
**As measured at first booking.
Missing data reported only where present.

to experience ill health and development concerns. Child development and postnatal depression were the most unequally distributed in both relative and absolute terms. Relative and absolute inequalities tended to be greater by parents' relationship status and maternal occupation than for area-level deprivation. Inequalities according to the combined socio-economic score, which captures multiple aspects of disadvantage, were greater still. The degree of this variation differed across outcomes - potentially pointing towards the differing pathways through which SECs can influence health.

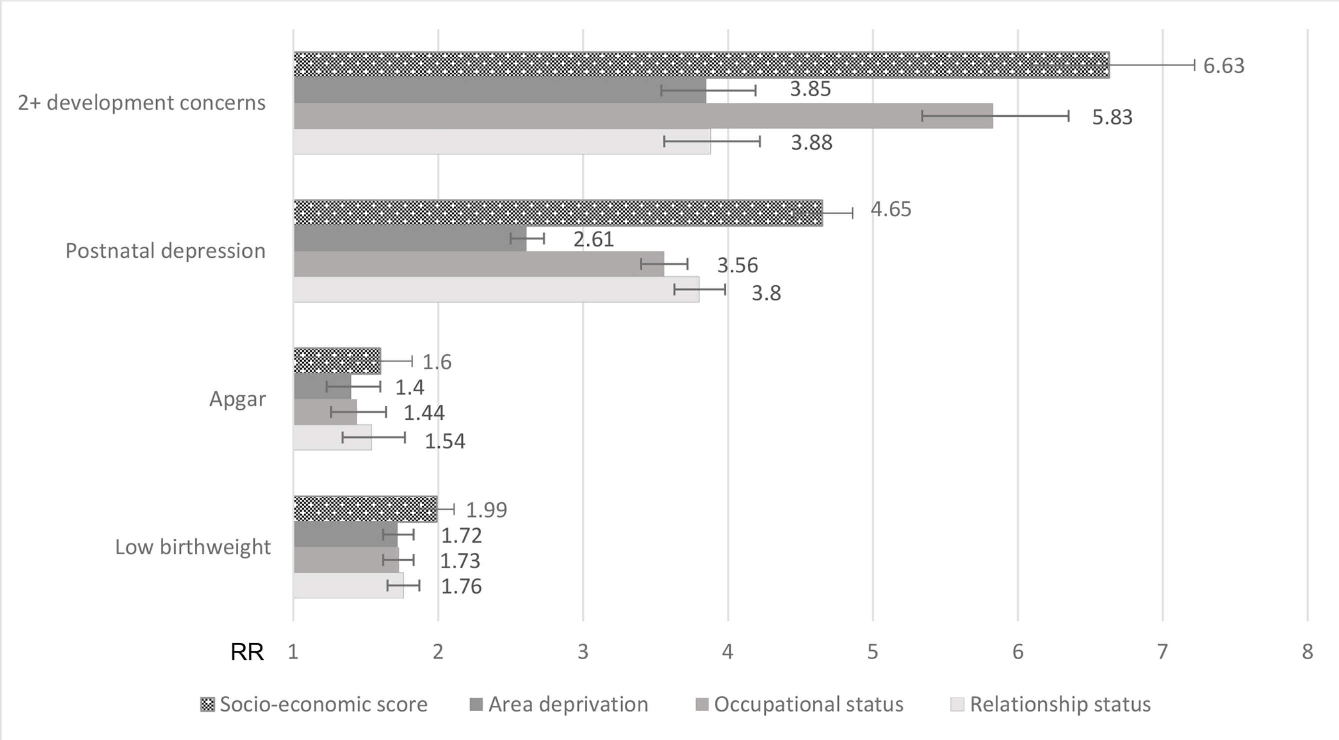
Socio-economic inequalities in longitudinal measures of maternal and child health

Table 4 presents inequalities in longitudinal health outcomes according to parental relationship status, mother's occupational class and area deprivation. Figure 4 presents estimates of relative inequality according to a socio-economic score combining the three (ranging from 0 – no adversities, to 6 – all three adversities). Underlying prevalences and numbers for Figure 4 are provided in Supplementary Appendix 2.

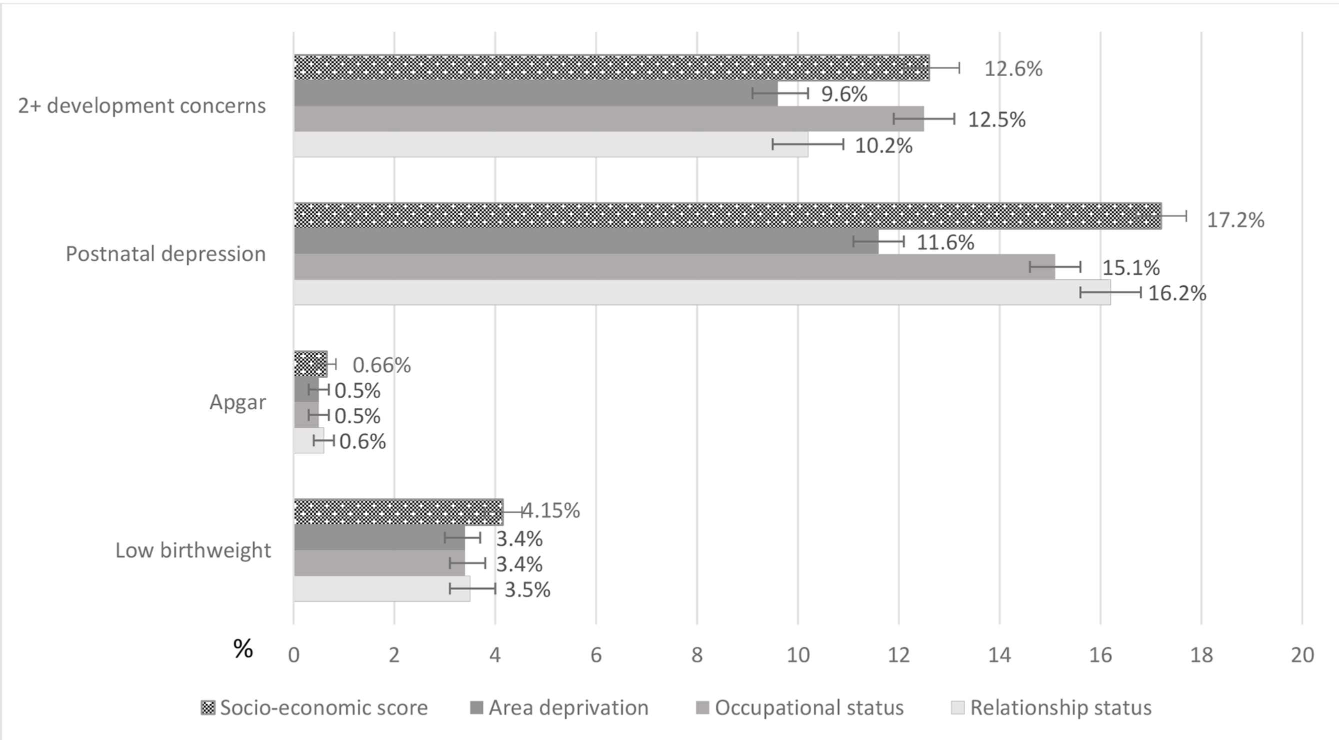
The largest inequalities were seen for tobacco smoke exposure across from pregnancy to age 27-30 months. Less than one-in-two children had not been exposed to smoke if their mother was economically inactive or if they lived in the most deprived quintile of areas, compared to around 84% in those whose mothers were from managerial and professional backgrounds or who lived in the least deprived quintile of areas. Inequalities were greater still when comparing across the socio-economic score – just 29% of children experiencing all three adversities were never exposed to tobacco smoke compared

Figure 3: Inequalities in selected outcomes according to three measures of socio-economic circumstances

a) Relative inequalities (relative index of inequality, risk ratios)



b) Absolute inequalities (slope index of inequality, risk differences)



– concerns recorded on 2 or more developmental domains vs 0-1 (Toddler sample);
– some medication dispensations for anxiety and depression vs none (Event sample);
– moderate/severe Apgar vs. normal (Birth sample);
– low birthweight vs. healthy birthweight (Birth sample).

Table 3: Infant and child health outcomes for the different samples (where relevant), % (n)

	Birth sample n = 198,483/ Event sample 188,010*	Infant sample, n = 182,666/ 173,665*	Toddler sample, n = 103,596/ 101,406*	Child sample, n = 164,640
Weight, kg (6–8 w)				
Mean (SD)		5.03 (0.73)	5.01 (0.73)	5.03 (0.73)
Missing		(4,957)	(7,773)	(16,205)
Length, cm (6-8 w)				
Mean (SD)		56.9 (2.77)	56.9 (2.74)	56.9 (2.76)
Missing		(8,918)	(9,871)	(19,445)
Motor developmental concerns at 27-30 m^				
Yes			3.4 (3,277)	2.9 (2,405)
Missing			(5,904)	(82,891)
Socio-emotional concerns at 27-30m^				
Yes			9.3 (9,172)	8.6 (7,137)
Missing			(4,696)	(81,844)
Speech development concerns at 27-30m^				
Yes			13.6 (13,508)	12.7 (10,573)
Missing			(3,971)	(81,197)
Sensory development concerns at 27-30m^				
Yes			4.0 (3,755)	3.8 (2,934)
Missing			(10,645)	(86,571)
BMI status at 27m				
Thinness			7.9 (5,804)	7.9 (5,045)
Healthy			72.1 (53,308)	72.1 (46,270)
Overweight			16.0 (11,853)	16.1 (10,361)
Obesity			4.0 (2,934)	3.9 (2,526)
Missing			(29,697)	(100,438)
BMI status at 4-5y**				
Thinness				5.5 (8,909)
Healthy				75.7 (122,975)
Overweight				13.8 (22,410)
Obesity				5.1 (8,218)
Missing				(2,128)
Any unintentional injury (age 0-4)*				
Yes	5.1 (9,570)	5.1 (8,795)	5.1 (5,187)	5.1 (8,346)
Severe injury (age 0-4)*				
Yes	0.3 (484)	0.3 (437)	0.3 (266)	0.3 (412)
Dispensed prescriptions for ADHD (age 0-7)*				
Yes	0.4 (749)	0.4 (645)	0.15 (149)	0.4 (696)

Kg: Kilograms; SD: Standard deviation; M: Months; Y: Years; BMI: Body mass index; ADHD: Attention deficit hyperactivity disorder.
*Excluding children whose mothers left Scotland before the end of the study period.
**Defined using age and sex adjusted International Obesity TaskForce cut-offs for BMI.
^Identified using approved tools which varied locally but mostly included the Ages and Stages Questionnaire and the Strengths and Difficulties Questionnaire.
Missing data reported only where present.

to 93% of those experiencing no adversities (Table Appendix 2). Whereas 41.4% of the most disadvantaged were exposed to tobacco smoke across the period, compared to 0.7% in the least disadvantaged children. These differences produced a relative difference in the likelihood of continuous smoke exposure, as opposed to never being exposed, of 141.7 (RRR

[125.9, 159.5]) when comparing the most to the least deprived groups (Figure 4).

Overall, the proportions of mothers who never breastfed were similar to proportions still breastfeeding by the 6-8 week visit (36-37%, Table 4). However, when broken down by socioeconomic circumstances, the majority of mothers in the most disadvantaged groups had never breastfed whereas the majority of women in the least disadvantaged groups breastfed to 6-8 weeks (Tables 4, Appendix 2 Table). There was a greater than forty-fold difference in the likelihood of never breastfeeding, as opposed to still breastfeeding by 6-8 weeks, when comparing the most to the least deprived groups on the combined socio-economic score (RRR: 41.7 [39.8, 43.8], Figure 4).

Over half (58%) of mothers had no recorded dispensed prescriptions for depression and anxiety across the period under study (from pre-pregnancy until 5 years), with 12% from infancy/toddlerhood onwards, and 7.6% across the majority of the time period. The biggest inequalities were seen for more persistent durations of dispensed prescriptions. Mothers experiencing all three aspects of deprivation were ten times more likely to have dispensed medications for anxiety and depression across most or the whole period, as opposed to never, when compared to mothers not experiencing any aspects of deprivation (10.3 [9.6,11.0], Figure 4).

The majority (94%) of children were fully immunised with the relevant immunisations by the end of the study period. Very few children were completely unimmunised ($<$1% across all social groups) (Table 4). However, there were relatively large inequalities in the proportions partially immunised. Children from the most disadvantaged groups (based on the socio-economic score) were three times as likely to be partially, as opposed to fully, immunised than their more advantaged peers (RRR 3.0 [2.8, 3.3]) (Figure 4).

Across all socio-economic measures, children from more disadvantaged backgrounds were more likely to experience overweight (including obesity) at age 27–30 months and 4–5 years, or to have become overweight. Those from more advantaged backgrounds were most likely to have been healthy weight at both time points (Table 4). Children with the highest levels of adversity on the combined socio-economic score were around twice as likely to experience overweight at both time points (RRR: 1.9 [1.7, 2.1]) or to have become overweight (RRR: 2.1 [1.9, 2.3]) (Figure 4).

Inequalities in the number of unintentional injuries experienced across the early years were also present. The biggest differences were seen when comparing those who experienced all three socio-economic adversities (9.4% injured at least once) to those who experienced none (4.8% injured at least once, Appendix 2 Table). In relative terms, children living in the most deprived socio-economic circumstances were twice as likely to have been hospitalised for an unintentional injury once (RRR: 1.9 (1.7, 2.0), and three times as likely to have been injured twice or more (RRR: 3.3 [2.5, 4.2]).

Despite the five samples appearing relatively representative, bias may be introduced when analysing variables which draw on information across multiple samples and will therefore have even higher levels of missingness. Sensitivity analyses using imputed data for the outcomes most susceptible to missingness showed extremely similar patterning of results to those reported here (see Supplementary Appendix 3).

Discussion

Summary of findings to date

Through linking different administrative records across the early years period, we have shown socio-economic inequalities in health at the family-level as well as using more commonly available measures of neighbourhood deprivation. This shows that for many aspects of health, inequalities by area-level deprivation are smaller than those by family-level socio-economic circumstances. Linking across multiple time points provides deeper insights. We observed a large drop off in breastfeeding and extreme levels of socio-economic inequality in both continuous tobacco smoke exposure in children and dispensation of medications for anxiety and depression in mothers. The large samples also afford the documentation of inequalities in very rare or severe outcomes, such as being hospitalised multiple times for unintentional injuries.

Different analytic samples are used for different outcomes, although comparisons of their characteristics and imputation for outcomes relying on multiple samples suggest that representativeness is maintained. If attrition across the different cohorts is indicative of levels of engagement with child health services, then these findings may suggest that inequalities in engagement throughout the early years are relatively small, although this requires direct investigation before drawing conclusions.

Published analyses of these data that have preceded this cohort profile paper include an examination of inequalities in hospitalisations for unintentional injuries [40]. The data allowed for comparisons across a range of outcomes including injury type, severity and frequency, the ages at which injuries occurred and whether the injury was recorded as occurring at home or elsewhere. In this comprehensive analysis, a reverse social gradient in unintentional injuries during infancy was identified, when measured according to area deprivation (but not family-level SECs, where typical social gradients were observed). This Scottish administrative cohort has also been used to investigate inequalities in ADHD-relevant symptoms and dispensed prescriptions of ADHD medications in more depth. Children from less-advantaged backgrounds were more likely to experience both of these outcomes. Additionally, among those with ADHD-relevant symptoms, the least advantaged were three times more likely to go on to have ADHD medications dispensed, adding to our understanding of how inequalities manifest along the ‘patient journey’ [41]. An indirect comparison with ADHD inequalities in the UK Millennium Cohort Study highlighted how the administrative cohort provided greater power to look at this relatively rare outcome. Inequalities appeared greater than in the cohort, potentially due to better representation or more objective measures. On the other hand, the cohort survey offered greater granularity of the experiences of families across childhood, with multiple reports of children’s symptoms and the daily impacts of these, from the parent’s perspective [41].

A scoping review of child health inequalities in the UK compared the use of social disadvantage measures across

Table 4: Inequalities in longitudinal outcomes (most and least disadvantaged groups only), complete case samples, %(n)

	Total	Relationship status of parents		NS-SEC of mother		SIMD	
	(n)	Sole	Married	Economically inactive	Managerial & professional	Quintile 1	Quintile 5
Exposure to smoking [1], n = 57,910							
None over period	66.2 (38,324)	36.8 (951)	82.7 (23,522)	45.9 (5,250)	84.3 (15,857)	49.3 (7,614)	84.9 (7,735)
Pregnancy only	4.3 (2,504)	9.7 (250)	2.1 (596)	6.4 (732)	2.3 (423)	6.1 (947)	1.9 (177)
Post birth	15.7 (9,073)	18.3 (473)	11.2 (3,171)	19.8 (2,267)	9.8 (1,815)	19.8 (3,055)	9.7 (881)
Pregnancy & post-birth	13.8 (8,009)	35.3 (912)	4.0 (1,143)	27.9 (3,185)	3.6 (676)	24.7 (3,817)	3.5 (318)
Breastfeeding [2], n = 169,745							
Never	36.8 (62,441)	61.4 (5,375)	22.7 (18,821)	49.3 (17,742)	18.5 (9,891)	53.2 (23,812)	17.7 (4,637)
Initiated only	15.4 (26,164)	15.2 (1,330)	13.7 (11,405)	13.7 (4,915)	14.0 (7,457)	15.6 (6,964)	13.3 (3,510)
Breastfeeding, 1st visit	11.5 (19,442)	9.0 (790)	12.1 (10,027)	9.8 (3,525)	12.4 (6,639)	9.6 (4,283)	12.5 (3,294)
Breastfeeding, 6-8w	36.4 (61,698)	14.4 (1,260)	51.5 (42,806)	27.2 (9,782)	55.1 (29,463)	21.7 (9,706)	56.5 (14,925)
Dispensed prescriptions for anxiety/depression in the mother [3], n = 188,010							
Never	58.3 (109,516)	38.2 (3,819)	67.7 (60,908)	48.1 (19,043)	69.8 (40,576)	49.1 (24,475)	70.0 (20,059)
Pre-preg-toddler only	8.0 (15,099)	9.8 (984)	6.6 (5,906)	9.1 (3,621)	6.3 (3,669)	9.3 (4,641)	6.3 (1,792)
Childhood only	9.3 (17,545)	10.7 (1,070)	8.3 (7,463)	10.1 (3,981)	8.0 (4,641)	10.3 (5,125)	8.1 (2,335)
Infant-toddler onwards	11.8 (22,089)	17.4 (1,736)	9.1 (8,138)	14.3 (5,674)	8.3 (4,849)	14.3 (7,132)	8.5 (2,434)
Pre-preg + relapse	5.1 (9,520)	8.3 (832)	3.3 (2,989)	6.8 (2,689)	3.1 (1,827)	6.7 (3,313)	2.9 (826)
Across most or whole period	7.6 (14,241)	15.6 (1,563)	5.0 (4,530)	11.6 (4,600)	4.4 (2,545)	10.3 (5,101)	4.3 (1,228)
Immunisation status by end of study period [1], n = 188,010							
Fully immunised	93.6 (175,940)	90.2 (9,026)	94.3 (84,823)	89.7 (35,537)	95.4 (55,442)	92.4 (46,010)	94.9 (27,217)
Partially immunised	5.8 (10,860)	8.9 (886)	4.9 (4,445)	9.4 (3,728)	4.0 (2,298)	7.1 (3,525)	4.4 (1,267)
Unimmunised	0.6 (1,210)	0.9 (92)	0.7 (666)	0.9 (343)	0.6 (367)	0.5 (252)	0.7 (190)
Experiences of overweight (& obesity) at 27-30m and 4-5y [4], n = 63,222							
Healthy at both	71.1 (44,977)	65.4 (1,963)	73.5 (23,060)	69.1 (8,083)	73.5 (15,459)	69.1 (10,309)	74.1 (7,450)
Became overweight	8.9 (5,595)	11.8 (354)	7.8 (2,437)	9.7 (1,138)	7.6 (1,598)	10.5 (1,568)	6.7 (671)
Became healthy	9.9 (6,232)	10.3 (309)	9.8 (3,069)	10.0 (1174)	10.2 (2,139)	9.1 (1,355)	11.1 (1,117)
Overweight at both	10.2 (6,418)	12.6 (378)	9.0 (2,826)	11.2 (1,310)	8.7 (1,825)	11.3 (1,688)	8.1 (813)
Hospitalisation for unintentional injury, number [1], n = 188,010							
0	93.7 (176,220)	91.7 (9,173)	94.6 (85,102)	92.4 (36,597)	94.7 (55,027)	92.7 (46,136)	94.8 (27,173)
1	5.9 (11,016)	7.6 (763)	5.1 (4,560)	7.0 (2,786)	5.0 (2,910)	6.8 (3,396)	5.0 (1,427)
2+	0.4 (774)	0.7 (68)	0.3 (272)	0.6 (225)	0.3 (170)	0.5 (255)	0.3 (74)

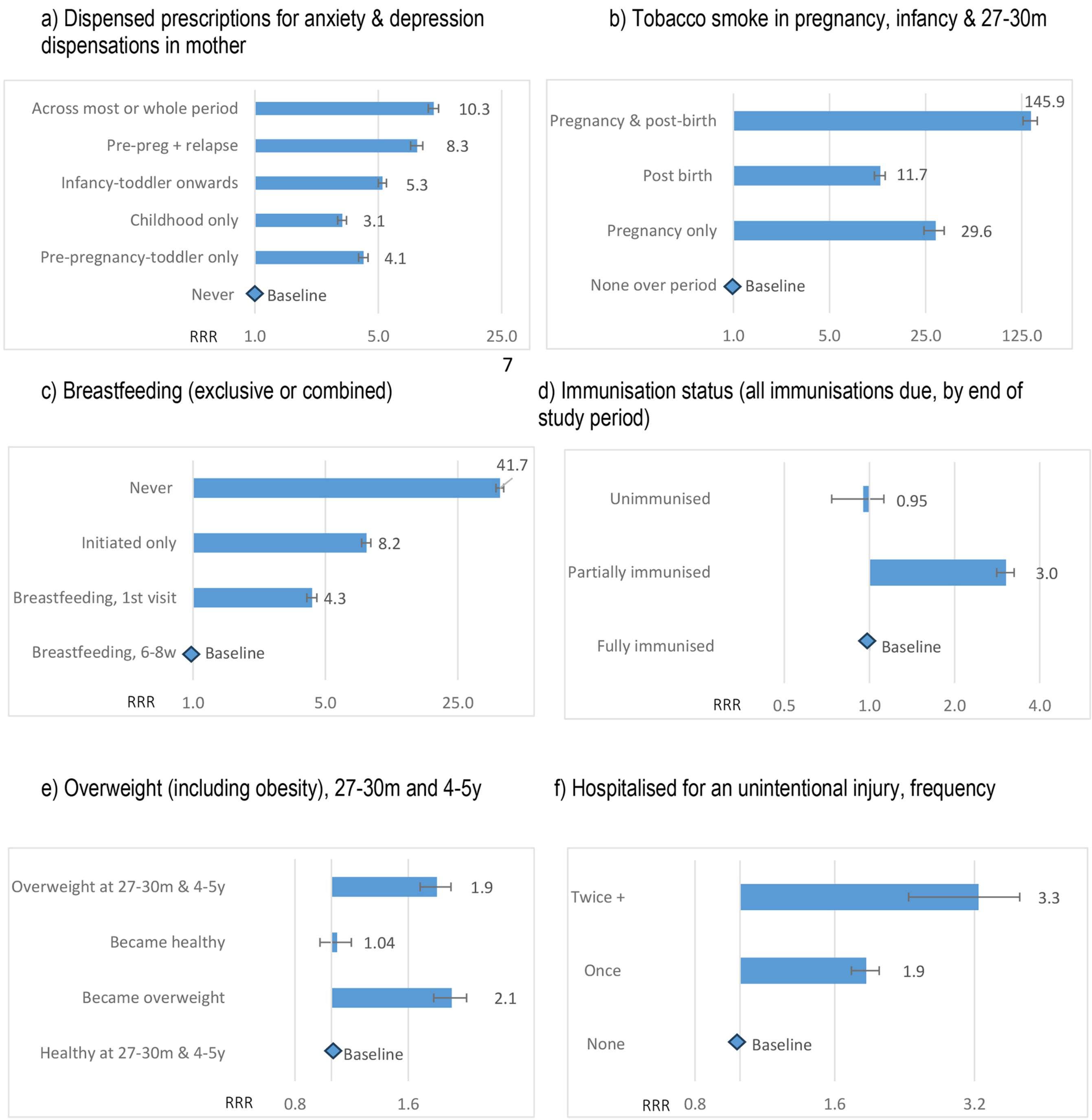
NS-SEC: National Statistics Socio-economic Classification; SIMD: Scottish Index of Multiple Deprivation; N: Number; M: Month; Y: Year; Preg: Pregnancy.
Requires outcome data from: [1] the birth, infant and toddler samples; [2] the birth and infant samples; [3] the event- based sample; [4] the toddler and child sample.

research in surveys, cohorts and administrative datasets. It found that area-level measures of disadvantage underestimate the impacts of social disadvantage [22]. Since research using administrative data is most likely to rely on these area-level measures, the authors call for better access to individual-level measures in administrative data. In the current study we have confirmed this to be the case – area-level measures of SECs can provide only part of the picture. To our knowledge, this is the first comprehensive description of child health inequalities provided from linked administrative data in the UK, although there are other platforms that could have the potential to do this [49–52].

Strengths and limitations

This is the first population-wide cohort created from administrative records in Scotland, currently following children up until 6 years, with potential to extend this as they age. Through linking across records, and specifically to occupational and relationship status of parents at birth,

Figure 4: Relative risk ratios (RRRs) for selected longitudinal health-related outcomes, according to a combined socio-economic score (comparing the notionally most to the notionally least deprived, using the relative index of inequality). (sample sizes correspond to Table 4, shown on the log scale)



W: Week; M: Month; Y: Year.

we have been able to extend our knowledge about health inequalities in Scottish children at the family-level. These findings show that inequalities, across a range of health outcomes, do vary according to the socio-economic measure under study and that area-level measures, such as the Scottish Index of Multiple Deprivation, can in some cases underestimate inequalities experienced across family socio-economic circumstances. Nevertheless, the family-level SEC measures may also underestimate inequalities; for example, because the NS-SEC economically inactive group combines students with long-term unemployed groups.

There are numerous, widely acknowledged strengths of administrative data relating to costs, burden on research participants, objective measurements, and greater statistical power. Here we focus on some of the particular strengths for life course and health inequalities research. First, administrative data cover almost the whole of the population and do not suffer from the same degree of loss-to-follow-up that cohort surveys do. This is especially important for understanding inequalities in health since the most disadvantaged and sickest in society are often under-represented in household surveys. Second, linkages can go back as far as the administrative data exist; they can be used to research the effects of pre-pregnancy exposures on child health and childhood exposures on healthy ageing, for example. Third, they provide the statistical power and time series data required to examine the impacts of historical policies on outcomes and, importantly, inequalities in those outcomes.

Administrative data are open to biases occurring at the recording/registration stage, during linkage, and in the coding, cleaning, and categorisation of variables [53]. Here we briefly consider those most likely to introduce bias in the case of this administrative, child health cohort.

The later child health checks had the lowest linkage rates, of around 83–85%. Not all children receive these health checks in Scotland, but rates were potentially lower than expected based on national reports of coverage. Generally speaking, the characteristics of these children were comparable to the base sample, although ethnic minority groups and mothers born outside Scotland were potentially under-represented. Due to changes in the child health system programme, only cohort children born from January 2011 were eligible to be included in the 27-30 month check, necessitating a separate analytic sample for research questions focussed on exposures or outcomes at this point in the early years. This highlights one of the generic shortcomings of administrative data – the questions researchers can answer, and how well, are bounded by service structures and procedures.

Related to the above, the variables available were not designed for research purposes, potentially introducing measurement error or requiring researchers to ask slightly different questions from those they would preferably investigate. For example, in the current paper, we have examined inequalities in dispensed ADHD medications in children. Ideally, we would have also been able examine ADHD diagnosis and alternative treatments. Similarly, we have only been able to explore unintentional injuries which ended in a hospital admission and the propensity of a child to be admitted following an injury may vary by hospitals, health care professionals and family characteristics. Mother's age at first live birth was estimated from previous birth records. This is imperfect and should be treated with caution — especially for non-Scottish ethnic groups (where the mother is potentially more likely to have had children in other countries). For some outcomes, which relied on an event being recorded, such as an immunisation or a hospitalisation, the absence of a record was assumed to represent an absence of the outcome itself. While we used the CHI record to identify mothers who had been removed from the register (for example because they had moved outside Scotland), this is an imperfect system.

This bespoke cohort was created to understand socio-economic inequalities in children and their mothers throughout the early years of childhood and excludes pregnancies that ended in a spontaneous early pregnancy loss or termination. Examination of these outcomes has recently been made possible by the establishment of the Scottish Linked Pregnancy and Baby Dataset (SLiPBD) [49], which links birth and maternity hospital records with termination of pregnancy notifications, statutory stillbirth registrations and infant mortality data. The availability of SLiPBD will make the replication of the linkages reported in this paper easier and could help to address some of the limitations. For example, it will provide more reliable measures of some variables, such as parity, and identification of attrition due to infant mortality.

The timing of measurements can also be restricted. For example, BMI is last routinely measured at the start of school in Scotland, meaning we cannot investigate overweight into mid-childhood and adolescence in administrative data. Mothers and children are linked via hospital birth records, whereas linkage to father's own health records is not currently possible. Therefore, we are limited to information contained in birth certificates (e.g. father's occupational status) and even this information is not available for sole registrations by mothers.

There were very high levels of missing ethnicity data and these are unlikely to be missing at random or completely at random, although changes to recording guidance and requirements in Scotland since are leading to more complete data [54]. Finally, this cohort cannot be used to examine the health of children not born in Scotland. Children born outside Scotland are captured by child health checks, hospital and immunisation records, and prescribing data, but we only have birth information for children born in Scotland and only these children can be linked to their mothers. Therefore, the health of the cohort children described here will deviate from their wider peer group and increasingly so as they get older.

Future plans and potential

Next steps for our research using this cohort include using the wide range of variables available to consider how well we can predict which children are most likely to go on to experience ill health in Scotland and what this might mean for targeting of future policies and interventions. We plan to carry out causal mediation analysis to consider the extent to which intervening on amenable pathways (such as maternal mental health or parental health behaviours) could reduce child health inequalities, in contrast to more upstream approaches which focus on altering families' socio-economic circumstances. Beyond our planned work, there is realistic potential to extend these linkages across multiple cohorts, into childhood and adolescence, to other family members (siblings and grandmothers), and to administrative data from other sectors such as education and employment. If such linkages were to become routine, then this would offer huge potential for evaluating natural policy experiments, allowing us to build highly sought-after evidence around the policies and interventions which are likely to be most successful in reducing health inequalities but are the hardest to evaluate since they cannot be tested in trial settings [55].

Data access

All of the datasets used in this project are accessible to researchers via the access procedures explained previously. At the time of writing fees ranged from £5,000-55,000 depending on the sector (NHS, Academia/Charities, Commercial/Industry) and the study size and complexity. More information on the datasets and application procedures can be found on the Public Health Scotland website: https://publichealthscotland.scot.

Conclusion

This is Scotland's first 'administrative child cohort', following all children (and their mothers), born in Scotland Oct 2009-end of March 2013, up until the average age of 6 years ($n \sim 200{,}000$). It harnesses birth registrations and various health records, linked across the early years, to consider multiple aspects of social disadvantage and how health changes and accumulates across the early years. Analyses presented here add to our understanding of health inequalities, showing that routine reports of inequalities which rely on area-level deprivation and snapshots of health may underestimate the detrimental impacts of social disadvantage on child and parental health in Scotland. Future work should extend linkages to examine health across the entire life course and to evaluate the impacts of upstream natural policy experiments on health inequalities and life chances.

Acknowledgements

The authors would like to acknowledge the support of the eDRIS Team (Public Health Scotland) for their involvement in obtaining approvals, provisioning and linking data and the use of the secure analytical platform within the National Safe Haven. We'd also like to extend our thanks to Rachael Wood, at Public Health Scotland, for her advice during the data application stage and when interpreting of some aspects of the data.

Ethics

The eData Research and Innovation Service (eDRIS) within Public Health Scotland provided access to the data, after ethical and data control criteria had been satisfied via a Public Benefit and Privacy Panel for Health and Social Care (PBPP) application (ref: 1617-0152). No further ethical approval was required for the analyses reported in this paper.

Data availability

All of the datasets used in this project are accessible to researchers via access procedures described previously. Code used to link the datasets and create the variables described in this paper can be made available upon request.

Funding

PMH and AP received funding from a Wellcome Trust Award to AP [205412/Z/16/Z]. PMH, RD, SVK, AL, and AP were supported by funds from the Medical Research Council [MC_UU_12022/2] and the Scottish Government Chief Scientist Office [SPHSU17]. AP, RD, AL receive support from the UK Prevention Research Partnership Maternal and Child Health Network (MR/S037608/1).

Conflicts of interest

The authors have no conflicts of interest to declare except for the funding noted above.

Author contributions

All authors contributed to the design of the study. PMH cleaned and structured the data; PMH and AP analysed the data. All authors provided advice on the analysis and interpreted the data. AP and PMH drafted the manuscript. All authors critically revised the manuscript. AP obtained funding for the study.

Rights retention statement

References

1. Center on the Developing Child at Harvard University. The Foundations of Lifelong Health Are Built in Early Childhood. Harvard University: Center on the Developing Child; 2010. https://developingchild.harvard.edu/resources/the-foundations-of-lifelong-health-are-built-in-early-childhood/.

2. Allen G. Early Intervention: Smart Investment, Massive Savings. The Second Independent Report to Her Majesty's Government London: HM Government; 2011. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/61012/earlyintervention-smartinvestment.pdf.

3. Power C, Atherton K, Strachan DP, Shepherd P, Fuller E, Davis A, et al. Life-course influences on health in British adults: effects of socio-economic position in childhood and adulthood. International journal of epidemiology. 2007;36:532–9.

4. Law C. Lifecourse influences on children's futures. In: Graham H, editor. Understanding Health Inequalities: Open University Press; 2009.

5. Heckman JJ, Stixrud J, Urzua S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics. 2006;24(3):411–82. https://doi.org/10.1086/504455

6. Leadsom A, Field F, Burstow P, Lucas C. 1001 Critical Days: The Importance of the Conception to Age Two Period. A cross party manifesto Wave Trust; 2014. https://www.wavetrust.org/1001-critical-days-the-importance-of-the-conception-to-age-two-period.

7. Howe LL, DA; Propper,C. Trajectories of socioeconomic inequalities in health, behaviours

and academic achievement across childhood and adolescence. Journal of Epidemiology and Community Health. 2013;Apr;67(4):358–64. https://doi.org/10.1136/jech-2012-201892

8. Nicholson J, Lucas,N, Berthelsen,D, Wake,M. Socioeconomic inequality profiles in physical and developmental health from 0-7 years: Australian National Study. Journal of Epidemiology and Child Health. 2010;Jan;66(1). https://doi.org/10.1136/jech.2009.103291

9. Scottish Government. Getting it right for every child (GIRFEC): policy statement. Scottish Government; 2022. https://www.gov.scot/publications/getting-right-child-girfec-policy-statement/documents/.

10. Rougeaux E, Hope S, Law C, Pearce A. Have health inequalities changed during childhood in the New Labour generation? Findings from the UK Millennium Cohort Study. BMJ Open. 2017;7(1):e012868. https://doi.org/10.1136/bmjopen-2016-012868

11. Wickham S, Whitehead M, Taylor-Robinson D, Barr B. The effect of a transition into poverty on child and maternal mental health: a longitudinal analysis of the UK Millennium Cohort Study. The Lancet Public Health. 2017;2(3):e141–e8.

12. Lai ETC, Wickham S, Law C, Whitehead M, Barr B, Taylor-Robinson D. Poverty dynamics and health in late childhood in the UK: evidence from the Millennium Cohort Study. Archives of Disease in Childhood. 2019;104(11):1049–55. https://doi.org/10.1136/archdischild-2018-316702

13. Hawkins SS, Cole TJ, Law C, the Millennium Cohort Study Child Health G. An ecological systems approach to examining risk factors for early childhoood overweight: findings from the UK Millennium Cohort Study. Journal of Epidemiology and Community Health. 2008;63:147–55.

14. Caryl F, Shortt NK, Pearce J, Reid G, Mitchell R. Socioeconomic inequalities in children's exposure to tobacco retailing based on individual-level GPS data in Scotland. Tobacco Control. 2020;29(4):367–73. https://doi.org/10.1136/tobaccocontrol-2018-054891

15. Pearce A, Marshall H, Bedford H, Lynch J. Barriers to childhood immunisation: Findings from the Longitudinal Study of Australian Children. Vaccine. 2015;33(29):3377–83. https://doi.org/10.1016/j.vaccine.2015.04.089

16. Fierloos IN, Windhorst DA, Fang Y, Bannink R, Stam M, Slijkerman CAA, et al. A prospective study on the association between social support perceived by parents of children aged 1-7 years and the use of community youth health care services. Front Public Health. 2022;10:950752. https://doi.org/10.3389/fpubh.2022.950752

17. Gialamas A, Mittinty MN, Sawyer MG, Zubrick SR, Lynch J. Social inequalities in childcare quality and their effects on children's development at school entry: findings from the Longitudinal Study of Australian Children. Journal of Epidemiology and Community Health. 2015. https://doi.org/10.1136/jech-2014-205031

18. Geoffroy MC, Côté SM, Giguère C, Dionne G, Zelazo PD, Tremblay RE, et al. Closing the gap in academic readiness and achievement: the role of early childcare. J Child Psychol Psychiatry. 2010;51(12):1359–67. https://doi.org/10.1111/j.1469-7610.2010.02316.x

19. Delpierre C, Kelly-Irving M. Big Data and the Study of Social Inequalities in Health: Expectations and Issues. Front Public Health. 2018;6:312. https://doi.org/10.3389/fpubh.2018.00312

20. Gorman E, Leyland AH, McCartney G, Katikireddi SV, Rutherford L, Graham L, et al. Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. Addiction. 2017;112(7):1270–80. https://doi.org/10.1111/add.13797

21. Stewart E, Brophy S, Cookson R, Gilbert R, Given J, Hardelid P, et al. Using administrative data to evaluate national policy impacts on child and maternal health: a research framework from the Maternal and Child Health Network (MatCHNet). Journal of Epidemiology and Community Health. 2023;77(11):710–3. https://doi.org/10.1136/jech-2023-220621

22. Clery A, Grant C, Harron K, Bedford H, Woodman J. Measuring disadvantage in the early years in the UK: A systematic scoping review. SSM Popul Health. 2022;19:101206. https://doi.org/10.1016/j.ssmph.2022.101206

23. Miall N, Fergie G, Pearce A. Health Inequalities in Scotland: trends in deaths, health and wellbeing, health behaviours, and health services since 2000. . https://gla.ac.uk/healthinequalities2022: University of Glasgow; 2022.

24. Elstad JI, Heggebø K, Dahl E. Nordic research on health inequalities: A scoping review of empirical studies published in Scandinavian Journal of Public Health 2000–2021. Scandinavian Journal of Public Health. 2022;50(7):843–51. https://doi.org/10.1177/14034948221101304

25. Royal College of Paediatrics and Child Health. State of Child Health 2020. Royal College of Paediatrics and Child Health; 2020. https://stateofchildhealth.rcpch.ac.uk/wp-content/uploads/sites/2/2020/03/SOCH-ENGLAND-02.03.20.pdf.

26. National Records of Scotland. Births Time Series Data. Table BT.3: Births registered in Scotland, by month of registration, 1954-2022. Edinburgh: National Records of Scotland. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/births/births-time-series-data.

27. Sengoelge M, Leithaus M, Braubach M, Laflamme L. Are There Changes in Inequalities in Injuries? A Review of Evidence in the WHO European Region. International Journal of Environmental Research and Public Health. 2019;16(4):653. https://www.mdpi.com/1660-4601/16/4/653.

28. Russell A, Ford T, Williams R, Russell G. The Association Between Socioeconomic Disadvantage and Attention Deficit/Hyperactivity Disorder (ADHD): A Systematic Review. Child Psychiatry Hum Dev. 2016;47(3):440–58. https://doi.org/10.1007/s10578-015-0578-3

29. Wright N, Moldavsky M, Schneider J, Chakrabarti I, Coates J, Daley D, et al. Practitioner Review: Pathways to care for ADHD - a systematic review of barriers and facilitators. J Child Psychol Psychiatry. 2015;56(6):598-617. https://doi.org/10.1111/jcpp.12398

30. Research Data Scotland. https://www.researchdata.scot/ 2024 [Available from: https://www.researchdata.scot/.

31. Health Data Research UK. Gateway to health data and tools for research https://www.healthdatagateway.org/2024.

32. University of Glasgow. Could you be a Data Detective? https://datadetective.sphsu.gla.ac.uk/2021 [Available from: https://datadetective.sphsu.gla.ac.uk/.

33. Pearce A, Dundas R, Whitehead M, Taylor-Robinson D. Pathways to inequalities in child health. Archives of Disease in Childhood. 2019;104(10):998–1003. https://doi.org/10.1136/archdischild-2018-314808

34. Scottish Government. Scottish Government. SIMD 2020 technical notes. Edinburgh: Scottish Government; 2020.

35. Scottish Government. Long-term Monitoring of Health Inequalities March 2023 report. Edinburgh: Scottish Government; 2023.

36. Scottish Government. Growing up in Scotland: health inequalities in the early years. Edinburgh: Scottish Government; 2010.

37. Public Health Scotland. Child poverty in Scotland: priority groups – lone-parent families. Edinburgh: Public Health Scotland; 2020.

38. Pearce A, Lewis H, Law C. The role of poverty in explaining health variations in 7-year-old children from different family structures: findings from the UK Millennium Cohort Study. Journal of Epidemiology and Community Health. 2013;67(2):181–9. https://doi.org/10.1136/jech-2012-200970

39. Crawford C, Goodman A, Greaves E. Cohabitation, marriage, relationship stability and child outcomes. IFS Report; 2013. Report No.: 190946323X,

40. Henery PM, Dundas R, Katikireddi SV, Leyland A, Wood R, Pearce A. Social inequalities and hospital admission for unintentional injury in young children in Scotland: A nationwide linked cohort study. Lancet Reg Health Eur. 2021;6:100117. https://doi.org/10.1016/j.lanepe.2021.100117

41. Pearce A, Henery P, Katikireddi SV, Dundas R, Leyland AH, Nicholls D, et al. Childhood attention-deficit hyperactivity disorder: socioeconomic inequalities in symptoms, impact, diagnosis and medication. Child Adolesc Ment Health. 2024. https://doi.org/10.1111/camh.12707

42. Fairley L, Leyland AH. Social class inequalities in perinatal outcomes: Scotland 1980–2000. Journal of Epidemiology and Community Health. 2006;60(1):31–6. https://doi.org/10.1136/jech.2005.038380

43. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes. 2012;7(4):284-94. https://doi.org/10.1111/j.2047-6310.2012.00064.x

44. Public Health Scotland. Body Mass Index of Primary 1 Children in Scotland: School Year 2019/20. Edinburgh: Public Health Scotland; 2021.

45. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology: Glossary. Journal of Epidemiology and Community Health. 2003;57:778-83.

46. Mackenbach J, Kunst A. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Social Science and Medicine. 1997;44(6):757–71.

47. Information Services Division. Child Health 27-30 Month Review Statistics: Technical Report, Scotland 2016/17. Edinburgh: NHS National Services Scotland; 2018.

48. Scotland's Census 2011. Table DC2103SC - Country of birth by age. All people. Scottish Council Area 2011 by Age by Country of birth (Flat) by Term-time Address (Indicator). Edinburgh: National Records of Scotland, https://www.scotlandscensus.gov.uk/webapi/jsf/tableView/tableView.xhtml.

49. Lindsay L, Mark K, Moore E, Carruthers J, Hopkins L, Jennings D, et al. Data resource profile: Scottish Linked Pregnancy and Baby Dataset (SLiPBD). International Journal of Population Data Science. 2024;9(2). https://doi.org/10.23889/ijpds.v9i2.2390

50. Allik M, Brown D, Taylor Browne Lūka C, Macintyre C, Leyland AH, Henderson M. Cohort profile: The 'Children's Health in Care in Scotland' (CHiCS) study—a longitudinal dataset to compare health outcomes for care experienced children and general population children. BMJ Open. 2021;11(9):e054664. https://doi.org/10.1136/bmjopen-2021-054664

51. SAIL Databank. SAIL Databank is a rich and trusted population databank: SAIL Databank; 2024 [Available from: https://saildatabank.com/.

52. London UC. ECHILD: Education and Child Health Insights from Linked Data University College London 2024 [Available from: https://www.echild.ac.uk/].

53. Shaw RJ, Harron KL, Pescarini JM, Pinto Junior EP, Allik M, Siroky AN, et al. Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses. European Journal of Epidemiology. 2022;37(12):1215–24. https://doi.org/10.1007/s10654-022-00934-w

54. ScotPHO. Ethnic minorities: key data sources Edinburgh: Public Health Scotland; 2023 [Available from: https://www.scotpho.org.uk/population-groups/ethnic-minorities/key-data-sources/].

55. Stewart E, Pearce A, Given J, Gilbert R, Brophy S, Cookson R, et al. Identifying opportunities for upstream evaluations relevant to child and maternal health: a UK policy-mapping review. Archives of Disease in Childhood. 2023;108(7):556–62. https://doi.org/10.1136/archdischild-2022-325219

Abbreviations

ADHD: Attention deficit hyperactivity disorder
BMI: Body Mass Index
CHSP: Child Health Systems Programme
CHI: Community Health Index
eDRIS: eData Research and Innovation Service
HPI: Health Plan Indicator
ID: Identification
Kg: Kilograms
MMR: Measles, Mumps and Rubella
MMR1: Measles, Mumps and Rubella first dose
MMR2: Measles, Mumps and Rubella second dose
M: Months
NRS: National Records of Scotland
NS-SEC: National Statistics Socio-economic Classification
n: Number
PIS: Prescribing Information System
PBPP: Public Benefit and Privacy Panel for Health and Social Care
PHS: Public Health Scotland
RII: Relative index of inequality
RRR: Relative risk ratio
RR: Risk Ratio
SBR: Scottish Birth Record
SIRS: Scottish Immunisation and Recall System
SIMD: Scottish Index of Multiple Deprivation
SLiPBD: Scottish Linked Pregnancy and Baby Dataset
SMR01: Scottish Morbidity Record 01
SMR02: Scottish Morbidity Record 02
SD: Standard deviation
SII: Slope index of inequality
SECs: Socio-economic circumstances
UK: United Kingdom
W: Weeks
Y: Years

Supplementary Appendix 1: Description of the datasets, main variables and coverage of the administrative cohort

Table Appendix 1: Description of the datasets, main variables and coverage of the administrative cohort

Dataset	Brief Description	Main variables	Period of data inclusion
National Records of Scotland (NRS) Births	Registration of all births registered in Scotland since 1975	• Scottish Index of Multiple Deprivation (SIMD) • National Statistics Socio-economic Classification (NS-SEC) of the mother and father (where applicable) • Relationship status of the parents	Births between October 2009 and March 2013
Scottish Morbidity Record 02 (SMR02)	All maternity and infant inpatient and day case episodes in Scotland. Around 50% episodes relate to births and it was these records that were requested for the purposes of the cohort. Includes data collected during the ante-natal booking (∼8–12 weeks of pregnancy)	• Child sex • Ethnicity • Number of births this pregnancy • Mother's country of birth • Mother's age (in years) at cohort birth • Mother's height and weight at booking • Smoking in pregnancy (never, former, current) • Number of previous live births • Delivery mode • Apgar score • Birthweight • Gestational age (weeks) • Breastfeeding initiation and infant feeding upon discharge • Health Board • Local authority • Scottish Index of Multiple Deprivation (SIMD)	Hospital births between October 2009 and March 2013
The Scottish Birth Record (SBR)	A web-based system used to record all infant neonatal care in Scotland, including home births. Used to supplement SMR02 (hospital births).	As for SMR02 above, only for home births	Home births between October 2009 and March 2013
Scottish Morbidity Record 01 (SMR01)	Episode-based patient record for all non-obstetric and non-psychiatric admissions to general hospitals. For the purposes of this study only admissions for unintentional injury among children were requested	Unintentional injuries (ICD codes V00-X59; Y85-Y86), allowing identification of: • Any injury • Injury type (falls, strikes/cuts/piercings/crushes, scalds, poisonings, transport-related, drowning/submersion, threats to breathing, smoke/fire/flame) • Severity (requiring transfer to another ward or death) • Location (home vs. elsewhere) • Scottish Index of Multiple Deprivation (SIMD)	Relevant episodes occurring to cohort children between October 2009 and April 2018

Continued

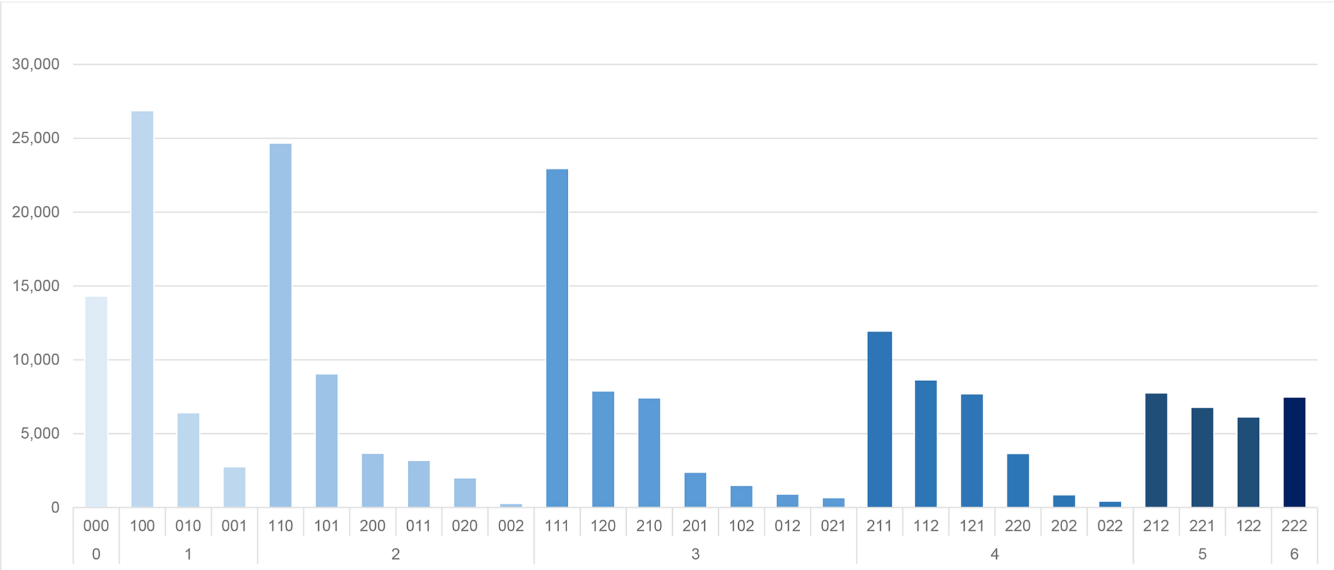
Table Appendix 1: Continued

Dataset	Brief Description	Main variables	Period of data inclusion
The Prescribing Information System (PIS)	National database of all medicines prescribed and dispensed in the community. For this cohort we requested drug dispensations that could relate maternal mental health and childhood attention deficit and hyperactivity disorder (ADHD)	• Maternal anxiety/depression (receipt of Hypnotics and Anxiolytics, Anti-depressants, Psychoses and related disorders) • Child ADHD (in receipt of atomoxetine, dexamfetamine sulphate, guanfacine, lisdexamfetamine mesilate, methylphenidate hydrochloride) • Scottish Index of Multiple Deprivation (SIMD)	Month of relevant dispensations occurring to cohort children and their mothers between January 2009 (mothers) / October 2009 (children) and April 2018. Any dispensation was captured for various periods. Conception was estimated from month of birth and length of gestation to identify the pregnancy period. Pre-pregnancy period was as defined 12 months prior to estimated conception date.
Scottish Immunisation Recall System (SIRS)	SIRS calls and recalls children for immunisation according to the UK childhood immunisation schedule and records details and dates of administered immunisations up until the age of six	• Primary immunisations1st, 2nd and 3rd doses [1] • Measles, mumps and rubella, 1st dose, booster • Preschool booster [2] • Scottish Index of Multiple Deprivation (SIMD) [1] DTaP/IPV/Hib (pertussis, diphtheria, tetanus, polio, Haemophilus influenzae type b) x 3; PCV (pneumococcal) x 2; MenC (Meningitis C) x 1*). [2] dTaP/IPV (pertussis, diphtheria, tetanus, and polio). *MenC2 was phased out in June 2013	Age at immunisation for all calls and recalls to cohort children between October 2009 and April 2018

Child Health Systems Programme (CHSP)	This system automatically calls and recalls children for scheduled child health according to the Healthy Child Programme and records data collected during those reviews. For the period covered by the cohort, the reviews comprised:
	1. CHSP 1st visit (first fortnight)
	2. CHSP 6-8w visit
	3. CHSP 27-30m visit
	4. CHSP Primary 1 visit (4-5 years)
		• Infant feeding (first and 6-8wk)
		• Smoking behaviours of caregivers (all reviews)
		• Developmental concerns (new or existing at 27-30m) [1]:
		  ◦ Cognitive (speech and language)
		  ◦ Vision & Eyesight
		  ◦ Gross and fine motor skills
		  ◦ Socio-emotional wellbeing
		• Height, weight (27-30mth and Primary 1) [2]
		• Scottish Index of Multiple Deprivation (SIMD) (all reviews)
		• Health plan indicator (all reviews) [3]
		[1] Recorded by health visitors using an approved instrument (these varied locally, but most commonly used were the Ages and Stages Questionnaire and the Strengths and Difficulties Questionnaire).
		[2] used to classify children as experiencing thinness, healthy weight, overweight or obesity using age and sex standardised International Obesity TaskForce (IOTF) cut-offs for 2-18 year olds[43]
		[3] used by health practitioners to allocate children to core, additional or intensive programme of support, depending on assessed level of need
			Health reviews taking place for cohort children, between October 2009 and April 2018.

Supplementary Appendix 2: Prevalences and numbers for the socio-economic score and inequalities in longitudinal outcomes, by the socio-economic score

Figure Appendix 2: Prevalence of different combinations of area deprivation, occupational status and relationship status, according to the combined socio-economic score



Key for socio-economic combinations; there are three digits, shown in order of: area deprivation, occupational status, relationships status:	
0	least disadvantaged group for that socio-economic measure (SIMD quintile 5, managerial & professional, married)
1	middle group(s) for that socio-economic measure (SIMD quintiles 2-4, routine/managerial/intermediate, cohabiting)
2	most disadvantaged group for that socio-economic measure (SIMD quintile 1, economically inactive, lone [sole or separated]
For example:	
000	most advantaged SIMD quintile, managerial & professional occupation, married parents
111	middle three SIMD quintiles, intermediate/routine & manual occupations, cohabiting parents
222	most disadvantaged SIMD quintile, unemployed/other occupation, lone parent

Table Appendix 2: Prevalences and numbers for inequalities in longitudinal outcomes by the socio-economic score (most and least deprived groups only), % (*n*)

Health outcome		Socio-economic score	
	Category of outcome	6*	0*
Exposure to smoking, $n=57{,}910$	Never exposed (over period)	28.5 (607)	92.7 (4,002)
	Pregnancy only	8.5 (181)	1.0 (41)
	Post birth	21.6 (460)	5.6 (242)
	Pregnancy and post-birth	41.4 (881)	0.7 (32)
Breastfeeding, $n=169{,}745$	Never	72.6 (4,636)	10.4 (1,289)
	Initiated only	13.1 (835)	10.9 (1,349)
	Breastfeeding at first visit	5.8 (373)	11.5 (1,419)
	Breastfeeding at 6-8 weeks	8.5 (543)	67.2 (8,320)
Mother prescribed medication for anxiety/depression, $n=188{,}010$*	Never	39.6 (2,874)	76.6 (10,133)
	Pre-pregnancy-toddler only	10.2 (737)	5.0 (663)
	Childhood only	11.2 (811)	7.1 (932)
	From infancy-toddler onwards	16.9 (1,226)	6.5 (862)
	Pre-preg + relapse	8.2 (594)	2.0 (267)
	Across most or whole period	13.9 (1,009)	2.8 (367)
Immunisation status by start of school, $n=188{,}010$	Fully immunised	91.0 (6,595)	96.0 (12,696)
	Partially immunised	8.6 (625)	3.5 (456)
	Unimmunised	0.4 (31)	0.5 (72)
Overweight (& obesity), $n=63{,}222$	Healthy at 27-30m & 4-5y	69.0 (1,359)	75.2 (3,563)
	Became overweight	9.2 (182)	5.7 (269)
	Became healthy	9.8 (192)	11.8 (558)
	Overweight at 27-30m & 4-5y	12.0 (237)	7.3 (346)
Hospitalisation due to unintentional injury (count), $n=188{,}010$*	0	90.8 (6,587)	95.2 (12,587)
	1	8.6 (625)	4.6 (609)
	2+	0.5 (39)	0.2 (28)
New/existing development concerns (27-30m), $n=91{,}139$	0	66.5 (2,187)	90.4 (5,964)
	1	16.5 (544)	6.9 (5454)
	2+	17.0 (560)	2.7 (179)

Preg: Pregnancy; M: Month; Y: Year.
*6 = most disadvantaged SIMD quintile, unemployed/other occupation, lone parent.
0 = most advantaged SIMD quintile, managerial & professional occupation, married parents.

Supplementary Appendix 3: Details of the imputation and inequalities in the longitudinal outcomes, post-imputation

Details of imputation

Multiple imputation by chained equations was employed in Stata 16.0 under a missing at random assumption, to create twenty imputation datasets. Imputation specific estimates were combined using Rubin's rules. The augment function used to overcome the presence of empty cells (so that any empty cells in the observed data remained empty in the imputed data).

Three imputed datasets were produced, corresponding to three different samples required for the following longitudinal outcomes:

1. Smoking: exposure to tobacco smoke during pregnancy through to 27-30 months of age. This required data from the birth, infant and toddler samples. The complete case sample was 57,910. After imputing item missingness, the imputed sample was 98,323.
    a. Information was imputed for the following variables, listed according to model:
        i. Binary logistic regression: low birthweight, smoking during pregnancy, moderately or severely depressed Apgar score, exposure to second hand smoke at first visit, exposure to second hand smoke smoking at 6-8 week visit, exposure to second hand smoke at 27-30 month visit;
        ii. Ordered (proportional odds) logistic regression: Scottish Index of Multiple Deprivation (SIMD) quintile, gestational age;
        iii. Multinomial logistic regression: Scottish region of residence, mother's country of birth, mode of delivery.
    b. The following regular variables were included in the imputation model as predictor variables:
        i. Regular: child sex, mother's occupational status, parents' relationship status, number of births, mother's age at current birth.
2. Infant feeding: infant feeding from birth through to the 6-8 week visit required data in the birth and infant sample. The complete case sample was 169,745; after imputing item missingness, the imputed sample was 182,666.
    a. Information was imputed for the following variables, listed according to model:
        i. Binary logistic regression: low birthweight, initiation of breastfeeding, breastfeeding at first visit, breastfeeding at 6-8 week visit, moderately or severely depressed Apgar score;
        ii. Ordered (proportional odds) logistic regression: SIMD quintile, gestational age;
        iii. Multinomial logistic regression: Scottish region of residence, mother's country of birth.
    b. The following regular and auxiliary variables were included in the imputation model as predictor variables:
        i. Regular: child sex, mother's occupational status, parents' relationship status, number of births.
3. Overweight and obesity: overweight and obesity status at 27-30 months and 4-5 years; this required data from the birth, toddler and child samples. The complete case sample was 63,222, the imputed sample, imputing only for item missingness, was 86,726.
    a. Information was imputed for the following variables, listed according to model:
        i. Binary logistic regression: moderately or severely depressed Apgar score, BMI status at 27-30 months, BMI status at primary 1 check, gestational age,
        ii. Ordered (proportional odds) logistic regression: SIMD quintile, gestational age,
        iii. Multinomial logistic regression: Scottish region of residence, mother's country of birth.
    b. The following regular and auxiliary variables were included in the imputation model as predictor variables:
        i. child sex, mother's occupational status, parents' relationship status, number of births, mother's age at current birth.

Table Appendix 3: Prevalence of the longitudinal outcomes overall and according to the least and most advantaged socio-economic groups: imputed results

		Total	Relationship status of parents		NS-SEC of mother		SIMD	
			Sole registration	**Married**	**Economically inactive**	**Managerial & professional**	**SIMD Q1**	**SIMD Q5**
Exposure to smoking [1], $n = 98{,}323$	Never exposed (over period)	65.2	36.6	82.1	44.9	84.0	48.3	84.4
	Pregnancy only	4.7	10.6	2.3	6.8	2.3	6.8	2.1
	From infancy/toddlerhood	16.0	18.2.	11.5	20.0	10.1	19.7	10.3
	Across all time points	14.1	34.6	4.1	28.3	3.4	25.3	3.2
Breastfeeding [2], $n = 182{,}666$	Never	36.8	61.4	22.7	49.3	18.5	53.2	17.7
	Initiated only	15.4	15.3	13.8	13.7	14.0	15.6	13.3
	Breastfeeding at first visit	11.6	9.0	12.2	9.9	12.5	9.7	12.6
	Breastfeeding at 6-8 weeks	36.2	14.2	51.3	27.0	55.0	21.5	56.3
Overweight (& obesity) [3], $n = 86{,}726$	Healthy at 27-30m & 4-5y	71.1	66.0	73.5	69.0	73.7	69.3	74.1
	Became overweight	8.9	11.2	7.8	9.9	7.5	10.4	6.7
	Became healthy	9.8	10.0	9.8	9.7	10.3	8.9	11.2
	Overweight at 27-30m & 4-5y	10.2	12.8	8.9	11.4	8.6	11.4	8.1

NS-SEC: National Statistics Socio-economic Classification; SIMD: Scottish Index of Multiple Deprivation; Q: quintile; M: Month; Y: Year.

Requires inclusion in the birth sample and: [1] the infant and toddler samples, complete case sample was 57,910; [2] the infant sample, complete case sample was 169,745; [3] the toddler and child sample, complete case sample was 63,222.