

# Analysis of Risk Factors Progression of Preterm Delivery Using Electronic Health Records

**Zeineb Safi**

Sidra Medicine

**Neethu Venugopal**

Sidra Medicine

**Haytham Ali**

Sidra Medicine, Division of Neonatology

**Michel Makhoul**

Sidra Medicine

**Sabri Boughorbel** (✉ [sboughorbel@sidra.org](mailto:sboughorbel@sidra.org))

Sidra Medicine Department of Emergency Medicine <https://orcid.org/0000-0003-2734-3356>

---

## Research

**Keywords:** preterm, pregnancy, EHR, epidemiology, risk factors, progression, temporal analysis, precision medicine, predictive models

**Posted Date:** September 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-78033/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## RESEARCH

# Analysis of Risk Factors Progression of Preterm Delivery using Electronic Health Records

Zeineb Safi<sup>1</sup>  
 , Neethu Venugopal<sup>1</sup>  
 , Haytham Ali<sup>2</sup>  
 , Michel Makhlouf<sup>3</sup>  
 and Sabri Boughorbel<sup>1\*</sup>

\* Correspondence:

[sboughorbel@sidra.org](mailto:sboughorbel@sidra.org)

<sup>1</sup>Department of Systems Biology,  
 Sidra Medicine, Doha, Qatar

Full list of author information is  
 available at the end of the article

† Equal contributor

## Abstract

**Background:** Preterm deliveries have many negative health implications on both mother and child. Identifying the population level factors that increase the risk of preterm deliveries is an important step in the direction of mitigating the impact and reducing the frequency of occurrence of preterm deliveries. The purpose of this work is to identify preterm delivery risk factors and their progression throughout the pregnancy from a large collection of Electronic Health Records (EHR).

**Results:** The study cohort includes more than 60,000 deliveries in the USA with the complete medical history from EHR for diagnoses, medications, procedures and demographics. We propose a temporal analysis of risk factors by estimating and comparing risk ratios at different time points prior to the delivery event. We selected the following time points before delivery: 9, 6, 3 and 1 month(s). We did so by conducting a retrospective cohort study of patient history for a selected set of mothers who delivered preterm and a control group of mothers that delivered full-term. We analyzed the extracted data using a logistic regression model. The results of our analyses showed that the highest risk ratio corresponds to history of previous preterm delivery. Other risk factors were identified, some of which are consistent with those that are reported in the literature, others need further investigation.

**Conclusions:** The comparative analysis of the risk factors at different time points showed that risk factors in the early pregnancy related to patient history and chronic condition, while the risk factors in late pregnancy are specific to the current pregnancy. Our analysis unifies several previously reported studies on preterm risk factors. It also gives important insights on the changes of risk factors in the course of pregnancy.

## Keywords:

preterm; pregnancy; EHR; epidemiology; risk factors; progression; temporal analysis; precision medicine; predictive models

## 1 Introduction

The WHO defines prematurity as births before 37 weeks of completed gestation [1]. Over the recent years there is a global increase in the rate of prematurity, ranging from 9-12% of all birth. In 2016, prematurity affected around 15 million live-born babies worldwide. [2, 2] The reporting of the short and long-term outcomes of prematurity attracted the interest of scientists, clinicians, as well as policymakers. Of

particular concern is the high rate of prematurity within the low socioeconomic class [3]. Extreme prematurity, defined as birth before 28 weeks of gestation, poses life-long consequences on the health, education, and the social life of children [4]. From the societal perspective, almost one-third of extremely premature infants required support in special education systems. The majority had poor educational attainment at school age [5]. It was found to be one the six major causes that were attributed to 73% of yearly deaths of children under the age of 5 years, where 10% of the deaths were due to preterm birth [6]. Children that are born preterm are at a higher risk of a plethora of psychological and physiological health implications, especially those with a low birth weight. A quarter of the very low birth weight children develop severe or multiple psychological problems, and another quarter develop moderate to mild problems. The psychological development is measured in different domains, namely cognitive development, behavioral and emotional status, social functioning and school adaptation [7]. Children that are very preterm have abnormal brain morphology when compared to those born full-term at seven years of age [8]. A Significant proportion of the children that are born very premature, born before 27 week gestation, are found to have difficulties in motor and academic skills in early school years [9]. They are also more likely to develop respiratory problems [10]. Preterm deliveries have many negative implications on the mother's and child's health alike. Mothers that deliver preterm babies, are at a greater risk to suffer from complications, particularly those who undergo a Cesarean delivery. The complications include hemorrhage, infection, ICU admission and death [11].

Identifying mothers that are at a higher risk through quantifying risk factors of preterm delivery at a population level helps clinicians to take preventive measures and mitigate the risks [12]. Traditionally, the identification of such risk factors is done through prospective studies. This method presents challenges, some of which are difficulty in recruiting trial participants and the tediousness of the data collection process. With the increased adoption of Electronic Health Records (EHR) in hospitals and health care facilities around the world, some of these challenges have been alleviated. EHR data is being extensively used in clinical research, even though it presents some challenges of its own, it provides fast and easy access to large amounts of data, that is more representative of the general population than data collected from clinical trials. Some of the research application of EHR data are epidemiology and observational research, safety surveillance and regulatory use, and prospective clinical research [13].

Most studies are also concerned with extracting risk factors from a limited set of suspected attributes, while in our analyses we extract risk factors from all possible attributes that are available to us in Electronic Health Records. We included diagnosis information (ICD-9), medication information (NDC codes), laboratory orders (ICD-9) and demographic information.

The purpose of this work is to perform a retrospective cohort study using a large EHR dataset to identify risk factors of preterm delivery. We hypothesize that risk factors are dynamics and change in the course of pregnancy. In the early phase of pregnancy, genetic predisposition might dominate the risk factors while in advanced stages of the pregnancy, other factors could play more important roles [12]. The analysis of the variation of risk factors as a function of time could help in building models for the early prediction of preterm delivery.

In this work, we identify the risk factors that are associated with preterm delivery in the literature, and present a few examples of the use of EHR data to perform epidemiology studies in section 2. We present the methodology adopted for extracting relevant preterm and full-term deliveries from the EHR database and statistical analysis of the data in section 3. We present our results and discuss our findings in section 4. Risk factors obtained from association studies should not be understood as causal factors. Among the risk factors that are statistically significant, we curated the ones related to spontaneous preterm deliveries from the medically induced ones. Section 5 concludes our findings and presents prospects for future work.

## 2 Related Work

### 2.1 Preterm Delivery Risk Factors

The identification of risk factors of preterm deliveries is a well researched problem. The history of previous preterm deliveries is one of the main risk factors that have been identified by different research studies [14, 15, 16, 17]. Maternal age is also a risk factor, adolescent pregnancies and advanced maternal age are associated with increased risk of preterm deliveries [14, 15]. Other factors are low maternal BMI [14, 17], obesity [15], women of African race [17, 16, 15], and short inter-pregnancy intervals [16, 15]. The mother anatomy, such as short cervical length [17, 16, 15], and uterine anomalies [16] are also well known risk factors. Existing conditions in the mother whether they are infectious conditions (HIV, chlamydia, and urinary tract infection) or other conditions (pre-eclampsia, low maternal vitamin D, pregestational and gestational diabetes [15], and anemia [18]) also increase the risk of preterm delivery. Another risk factor is multiple gestation [17] or singleton and twin pregnancies formed through in vitro fertilization [15], in addition to some of the mother's behaviors such as smoking and drug use during pregnancy [15]. Some work also investigated the role of environmental factors on preterm birth [19], where an increase in  $NO_2$  concentration was shown to have an association with preterm birth.

### 2.2 Risk Factors from Electronic Health Records

With the increasing worldwide adoption of Electronic Health Records, many research groups made use of the availability of such large data volumes for different medical research purposes. Epidemiology studies are use cases that are becoming increasingly popular. An example application is finding associations between late-preterm birth and persistent asthma in young children, inhaled corticosteroid use and more acute respiratory visits, by conducting a retrospective cohort analyses [10]. EHR data was also used to examine the association between the exposure to psychotherapy during pregnancy and the risk reduction of adverse pregnancy outcomes among women with history of major depressive disorder and obstetric outcomes [20]. Other examples include associating glucocorticoid use to predicting fracture risk [21], using EHR data to predict the epidemiology of disease biomarkers [22], and identifying the risk factors of Angiotensin -converting enzyme inhibitors (ACEIs) by conducting a retrospective cohort study from EHR data [23].

### 3 Methodology

In order to identify the variation of risk factors in the course of pregnancy, we performed association studies at four time points during pregnancy. Since the gestational age is not available in our dataset, we used delivery date as an anchoring time point for our analyses. These time points are 9, 6, 3 and 1 month(s) before delivery. At each time point, only the past medical history with respect to the time point is included for each subject. For example, if the time point of analysis is 6 month before delivery, all the data between 6 months and the delivery event are discarded. This mimics a realistic scenario where future data is not available at each time point of analysis. Figure 1 illustrates the data timeline used in the analysis. The inclusion of the time dimension in the analysis gives us a predictive tool. The risk factors at different time points can be interpreted as risk predictors for preterm delivery. The association study is based on multivariate logistic regression model.

#### 3.1 Data Extraction

We used Health Facts<sup>©</sup> Cerner EHR database to extract the dataset that is used in our analyses. The dataset consists of a retrospective and longitudinal cohort of full-term and preterm deliveries in the period 2001-2014. The dataset of our cohort is gathered from 120 hospitals in the USA. Gestational ages of the pregnancies were not available in the data, for that reason, we relied on ICD-9 diagnosis codes to identify preterm and full-term pregnancies. The ICD-9 code that was used to identify preterm deliveries are 644.2x, and the ICD-9 codes that were used to identify full-term deliveries are 645.xx, 649.8, 650 and 652.5. To evaluate the risk factors at different time intervals prior to delivery event, we excluded patient history from the delivery event until a defined time point. Figure 1 illustrates the data selection process. Each line in the figure depicts a mother's history timeline. The dots on the lines represent delivery events. Green dots represent full-term deliveries, while red dots represent preterm deliveries. The post delivery visits were excluded from the data in the analyses. The dark gray portion before the delivery event is the time gap between the analysis time point and delivery. The data in this time interval were also omitted from the mother's history. We first used a time gap of 270 days, and selected mothers that had at least two hospital visits in their history prior to the time gap. Mother E in figure 1 would not be selected in this case, as not enough history data is available for the analyses. The datasets for the 30 days, 90 days, and 180 days time gaps were constructed the same way but by including only the mothers that were selected in the 270 days dataset.

The medical history of each patient encompasses the following information: drug orders (using NDC encoding), procedures (using ICD-9, CPT4, and HCPCS encoding), diagnosis codes (using ICD-9 encoding) and lab orders (using LOINC encoding). In addition to these, the dataset includes demographic information of the patients. The latter included are race, marital status, age, medical specialty and hospital ID. The demographic information are included to reduce the effects of confounding factors.

### 3.2 Risk Factors Identification

We used logistic regression as shown in equation (1) [24] for risk factor identification using four datasets, each containing the same set of patients, with a portion of the data omitted according to the previously mentioned time gaps.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (1)$$

In equation (1),  $\beta_i$  are the coefficients to be estimated and  $x_i$  values are the corresponding independent variables that represent medications, diagnosis, lab orders, procedures and demographics. The value of  $i$  ranges from 1 to  $p$ , where  $p$  is the total number of dependent variables.

The dependent variable in the analysis  $y$  indicates whether the delivery is preterm  $y_i = 1$  or full-term  $y_i = 0$ . We used Python's scikit-learn<sup>[1]</sup> implementation of logistic regression, which uses maximum likelihood for parameter estimation.

After estimating the multiple logistic regression coefficients, we performed a t-test in order to identify the statistically significant variables in the analysis. The covariates with the significant p-values were selected as risk factors in the corresponding time gap. The p-value cut-off was chosen to be 0.01 to retain the statistical significant variables. We report the risk ratios as well as their 95 % confidence intervals corresponding to each identified statistically significant risk factor.

The rate of preterm birth is reported to be 1:10 in the US according to CDC [2]. In our dataset the preterm birth rate is 1:3, for that reason, the odds ratios in our analysis exaggerates the real risk ratios [25]. We use equation (2) to calculate the corrected risk ratio, where OR is the odds ratio and  $P_0$  is the incidence of the covariate in the full-term group.

$$RR = \frac{OR}{(1 - P_0) + (P_0 \times OR)} \quad (2)$$

## 4 Results and Discussion

### 4.1 Dataset

The final dataset that was used in the experiment contained a total of 60,965 delivery events, 48,448 of which are full-term deliveries and 12,517 are preterm deliveries. The logistic regression covariates in this experiment are the different medications, procedures, lab orders, and diagnosis that occurred more than 100 times across all patient visits at each time gap. The number of covariates changes across the datasets, as some of them that appeared in a specific time gap from the delivery would disappear as the time gap between the delivery event and the start of considered history is made bigger. A summary description of the covariates for the different clinical modalities is given in Table 1.

<sup>[1]</sup><https://scikit-learn.org>

<sup>[2]</sup><https://www.cdc.gov/features/prematurebirth/index.html>

## 4.2 Logistic Regression Model Evaluation

To evaluate the goodness-of-fit of the logistic regression model we used the Area Under the receiver operating characteristic Curve (AUC) of the different datasets. The AUC measures how well the logistic regression model can classify full-term and preterm groups. The different AUC values are shown in Table 2.

## 4.3 Identified Risk Factors

Tables 3, 4, 5, 6, 7, and 8 summarize the significant risk factors identified. The risk factors that are identified in each time gap are those with a p-value  $\leq 0.01$ . To follow the progression of risk factors before and during the pregnancy at different stages, we group the risk factors that are common to all four time gaps (Table 3), three time gaps (Table 4), and two time gaps (Table 5), as well as those that are specific to the 30 days time gap (Table 6), 90 days time gap (Table 7), 180 days time gap (Table 8) and 270 days time gap (Table 9). The tables show the medication, diagnosis, medical procedure and lab orders codes preceded by M\_, D\_, P\_ and L\_ respectively in the first columns of the table. The corresponding descriptions of these codes are also included in the table. The description of medication codes consist of brand names and generic names of the medication separated by "/". Risk factors with a risk ratio between 0.9 and 1.1 are excluded from the analysis, as the coefficients of these risk ratios are very close to 1, and are found to not have a significant impact.

### 4.3.1 Risk Factors Common to all Time Gaps

The risk factors that are common to all time gaps, meaning, they indicate a high risk pregnancy before and during the current pregnancy, are mainly the existence of a previous preterm delivery, possible pregnancy complications of previous pregnancy and having a pre-existing hypertension. The previous preterm delivery is indicated by the diagnosis code D\_644.21, and the medication Celestone Soluspan (Betamethasone), which is a prenatal corticosteroid that is administered in high risk pregnancies in order to reduce the risk of preterm infants having lung problem [26]. We clarify that Betamethasone as mentioned in Table 3 should not be interpreted as a risk factor for preterm birth but it is rather a proxy indicating the presence of previous pregnancy complications in the clinical history treated using Betamethasone. Pre-existing hypertension is identified as a risk factor for preterm delivery at different stages, this finding is supported by previous research [27]. Our findings confirm the existence of an association between a previous Cesarean section and future preterm deliveries [28].

### 4.3.2 Risk Factors Common to Three Time Gaps

The significant factors common to three time gaps are summarised in Table 4. Previous twin pregnancies are associated with a high risk of preterm delivery in the 30, 90 and 180 days time gaps. This is supported by findings in the literature [17]. The diagnosis code D.V27.2 indicates a previous twin delivery, in which case, the risk ratio for preterm delivery is significantly low. The results indicate that a subsequent pregnancy has a higher chance of being full term. Literature findings suggest that following a twin preterm delivery, there is a significantly higher risk

of a preterm birth in a subsequent singleton pregnancy [29]. Further analysis needs to be done to either confirm or deny these findings, as the diagnosis code does not indicate whether the current pregnancy is singleton or a multiple pregnancy, or if the previous twin pregnancy was a preterm or a full-term delivery. Our findings indicate that a term delivery increases the chances of subsequent term deliveries.

#### 4.3.3 Risk Factors Common to two Time Gaps

These factors are given in Table 5. A previous high risk pregnancy is reflected by the diagnosis code D\_644.03 threatened premature labor, presents a high risk for the current pregnancy to be of high risk. Multiple gestation of triplets is a common risk factor in the 30 and 90 days time gaps. Repair of internal cervical os and previous obstetrical injury to pelvic organs common to 30 and 90 days, and 90 and 180 days respectively. It is established that weakness of the cervix triggers preterm birth [30]. The diagnosis code D\_V89.02 indicates a medical follow-up for a suspected problem that turned out to be negative at later stages of the pregnancy. This code is shown to increase the chances of a full term delivery.

#### 4.3.4 Risk Factors at 30 Days

The results summary can be found in table 6. A history of previous preterm delivery is identified as a significant risk of preterm delivery in subsequent pregnancies at the 30 days history gap. This is reflected through the diagnosis code D\_V23.41, and the medication Terbutaline Sulfate. The latter is a  $\beta_2$  adrenergic receptor agonist, which works as anti contraction agent to reduce uterine contraction and prevent preterm labor [31, 32].

Patients with placenta previa or premature separation of placenta, with ICD9 codes D\_641.13 and D\_641.23 have a higher rate of preterm, and a higher rate of reoccurrence of preterm birth [33]. Women with congenital uterine abnormalities are at a higher risk of preterm birth [34]. The medication Macrobid, in Table 6, is used for treating urinary tract infections. Urinary tract infection is a risk factor of preterm delivery as confirmed by the literature [35].

The link between some viral infection such as HIV infections, Hepatitis B and preterm birth has been established in the literature [36, 37]. This is reflected in our findings in the lab orders L\_41461-5 (DNA PCR Misc Virus) and L\_31201-7 (HIV 1+2 Antibody Qualitative Serum or Plasma Immunoassay), in table reference , which are tests for such viral infections. Rhesus isoimmunization is also identified as a risk factor for preterm delivery, which is also validated by the literature [38].

Our results show that lab orders with Streptococcus B is associated with have higher chances of a full-term delivery. However our data did not include whether the outcome of test was positive or negative. Previous work the literature give more details on the association of Streptococcus B with preterm birth [39]. This study concluded that presence of Streptococcus B is rather a risk factor for preterm delivery.

The diagnosis code D\_V74.9 indicates that a test for a bacterial or spirochetal disease is conducted, but it does not show whether or not the tests are positive. We can speculate that since the risk ratio is low it indicates negative results of the test, and that if a bacterial disease is non existent, chances of a full-term delivery are higher. This needs to be confirmed with additional data on the lab test results.

#### 4.3.5 Risk Factors at 90 Days

For this section, table 7 summarises the results. A previous delivery by a midwife is found to increase the risk of preterm delivery in the 90 days time gap. Literature findings indicate that continuous midwife care lowers the risk of preterm delivery when compared to other models of care [40]. Certified nurse midwives are also found to have excellent birth outcomes when compared to physician deliveries [41]. The findings are not supported by literature, this could be due to the small number of previous deliveries by midwives in our study, where the number of full-term and preterm midwife deliveries are 37 and 14 respectively. Further analysis is needed to make a conclusion regarding this issue.

The medication Brethine (Terbutaline) is administered to suppress contractions in women with preterm labors [42]. The identification of this medication to be associated with preterm birth should be read as a correlation with previous preterm delivery as a risk factor in the 90 days time gap. Hence previous preterm-birth are risk factors for preterm birth.

The glucose tolerance test in pregnancy is performed in two steps. In the first step the patient drinks a liquid that contains glucose, and a blood test is done one hour after to check the blood glucose level. This is indicated by the lab order Glucose GTT, Serum 1 hr. This lab order is shown to increase the chances of full term deliveries. If the blood glucose is high after the first test, a second test is performed. For the second test, the patient is asked to fast for 8 to 14 hours, and drink a glucose liquid afterwards. Blood is drawn before drinking the liquid and every 60 minutes afterwards. This test is indicated by the lab order Glucose GTT, Serum 3 hr and is shown to increase the risk of preterm delivery. Since the test is only performed if the blood glucose is high after the first test, it indicates the presence of diabetes condition. Type 1 and type 2 diabetes are both known risk factors for preterm delivery [43, 44].

#### 4.3.6 Risk Factors at 180 Days

Risk factors that are identified at 180 or 270 days before delivery as shown in Tables 8 and 9 respectively are most likely not relevant to the current pregnancy. They reflect preexisting chronic conditions in the mother's history and general predisposition to deliver on preterm. Infertility problems are some of the factors that increase the risk of preterm delivery [45]. Women of Asian race are more likely to deliver full term [17].

#### 4.3.7 Risk Factors at 270 Days

The only risk factor that is specific to the 270 days time gap is 'poor fetal growth affecting management of mother'. This is indicated in Table 9. This diagnosis code applies to newborns that are small or light for gestational age in a previous pregnancy.

### 4.4 Limitations

While our study supports the importance EHR data in the prediction of preterm delivery, it highlighted several limitations with this method. The accuracy of selection of the appropriate diagnosis or procedure codes is questioned. Similarly, in order

for the laboratory values to make sense, they need to be stratified into positive or negative values. A more clinically relevant prediction model will be able to combine the risk factors and produce a risk index.

## 5 Conclusion

In this work, we successfully used electronic health records data to extract risk factors of preterm delivery. We also explored the progression of these risk factors across different stages of the pregnancy. Risk factors before and in early pregnancy related to chronic condition and general predisposition such as hypertension, diabetes and anatomy. Risk factors at advanced stage of pregnancy are mostly related to the current pregnancy and reflect acute condition. Follow-up visits and Lab tests are identified to have preventive effect on preterm delivery. Some of the risk factors that are known in the literature were not identified using this method. This could be explained by the stringency of the significance level of the p-value (0.01) used. A higher p-value would lead to the identification of other risk factors, but will also result in extracting false positives. The results from this experiment are easily reproducible to other diseases. Some risk factors can change over time with the change in the society such as lifestyle, income level etc. Re-evaluating the risk factors will be easy using electronic health records. As future work, we plan to extend this analysis by adding other modalities (microbiology, physiological measurements) and test results of lab order to better refine the obtained results.

### Acknowledgements

We would like to thank Haithum Elhadi for his support on this project.

### Author's contributions

ZS implemented the method, conducted the experiments, and drafted the manuscript. SB prepared the data, implemented part of the method, supervised the work, and edited the manuscript. HA gave feedback and edited the manuscript. MM supervised the work. All authors read and approved the final manuscript.

### Funding

This work is funded by Sidra Internal Research Funds: SDR200033

### Availability of data and materials

The data is part of Cerner Health Facts. It is protected by copyright and obtained from Cerner through a Data Use Agreement.

### Ethics approval and consent to participate

The project obtained from Sidra IRB office an Exempt IRB protocol # 1807026901

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Systems Biology, Sidra Medicine, Doha, Qatar. <sup>2</sup>Division of Neonatology, Sidra Medicine, Doha, Qatar. <sup>3</sup>Department of Maternal-Fetal Medicine, Sidra Medicine, Doha, Qatar.

### References

1. A. Dbstet, Who: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths, *Acta Obstet Gynecol Scand* 56 (3) (1977) 247–53.
2. S. Chawanpaiboon, J. P. Vogel, A.-B. Moller, P. Lumbiganon, M. Petzold, D. Hogan, S. Landoulsi, N. Jampathong, K. Kongwattanakul, M. Laopaiboon, et al., Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis, *The Lancet Global Health* 7 (1) (2019) e37–e46.
3. S. Chawanpaiboon, J. P. Vogel, A.-B. Moller, P. Lumbiganon, M. Petzold, D. Hogan, S. Landoulsi, N. Jampathong, K. Kongwattanakul, M. Laopaiboon, et al., Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis, *The Lancet Global Health* 7 (1) (2019) e37–e46.

4. S. Johnson, N. Marlow, Early and long-term outcome of infants born extremely preterm, *Archives of disease in childhood* 102 (1) (2017) 97–102.
5. S. Johnson, E. Hennessy, R. Smith, R. Trikić, D. Wolke, N. Marlow, Academic attainment and special educational needs in extremely preterm children at 11 years of age: the epicure study, *Archives of Disease in Childhood-Fetal and Neonatal Edition* 94 (4) (2009) F283–F289.
6. J. Bryce, C. Boschi-Pinto, K. Shibuya, R. E. Black, WHO estimates of the causes of death in children, *The Lancet* 365 (9465) (2005) 1147–1152. doi:10.1016/S0140-6736(05)71877-8.
7. D. Wolke, Psychological development of prematurely born children, *Archives of Disease in Childhood* 78 (6) (1998) 567–570. doi:10.1136/adc.78.6.567.
8. Y. Zhang, T. E. Inder, J. J. Neil, D. L. Dierker, D. Alexopoulos, P. J. Anderson, D. C. Van Essen, Cortical structural abnormalities in very preterm children at 7 years of age, *NeuroImage* 109 (2015) 469–479. doi:10.1016/j.neuroimage.2015.01.005.
9. C. Wocadlo, I. Rieger, Motor impairment and low achievement in very preterm children at eight years of age, *Early Human Development* 84 (11) (2008) 769–776. doi:10.1016/j.earlhumdev.2008.06.001.
10. N. K. Goyal, A. G. Fiks, S. A. Lorch, Association of late preterm birth with asthma in young children: Practice based study, *Pediatrics* 128 (4) (2011) e830–e838. doi:10.1542/peds.2011-0809.
11. U. M. Reddy, M. M. Rice, W. A. Grobman, J. L. Bailit, R. J. Wapner, M. W. Varner, J. M. Thorp, K. J. Leveno, S. N. Caritis, M. Prasad, A. T. N. Tita, G. R. Saade, Y. Sorokin, D. J. Rouse, S. C. Blackwell, J. E. Tolosa, C. Spong, S. Tolivaisa, M. Talucci, M. Zylfijaj, Z. Reid, R. Leed, J. Benson, S. Forester, C. Kitto, S. Davis, M. Falk, C. Perez, K. Hill, A. Sowles, J. Postma, S. Alexander, G. Andersen, V. Scott, V. Morby, K. Jolley, J. Miller, B. Berg, K. Dorman, J. Mitchell, E. Kaluta, K. Clark, K. Spicer, S. Timlin, K. Wilson, L. Moseley, M. Santillan, J. Price, K. Buentipo, V. Bludau, T. Thomas, L. Fay, C. Melton, J. Kingsbery, R. Benezue, H. Simhan, M. Bickus, D. Fischer, T. Kamon, D. DeAngelis, B. Mercer, C. Milluzzi, W. Dalton, T. Dotson, P. McDonald, C. Brezine, A. McGrail, C. Latimer, L. Guzzo, F. Johnson, L. Gerwig, S. Fyffe, D. Loux, S. Frantz, D. Cline, S. Wylie, J. Iams, M. Wallace, A. Northen, J. Grant, C. Colquitt, D. Rouse, W. Andrews, G. Mallett, M. Ramos-Brinson, A. Roy, L. Stein, P. Campbell, C. Collins, N. Jackson, M. Dinsmoor, J. Senka, K. Paychek, A. Peaceman, J. Moss, A. Salazar, A. Acosta, G. Hankins, N. Hauff, L. Palmer, P. Lockhart, D. Driscoll, L. Wynn, C. Sudz, D. Dengate, C. Girard, S. Field, P. Breault, F. Smith, N. Annunziata, D. Allard, J. Silva, M. Gamage, J. Hunt, J. Tillinghast, N. Corcoran, M. Jimenez, F. Ortiz, P. Givens, B. Rech, C. Moran, M. Hutchinson, Z. Spears, C. Carreno, B. Heaps, G. Zamora, J. Seguin, M. Rincon, J. Snyder, C. Farrar, E. Lairson, C. Bonino, W. Smith, K. Beach, S. Van Dyke, S. Butcher, E. Thom, Y. Zhao, P. McGee, V. Momirova, R. Palugod, B. Reamer, M. Larsen, T. Spangler, V. Bhandaru, J. P. VanDorsten, Serious maternal complications after early preterm delivery (24–33 weeks' gestation), *American Journal of Obstetrics and Gynecology* 213 (4) (2015) 538.e1–538.e9.
12. M. K. Menard, Toward achieving risk-appropriate maternity care: Maternal morbidity prediction, *Obstetrics & Gynecology* 134 (2) (2019) 213–215.
13. M. R. Cowie, J. I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay, A. Michel, S. Ong, J. P. Pell, M. R. Southworth, W. G. Stough, M. Thoenes, F. Zannad, A. Zalewski, Electronic health records to facilitate clinical research, *Clinical Research in Cardiology* 106 (1) (2017) 1–9. doi:10.1007/s00392-016-1025-6.
14. M. Ip, E. Peyman, V. Lohsoonthorn, M. A. Williams, A case-control study of preterm delivery risk factors according to clinical subtypes and severity, *The journal of obstetrics and gynaecology research* 36 (1) (2010) 34–44. doi:10.1111/j.1447-0756.2009.01087.x.
15. J. P. Vogel, S. Chawanpaiboon, A.-B. Moller, K. Watananirun, M. Bonet, P. Lumbiganon, The global epidemiology of preterm birth, *Best Practice & Research Clinical Obstetrics & Gynaecology* 52 (2018) 3–12. doi:10.1016/j.bpobgyn.2018.04.003.
16. S. E. Purisch, C. Gyamfi-Bannerman, Epidemiology of preterm birth, *Seminars in Perinatology* 41 (7) (2017) 387–391. doi:10.1053/j.semperi.2017.07.009.
17. R. L. Goldenberg, J. F. Culhane, J. D. Iams, R. Romero, Epidemiology and causes of preterm birth, *The Lancet* 371 (9606) (2008) 75–84. doi:10.1016/S0140-6736(08)60074-4.
18. Q. Zhang, C. V. Ananth, Z. Li, J. C. Smulian, Maternal anaemia and preterm birth: a prospective cohort study, *International Journal of Epidemiology* 38 (5) (2009) 1380–1389. doi:10.1093/ije/dyp243.
19. X. Ji, X. Meng, C. Liu, R. Chen, Y. Ge, L. Kan, Q. Fu, W. Li, L. A. Tse, H. Kan, Nitrogen dioxide air pollution and preterm birth in Shanghai, China, *Environmental Research* 169 (2019) 79–85. doi:10.1016/j.envres.2018.11.007.
20. L. A. Snapper, K. L. Hart, K. K. Venkatesh, A. J. Kaimal, R. H. Perlis, Cohort study of the relationship between individual psychotherapy and pregnancy outcomes, *Journal of Affective Disorders* 239 (2018) 253–257. doi:10.1016/j.jad.2018.05.083.
21. D. E. Robinson, T. P. v. Staa, E. M. Dennison, C. Cooper, W. G. Dixon, The limitations of using simple definitions of glucocorticoid exposure to predict fracture risk: A cohort study, *Bone* 117 (2018) 83–90. doi:10.1016/j.bone.2018.09.004.
22. J. D. Mosley, Q. Feng, Q. S. Wells, S. L. V. Driest, C. M. Shaffer, T. L. Edwards, L. Bastarache, W.-Q. Wei, L. K. Davis, C. A. McCarty, W. Thompson, C. G. Chute, G. P. Jarvik, A. S. Gordon, M. R. Palmer, D. R. Crosslin, E. B. Larson, D. S. Carrell, I. J. Kullo, J. A. Pacheco, P. L. Peissig, M. H. Brilliant, J. G. Linneman, B. Namjou, M. S. Williams, M. D. Ritchie, K. M. Borthwick, S. S. Verma, J. H. Karnes, S. T. Weiss, T. J. Wang, C. M. Stein, J. C. Denny, D. M. Roden, A study paradigm integrating prospective epidemiologic cohorts and electronic health records to identify disease biomarkers, *Nature Communications* 9 (1) (2018) 3522. doi:10.1038/s41467-018-05624-4.
23. A. Banerji, K. G. Blumenthal, K. H. Lai, L. Zhou, Epidemiology of ACE Inhibitor Angioedema Utilizing a Large Electronic Health Record, *The Journal of Allergy and Clinical Immunology: In Practice* 5 (3) (2017) 744–749. doi:10.1016/j.jaip.2017.02.018.
24. D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, *Applied logistic regression*, Vol. 398, John Wiley & Sons,

- 2013.
25. J. Zhang, K. F. Yu, What's the Relative Risk?: A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes, *JAMA* 280 (19) (1998) 1690–1691. doi:10.1001/jama.280.19.1690.
  26. C. A. Crowther, C. J. McKinlay, P. Middleton, J. E. Harding, Repeat doses of prenatal corticosteroids for women at risk of preterm birth for improving neonatal health outcomes, *The Cochrane database of systematic reviews* (6) (2011) CD003935. doi:10.1002/14651858.CD003935.pub3.
  27. J. M. Catov, E. A. Nohr, J. Olsen, R. B. Ness, **Chronic Hypertension Related to Risk for Preterm and Term Small-for-Gestational-Age Births**, *Obstetrics and gynecology* 112 (2 Pt 1) (2008) 290–296. doi:10.1097/AOG.0b013e31817f589b.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2596352/>
  28. C. M. Williams, I. Asaolu, N. R. Chavan, L. H. Williamson, A. M. Lewis, L. Beaven, K. B. Ashford, Previous cesarean delivery associated with subsequent preterm birth in the United States, *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 229 (2018) 88–93. doi:10.1016/j.ejogrb.2018.08.013.
  29. M. K. Menard, R. B. Newman, A. Keenan, M. Ebeling, **Prognostic significance of prior preterm twin delivery on subsequent singleton pregnancy**, *American Journal of Obstetrics and Gynecology* 174 (5) (1996) 1429–1432. doi:10.1016/S0002-9378(96)70584-7.  
URL <http://www.sciencedirect.com/science/article/pii/S0002937896705847>
  30. J. P. Nott, E. A. Bonney, J. D. Pickering, N. A. B. Simpson, **The structure and function of the cervix during pregnancy**, *Translational Research in Anatomy* 2 (2016) 1–7. doi:10.1016/j.tria.2016.02.001.  
URL <http://www.sciencedirect.com/science/article/pii/S2214854X1530008X>
  31. J. F. King, A. Grant, M. J. Keirse, I. Chalmers, Beta-mimetics in preterm labour: an overview of the randomized controlled trials, *British Journal of Obstetrics and Gynaecology* 95 (3) (1988) 211–222.
  32. F. Lam, J. Elliott, J. S. Jones, M. Katz, R. A. Knuppel, J. Morrison, R. Newman, J. Phelan, R. Willcourt, Clinical issues surrounding the use of terbutaline sulfate for preterm labor, *Obstetrical & Gynecological Survey* 53 (11 Suppl) (1998) S85–95.
  33. O. Erez, L. Novack, V. Klaitman, I. Erez-Weiss, R. Beer-Weisel, D. Dukler, M. Mazor, **Early preterm delivery due to placenta previa is an independent risk factor for a subsequent spontaneous preterm birth**, *BMC Pregnancy and Childbirth* 12 (2012) 82. doi:10.1186/1471-2393-12-82.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3489587/>
  34. M. Hua, A. O. Odibo, R. E. Longman, G. A. Macones, K. A. Roehl, A. G. Cahill, **Congenital uterine anomalies and adverse pregnancy outcomes**, *American Journal of Obstetrics and Gynecology* 205 (6) (2011) 558.e1–558.e5. doi:10.1016/j.ajog.2011.07.022.  
URL <http://www.sciencedirect.com/science/article/pii/S0002937811009380>
  35. I. Verma, K. Avasthi, V. Berry, **Urogenital Infections as a Risk Factor for Preterm Labor: A Hospital-Based Case–Control Study**, *Journal of Obstetrics and Gynaecology of India* 64 (4) (2014) 274–278. doi:10.1007/s13224-014-0523-6.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4126939/>
  36. X. Ma, D. Sun, C. Li, J. Ying, Y. Yan, **Chronic hepatitis b virus infection and preterm labor(birth) in pregnant women—an updated systematic review and meta-analysis**, *Journal of Medical Virology* 90 (1) (2018) 93–100. doi:10.1002/jmv.24927.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.24927>
  37. P.-L. Xiao, Y.-B. Zhou, Y. Chen, M.-X. Yang, X.-X. Song, Y. Shi, Q.-W. Jiang, **Association between maternal hiv infection and low birth weight and prematurity: a meta-analysis of cohort studies**, *BMC Pregnancy and Childbirth* 15 (1) (2015) 246. doi:10.1186/s12884-015-0684-z.  
URL <https://doi.org/10.1186/s12884-015-0684-z>
  38. J. Fan, B. K. Lee, A. T. Wikman, S. Johansson, M. Reilly, **Associations of rhesus and non-rhesus maternal red blood cell alloimmunization with stillbirth and preterm birth**, *International Journal of Epidemiology* 43 (4) (2014) 1123–1131. doi:10.1093/ije/dyu079.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4258779/>
  39. F. Bianchi-Jassir, A. C. Seale, M. Kohli-Lynch, J. E. Lawn, C. J. Baker, L. Bartlett, C. Cutland, M. G. Gravett, P. T. Heath, M. Ip, K. Le Doare, S. A. Madhi, S. K. Saha, S. Schrag, A. Sobanjo-ter Meulen, J. Vekemans, C. E. Rubens, **Preterm birth associated with group b streptococcus maternal colonization worldwide: Systematic review and meta-analyses**, *Clinical Infectious Diseases* 65 (suppl.2) (2017) S133–S142. doi:10.1093/cid/cix661.  
URL [https://academic.oup.com/cid/article/65/suppl\\_2/S133/4589591](https://academic.oup.com/cid/article/65/suppl_2/S133/4589591)
  40. J. Sandall, H. Soltani, S. Gates, A. Shennan, D. Devane, **Midwife-led continuity models versus other models of care for childbearing women**, *Cochrane Database of Systematic Reviews* (8). doi:10.1002/14651858.CD004667.pub3.  
URL <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004667.pub3/abstract>
  41. M. F. MacDorman, G. K. Singh, **Midwifery care, social and medical risk factors, and birth outcomes in the USA**, *Journal of Epidemiology and Community Health* 52 (5) (1998) 310–317.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756707/>
  42. S. Motazedian, F. Ghaffarpassand, K. Mojtahedi, N. Asadi, **Terbutaline versus salbutamol for suppression of preterm labor: a randomized clinical trial**, *Annals of Saudi Medicine* 30 (5) (2010) 370–375. doi:10.4103/0256-4947.67079.  
URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2941249/>
  43. M. Bashir, E. Naem, F. Taha, J. C. Konje, A.-B. Abou-Samra, **Outcomes of type 1 diabetes mellitus in pregnancy; effect of excessive gestational weight gain and hyperglycaemia on fetal growth**, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 13 (1) (2019) 84–88. doi:10.1016/j.dsx.2018.08.030.  
URL <http://www.sciencedirect.com/science/article/pii/S1871402118303448>
  44. M. Bashir, Z. Dabbous, K. Baagar, F. Elkhatib, A. Ibrahim, S.-A. Brich, M. E. Abdel-Rahman, J. C. Konje, A.-B. Abou-Samra, **Type 2 diabetes mellitus in pregnancy: The impact of maternal weight and early glycaemic**

control on outcomes, European Journal of Obstetrics & Gynecology and Reproductive Biology 233 (2019) 53–57. doi:10.1016/j.ejogrb.2018.12.008.

URL <http://www.sciencedirect.com/science/article/pii/S0301211518311333>

45. M. Filicori, G. E. Cognigni, E. Gamberini, E. Troilo, L. Parmegiani, S. Bernardi, [Impact of medically assisted fertility on preterm birth](#), BJOG: An International Journal of Obstetrics & Gynaecology 112 (s1) (2005) 113–117. doi:10.1111/j.1471-0528.2005.00598.x.

URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-0528.2005.00598.x>

## Figures

**Figure 1** Timeline of the clinical history for different pregnancies included in the cohort.

## Tables

**Table 1** Covariate Counts by Category.

Category	30 days	90 days	180 days	270 days
Diagnosis	7677	7541	7362	7222
Medication	11310	10980	10727	10504
Procedure	2581	2489	2386	2262
Lab orders	1959	1917	1894	1886
Marital Status	7	7	7	7
Race	11	11	11	11
Medical Specialty	116	116	115	114
Hospital ID	200	199	197	196
Total	23863	23262	22701	22204

**Table 2** AUC scores.

Number of days	30	90	180	270
AUC	76.00%	74.57%	72.95%	71.72%

**Table 3** Common risk factor across all time gaps.

Code	description	Risk Ratios(95% CI)			
		30 days	90 days	180 days	270 days
D_644.21	Early Onset of Delivery, Delivered, with or without Mention of Antepartum Condition	1.91(1.84-1.98)	1.98(1.91-2.04)	1.94(1.88-2)	1.88(1.82-1.94)
M_85056605.0	Celestone Solus-pan/betamethasone	1.44(1.35-1.53)	1.34(1.21-1.48)	1.43(1.30-1.56)	1.43(1.29-1.57)
D_654.21	Previous Cesarean Section, Delivered, with or without Mention of Antepartum Condition	1.33(1.22-1.44)	1.34(1.23-1.45)	1.43(1.32-1.54)	1.39(1.28-1.51)
P_74.1	Low Cervical Cesarean Section	1.31(1.18-1.45)	1.41([1.29-1.54)	1.46([1.35-1.57)	1.41([1.29-1.52)
D_401.9	Unspecified Essential Hypertension	1.13(1.08-1.18)	1.16(1.10-1.21)	1.22(1.17-1.27)	1.22(1.16-1.27)

**Table 4** Common risk factor across at least 3 time gaps.

Code	description	Risk Ratios(95% CI)			
		30 days	90 days	180 days	270 days
D_651.03	Twin Pregnancy, Antepartum Condition or Complication	1.82(1.75-1.89)	1.96(1.89-2.04)	1.15(1.08-1.22)	
D_V27.2	Mother with Twins, Both Liveborn	0.04(-2.41-2.49)	0.05(-1.87-1.96)	0.15(-1.53-1.84)	
D_645.11	Post Term Pregnancy, Delivered, with or without Mention of Antepartum Condition		0.88(0.84-0.92)	0.89(0.86-0.93)	0.88(0.85-0.92)

**Table 5** Common risk factor across at least 2 time gaps.

Code	description	Risk Ratios(95% CI)			
		30 days	90 days	180 days	270 days
D_651.13	Triplet Pregnancy, Antepartum Condition or Complication	7.68(6.43-8.92)	10.67(9.21-12.12)		
P_67.59	Other Repair of Internal Cervical Os	1.88(1.62-2.14)	1.88(1.62-2.15)		
D_V89.02	Suspected Placental Problem Not Found	0.84(0.69-0.99)	0.85(0.67-1.02)		
D_665.51	Other Obstetrical Injury to Pelvic Organs, Delivered, with or without Mention of Antepartum Condition		1.16(1.08-1.24)	1.13(1.05-1.21)	
D_644.03	Threatened Premature Labor, Antepartum Condition or Complication			1.11(1.08-1.14)	1.10(1.06-1.14)

**Table 6** Risk factors unique to 30 days.

code	description	Risk Ratios(95% CI)
D_641.13	Hemorrhage from Placenta Previa, Antepartum Condition or Complication	1.32(1.20-1.44)
D_654.03	Congenital Abnormalities of Uterus, Antepartum Condition or Complication	1.28(1.16-1.39)
D_641.23	Premature Separation of Placenta, Antepartum Condition or Complication	1.27(1.12-1.43)
M_149071001	Macrobid/nitrofurantoin	1.17(1.08-1.26)
M_55390019310	Terbutaline Sulfate/terbutaline	1.14(1.06-1.22)
L_41461-5	DNA PCR Misc Virus	1.14(1.07-1.21)
L_31201-7	HIV 1+2 Antibody Qualitative Serum or Plasma Immunoassay	1.14(1.08-1.19)
D_656.11	Rhesus Isoimmunization Affecting Management of Mother, Delivered, with or without Mention of Antepartum Condition	1.13(1.08-1.18)
D_V23.41	Pregnancy with History of Pre-Term Labor	1.11(1.06-1.15)
L_19023-1	Granulocyte Percent	0.89(0.80-0.97)
D_V28.6	Encounter for Antenatal Screening for Streptococcus B of Mother	0.77(0.71-0.83)
D_V74.9	Screening Examination for Unspecified Bacterial and Spirochetal Diseases	0.74(0.55-0.94)
L_11266-4	Streptococcus B Antigen	0.61(0.21-1.02)

**Table 7** Risk factors unique to 90 days.

code	description	Risk Ratios(95% CI)
medsp_Midwife_		1.26(1.18-1.34)
M_28750701	Brethine/terbutaline	1.18(1.08-1.29)
L_19105-6	Glucose GTT, Serum 3 hr	1.13(1.11-1.15)
D_654.23	Previous Cesarean Section, Antepartum Condition or Complication	1.11(1.06-1.16)
L_50583-4	Glucose Gestational	0.84(0.76-0.92)
L_10449-7	Glucose GTT, Serum 1 hr	0.84(0.80-0.87)

**Table 8** Risk factors unique to 180 days.

code	description	Risk Ratios(95% CI)
D_V26.21	Fertility Testing	1.12(1.05-1.18)
D_V28.82	Encounter for Screening for Risk of Pre-Term Labor	0.89(0.73-1.04)
race_Asian		0.81(0.61-1.02)
M_6335753	Mefoxin/cefloxitin	0.61(0.15-1.06)

**Table 9** Risk factors unique to 270 days.

code	description	Risk Ratios(95% CI)
D_656.51	Poor Fetal Growth Affecting Management of Mother, Delivered, with or without Mention of Antepartum Condition	1.20(1.09-1.31)

## Figures



Figure 1

Timeline of the clinical history for different pregnancies included in the cohort.