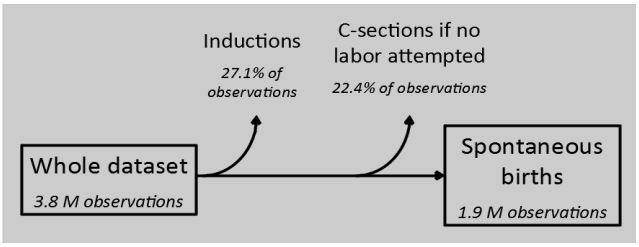
Natality Project: Key findings

Berenice Dethier – Springboard Capstone project

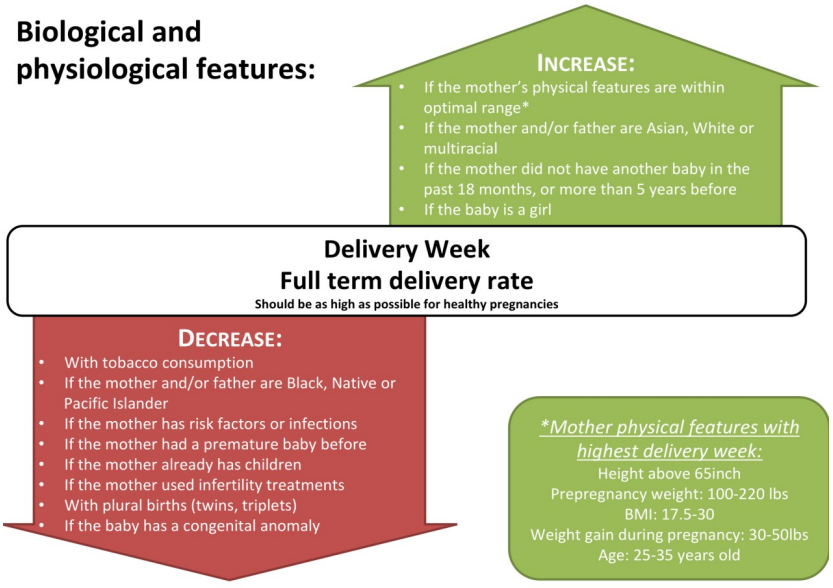
We analyzed the CDC 2018 Natality public file to identify factors affecting the term of birth of women in the United States. Preterm babies are born at 36 weeks of gestation or earlier. The biological, socioeconomic and other factors impacting the term of birth are summarized in this document.



Method

Spontaneous births were the only observations considered (about 1.9 million records). Scheduled c-sections and inductions were filtered out (one half of 3.8 million births in 2018).

Data collection and cleaning was performed with R Studio. Exploratory data analysis was done in two parts: visual EDA with Tableau and statistical analysis in Python Jupyter notebook. A full report of the work, code and details are available on github. Interactive visualizations are available on Tableau Public.



Biological factors

There was an optimal range for most physical features in which women would have a lower risk of having a premature baby.

The race of the parents had a significant impact: Asian and White parents have the longest pregnancies.

Female newborns have lower prematurity rates, as well as first born children.

Previous premature births, and use of infertility treatments were correlated with more premature births.

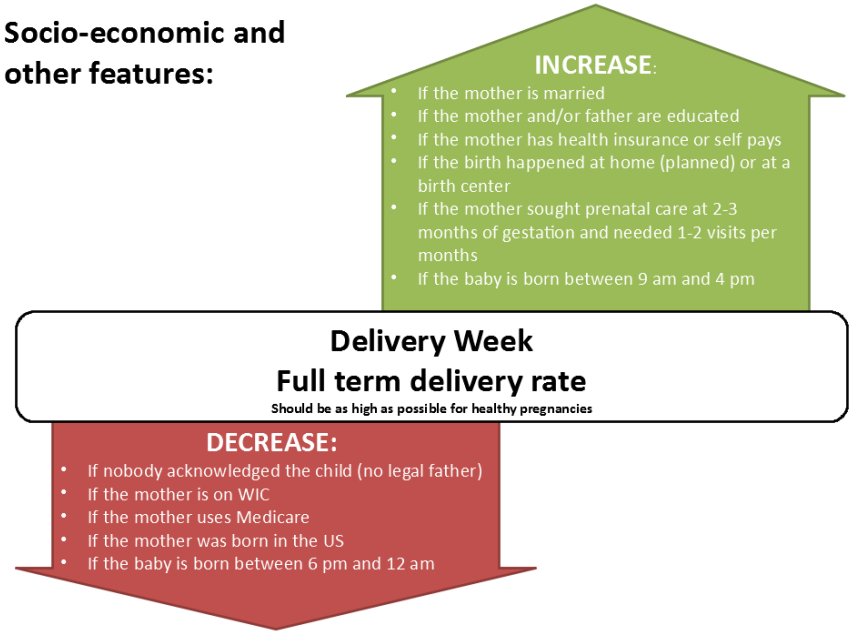
Unsurprisingly, expecting multiple babies, risk factors, infections, and tobacco also increase the prematurity rates.

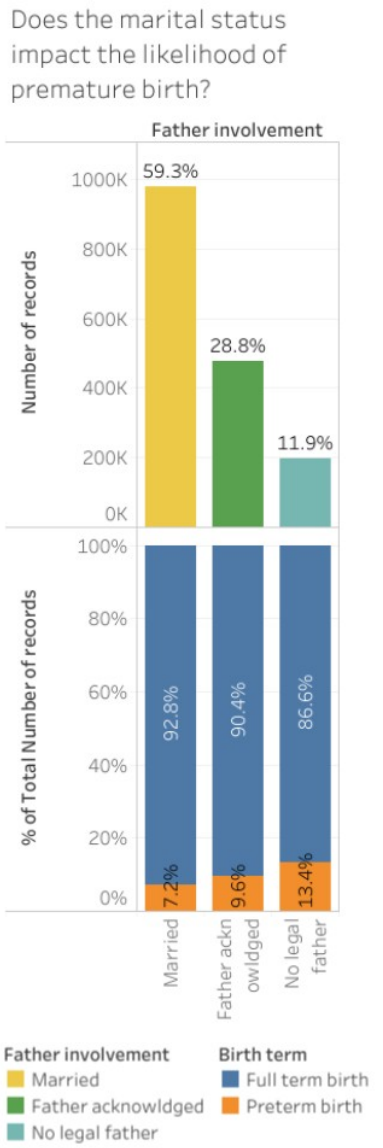
Only one biological factor was not significant in predicting the delivery week: the age of the father.

Socioeconomic factors

Factors associated with higher economic status are correlated with fewer premature births: education of the parents, health insurance, not on WIC.

One interesting correlation was the marital status of the mother: a single mother has almost twice as much risk to deliver prematurely than a married mother, even when controlling for age or race.





Other findings

The place of birth of the mother (in the US or not) also impacts the prematurity rates.

Surprisingly, the month of birth and time of birth are correlated with the rate of premature births. Babies born between 6 pm and 12 am are more likely to be premature.

The outcomes for the baby and mother were analyzed. The birth weight, maternal morbidity and breastfeeding rates were on average lower for premature babies. The rate of abnormal conditions on the other hand was higher for premature babies.

Conclusions

The differences between premature and full term babies presented in this document have been confirmed to be of statistical significance.

The project intends to identify the described trends in order to use them later for prediction. Our goal was not to explain them.