

Improving case ascertainment of congenital anomalies: findings from a prospective birth cohort with detailed primary care record linkage

Chrissy Bishop,¹ Neil Small,¹ Dan Mason,² Peter Corry,² John Wright,² Roger C Parslow,^{2,3} Alan H Bittles,^{4,5} Eamonn Sheridan⁶

To cite: Bishop C, Small N, Mason D, *et al.* Improving case ascertainment of congenital anomalies: findings from a prospective birth cohort with detailed primary care record linkage. *BMJ Paediatrics Open* 2017;1:e000171. doi:10.1136/bmjpo-2017-000171

Received 27 June 2017

Revised 12 October 2017

Accepted 18 October 2017



CrossMark

¹Faculty of Health Studies, University of Bradford, Bradford, UK

²Bradford Institute for Health Research, Bradford Royal Infirmary, Bradford, UK

³Division of Epidemiology and Biostatistics, University of Leeds, Leeds, UK

⁴Centre for Comparative Genomics, Murdoch University, Perth, Western Australia, Australia

⁵School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia

⁶Institute of Biomedical and Clinical Sciences, University of Leeds, Leeds, UK

Correspondence to

Chrissy Bishop; c.bishop1@bradford.ac.uk

ABSTRACTS

Background Congenital anomalies (CAs) are a common cause of infant death and disability. We linked children from a large birth cohort to a routine primary care database to detect CA diagnoses from birth to age 5 years. There could be evidence of underreporting by CA registries as they estimate that only 2% of CA registrations occur after age 1 year.

Methods CA cases were identified by linking children from a prospective birth cohort to primary care records. CAs were classified according to the European Surveillance of CA guidelines. We calculated rates of CAs by using a bodily system group for children aged 0 to <5 years, together with risk ratios (RRs) with 95% CIs for maternal risk factors.

Results Routinely collected primary care data increased the ascertainment of children with CAs from 432.9 per 10 000 live births under 1 year to 620.6 per 10 000 live births under 5 years. Consanguinity was a risk factor for Pakistani mothers (multivariable RR 1.87, 95% CI 1.46 to 2.83), and maternal age >34 years was a risk factor for mothers of other ethnicities (multivariable RR 2.19, 95% CI 1.36 to 3.54). Education was associated with a lower risk (multivariable RR 0.78, 95% CI 0.62 to 0.98).

Conclusion 98% of UK CA registrations relate to diagnoses made in the first year of life. Our data suggest that this leads to incomplete case ascertainment with a further 30% identified after age 1 year in our study. Risk factors for CAs identified up to age 1 year persist up to 5 years. National registries should consider using routine data linkage to provide more complete case ascertainment after infancy.

INTRODUCTION

Around 93% of children with a congenital anomaly (CA) survive infancy and will require varying levels of support from health services.¹ Most registrations reported by CA registers in the UK occur very early in life, with less than 2% of registrations after 1 year of age.² Out of the 36 CA registries in Europe, seven reported that more than 2% of cases were diagnosed after age 1 year. These later registrations amounted to between 5% and 10% of

What is already known

- ▶ Interrogating primary care data for Read medical codes is a valid and useful source for ascertaining disease prevalence in research studies.
- ▶ Primary care data linked to prospective birth cohort studies allow for detailed, longitudinal analysis.

What this study adds

- ▶ Continuing case ascertainment to age 5 years increased the prevalence of congenital anomalies (CAs) in our population, from 432.9 per 10 000 live births at age 1 year to 620.6 at age 5 years.
- ▶ Without accurate case ascertainment, it is likely that both general and specialist services for CAs will be underresourced.

total CA registrations.³ Comparisons between registries is made more difficult by inconsistencies in the definitions and data variables between different registers and by an absence of follow-up data.^{4–8}

This study uses detailed sociodemographic and clinical information from the Born in Bradford (BiB) prospective birth cohort. The incidence of CA in Bradford is high, previously reported at 306 per 10 000 live births, compared with a national average of 227 per 10 000 live births.^{2 9 10} The Bradford rate, however, is based on case ascertainment up to the child's first birthday, as are 98% of cases used to calculate the national rate. We suspect this may lead to underascertainment of CA and a disparity in both need and demand for children's healthcare. For example, delayed diagnoses have been reported in 10% of congenital heart defects,¹¹ one of the most frequently diagnosed subgroups of CA nationally.^{2 10 12} Interrogation of diagnoses recorded

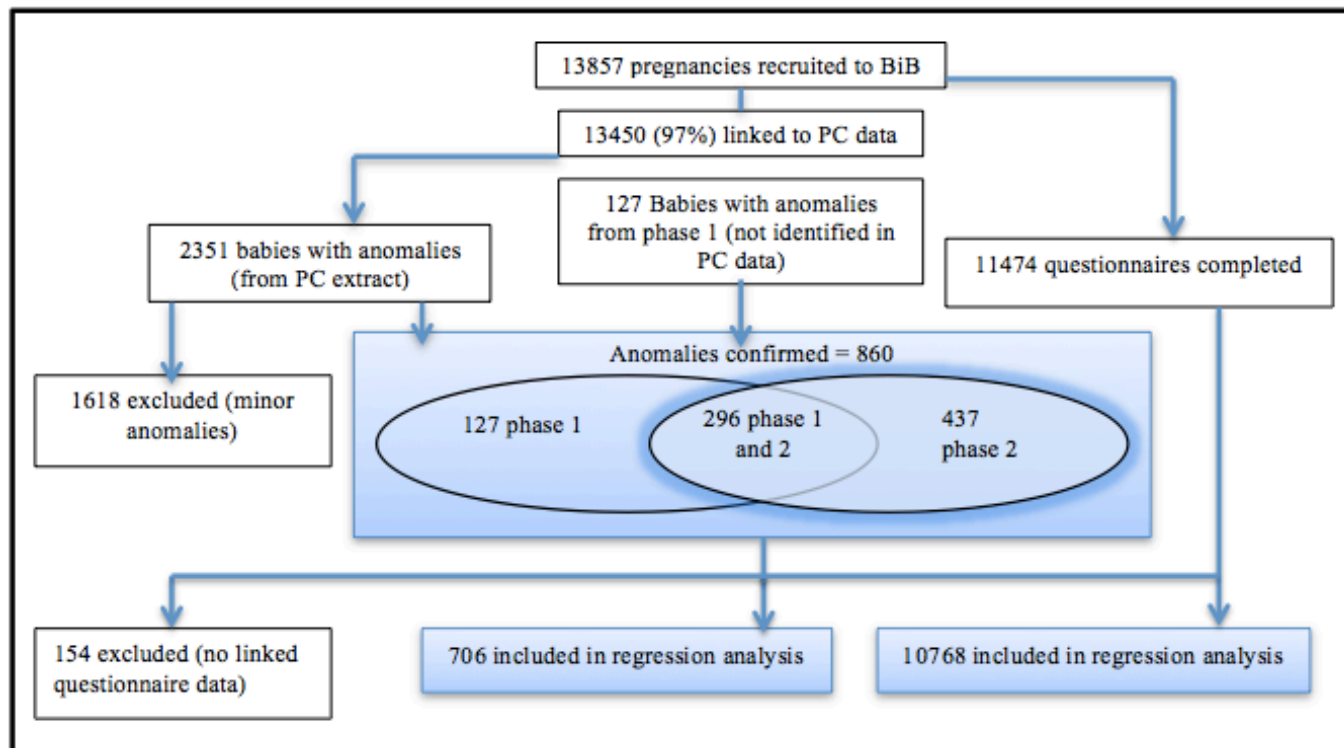


Figure 1 Flow diagram of steps in the analysis. BiB, Born in Bradford; PC, primary care.

in the General Practice Research Database (GPRD) also revealed the prevalence of heart defects to be nearly 50% higher than in European Registers of Congenital Anomalies (EUROCAT) projections, with a similar increase in prevalence in other system-specific subgroups also recorded.¹³

Higher overall CA prevalence rates have been recorded from primary care data sources in comparison with national databases, suggesting the utility of primary care data to serve as a more complete source of background prevalence.^{12 13} Several other studies report the potential of routine primary care databases for CA case ascertainment and diagnoses after age 1 year.^{5 12-17} These databases generally identify more cases than national registers and also have the advantage (over national registers) of including information on potential parental and child risk factors for CA and clinical information on the child across their life course, both of which can be used to improve understanding of the causes and consequences of CA.^{5 12-17} However, primary care databases may lack information on all diagnoses and will not have systematically collected data on potential risk factors using the same methods for all children.

OBJECTIVE

Our aims were to compare case ascertainment of CA from birth to under 5 years between national CA rates and a pregnancy/birth cohort linked to primary care data. Our a priori hypotheses are that more cases will be ascertained from the primary care-linked birth cohort than national

rates, and the detailed data in the birth cohort will allow us to determine whether magnitudes of association for risk factors persisted. We compare CA detection between the primary care database and the clinically diagnosed CA from medical records to determine the accuracy of the primary care diagnoses information.

DESIGN Setting

We used data from BiB, an ongoing prospective birth cohort study, which recruited 12450 pregnant women between 2007 and 2011. The BiB methods are reported in detail elsewhere.¹⁸

Case ascertainment and coding methods

BiB recruits gave their consent to access electronic primary care records held on SystmOne,¹⁹ the patient contact single-source system that currently has complete coverage in Bradford and is linked to BiB baseline questionnaire data. This linkage provides a unique data set comprising detailed social, environmental and clinical data on the mothers and children in the study. Primary care data were extracted for each child when there was an exact match for the National Health Service (NHS) number, surname, date of birth and gender between SystmOne¹⁹ and BiB. Of 13857 recruits, 97% were matched to primary care data, forming the study population. In all, there were 74386 person-years of data, and the average time over which data were recorded in the primary care record was 5.5 years, with a maximum of 7.6 years. Not

all children had reached age 7 years, so we censored our follow-up of these cases to under age 5 years. SystmOne¹⁹ assigns any diagnosis a CTV3 Read medical code. We mapped Read codes to the *International Classification of Diseases, 10th Edition* (ICD-10) codes to allow anomaly classification, with assignment to an anomaly group based on the system affected and syndrome (where applicable). We followed the EUROCAT guidelines,²⁰ using the British Isles National Organisation of CA Registers (BINOCAR) methodology.² A codebook of both major and minor CAs was mapped to Read codes extracted from the primary care database using cross mapping.²¹ Classifications were reviewed by a clinical geneticist (ES).

Sensitivity analyses

We validated the original mapping of CTV3 Read codes to ICD-10 by repeating the mapping process with different members of the research team. We recorded all cases up to the date the child left the primary care practice (ie, had died or moved away), the date the practice stopped recording primary care appointment data or the date of the child's last appointment at time of data extraction (July 2016). A child was classified as having a CA if one or more Read codes for CA were recorded in the child's primary care record at any time during which the child was registered at the practice. From combining CAs reported in the previous BiB data set¹⁰ (phase 1 'notifications'), which were children identified by a standard hospital notification system ages 0 to <1 year and confirmed by a clinical review, and CA identified in primary care data (phase 2 'data linkage') ages 0 to <5 years, we found 296 CAs reported by both phase 1 and phase 2 methodologies. We were able to validate whether phase 1 and phase 2 diagnoses matched, an important step in determining whether clinical diagnoses made in hospital matched Read code entry into primary care databases (figure 1). We reached 83% agreement. We then calculated the prevalence of CA overall and for bodily system-specific subgroups for children diagnosed between ages 0 and <5 years. We also calculated prevalence up to age 1 year for comparison with EUROCAT registries. We also found 127 CAs (17%), which did not match between phase 1 and phase 2 methodologies. On further inspection, these cases had ICD-10 codes outside of the 27th June CA chapter as recommended by EUROCAT. The clinicians responsible for the phase 1 study explained this was due to some conditions being so rare they could not find an appropriate code within the recommended CA ICD-10 chapter. Other reasons included the death or moving primary care practice more than one time, causing potential errors in their diagnoses records.

Risk factors

We reviewed the following maternal risk factors for CA: ethnic origin (White British, Pakistani and Other); age of mother (<20, 20–34 and >34 years); educational attainment (less than five General Certificate of Secondary Education (GCSE) equivalents; five or more

GCSE equivalents at grades A–C; two Advanced Level equivalents; diploma, degree or higher degrees; other; unknown; and foreign unknown); socioeconomic status (Index of Multiple Deprivation 2010 (IMD)²²); smoking (whether mother smoked during pregnancy or not); alcohol consumption (drank alcohol during pregnancy or 3 months before pregnancy (yes or no)); and consanguinity (first cousin, second cousin, other blood relation (less than second cousin) or non-consanguineous). We categorised results for body mass index (BMI) and oral glucose tolerance test in accordance with WHO guidelines.^{23 24}

Statistical analysis

For risk factors, we estimated univariate risk ratios (RRs) and 95% CIs for the occurrence of an anomaly with Poisson regression and robust error variance. We calculated risks for all ethnic groups and separately for White British, Pakistani and other groups. The CIs for smoking and BMI included one in univariate analyses and so were not included in multivariable analyses. We performed a test for interaction between consanguinity and IMD score. To address issues of multiple testing, all models were rerun using 99.9% CIs (data not shown). All analyses were performed in Stata (version 13).

RESULTS

An additional 437 children with a CA were identified using the primary care database. Comparison of rates with BINOCAR was based on 1408 CAs noted in the total of 860 children remaining after the exclusion of minor CA. Table 1 compares the prevalence of anomalies in Bradford with National data reported by BINOCAR.² In 2014, BINOCAR reported a prevalence (based on 2012 data) for CA, excluding chromosomal disorders, of 184 per 10 000 live births (table 1).² Up to age 1 year, we report a total CA rate of 376 per 10 000 live births and a rate of 571.6 per 10 000 live births, including cases ascertained in all children under 5 years. Both these rates specifically exclude children with chromosomal or metabolic disorders. This difference in exclusions is due to BINOCAR registering children with metabolic disorders only if they also have a structural anomaly. BiB includes children with metabolic disorders whether or not they had a structural anomaly. Despite the large difference in BINOCAR rates at age 1 year and BiB rates at age 1 year, the phase 1 study¹⁰ found a similar 305.74 per 10 000 live births, helping to explain the influence of the Bradford demographics on the high numbers, before additional cases are added post age 1 year using primary care records. Regression analyses were based on 706 children with CA, including metabolic and chromosomal conditions for whom BiB questionnaire data were available. The comparison group was the 10 768 cases without CA but for whom questionnaire data were available.

Table 2 shows the characteristics of mothers in the BiB study who gave birth to children with or without CA,

Table 1 Comparison of CA prevalence per 10000 live births comparing BiB and BINOCAR data

	All live births in the BiB cohort (n=13857)				BINOCAR			
	Total*		Total excluding chromosomal and metabolic anomalies*		Total	Live births only (259714)	Live births and stillbirths (260927)	Total excluding chromosomal anomalies
	≤1 year old	Age 0–<5 years	≤1 year old	Age 0–<5 years				
Any CA	432.9	620.6	376.0	571.6	226.5	168.2	171.5	183.6
Nervous system	93.8	215.1	73.6	192.7	26.4	10.2	10.9	23.6
Heart	91.7	122.7	90.9	116.2	60.2	50.7	51.8	51.4
Respiratory	11.5	16.6	10.1	14.4	9.8	6.3	6.6	8.9
Orofacial clefts	25.3	27.4	25.3	25.9	14.3	12.6	12.9	13.2
Digestive system	39.7	48.4	35.4	37.5	18.1	15.3	15.7	16.2
Abdominal wall defects	2.2	2.9	2.2	2.9	9.1	5.3	5.4	7.3
Urinary	49.1	64.2	44.0	49.8	27.8	22.1	22.5	26.7
Genital	37.5	51.2	37.5	44.7	17.9	16.9	17.1	17.4
Limb	44.7	57.7	44.0	53.4	34.9	28.9	29.6	32.2
Other/syndromes					6.5	4.7	4.8	6.06
Skeletal dysplasia	5.8	18.0	5.1	12.3	1.5	1.0	1.0	1.5
Genetic syndromes and microdeletions	13.0	17.3	11.5	14.4	4.9	4.0	4.0	4.6
Sequences	7.9	10.8	7.9	10.1	2.3	1.3	1.4	–
Chromosomal	25.3	32.5	0.00	0.00	43.0	14.9	16.6	–

BiB, Born in Bradford; BINOCAR, British Isles National Organisation of CA Registers; CA, congenital anomaly.

combining cases from phases 1 and 2. [Table 3](#) shows the univariate and multivariable analyses of the risk factors included in [table 2](#). The cohort was multiethnic, with 40% White British, 45% Pakistani and 15% reporting different ethnicities we refer to as ‘Other’ ([table 2](#)). In keeping with rates from phase 1,¹⁰ the proportion of children with anomalies born to Pakistani mothers was higher than those born to mothers in the cohort overall (53% for Pakistani children and 47% for White British and Other combined). [Figure 2](#) demonstrates the age of the child when they received their first CA diagnoses. Without the additional cases from primary care data, this plot would only show diagnoses up to age 1 year, a total of 600 children. Primary care data add a further 260 cases (30%).

Less than 1% of children of White British origin with CA were the offspring of first cousin unions compared with 49% of children with CA in the first cousin Pakistani subgroup. There was a positive step-wise association between the CA prevalence and the degree of consanguinity in the Pakistani subgroup: 9.5% of first cousin progeny, 7.7% of second cousin progeny, 7.5% of beyond second cousins and 4.8% in

non-consanguineous progeny. Sixty-six per cent of the BiB cohort who had completed questionnaires lived in areas defined by the IMD as the most deprived fifth of England ([table 2](#)). The adjusted rates show an excess risk to children born to mothers in the least deprived fifth overall, which was also shown in phase 1.¹⁰ However, the numbers are very small and thus should be treated with caution ([table 3](#)). Consanguinity was found to be a major risk factor for CA in Pakistani mothers in both adjusted and unadjusted rates (multivariable RR 1.87, 95% CI 1.46 to 2.38) and maternal age >34 years for mothers of other ethnicities (multivariable RR 2.19, 95% CI 1.36 to 3.54). Conversely, a higher level of education was associated with a low risk of CA (multivariable RR 0.78, 95% CI 0.62 to 0.98) ([table 3](#)).

DISCUSSION

In the most recent BINOCAR report of data for 2012, less than 2% of live births were diagnosed after 1 year of age.² Using primary care data on children aged 0–<5 years revealed an additional 437 children with CAs, almost two times of those previously reported in

Table 2 Demographics, lifestyle and clinical information of Born in Bradford mothers who gave birth to children with or without a CA by ethnic group

	All		White British		Pakistani		Other	
	No anomaly	Anomaly	No anomaly	Anomaly	No anomaly	Anomaly	No anomaly	Anomaly
Ethnic origin	10 768 (94%)	706 (6%)	4 288 (95%)	245 (5%)	4 804 (93%)	371 (7%)	1 653 (95%)	90 (5%)
Age (years)								
20–34	8 716 (81%)	544 (78%)	3 231 (75%)	180 (73%)	4 099 (85%)	312 (84%)	1 366 (83%)	62 (69%)
<20	776 (7%)	51 (7%)	536 (13%)	29 (12%)	148 (3%)	15 (4%)	92 (6%)	7 (8%)
>34	1 276 (12%)	101 (14%)	521 (12%)	36 (15%)	557 (12%)	44 (12%)	195 (12%)	21 (23%)
Missing	0	0	0	0	0	0	0	0
Education								
<5GCSE equivalents	2 304 (21%)	172 (24%)	856 (20%)	52 (21%)	1 230 (26%)	106 (29%)	212 (13%)	14 (16%)
≥5GCSE equivalents at grades A–C	3 281 (30%)	229 (32%)	1 457 (34%)	85 (35%)	1 488 (31%)	124 (33%)	332 (20%)	20 (22%)
2 Advanced level equivalents	1 561 (15%)	96 (14%)	731 (17%)	37 (15%)	604 (13%)	49 (13%)	224 (14%)	10 (11%)
Diploma, degree or higher degrees	2 775 (26%)	145 (21%)	826 (19%)	45 (18%)	1 265 (26%)	70 (19%)	677 (41%)	30 (33%)
Other	590 (5%)	45 (6%)	372 (9%)	22 (9%)	144 (3%)	17 (5%)	74 (4%)	6 (7%)
Not known	120 (1%)	9 (1%)	40 (1%)	3 (1%)	55 (1%)	4 (1%)	25 (2%)	2 (2%)
Foreign unknown	110 (1%)	8 (1%)	3 (<1%)	0	6 (<1%)	0 (<1%)	101 (6%)	8 (9%)
Missing	27 (<1%)	2 (<1%)	3 (<1%)	1 (<1%)	12 (<1%)	1 (<1%)	0	0
IMD 2010 score (fifths)								
1 (most deprived)	7 124 (66%)	491 (70%)	2 181 (51%)	124 (51%)	3 812 (79%)	299 (81%)	1 115 (67%)	68 (76%)
2	1 949 (18%)	116 (16%)	925 (22%)	54 (22%)	688 (14%)	49 (13%)	332 (20%)	13 (14%)
3	1 199 (11%)	64 (9%)	761 (18%)	40 (16%)	271 (6%)	18 (5%)	165 (10%)	6 (7%)
4	316 (3%)	19 (3%)	267 (6%)	15 (6%)	24 (<1%)	2 (<1%)	24 (1%)	2 (2%)
5 (least deprived)	177 (2%)	16 (2%)	152 (4%)	12 (5%)	8 (<1%)	3 (1%)	17 (1%)	1 (1%)
Missing	3 (<1%)	0	2 (<1%)	0	1 (<1%)	0	0	0
Smoking								
No	8 963 (83%)	604 (86%)	2 842 (66%)	160 (65%)	4 624 (96%)	365 (98%)	1 479 (89%)	79 (88%)
Yes	1 785 (17%)	101 (14%)	1 444 (34%)	84 (34%)	167 (3%)	6 (2%)	172 (10%)	11 (12%)
Missing	20 (<1%)	1 (<1%)	2 (<1%)	1 (<1%)	13 (<1%)	0	2 (<1%)	0
Alcohol								
Yes	3 311 (31%)	184 (26%)	2 878 (67%)	164 (67%)	16 (<1%)	0	410 (25%)	20 (22%)
No	7 426 (69%)	518 (73%)	1 403 (33%)	81 (33%)	4 773 (99%)	369 (99%)	1 238 (75%)	68 (76%)
Missing	31 (<1%)	4 (1%)	7 (<1%)	0 (<1%)	15 (<1%)	2 (<1%)	5 (<1%)	2 (2%)
Consanguinity *								

Continued

Table 2 Continued

	All		White British		Pakistani		Other	
	No anomaly	Anomaly	No anomaly	Anomaly	No anomaly	Anomaly	No anomaly	Anomaly
Non-consanguineous	7850 (73%)	424 (60%)	4284 (99%)	244 (99%)	2008 (42%)	102 (27%)	1538 (93%)	78 (87%)
First cousin	1834 (17%)	192 (27%)	1 (<1%)	1 (<1%)	1753 (36%)	183 (49%)	79 (5%)	8 (9%)
Second cousin	637 (6%)	55 (8%)	0	1 (<1%)	611 (13%)	51 (14%)	25 (2%)	4 (4%)
Other blood	447 (4%)	35 (5%)	3 (<1%)	0	432 (9%)	35 (9%)	11 (1%)	0
Missing	0	0	0	0	0	0	0	0
BMI								
Normal	4529 (42%)	296 (42%)	1738 (41%)	87 (36%)	2018 (42%)	165 (44%)	762 (46%)	44 (49%)
Overweight	2882 (27%)	199 (28%)	1141 (27%)	65 (27%)	1341 (28%)	114 (31%)	399 (24%)	20 (22%)
Obese	2126 (20%)	122 (17%)	992 (23%)	54 (22%)	856 (18%)	51 (14%)	273 (17%)	17 (19%)
Underweight	439 (4%)	27 (4%)	103 (2%)	5 (2%)	256 (5%)	21 (6%)	78 (5%)	1 (1%)
Missing	792 (7%)	62 (9%)	314 (7%)	34 (14%)	333 (7%)	20 (5%)	141 (9%)	8 (9%)
OGTT								
Normal	9850 (91%)	637 (90%)	4062 (95%)	227 (93%)	4261 (89%)	328 (88%)	1507 (91%)	82 (91%)
Impaired fasting glucose	33 (<1%)	2 (<1%)	6 (<1%)	1 (<1%)	18 (<1%)	1 (<1%)	8 (<1%)	0
Impaired glucose tolerance	714 (7%)	56 (8%)	187 (4%)	16 (7%)	421 (9%)	34 (9%)	104 (6%)	6 (7%)
Diabetes	129 (1%)	8 (1%)	15 (<1%)	1 (<1%)	91 (2%)	5 (1%)	23 (1%)	2 (2%)
Missing	42 (<1%)	3 (<1%)	18 (<1%)	0	13 (<1%)	3 (1%)	11 (<1%)	0

*Coefficient of inbreeding values: $F=0$ for non-consanguineous; $F=0.0625$, 0.0156 and <0.0156 for first cousin, second cousin and other blood, respectively.³⁵

BMI, body mass index; CA, congenital anomaly; GCSE, General Certificate of Secondary Education; IMD, Index of Multiple Deprivation 2010; OGTT, oral glucose tolerance test.

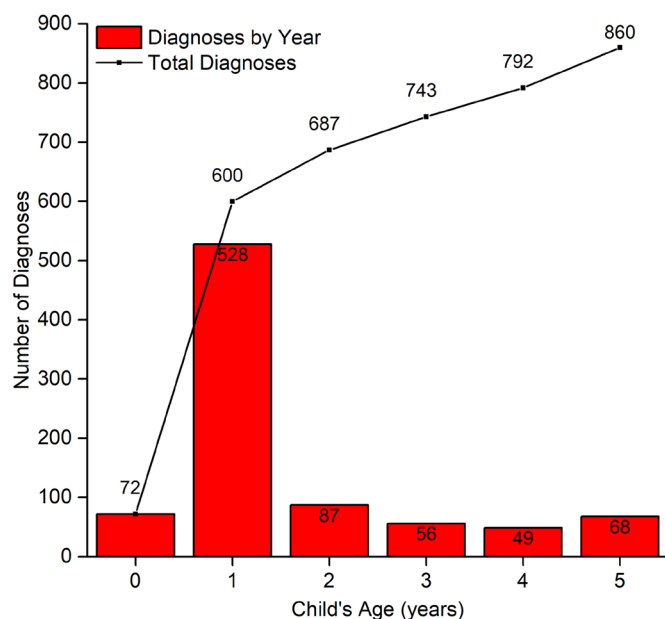


Figure 2 Total number of CA diagnoses per year, and additional CA diagnoses per year using phase 1 and phase 2 data combined.

phase 1.¹⁰ Only 70% of diagnoses in the present study were made prior to the children's first birthday. Similar results have been reported by others and confirm the value of primary care data as a source to investigate CA.^{5 12–17} One review of CA reported in a UK database of primary care records (The Health Improvement Network) revealed that 72% of cases were diagnosed up to age 1 year, and after including diagnoses made up to age 5 years, rates increased from 198 per 10 000 to 277 per 10 000.¹² The overall prevalence of heart defects was reported as lower for infants diagnosed up to age 1 year than infants diagnosed up to age 6 years in a further primary care database study.¹⁴ Late detection of heart CA could be attributable to some cases being missed at antenatal screening due to detection being difficult.²⁵ Two studies using the GPRD found an increase in CA diagnoses when post-age 1 year CA diagnoses were included.^{14 17} Excluding chromosomal and metabolic conditions and considering diagnoses for age 0–5 years, we found the profile of disorders in terms of bodily system categories to be consistent with those reported previously.^{2 10 12 26} The only exception is nervous system disorders, seen as the most common group in our study and in CA research specific to the UK.^{2 10} Considering the percentage increase by bodily system group, skeletal dysplasias increased considerably from age 1 year to age 5 years (210%), primarily due to diagnoses of short stature, followed by nervous system (77%), due to an inflation of hearing loss in Bradford, and respiratory (44%) disorders that are known to be high in Bradford²⁷ (table 1). Some of the conditions in these subgroups are not expected to be detected in the prenatal period,⁸ but our data suggest that they may be taking longer to diagnose than previously thought, which has significant clinical implications. Delayed

diagnoses are seen to create increased complications with care coordination and create a reliance on emergency care.^{28–31} We also assessed the effect of improved ascertainment on the point estimates and statistical significance of the risk factors for CA. We found no substantial change in these risk factors, even with a slightly different CA profile. Changes to statistical significance of risk factors would have had implications for comparative analyses between registries with different ascertainment methods.

Our findings, combined with other primary care database studies for CA ascertainment, therefore suggest that there are more than 2% of CA diagnoses being made after age 1 year, and registries may need to be more specific about their data collection methods for later diagnoses. In England, the recently established National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)³² specify that they collect data, including risk factors, via notifications after age 1 year. The service is new, and the longer-term picture in terms of comprehensive ascertainment is not yet known. Our study demonstrates that ascertainment is strengthened with the use of routine NHS data sources. Consequently, the increased prevalence of children living with CAs in the community may require additional specialist resources for paediatric, obstetric and genetic care for this, as-yet, underascertained cohort. The results could also have implications for transition to adult services.

There are limitations to our study. The time of diagnoses based on Read code entry into primary care systems has been reported as later, on average, than dates of actual diagnosis by general practitioners.¹⁷ The cross mapping of Read to ICD-10 is vulnerable to discrepancies due to multiple Read codes matching one ICD-10 code. We accounted for this by performing a clinical review, assigning the most appropriate Read code to ICD-10 match. CA cases that are stillborn or diagnosed antenatally, resulting in termination, are not well recorded in women's primary care records, thus underestimating CA cases. NCARDRS highlights that 71% of CA are detected antenatally, and 42% of CA diagnosed antenatally resulted in termination.³² Another CA study found fetuses diagnosed with a major CA had a high likelihood of termination of pregnancy, at 50% for consanguineous unions and at 60.9% for non-consanguineous pregnancies.³³ BiB does not report terminations of pregnancy or miscarriages because recruitment is at 26–28 weeks gestation. This current, unpublished work on the BiB cohort has revealed a high level of concordance between self-report of consanguineous marriage and genetic relatedness. There is also a suggestion that the level of genetic relatedness exceeds that expected from self-report, and this may be a function of traditional male socio-occupational groupings. These aspects are being investigated in the BiB cohort as part of an ongoing programme of research.³⁴

Table 3 Univariate and multivariable RRs of CAs related to demographic, lifestyle and clinical risk factors in the Born in Bradford cohort by ethnic group

	Multivariable												
	All			White British			Pakistani			Other			
	RR (95% CI)	P	RR (95% CI)	RR (95% CI)	P	RR (95% CI)	RR (95% CI)	P	RR (95% CI)	RR (95% CI)	P	RR (95% CI)	
Ethnic origin	-	-	1	1	<0.0001	0.96 (0.76 to 1.21)	1	1	1	1	0.7	-	-
Age (years)													
20-34	1	1	1	1	1	1	1	1	1	1	1	1	1
<20	1.01 (0.76 to 1.34)	0.41	0.95 (0.65 to 1.40)	0.81	0.43	1.63 (0.77 to 3.48)	1.23 (0.74 to 2.05)	0.2	1.08 (0.81 to 1.44)	0.6	0.93 (0.63 to 1.37)	0.71	1.23 (0.73 to 2.05)
>34	1.15 (0.95 to 1.39)	0.16	1.04 (0.74 to 1.45)	0.83	0.61	1.99 (1.19 to 3.34)	1.07 (0.82 to 1.41)	0.005	1.17 (0.96 to 1.43)	0.12	1.01 (0.71 to 1.42)	0.98	1.08 (0.82 to 1.44)
Education													
<5GCSE equivalents	1	-	1	-	-	1	1	-	1	-	1	-	1
>5GCSE equivalents	0.94 (0.78 to 1.14)	0.52	0.96 (0.69 to 1.35)	0.82	0.81	0.92 (0.47 to 1.78)	0.97 (0.76 to 1.24)	0.8	1.00 (0.82 to 1.21)	0.97	0.97 (0.69 to 1.36)	0.84	1.00 (0.77 to 1.28)
Advanced level equivalents	0.83 (0.65 to 1.06)	0.14	0.84 (0.56 to 1.27)	0.41	0.74	0.69 (0.31 to 1.52)	0.95 (0.68 to 1.31)	0.36	0.92 (0.71 to 1.18)	0.49	0.81 (0.54 to 1.22)	0.32	1.00 (0.71 to 1.41)
Diploma, degree or higher degrees	0.71 (0.58 to 0.89)	0.002	0.90 (0.61 to 1.33)	0.6	0.01	0.68 (0.37 to 1.27)	0.66 (0.49 to 0.89)	0.23	0.78 (0.62 to 0.98)	0.03	0.86 (0.56 to 1.33)	0.5	0.74 (0.54 to 1.00)
Other	1.02 (0.74 to 1.40)	0.9	0.98 (0.60 to 1.58)	0.92	0.25	1.21 (0.48 to 3.04)	1.33 (0.82 to 2.16)	0.68	1.17 (0.85 to 1.61)	0.35	0.98 (0.60 to 1.60)	0.94	1.36 (0.83 to 2.22)
Do not know	1.00 (0.53 to 1.92)	0.99	1.22 (0.40 to 3.75)	0.73	0.75	1.20 (0.29 to 4.98)	0.85 (0.33 to 2.24)	0.81	1.01 (0.53 to 1.93)	0.98	1.28 (0.41 to 3.94)	0.67	0.80 (0.31 to 2.12)
Foreign unknown	0.98 (0.49 to 1.93)	0.94	0	<0.0001	<0.0001	1.18 (0.51 to 2.74)	0	0.69	1.23 (0.62 to 2.45)	0.56	0	0	<0.0001
Alcohol													
No	1	-	1	-	-	1	1	-	1	-	1	-	1
Yes	0.81 (0.69 to 0.95)	0.01	0.99 (0.76 to 1.28)	0.93	<0.0001	0.89 (0.55 to 1.45)	0	0.65	1.01 (0.83 to 1.23)	0.91	0.98 (0.76 to 1.27)	0.86	0
Smoking													
No	1	-	1	-	-	1	1	-	1	-	1	-	1
Yes	0.85 (0.69 to 1.04)	0.12	1.03 (0.80 to 1.33)	0.81	0.07	1.19 (0.64 to 2.19)	0.47 (0.21 to 1.05)	0.59	-	-	-	-	-
Missing													
OGTT													
Normal	1	-	1	-	-	1	1	-	1	-	1	-	1
Impaired fasting glucose	0.94 (0.24 to 3.62)	0.93	2.70 (0.44 to 16.65)	0.28	0.75	0	0.74 (0.11 to 4.98)	<0.0001	0.46 (0.07 to 3.12)	0.43	1.56 (0.54 to 4.51)	0.42	0
Impaired glucose tolerance	1.20 (0.92 to 1.56)	0.18	1.49 (0.91 to 2.42)	0.11	0.8	1.06 (0.47 to 2.37)	1.05 (0.74 to 1.47)	0.89	1.12 (0.86 to 1.46)	0.41	1.50 (0.92 to 2.46)	0.1	1.01 (0.72 to 1.43)

Continued

Table 3 Continued

	Multivariable															
	All			White British			Pakistani			Other						
	RR (95% CI)	P		RR (95% CI)	P		RR (95% CI)	P		RR (95% CI)	P					
Diabetes	0.96 (0.49 to 1.89)	0.91	1.18 (0.18 to 7.91)	0.86	0.73 (0.31 to 1.72)	0.47	1.55 (0.40 to 5.96)	0.52	0.85 (0.43 to 1.69)	0.65	1.21 (0.18 to 8.17)	0.84	0.72 (0.30 to 1.72)	0.46	1.14 (0.29 to 4.49)	0.86
IMD 2010 score																
1 (most deprived)	1	-	1	-	1	-	1	-	1	-	1	-	1	-	1	-
2	0.87 (0.72 to 1.06)	0.17	1.03 (0.75 to 1.40)	0.88	0.91 (0.68 to 1.22)	0.55	0.66 (0.37 to 1.17)	0.15	0.96 (0.78 to 1.18)	0.7	1.02 (0.74 to 1.40)	0.93	0.99 (0.74 to 1.33)	0.95	0.66 (0.37 to 1.21)	0.2
3	0.79 (0.61 to 1.01)	0.06	0.93 (0.66 to 1.31)	0.67	0.86 (0.54 to 1.36)	0.51	0.61 (0.27 to 1.39)	0.24	0.89 (0.68 to 1.17)	0.41	0.92 (0.63 to 1.34)	0.67	0.99 (0.62 to 1.58)	0.97	0.41 (0.14 to 1.16)	0.94
4	0.88 (0.56 to 1.37)	0.57	0.99 (0.59 to 1.67)	0.97	1.06 (0.28 to 4.02)	0.93	1.34 (0.35 to 5.17)	0.67	1.11 (0.70 to 1.76)	0.66	1.02 (0.59 to 1.75)	0.95	1.64 (0.45 to 6.02)	0.45	1.30 (0.39 to 4.34)	0.67
5 (least deprived)	1.29 (0.80 to 2.07)	0.3	1.36 (0.77 to 2.41)	0.29	3.75 (1.42 to 9.90)	0.01	0.97 (0.14 to 6.59)	0.97	1.66 (1.01 to 2.71)	0.05	1.41 (0.77 to 2.57)	0.26	4.93 (1.77 to 13.73)	0.002	0.97 (1.15 to 6.45)	0.98
Consanguinity																
Non-consanguineous	1	-	1	-	1	-	1	-	1	-	1	-	1	-	1	-
First cousin	1.85 (1.57 to 2.18)	<0.0001	9.28 (2.31 to 37.30)	0.002	1.96 (1.55 to 2.47)	<0.0001	1.91 (0.95 to 3.82)	0.07	1.82 (1.50 to 2.20)	<0.0001	6.87 (1.68 to 23.18)	0.01	1.87 (1.46 to 2.38)	<0.0001	1.79 (0.86 to 3.74)	0.12
Second cousin	1.55 (1.18 to 2.03)	0.001	1	-	1.59 (1.15 to 2.20)	0.005	2.86 (1.12 to 7.28)	0.028	1.58 (1.19 to 2.10)	0.002	0	-	1.54 (1.11 to 2.15)	0.009	3.33 (1.27 to 8.73)	0.015
Other blood	1.42 (1.02 to 1.98)	0.04	0	<0.0001	1.55 (1.07 to 2.25)	0.02	0	<0.0001	1.44 (1.03 to 2.03)	0.35	0	<0.0001	1.53 (1.06 to 2.21)	0.024	0	<0.0001
BMI																
Normal	1	-	1	-	1	-	1	-	1	-	1	-	1	-	1	-
Overweight	1.05 (0.88 to 1.25)	0.56	1.13 (0.83 to 1.55)	0.44	1.04 (0.82 to 1.30)	0.76	0.87 (0.52 to 1.46)	0.61	-	-	-	-	-	-	-	-
Obese	0.88 (0.72 to 1.09)	0.24	1.08 (0.78 to 1.51)	0.64	0.74 (0.55 to 1.01)	0.06	1.07 (0.62 to 1.85)	0.8	-	-	-	-	-	-	-	-
Underweight	0.94 (0.64 to 1.38)	0.77	0.97 (0.40 to 2.34)	0.95	1.00 (0.65 to 1.55)	0.99	0.23 (0.03 to 1.66)	0.150.15	-	-	-	-	-	-	-	-
Missing																

BMI, body mass index; CA, congenital anomaly; GCSE, General Certificate of Secondary Education; IMD, Index of Multiple Deprivation 2010; OGTT, oral glucose tolerance test; RR, risk ratio.

CONCLUSION

We have combined personal and clinical information from a large cohort study with routine primary care data to produce a more comprehensive assessment of the burden of CA in live births. We have demonstrated that more complete case ascertainment of CA can be achieved by linking to primary care data, and by using these, data we are able to detect later diagnoses up to the age of 5 years. Our study also reaffirms consanguinity as a major risk factor for CA in the Bradford Pakistani community.

Acknowledgements This paper presents independent research by a PhD candidate. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health. We thank the families who took part in the Born in Bradford study, the midwives for their help in recruitment, the paediatricians and health visitors, the Born in Bradford team, which included interviewers, data managers, laboratory staff, clerical workers, research scientists, volunteers and managers.

Contributors CB, RP, JW and NS conceived the idea and designed the protocol, with advice from DM on primary care database linkage and conversion of Read codes. ES, RP and PC reviewed all anomalies reported. CB did the statistical analysis, which was overseen by RP with additional interpretation by AHB. All authors contributed to and have approved the final analyses.

Funding This work was supported by a Bradford University studentship, in conjunction with the White Rose Consortium, and the National Institute for Health Research, Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber programme 'Healthy Children Healthy Families Theme' (IS-CLA-0113-10020). The sponsors of this study had no role in the study design, data collection, analysis or interpretation or writing of the report.

Competing interests None declared.

Ethics approval Ethics approval for the cohort study was provided by Bradford Local Research Ethics Committee (reference 06/Q1202/48).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Additional information is available on request from the corresponding author c.bishop1@bradford.ac.uk.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Department of Health. Chief Medical Officer's annual report 2012: our children deserve better: prevention pays. 2013 <https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2012-our-children-deserve-better-prevention-pays> (accessed 31 Oct 2016).
- British Isles Network of CA Registers. CA statistics 2012. 2014 http://www.binocar.org/content/Annual%20report%202012_FINAL_nologo.pdf (accessed 31 Oct 2016).
- Greenlees R, Garne E. Congenital anomalies diagnosed after age 1 year and their impact on prevalence. 2009 <http://www.eurocat-network.eu/content/EUROCAT-Short-Report-Cases-Diagnosed-After-1-Year.pdf> (accessed 31 Oct 2016).
- Boyd PA, Armstrong B, Dolk H, *et al.* Congenital anomaly surveillance in England-ascertainment deficiencies in the national system. *BMJ* 2005;330:27.
- Charlton RA, Weil JG, Cunningham MC, *et al.* Identifying major congenital malformations in the UK General Practice Research Database (GPRD): a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010;33:741-50.
- Scheuerle A, Vannappagari VX, Miller MK. Measurements of birth defect prevalence: which is most useful as a comparator group for pharmaceutical pregnancy registries? *Birth Defects Res A Clin Mol Teratol* 2009;85:611-20.
- Loane M, Dolk H, Kelly A, *et al.* Paper 4: EUROCAT statistical monitoring: identification and investigation of ten year trends of congenital anomalies in Europe. *Birth Defects Res A Clin Mol Teratol* 2011;91(Suppl 1):S31-43.
- Garne E, Dolk H, Loane M, *et al.* Paper 5: Surveillance of multiple congenital anomalies: implementation of a computer algorithm in European registers for classification of cases. *Birth Defects Res A Clin Mol Teratol* 2011;91(Suppl 1):S44-50.
- Bradford Safeguarding Children Board. Child death overview panel annual report. 2015 [http://www.bradford-scb.org.uk/PDF/CDOP%20Annual%20Report%20201415%20Final%20\(pdf\)](http://www.bradford-scb.org.uk/PDF/CDOP%20Annual%20Report%20201415%20Final%20(pdf))
- Sheridan E, Wright J, Small N, *et al.* Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. *Lancet* 2013;382:1350-9.
- Liberman RF, Getz KD, Lin AE, *et al.* Delayed diagnosis of critical congenital heart defects: trends and associated factors. *Pediatrics* 2014;134:e373-e381.
- Sokal R, Fleming KM, Tata LJ. Potential of general practice data for congenital anomaly research: comparison with registry data in the United Kingdom. *Birth Defects Res A Clin Mol Teratol* 2013;97:546-53.
- Wurst KE, Ephross SA, Loehr J, *et al.* Evaluation of the general practice research database congenital heart defects prevalence: comparison to United Kingdom national systems. *Birth Defects Res A Clin Mol Teratol* 2007;79:309-16.
- Wurst KE, Ephross SA, Loehr J, *et al.* The utility of the general practice research database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf* 2007;16:867-77.
- Devine S, West SL, Andrews E, *et al.* Validation of neural tube defects in the full featured-general practice research database. *Pharmacoepidemiol Drug Saf* 2008;17:434-44.
- Sokal R, Tata LJ, Fleming KM. Sex prevalence of major congenital anomalies in the United Kingdom: a national population-based study and international comparison meta-analysis. *Birth Defects Res A Clin Mol Teratol* 2014;100:79-91.
- Hammad TA, Margulis AV, Ding Y, *et al.* Determining the predictive value of Read codes to identify congenital cardiac malformations in the UK Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf* 2013;22:1233-8.
- Wright J, Small N, Raynor P, *et al.* Cohort profile: the born in Bradford multi-ethnic family cohort study. *Int J Epidemiol* 2013;42:978-91.
- The Phoenix Partnership. (TPP) SystmOne. 2016 <https://www.tpp-uk.com/products/systmone> (accessed 13 Mar 2017).
- EUROCAT Central Registry, University of Ulster. EUROCAT Guide 1.4: instruction for the registration of congenital anomalies. 2013 http://www.eurocatnetwork.eu/content/Full%20Guide%201%204%20v4-%2027_Oct2016.pdf (accessed 1 Nov 2016).
- Health and Social Care Information Centre. UK read code releases. 2016 <https://isd.hscic.gov.uk/trud3/user/guest/group/0/pack/9/subpack/19/releases> (accessed 31 Oct 2016).
- Department for Communities and Local Government. The english indices of deprivation. 2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf (accessed 31 Oct 2016).
- WHO. *Definition and diagnoses of diabetes mellitus and intermediate hyperglycemia*. Geneva: World Health Organization, 2006.
- WHO. *Global database on body mass index*. 206. Geneva: World Health Organization, 2006.
- Lee W, Comstock CH. Prenatal diagnosis of congenital heart disease: where are we now? *Ultrasound Clin* 2006;1:273-91.
- EUROCAT Prevalence data tables. Cases and prevalence (per 10,000 births) of all congenital anomaly subgroups for all registries, from 2011 - 2015. 2012 <http://www.eurocat-network.eu/newprevdata/showPDF.aspx?winx=1416&winy=729&file=allsubgroups.aspx> (accessed 1 Nov 2016).
- Bradford Joint Strategic Needs Assessment. Long term conditions. 2014 <https://jsna.bradford.gov.uk/documents/JSNA%20-%204.%20Children%20and%20Young%20People/4.4%20Long%20Term%20Conditions/4.4.01%20Long%20Term%20Conditions.pdf> (accessed 29 Sep 2016).
- Kim PCW, Walker M, Beduz MA, *et al.* The fetal alert network: an innovative program of access to care, surveillance, and education for birth defects in Ontario. *J Obstet Gynaecol Can* 2006;28:1099-102.

29. Chapman DA, Stampfel CC, Bodurtha JN, *et al.* Impact of co-occurring birth defects on the timing of newborn hearing screening and diagnosis. *Am J Audiol* 2011;20:132–9.
30. Ellison S, Lamb J, Haines A, *et al.* A guide for identification and continuing care of adult congenital heart disease patients in primary care. *Int J Cardiol* 2013;163:260–5.
31. Peterson C, Dawson A, Grosse SD, *et al.* Hospitalizations, costs, and mortality among infants with critical congenital heart disease: how important is timely detection? *Birth Defects Res A Clin Mol Teratol* 2013;97:664–72.
32. Public Health England. Guidance: the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS). 2015 <https://www.gov.uk/guidance/the-national-congenital-anomaly-and-rare-disease-registration-service-ncardrs> (accessed 27 Jun 2017).
33. Becker R, Keller T, Wegner RD, *et al.* Consanguinity and pregnancy outcomes in a multi-ethnic, metropolitan European population. *Prenat Diagn* 2015;35:81–9.
34. Small N, Bittles AH, Petherick ES, *et al.* Endogamy, consanguinity and the health implications of changing marital choices in the uk pakistani community. *J Biosoc Sci* 2017;49:435–46.
35. Bittles AH. *Consanguinity in context*. Cambridge: Cambridge University Press, 2012:6.