

## Weighting of risk factors for low birth weight: A linked routine data cohort study in Wales, UK.

Amrita Bandyopadhyay<sup>1\*</sup>, Hope Jones<sup>1</sup>, Michael Parker<sup>1</sup>, Emily Marchant<sup>1</sup>, Julie Evans<sup>2</sup>, Charlotte Todd<sup>2</sup>, Muhammad A. Rahman<sup>3</sup>, James Healy<sup>1,4</sup>, Tint Lwin Win<sup>1</sup>, Ben Rowe<sup>5</sup>, Simon Moore<sup>6,7</sup>, Angela Jones<sup>2</sup>, Sinead Brophy<sup>1</sup>.

- 1 National Centre for Population Health and Wellbeing Research, Swansea University Medical School, Wales, SA2 8PP, UK
- 2 Public Health Wales, Keir Hardie University Health Park, CF48 1BZ, Wales. UK
- 3 Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, CF5 2YB, Wales, UK
- 4 Office for National Statistics, Government Buildings, Cardiff Road, Duffryn, Newport, NP10 8XG, UK
- 5 National Police Chiefs' Council Lead for Mental Health and Age, 10 Victoria Street, London, SW1H 0NN, UK
- 6 School of Dentistry, Cardiff University, Cardiff, Wales, CF14 4XY, UK
- 7 Security, Crime, Intelligence Institute, Cardiff University, SPARK, Maindy Road, Cardiff, Wales, CF10 3AE, UK

\* Corresponding author

Current address: National Centre for Population Health and Wellbeing Research, Swansea University Medical School, Wales, SA2 8PP, UK

Email [a.bandyopadhyay@swansea.ac.uk](mailto:a.bandyopadhyay@swansea.ac.uk)

List of keywords: Low birth weight, Maternal health, Pregnancy interval, Data linkage, Cohort study

## Abstract

### Background

Globally 20 million children are born with a birth weight below 2,500 grams every year which is considered as a low birthweight (LBW) baby. This study investigates the contribution of modifiable risk factors to inform activities that reduce the rates of LBW in Wales.

### Method

The cohort (N = 693,377) was comprised of children born between 1<sup>st</sup> January 1998 and 31<sup>st</sup> December 2018 in Wales. A multivariable logistic regression model and a predictive model using decision tree were used to investigate the associations between the risk factors and LBW.

### Results

The study found that non-singleton children had the highest risk of LBW (adjusted odds ratio 21.74, 95% confidence interval (CI) 21.09 – 22.40), followed by pregnancy interval less than one year (2.92, 95% CI 2.70 – 3.15), maternal diabetes (2.03, 95% CI 1.81 – 2.28), maternal hospital admission for anaemia (1.26, 95% CI 1.16 – 1.36), depression (1.58, 95% CI 1.43 – 1.75), serious mental illness (1.46, 95% CI 1.04 – 2.05), anxiety (1.22, 95% CI 1.08 – 1.38) and use of anti-depressant medication during pregnancy (1.92, 95% CI 1.20 – 3.07). Additional maternal risk factors include smoking (1.80, 95% CI 1.76 – 1.84), alcohol-related hospital admission (1.60, 95% CI 1.30 – 1.97), substance misuse (1.35, 95% CI 1.29 – 1.41), living in areas of high deprivation, and evidence of domestic abuse.

### Conclusion

This work suggests that to address LBW, measures need to focus on improving maternal health, addressing pre-term births, promoting awareness of sufficient pregnancy interval, ensuring adequate support and resources for mother's mental health and wellbeing.

## What is already known on this topic

- Each year 6.9% of live births are identified as low birth weight (LBW) in the UK.
- LBW children are at risk of poor cognitive development, which is associated with developmental disabilities and poor academic achievement in later life.
- The progress to reduce the LBW prevalence in high income regions (including Europe) is not satisfactory to meet World Health Organisation (WHO) LBW target by 2025.

## What this study adds

- This work has built an e-cohort using data-linkage across multiple routinely collected administrative datasets to investigate the risk factors of LBW for the population of Wales.
- Non-singleton children had 22 times higher risk of LBW than singleton children.
- Findings suggested that the most important factors to address to reduce the risk of LBW are multiple births, underlying maternal physical (diabetes, anaemia) and mental health, maternal smoking, and substance use (alcohol/drugs), adequate pregnancy interval, higher deprivation, and domestic abuse during pregnancy.

## How this study might affect research, practice, or policy

- This finding suggests that to address the LBW, adequate support is needed to the mother and their family when they are planning for and during pregnancy.
- Multiagency (doctors, midwives, security department) can liaise between them and can take necessary mitigative action to reduce the prevalence of LBW.

## Introduction

The World Health Organisation (WHO) defines low birth weight (LBW) as infants weighing less than 2,500 grams (5.5 pounds) irrespective of gestational age [1,2]. Latest figures show that each year around 53,000 live births (6.9%) are identified as LBW in the UK [3]. LBW is the result of intra-uterine growth restriction (less than 10th centile of weight for sex and gestational age), prematurity (gestational age less than 37 weeks), or a combination of both [4]. LBW can impair the baby's cognitive development and lead to developmental disabilities and poor academic achievement [5]. Furthermore, LBW significantly increases the risk of perinatal and neonatal mortality and longstanding morbidity in early and later life [6]. Whilst there has been a reduction in mortality amongst preterm infants in the last two decades, the incidence of preterm birth has increased in many developed countries [6–8]. The increase is also associated with preterm delivery of multiple pregnancies, with medically indicated preterm birth 10 times more likely in multiple pregnancies than singleton births [9]. To address the global burden of LBW, the Sixty Fifth World Health Assembly Resolution 65.6 endorsed a comprehensive implementation plan to achieve a 30% reduction in LBW by 2025 [1]. A study conducted on the birth data from 148 countries of 195 United Nations' member states indicated that there had been a 2.9% reduction in the LBW prevalence in 2015, compared to 2000 worldwide. However, there has not been any change in the LBW prevalence in high income regions (including Europe) and the progress is slower than required to meet the WHO LBW target by 2025 [10].

Existing research has found factors linked with mothers, such as age, high deprivation, and low academic qualification, are associated with increased odds of LBW [11,12]. Modifiable risk factors for LBW include inter-pregnancy interval [13], maternal physical [14–17] and mental health [18,19], and environmental exposures during pregnancy [20]. Studies have also shown numerous health behaviours, particularly smoking during pregnancy [21,22], alcohol, in which there is a dose-response relationship with LBW [24], and/or illicit drug use [23] are modifiable risk factors. Indirect (negative maternal behaviours, inadequate nutrition or prenatal care, and increased stress) or direct (physical assault, sexual trauma) experience of intimate partner abuse during pregnancy can lead to adverse infant outcomes, including LBW [25,26].

It is important to gain an understanding of these risk factors, particularly modifiable risk factors, so that resources and interventions can be scheduled effectively. Moreover, the wide range of risk factors cannot be addressed in isolation. Most of the risk factors that are strongly independently associated with LBW are correlated. This study aimed to understand the contributions of risk factors

to the burden of LBW for the population of Wales, using traditional statistical methods and supervised machine learning models.

## Method

### Participants and linkage

The linked data cohort (N = 693,377) was comprised of children born in Wales between 1<sup>st</sup> January 1998 and 31<sup>st</sup> December 2018. The study population was identified in the National Community Child Health database (NCCHD), which is a local Child Health System database held by the National Health Service (NHS). The participants were linked to the Wales-wide administrative register, the Wales Demographic Service Dataset (WDS). Linkage was undertaken using an anonymised encrypted linkage key, the Anonymised Linking Field (ALF), in the Secure Anonymised Information Linkage (SAIL) Databank [27]. WDS provided the anonymised residential linking fields (RALFs), which is an encrypted residential address and its corresponding lower super output area (LSOA, small geographic areas with a population of approximately 1,500) when the child was born. LSOA was linked with the Welsh Index of Multiple Deprivation (WIMD) 2014, which is a measure of relative deprivation. The participants flow diagram is displayed in Supplementary Figure 1.

### Explanatory variables

The maternal variables related to a childbirth were obtained from NCCHD and Maternal Indicator Database (MID). The variables for maternal physical and mental health during pregnancy were obtained from primary care Welsh Longitudinal General Practice (WLGP) and hospital admissions dataset known as Patient Episode database in Wales (PEDW). The record of physical assault linked with mothers during pregnancy was obtained from PEDW. The substance misuse database (SMD) provided the information on alcohol and other drug abuse by the mother during pregnancy. Area type (urban/rural) and local authority (LA) under which they lived during the pregnancy and their overall and physical environment quantified in the WIMD were included in this study. The derived maternal variables include multiple birth flag (to distinguish between singleton and non-singleton), pregnancy interval, harmonised maternal smoking, and maternal weight. The variables and their sources have been described in Supplementary Table 1.

The impact of domestic abuse was examined using a subset of the study population (participants from Rhondda Cynon Taff born between June 2016 and 2018) with linked Public Protection Notification (PPN) dataset [28].

## Outcome variable

In this study a binary variable was created using the birth weight variable obtained from NCCHD.

- LBW = birth weight <2,500
- Not LBW (nLBW) = birth weight  $\geq$ 2,500

## Statistical Analysis

It is known that gestational age is highly correlated with LBW. However, as the gestational age is only obtained at the point of birth, making it a non-modifiable risk factor, this study has not considered it as a predictor variable. The models were stratified by the multiple birth as this is one of the main predictors of LBW. The missing records in birth weight variable were removed from the analysis. Since there was around 15% missing data in maternal weight variable, the variable was imputed by the simple random imputation method [29]. The missing data in the other explanatory variables (less than 10%) were recoded as 'Unknown'. The birth record for stillbirth and pregnancy interval of less than 22 weeks (as that is the minimum duration for a considerable gestation period) were also not considered for the statistical analysis. Data preparation including data linkage and data cleaning for this analysis was done on SAIL DB2 SQL platform. All statistical analyses were performed in R version 4.0.3.

A multivariable logistic regression (MLR) model was developed to identify the most important risk factors associated with LBW. The MLR model was built on the overall study population (whole Wales dataset) to identify the associations between all the explanatory and outcome variables.

A supervised machine learning classifier - decision tree (DT) model was developed to build a risk profile for LBW and test its predictive performance. Classification tree – DT models were constructed using RPART (Recursive Partitioning And Regression Trees) packages in R [30,31]. The algorithm recursively partitions the data into multiple sub-spaces to obtain the homogeneous final sub-space of predictor variables. For DT, the whole Wales data except for Rhondda Cynon Taff was used to train the model and prediction performance was evaluated on a test dataset which consisted of a sample of participants from the LA of Rhondda Cynon Taff. This LA was chosen because it had one of the highest rates of LBW in Wales and is an area which would benefit most from an accurate prediction model.

A separate data linkage was undertaken with a subset of the study population which was linked to the mother's domestic abuse record from PPN dataset (the latter was only available for Rhondda Cynon Taff). Another adjusted MLR model was developed on this linked data to investigate the risk association for LBW.

## Ethical approval

This study had been approved by the SAIL Databank independent Information Governance Review Panel (IGRP) (project number 0916 – WECC Phase 4).

## Results

The study population consisted of 693,377 children of those 54,214 were from Rhondda Cynon Taff and 639,163 were from other LAs. The children from Rhondda Cynon Taff, which was later used as a test set for DT were well representative of the Welsh population (see Table 1). In the overall study population, 51.26% were boys, 96.92% were singleton and 90.38% children were born to term (gestational age between 37 and 42 weeks). 49.85% children were born as the first child in the family. Mothers of 0.48% children were admitted to hospital for diabetes and 0.09% had a GP visit for diabetes, 1.27% had depression, 1.52% with anxiety and 0.02% were on antidepressant medication during pregnancy. There were 1.26% and 21.51% children whose mothers had alcohol-related substance misuse and smoking record during pregnancy respectively. The average maternal age at birth of child and maternal weight was 28 years and 70.82 kg (after imputation) respectively and 63.68% of them were living in densely populated urban areas. Overall, 7.1% (8.26% in test set and 7% in other LAs) of children were born as LBW.

**Table 1: Characteristics of the study population**

Variables	Overall training set (n = 639,163)		Overall test set (n = 54,214)		Overall (n = 693,377)	
Gender						
Girl	311,193	48.69%	26,689	49.23%	337,882	48.73%
Boy	327,920	51.30%	27,522	50.77%	355,442	51.26%
Unknown/NULL	50	0.01%	<5	-	-	-
Maternal age						
Less than 19	46,668	7.30%	5,156	9.51%	51,824	7.47%
20-24	133,792	20.93%	13,149	24.25%	146,941	21.19%
25-29	181,233	28.35%	16,454	30.35%	197,687	28.51%
30-34	170,957	26.75%	12,938	23.86%	183,895	26.52%
35 and above	105,869	16.56%	6,513	12.01%	112,382	16.21%
Unknown/NULL	644	0.10%	<5	-	-	-
Birth order						
1 <sup>st</sup> child	319,093	49.92%	26,552	48.98%	345,645	49.85%
2 <sup>nd</sup> child	212,155	33.19%	18,672	34.44%	230,827	33.29%
3 <sup>rd</sup> child	74,724	11.69%	6,407	11.82%	81,131	11.70%
4 <sup>th</sup> or above	33,191	5.19%	2,583	4.76%	35,774	5.16%

Pregnancy interval						
Only/First child	319,093	49.92%	26,552	48.98%	345,645	49.85%
less than 1 year	5,708	0.89%	526	0.97%	6,234	0.90%
1-2 years	67,986	10.64%	5,333	9.84%	73,319	10.57%
2-5 years	162,590	25.44%	13,519	24.94%	176,109	25.40%
5-7 years	41,060	6.42%	4,081	7.53%	45,141	6.51%
7-10 years	27,161	4.25%	2,707	4.99%	29,868	4.31%
Above 10 years	15,565	2.44%	1,496	2.76%	17,061	2.46%
Gestational age (week)						
1: Extremely pre-term: <28 week	2,361	0.37%	208	0.38%	2,569	0.37%
2: Very pre-term: 28-31	5,296	0.83%	565	1.04%	5,861	0.85%
3: Pre-term: 32-36	38,565	6.03%	3,664	6.76%	42,229	6.09%
4: term: 37-42	577,104	90.29%	49,540	91.38%	626,644	90.38%
5: Late term: 43-45	3,909	0.61%	91	0.17%	4,000	0.58%
Unknown/NULL	11,928	1.87%	146	0.27%	12,074	1.74%
Birth weight (gm)						
1: BW≤ 1,000	3,010	0.47%	246	0.45%	3,256	0.47%
2: BW 1,001 - 1,500	4,372	0.68%	445	0.82%	4,817	0.69%
3: BW 1,501 - 2,499	39,143	6.12%	3,923	7.24%	43,066	6.21%
4: BW 2,500- 4,000	521,172	81.54%	44,524	82.13%	565,696	81.59%
5: BW 4,001 - 4,500	61,658	9.65%	4,413	8.14%	66,071	9.53%
6: BW 4,501 - 5000	9,808	1.53%	663	1.22%	10,471	1.51%
Low Birth Weight (LBW)						
nLBW	594,408	93.00%	49,734	91.74%	644,142	92.90%
LBW	44,755	7.00%	4,480	8.26%	49,235	7.10%
Multiple birth flag						
Singleton	619,458	96.92%	52,583	96.99%	672,041	96.92%
Non-singleton	19,705	3.08%	1,631	3.01%	21,336	3.08%
Maternal smoking						
No	502,914	78.68%	41,344	76.26%	544,258	78.49%
Yes	136,249	21.32%	12,870	23.74%	149,119	21.51%
Welsh Index of Multiple Deprivation						
1 (most deprived)	147,204	23.03%	17,946	33.10%	165,150	23.82%
2	118,271	18.50%	17,711	32.67%	135,982	19.61%
3	117,242	18.34%	7,089	13.08%	124,331	17.93%
4	104,056	16.28%	3,646	6.73%	107,702	15.53%
5 (least deprived)	94,190	14.74%	6,242	11.51%	100,432	14.48%
Unknown/NULL	58,200	9.11%	1,580	2.91%	59,780	8.62%
Diabetes GP (mother)						
No	638,628	99.92%	54,149	99.88%	692,777	99.91%
Yes	535	0.08%	65	0.12%	600	0.09%
Diabetes PEDW (mother)						
No	636,104	99.52%	53,978	99.56%	690,082	99.52%
Yes	3,059	0.48%	236	0.44%	3,295	0.48%



Depression GP (mother)							
	No	631,230	98.76%	53,323	98.36%	684,553	98.73%
	Yes	7,933	1.24%	891	1.64%	8,824	1.27%
Depression PEDW (mother)							
	No	634,990	99.35%	53,950	99.51%	688,940	99.36%
	Yes	4,173	0.65%	264	0.49%	4,437	0.64%
Serious Mental Illness (mother)							
	No	638,887	99.96%	54,185	99.95%	693,072	99.96%
	Yes	276	0.04%	29	0.05%	305	0.04%
Anxiety GP (mother)							
	No	629,681	98.52%	53,131	98.00%	682,812	98.48%
	Yes	9,482	1.48%	1,083	2.00%	10,565	1.52%
Anxiety PEDW (mother)							
	No	635,910	99.49%	53,967	99.54%	689,877	99.50%
	Yes	3,253	0.51%	247	0.46%	3,500	0.50%
Anti-depressant medication (mother)							
	No	639,019	99.98%	54,194	99.96%	693,213	99.98%
	Yes	144	0.02%	20	0.04%	164	0.02%
Vitamin D (mother)							
	No	637,171	99.69%	54,170	99.92%	691,341	99.71%
	Yes	1,992	0.31%	44	0.08%	2,036	0.29%
FOLIC Acid (mother)							
	No	486,360	76.09%	37,663	69.47%	524,023	75.58%
	Yes	152,803	23.91%	16,551	30.53%	169,354	24.42%
Anaemia GP (mother)							
	No	621,276	97.20%	52,636	97.09%	673,912	97.19%
	Yes	17,887	2.80%	1,578	2.91%	19,465	2.81%
Anaemia PEDW (mother)							
	No	631,370	98.78%	53,883	99.39%	685,253	98.83%
	Yes	7,793	1.22%	331	0.61%	8,124	1.17%
Alcohol - GP (mother)							
	No	601,660	94.13%	50,836	93.77%	652,496	94.10%
	Yes	37,503	5.87%	3,378	6.23%	40,881	5.90%
Alcohol - PEDW (mother)							
	No	638532	99.90%	54177	99.93%	692,709	99.90%
	Yes	631	0.10%	37	0.07%	668	0.10%
Assault - PEDW (mother)							
	No	638,479	99.89%	54165	99.91%	692,644	99.89%
	Yes	684	0.11%	49	0.09%	733	0.11%
Substance misuse – any (mother)							
	No	607,433	95.04%	51,087	94.23%	658,520	94.97%
	Yes	31,730	4.96%	3,127	5.77%	34,857	5.03%
Substance misuse - alcohol (mother)							

No	631,148	98.75%	53,235	98.19%	684,383	98.70%
Yes	7,767	1.22%	975	1.80%	8,742	1.26%
Unknown/NULL	248	0.04%	<5	-	-	-
Substance misuse - other (mother)						
No	632,443	98.95%	53,381	98.46%	685,824	98.91%
Yes	6,199	0.97%	794	1.46%	6,993	1.01%
Unknown/NULL	521	0.08%	39	0.07%	560	0.08%
Mother weight (kg)						
Average (before imputation)	71		72.39		71.06	
Median (before imputation)					67.58	
Average (after imputation)					70.82	
Median (after imputation)					67.00	
Living area						
Town and Fringe - Less Sparse	82,435	12.90%	12,702	23.43%	95,137	13.72%
Town and Fringe - Sparse	22,106	3.46%	71	0.13%	22,177	3.20%
Urban > 10K - Less Sparse	403,591	63.14%	37,924	69.95%	441,515	63.68%
Urban > 10K - Sparse	13,441	2.10%	37	0.07%	13,478	1.94%
Village, Hamlet & Isolated Dwellings - Less Sparse	46,166	7.22%	1,464	2.70%	47,630	6.87%
Village, Hamlet & Isolated Dwellings - Sparse	48,314	7.56%	169	0.31%	48,483	6.99%
Unknown/NULL	23,110	3.62%	1,847	3.41%	24,957	3.60%
Local authority						
Blaenau Gwent					15,008	2.16%
Bridgend					28,018	4.04%
Caerphilly					40,418	5.83%
Cardiff					80,247	11.57%
Carmarthenshire					34,705	5.01%
Ceredigion					11,090	1.60%
Conwy					20,389	2.94%
Denbighshire					19,697	2.84%
Flintshire					32,471	4.68%
Gwynedd					23,249	3.35%
Isle of Anglesey					13,941	2.01%
Merthyr Tydfil					13,259	1.91%
Monmouthshire					14,899	2.15%
Neath Port Talbot					28,854	4.16%
Newport					35,153	5.07%
Pembrokeshire					23,929	3.45%
Powys					20,546	2.96%
Rhondda Cynon Taff					54,214	7.82%
Swansea					49,588	7.15%
Torfaen					20,500	2.96%
Vale of Glamorgan					25,657	3.70%
Wrexham					29,346	4.23%

Unknown/NULL					58,199	8.39%
--------------	--	--	--	--	--------	-------

### Factors associated with LBW: MLR results

Non-singleton children were at almost 22 times higher risk of LBW than singleton children (adjusted odds ratio (aOR) – 21.74 (95% confidence interval (CI) 21.09 – 22.40)). Mothers with diabetes-related GP visits (2.03 (1.81 – 2.28)) and hospital admission records of anaemia (1.26 (1.16 – 1.36)) during pregnancy were at very high risk of having LBW children. Poor mental health during pregnancy such as severe depression (1.58 (1.43– 1.75)), serious mental illness (1.46 (1.04 – 2.05)), severe anxiety (1.22 (1.08 – 1.38)) and antidepressant medications (1.92 (1.20– 3.07)) were risk factors for LBW. The other highly significant modifiable risk factors linked with pregnant mothers include maternal smoking (1.80 (1.76 – 1.84)), alcohol related hospital admissions (1.60 (1.30 – 1.97)) and any substance misuse (alcohol/other drugs) (1.35 (1.29 – 1.41)) during pregnancy. Higher maternal age was also associated with the risk of LBW. Though maternal age less than 19 was significantly associated with the risk of LBW in the univariable model, after adjusting all the other explanatory variables, this did not remain as a risk factor of LBW. The first child born was at higher risk of LBW than subsequent births, The odds of LBW for the 2<sup>nd</sup> child was 0.59 (0.57 – 0.60) compared to the first child. Mothers living in the least deprived and rural areas during pregnancy were at lower risk of having LBW children than others living in more deprived and urban areas. The statistically significant risk factors with their aOR and CI have been described in Table 2 visualised in Supplementary Figure 2.

**Table 2: Multivariable logistic regression model to identify the risk factors of LBW among the overall study population.**

Variable name in model (description)	OR	Lower CI	Upper CI
GENDER (Gender)			
Boy	1		
Girl	1.16	1.14	1.18
MOMSMOKE (Maternal smoking)			
No	1		
Yes	1.80	1.76	1.84
MOMAGE (Maternal age)			
Less than 19	0.94	0.90	0.97
20-24	1.00	0.97	1.03
25-29	1		
30-34	1.05	1.02	1.09
35 and above	1.24	1.20	1.29

BIRTHORDER (Birth order)				
	1 <sup>st</sup> child	1		
	2 <sup>nd</sup> child	0.59	0.57	0.60
	3 <sup>rd</sup> child	0.65	0.62	0.67
	4 <sup>th</sup> or above	0.84	0.80	0.88
PREGNANCY_INTERVAL (Pregnancy interval)				
	Less than 1 year	2.92	2.70	3.15
	1-2 years	1.13	1.09	1.18
	2-5 years	1		
	5-7 years	1.15	1.10	1.21
	7-10 years	1.30	1.24	1.37
	Above 10 years	1.60	1.51	1.70
MULTIPLE_BIRTH (Multiple birth flag)				
	Singleton			
	Non-singleton	21.74	21.09	22.40
WIMD (Welsh Index of Multiple Deprivation)				
	1 (most deprived)	1		
	2	0.91	0.88	0.94
	3	0.84	0.82	0.87
	4	0.78	0.75	0.81
	5 (least deprived)	0.70	0.67	0.72
WIMDENV (Welsh Index of Multiple Deprivation – Environment score)				
	1 (most deprived)	1		
	2	0.99	0.96	1.02
	3	1.01	0.98	1.04
	4	1.02	0.99	1.06
	5 (least deprived)	1.00	0.96	1.03
LA (Local authority)				
	Blaenau Gwent	1.23	1.15	1.32
	Bridgend	1.01	0.95	1.07
	Caerphilly	1.06	1.00	1.11
	Cardiff	1		
	Carmarthenshire	1.02	0.96	1.08
	Ceredigion	0.91	0.83	1.01
	Conwy	1.05	0.98	1.12
	Denbighshire	1.13	1.06	1.21
	Flintshire	1.08	1.02	1.14
	Gwynedd	1.04	0.97	1.12
	Isle of Anglesey	1.11	1.02	1.20
	Merthyr Tydfil	1.13	1.05	1.22
	Monmouthshire	1.02	0.94	1.10
	Neath Port Talbot	0.94	0.89	1.00
	Newport	1.13	1.08	1.19
	Pembrokeshire	1.04	0.97	1.11

Powys	1.01	0.93	1.09
Rhondda Cynon Taff	1.23	1.17	1.28
Swansea	0.97	0.93	1.02
Torfaen	1.09	1.02	1.16
Vale of Glamorgan	0.98	0.92	1.04
Wrexham	1.20	1.14	1.27
MOM_DIAB_GP (Diabetes GP (mother))			
No	1		
Yes	2.03	1.81	2.28
MOM_DIAB_PEDW (Diabetes PEDW (mother))			
No	1		
Yes	1.32	1.01	1.74
MOM_DEPRE_GP (Depression GP (mother))			
No	1		
Yes	1.24	1.14	1.34
MOM_DEPRE_PEDW (Depression PEDW (mother))			
No	1		
Yes	1.58	1.43	1.75
MOM_SeriousMentalIllness_GP (Serious Mental Illness (mother))			
No	1		
Yes	1.46	1.04	2.05
MOM_VITD_GP (Vitamin D (mother))			
No	1		
Yes	1.15	0.96	1.38
MOM_FOLIC_GP			
No	1		
Yes	1.09	1.06	1.11
MOM_ALCO_GP (Alcohol - GP (mother))			
No	1		
Yes	1.02	0.98	1.06
MOM_ALCO_PEDW (Alcohol -PEDW (mother))			
No	1		
Yes	1.60	1.30	1.97
MOM_ANXIETY_GP (Anxiety GP (mother))			
No	1		
Yes	1.10	1.02	1.19
MOM_ANXIETY_PEDW (Anxiety PEDW (mother))			
No			
Yes	1.22	1.08	1.38
MOM_ANTIDEP_MED (Anti-depressant medication (mother))			
No	1		
Yes	1.92	1.20	3.07
MOM_ANAEMIA_GP (Anaemia GP (mother))			
No	1		

	Yes	0.70	0.65	0.74
MOM_ANAEMIA_PEDW (Anaemia PEDW (mother))				
	No	1		
	Yes	1.26	1.16	1.36
MOM_ASSAULT (Assault - PEDW (mother))				
	No	1		
	Yes	1.16	0.91	1.47
MOM_SubstanceMisuse_Any (Substance misuse – any (mother))				
	No	1		
	Yes	1.35	1.29	1.41
MOM_SubstanceMisuse_Alcohol (Substance misuse - alcohol (mother))				
	No	1		
	Yes	1.27	1.17	1.38
MOM_SubstanceMisuse_Otherdrug (Substance misuse - other (mother))				
	No	1		
	Yes	1.14	1.04	1.24
MOMWEIGHT (Mother weight)				
		0.99	0.99	0.99
LIVINGAREA (Living area)				
	Town and Fringe - Less Sparse	1.02	0.95	1.09
	Town and Fringe - Sparse	1.03	0.96	1.12
	Urban > 10K - Less Sparse	1		
	Urban - Sparse	1.02	0.99	1.05
	Village, Hamlet & Isolated Dwellings - Less Sparse	0.95	0.91	0.99
	Village, Hamlet & Isolated Dwellings - Sparse	0.89	0.85	0.94

### Finding from the linked PPN data model

Linkage of the cohort with PPN gave a dataset of 5,854 mothers of those who had a PPN call during pregnancy 18% had a LBW child whereas those who did not have PPN call 8.7% had a child with LBW (see Table 3). Mothers with a PPN call during pregnancy had almost 2 times higher risk of having a LBW baby (1.98 (1.39 – 2.81)) than mothers without PPN call after adjusting for confounding factors (see Supplementary Figure 5).

**Table 3: Distribution of LBW and nLBW children for the subset who were linked with mother’s PPN record during pregnancy**

PPN record during pregnancy		n = 5,854	
No			
	nLBW	5,074	91.3%
	LBW	485	8.7%
Yes			

	nLBW	241	82%
	LBW	53	18%

### Predictive DT model

Since LBW were disproportionately more prevalent in non-singleton children (5.61% singleton vs 53.91% of the non-singleton children were LBW), two separate predictive models using DTs were developed.

#### Singleton children

There were 619,458 observations in the training model. The most important risk factors selected by the DT algorithm to develop the final tree were maternal smoking, maternal weight, pregnancy interval, birth order, maternal substance misuse record (any), maternal age, deprivation - WIMD score, maternal substance misuse record (other drug) and maternal substance misuse record (alcohol). Supplementary Figure 3 depicts the final tree with the branches including final 33 terminal nodes. For example, the model would predict a LBW baby if a) maternal smoking is positive (e.g., mum smokes during pregnancy) and b) maternal weight less than 60 kg. The number of women in this category who had a LBW child is 73% (see terminal node 4 in Supplementary Figure 3) and risk profile was found in 7% of the training model population (e.g., 7% of pregnant women were smokers who weighed less than 60 kg during pregnancy).

The test data was built on the 52,583 singleton children, which is 7.82% of the total singleton children in this study. The model performance is explained in a confusion matrix with 60.54% accuracy, 60.41% sensitivity, 60.55% specificity, 9.68% positive predictive values and 95.63% negative predictive value (see Table 4 and 5).

**Table 2: Confusion matrix/two by two table of the DT (singleton and non-singleton) models**

Prediction	Reference (Singleton) n = 52,583		Reference (Non-singleton) n = 1,631	
	LBW	nLBW	LBW	nLBW
LBW	2,077 (TP)	19,389 (FP)	716 (TP)	347 (FP)
nLBW	1,361 (FN)	29,756 (TN)	326 (FN)	242 (TN)

**Table 3: Prediction model performance (n=52,583 Singleton, n = 1,631 Non-singleton from test set)**

	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
<b>DT Singleton model</b>	60.54%	60.41%	60.55%	09.68%	95.63%
<b>DT Non-singleton model</b>	58.74%	68.71%	41.09%	67.36%	42.61%

### Non-singleton children

There were 19,705 children in the non-singleton training subset. The variables selected to generate the tree by the DT algorithm in the importance order were pregnancy interval, birth order, maternal weight, maternal age, gender, deprivation - WIMD score, maternal smoking, living area, deprivation - WIMD (environment) score and maternal substance misuse record (any). Figure 4 depicts the final tree with the branches including final 29 terminal nodes. For example, the model would predict a LBW baby if a) this is the first child or pregnancy interval is either above 10 years or less than 1 year and b) maternal weight less than 60 kg (terminal node 4).

The test set was built on the 1,631 non-singleton children, which is 7.64% of the total non-singleton children in this study. The model performance was measured as 58.74% accuracy, 68.71% sensitivity, 41.09% specificity, 67.36% positive predictive values and 42.61% negative predictive value (see Table 4 and 5).

## Discussion

7.1% of the overall study population in Wales was LBW between 1998 and 2018. Global trend of LBW is around 7.0% in both 2000 and 2015 for the developed regions (Europe, North America, Australia), which is consistent with our finding [2]. Findings from the Office for National Statistics (ONS) state a combined English and Welsh rate of LBW of 7.0% in 2016, unchanged from 2011 [32]. Our findings show that LBW is strongly associated with non-singleton pregnancy, and maternal health which includes a short pregnancy interval, non-optimal maternal body weight (e.g., low, or high weight), maternal smoking, diabetes, anaemia, mental illness and living in a deprived urban area and exposed to domestic abuse during pregnancy.

The findings of short and long pregnancy intervals being associated with increased odds of LBW has been reported previously [13]. However, Regan *et al.* highlighted many of the studies examining long inter-pregnancy interval are more prone to measurement error, with miscarriages and



abortions in this time period difficult to capture and suggest caution should be taken interpreting these findings [33]. In terms of putting this evidence in context, when considering advice over pregnancy intervals, it will be important to consider all the available evidence including the impact of pregnancy interval on preterm birth and maternal outcomes [34]. Modifiable risk factors found to be important in this study included smoking during pregnancy. A number of reviews have been carried out in the field of interventions to reduce smoking in pregnancy which suggest that psychosocial interventions (counselling, feedback and incentives) appear to be effective at supporting women to stop smoking in pregnancy and can reduce the proportion of babies born LBW [35]. However, they argue that the context of the intervention needs to be given consideration and that whilst evidence exists for potentially effective interventions which could be piloted through delivery of programmes locally, efforts should also be directed at population wide strategies to reduce smoking uptake in young women. This may be especially important given the clear difficulties those pregnant experience in giving up smoking [35]. With regards to our finding of maternal mental health increasing the risk of LBW, both severe depression and anxiety were associated with an increased odds of LBW in our study [36].

The singleton DT model correctly predicted 60.41% of all the true positive cases. However, the positive predictive value of 9.68% indicates that the model assigned a false positive 'LBW' classification for 89.32% cases. This model only includes singleton children and since non-singleton pregnancies are highly associated with LBW, removing this variable from the model has lessened its predictive capability. This is evidenced by the significantly improved positive predictive value (67.36%) for the non-singleton model (table 5). Previous machine learning models appear to show better prediction as they included non-singleton, gestational age (which is in terms of temporal association highly associated with LBW but occurs at the same time as the LBW can be measured) and preeclampsia in third trimester.

The strength of this study lies in using national datasets of all births in Wales across a large time. However, as this was a linked routine data study some of the lifestyle factors cannot be captured (diet, physical activity, stress, emotional state) which can be important in determining LBW [37,38]. This work can only identify the more severe cases which are recorded in the health care system, and undiagnosed cases that did not result in the system will be missed. The two different models (MLR and DT) used in this study found very similar findings suggesting that factors which are common and so are predictive (using DT methods) such as maternal smoking status and low weight mother could be targeted to address population risk. Factors which have a strong association with LBW (using

regression analysis) such as a mother with diabetes or mother on antidepressants, can be addressed to reduce individual risk for that mother/child.

## Conclusion

This study suggests that the most important factors to reduce the risk of LBW are to address multiple birth (e.g. in assisted reproduction practices), addressing factors associated with preterm births (previous history of preterm birth), addressing maternal health such as reducing smoking, investment in maternal mental health, addressing substance use (alcohol/drugs), treating underlying health conditions (diabetes/anaemia), and promoting planning of pregnancy to give an adequate pregnancy interval and healthy weight of mother especially for those in deprived urban areas.

## Acknowledgement

### Funding

This work was funded by Public Health Wales (PHW), grant number (105186). This work was supported by Nation Institute for Health Research (NIHR), grant number (NIHR133680). This research has been carried out as part of the ADR Wales programme of work. The ADR Wales programme of work is aligned to the priority themes as identified in the Welsh Government's national strategy: Prosperity for All. ADR Wales brings together data science experts at Swansea University Medical School, staff from the Wales Institute of Social and Economic Research, Data and Methods (WISERD) at Cardiff University and specialist teams within the Welsh Government to develop new evidence which supports Prosperity for All by using the SAIL Databank at Swansea University, to link and analyse anonymised data. ADR Wales is part of the Economic and Social Research Council (part of UK Research and Innovation) funded ADR UK (grant ES/S007393/1).

This work was also supported by the National Centre for Population Health and Well-Being Research (NCPHWR) which is funded by Health and Care Research Wales. This work was supported by Health Data Research UK which receives its funding from HDR UK Ltd (NIWA1) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust.

This work uses data provided by patients and collected by the NHS as part of their care and support. This study used anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who enable SAIL to make anonymised data available for research. We acknowledge the support provided by South Wales Police.

## Contributorship statement

All authors contributed to the study conception and design. The police PPN and MID data accusation was supported by Ben Rowe and Julie Evans respectively. The core dataset was prepared by Muhammad A. Rahman. Further data preparation and the full analysis was done by Amrita Bandyopadhyay. James Healy, and Michael Parker contributed to the analysis. The first draft of the manuscript was written by Amrita Bandyopadhyay and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conceptualization: Sinead Brophy, Angela Jones and Julie Evans and Amrita Bandyopadhyay; Methodology: Amrita Bandyopadhyay and Sinead Brophy; Formal analysis and investigation: Amrita Bandyopadhyay Writing - original draft preparation: Amrita Bandyopadhyay; Writing - review and editing: Charlotte Todd, Michael Parker, Julie Evans, Emily Marchant, Hope Jones, Muhammad A. Rahman, James Healy, Tint Lwin Win, Ben Rowe, Simon Moore, Angela Jones, and Sinead Brophy, Supervision: Sinead Brophy.

## Conflict of interest

The authors declare that they have no conflict of interest.

The views expressed in this paper are those of the authors and not necessarily those of the Office for National Statistics

## Refence

- 1 WHO | Global Nutrition Targets 2025: Low birth weight policy brief. WHO. [http://www.who.int/nutrition/publications/globaltargets2025\\_policybrief\\_lbwt/en/](http://www.who.int/nutrition/publications/globaltargets2025_policybrief_lbwt/en/) (accessed 19 Nov 2020).
- 2 UNICEF-WHO Low birthweight estimates: Levels and trends 2000–2015. <https://www.unicef.org/reports/UNICEF-WHO-low-birthweight-estimates-2019> (accessed 3 Mar 2022).

- 3 Johnson CD, Jones S, Paranjothy S. Reducing low birth weight: prioritizing action to address modifiable risk factors. *J Public Health Oxf Engl* 2017;**39**:122–31. doi:10.1093/pubmed/fdv212
- 4 Mohammed SG. Low birth weight in Omdurman maternity hospital. *Int J Sci Res Publ* 2014;**4**:1–13.
- 5 Breslau N, Paneth NS, Lucia VC. The Lingering Academic Deficits of Low Birth Weight Children. *Pediatrics* 2004;**114**:1035–40. doi:10.1542/peds.2004-0069
- 6 Ohlsson A, Shah P. *Determinants and prevention of low birth weight: a synopsis of the evidence*. Institute of Health Economics 2008.
- 7 Heaman MI, Sprague AE, Stewart PJ. Reducing the Preterm Birth Rate: A Population Health Strategy. *J Obstet Gynecol Neonatal Nurs* 2001;**30**:20–9. doi:https://doi.org/10.1111/j.1552-6909.2001.tb01518.x
- 8 Yuan W, Duffner AM, Chen L, et al. Analysis of preterm deliveries below 35 weeks' gestation in a tertiary referral hospital in the UK. A case-control survey. *BMC Res Notes* 2010;**3**:119. doi:10.1186/1756-0500-3-119
- 9 Blencowe H, Cousens S, Chou D, et al. Born Too Soon: The global epidemiology of 15 million preterm births. *Reprod Health* 2013;**10**:S2. doi:10.1186/1742-4755-10-S1-S2
- 10 Blencowe H, Krusevec J, Onis M de, et al. National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health* 2019;**7**:e849–60. doi:10.1016/S2214-109X(18)30565-5
- 11 Shi L, Macinko J, Starfield B, et al. Primary care, infant mortality, and low birth weight in the states of the USA. *J Epidemiol Community Health* 2004;**58**:374–80. doi:10.1136/jech.2003.013078
- 12 Silvestrin S, Silva CH da, Hiraakata VN, et al. Maternal education level and low birth weight: a meta-analysis. *J Pediatr (Rio J)* 2013;**89**:339–45. doi:10.1016/j.jpeds.2013.01.003
- 13 Conde-Agudelo A, Rosas-Bermúdez A, Kafury-Goeta AC. Birth Spacing and Risk of Adverse Perinatal Outcomes: A Meta-analysis. *JAMA* 2006;**295**:1809–23. doi:10.1001/jama.295.15.1809
- 14 Yu Z, Han S, Zhu J, et al. Pre-Pregnancy Body Mass Index in Relation to Infant Birth Weight and Offspring Overweight/Obesity: A Systematic Review and Meta-Analysis. *PLOS ONE* 2013;**8**:e61627. doi:10.1371/journal.pone.0061627
- 15 Daalderop LA, Wieland BV, Tomsin K, et al. Periodontal Disease and Pregnancy Outcomes: Overview of Systematic Reviews. *JDR Clin Transl Res* 2018;**3**:10–27. doi:10.1177/2380084417731097
- 16 Flynn CA, Helwig AL, Meurer LN. Bacterial vaginosis in pregnancy and the risk of prematurity. *J Fam Pract* 1999;**48**:885–92.
- 17 Figueiredo ACMG, Gomes-Filho IS, Silva RB, et al. Maternal Anemia and Low Birth Weight: A Systematic Review and Meta-Analysis. *Nutrients* 2018;**10**. doi:10.3390/nu10050601

- 18 Dadi AF, Miller ER, Bisetegn TA, *et al.* Global burden of antenatal depression and its association with adverse birth outcomes: an umbrella review. *BMC Public Health* 2020;**20**:173. doi:10.1186/s12889-020-8293-9
- 19 Lima SAM, Dib RPE, Rodrigues MRK, *et al.* Is the risk of low birth weight or preterm labor greater when maternal stress is experienced during pregnancy? A systematic review and meta-analysis of cohort studies. *PLOS ONE* 2018;**13**:e0200594. doi:10.1371/journal.pone.0200594
- 20 Fleischer Nancy L., Merialdi Mario, van Donkelaar Aaron, *et al.* Outdoor Air Pollution, Preterm Birth, and Low Birth Weight: Analysis of the World Health Organization Global Survey on Maternal and Perinatal Health. *Environ Health Perspect* 2014;**122**:425–30. doi:10.1289/ehp.1306837
- 21 Flower A, Shawe J, Stephenson J, *et al.* Pregnancy planning, smoking behaviour during pregnancy, and neonatal outcome: UK millennium cohort study. *BMC Pregnancy Childbirth* 2013;**13**:238. doi:10.1186/1471-2393-13-238
- 22 Jaddoe VWV, Troe E-JWM, Hofman A, *et al.* Active and passive maternal smoking during pregnancy and the risks of low birthweight and preterm birth: the Generation R Study. *Paediatr Perinat Epidemiol* 2008;**22**:162–71. doi:10.1111/j.1365-3016.2007.00916.x
- 23 dos Santos JF, de Melo Bastos Cavalcante C, Barbosa FT, *et al.* Maternal, fetal and neonatal consequences associated with the use of crack cocaine during the gestational period: a systematic review and meta-analysis. *Arch Gynecol Obstet* 2018;**298**:487–503. doi:10.1007/s00404-018-4833-2
- 24 Patra J, Bakker R, Irving H, *et al.* Dose–response relationship between alcohol consumption before and during pregnancy and the risks of low birthweight, preterm birth and small for gestational age (SGA)—a systematic review and meta-analyses. *BJOG Int J Obstet Gynaecol* 2011;**118**:1411–21. doi:10.1111/j.1471-0528.2011.03050.x
- 25 Hill A, Pallitto C, McCleary-Sills J, *et al.* A systematic review and meta-analysis of intimate partner violence during pregnancy and selected birth outcomes. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet* 2016;**133**:269–76. doi:10.1016/j.ijgo.2015.10.023
- 26 Donovan BM, Spracklen CN, Schweizer ML, *et al.* Intimate partner violence during pregnancy and the risk for adverse infant outcomes: a systematic review and meta-analysis. *BJOG Int J Obstet Gynaecol* 2016;**123**:1289–99. doi:https://doi.org/10.1111/1471-0528.13928
- 27 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;**9**:3. doi:10.1186/1472-6947-9-3
- 28 Constabulary © Her Majesty’s Inspectorate of, Fire. Police effectiveness 2015 (Vulnerability) – Dyfed-Powys Police. HMICFRS. <https://www.justiceinspectors.gov.uk/hmicfrs/publications/police-effectiveness-vulnerability-2015-dyfed-powys/> (accessed 3 Nov 2021).
- 29 Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press 2006.
- 30 Lewis RJ. An introduction to classification and regression tree (CART) analysis. In: *Annual meeting of the society for academic emergency medicine in San Francisco, California*. 2000.

- 31 Atkinson Beth. rpart function | R Documentation.  
<https://www.rdocumentation.org/packages/rpart/versions/4.1-15/topics/rpart> (accessed 14 Jan 2021).
- 32 ONS. Birth characteristics in England and Wales - Office for National Statistics. 2019. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2017> (accessed 14 Jul 2021).
- 33 Regan AK, Ball SJ, Warren JL, *et al.* A Population-Based Matched-Sibling Analysis Estimating the Associations Between First Interpregnancy Interval and Birth Outcomes. *Am J Epidemiol* 2019;**188**:9–16. doi:10.1093/aje/kwy188
- 34 Hutcheon JA, Nelson HD, Stidd R, *et al.* Short interpregnancy intervals and adverse maternal outcomes in high-resource settings: An updated systematic review. *Paediatr Perinat Epidemiol* 2019;**33**:O48–59. doi:10.1111/ppe.12518
- 35 Chamberlain C, O'Mara-Eves A, Porter J, *et al.* Psychosocial interventions for supporting women to stop smoking in pregnancy. *Cochrane Database Syst Rev* Published Online First: 2017. doi:10.1002/14651858.CD001055.pub5
- 36 Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. *World Psychiatry* 2020;**19**:313–27. doi:10.1002/wps.20769
- 37 Ghavi A, Fadakar Sogheh K, Niknamy M, *et al.* Investigating the Relationship between Maternal Lifestyle during Pregnancy and Low-Birth-Weight of Term Neonates. *Iran J Obstet Gynecol Infertil* 2012;**15**:14–24.
- 38 Xi C, Luo M, Wang T, *et al.* Association between maternal lifestyle factors and low birth weight in preterm and term births: a case-control study. *Reprod Health* 2020;**17**:93. doi:10.1186/s12978-020-00932-9