Data Section - 2015 U.S. Natality Data

Problem Set #1, MACS 30200

Chih-Yu Chiang

This research examines potential relations between parents' demographic backgrounds and newborns' health conditions by using the 2015 U.S. Natality Data[1]. The U.S. Natality Data consists of newborn's birth information, health conditions, and parents' geographic, demographic, and medical data of all States of the United States. It contains data of 4 million individual births occurred in 2015 within the United States, regardless parents' residential condition. The data is part of the National Vital Statistics System and is curated by National Center for Health Statistics (NCHS), affiliated to The Centers for Disease Control and Prevention, a federal agency under the Department of Health and Human Services. By federal laws, NCHS collects national birth data, dating back to 1985, by compiling information filed in birth certificates, which are compulsory for all childbirth in the U.S.

Based on Natality Data, NCHS publishes national statistics reports on teen birth (Ventura, Hamilton, & Matthews, 2014), infant mortality (MacDorman et al., 2014), and other health issues of the country[2]. The data is widely used in gynecology and obstetrics research as its primary source of birth data, in sub-fields like gestation (Zhang & Bowes Jr, 1995; Alexander et al., 1996; Davidoff et al., 2006), delivery (MacDorman, Menacker, & Declercq, 2008; Bettegowda, 2008; Menacker & Hamilton, 2010), and unintended pregnancy (Finer & Henshaw, 2006; Finer & Zolna, 2011; 2016). Other research combines NCHS's natality information with extra data for inter-disciplinary investigations such as in family planning policy (Dickert-Conlin & Chandra, 1999) and infant hospitalization (Holman, 2003). Another key area of this data's application is in newborn abnormal conditions (Oken et al., 2003; Boulet et al.,

[1] The 2015 U.S. Natality Data can be accessed at the website of Centers for Disease Control and Prevention (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Births) or, alternatively, through the website of The National Bureau of Economic Research (http://nber.org/data/vital-statistics-natality-data.html).

[2] Other publications by National Center for Health Statistics (NCHS) can be found on its website: https://www.cdc.gov/nchs/nvss/birth_products.htm.

2006; MacDorman et al., 2008). The U.S. Natality Data provides detailed records of newborn illness diagnosis of anencephaly, cleft lip, heart disease, and several others. In this research, these illness data will be compared between different demographic groups of parents, to gain insights into relationships between newborn health conditions and parent racial and educational backgrounds.

This study employs 3,289,382 individual newborn observations of the United States in 2015, with 7 measures of parent demographic backgrounds of race, age, education, and marital status (Table 1), 12 illnesses markers of specific diseases, and a 5-minute APGAR score (Table 2) that evaluates health summary of newborn children on a 0 to 10 scale; the higher the score, the better health condition of the baby.

Table 1. Variables of Parents' Demographic Background

	Mother's Age	Mother's Education	Father's Age	Father's Education	Marital Status
min	12	1	13	1	1
max	50	8	98	8	2
median	29	4	31	4	1
mean	28.9056	4.4304	31.3499	4.1911	1.3209
std.dev	5.6933	1.7441	6.7865	1.7551	0.4668

* For the education, 1 represents 8th grade or less; 8 represents doctorate or professional degree; 2 to 7 includes associate, bachelor, and other degree levels in between.

** For the marital status 1 represents married; 2 represents unmarried.

Table 2. Variables of Newborn's Health Condition

	5-min APGAR Score	Anencephaly	Meningomyelocele / Spina Bifida	Cyanotic Congenital Heart Disease	Congenital Diaphragmatic Hernia	Omphalocele	Gastroschisis
min	0	0	0	0	0	0	0
max	10	1	1	1	1	1	1
median	9	0	0	0	0	0	0
mean	8.79929	0.00010	0.00012	0.00059	0.00012	0.00009	0.00024
std.dev	0.78998	0.00975	0.01115	0.02428	0.01086	0.00966	0.01555

	Limb Reduction Defect	Cleft Lip	Cleft Palate alone	Down Syndrome	Suspected Chromosomal Disorder	Hypospadias
min	0	0	0	0	0	0
max	1	1	1	1	1	1
median	0	0	0	0	0	0
mean	0.00013	0.00053	0.00022	0.00051	0.00041	0.00059
std.dev	0.01151	0.02305	0.01479	0.02265	0.02021	0.02425

* For the diseases, 1 represents diagonosis; 0 represents non-diagnosis.

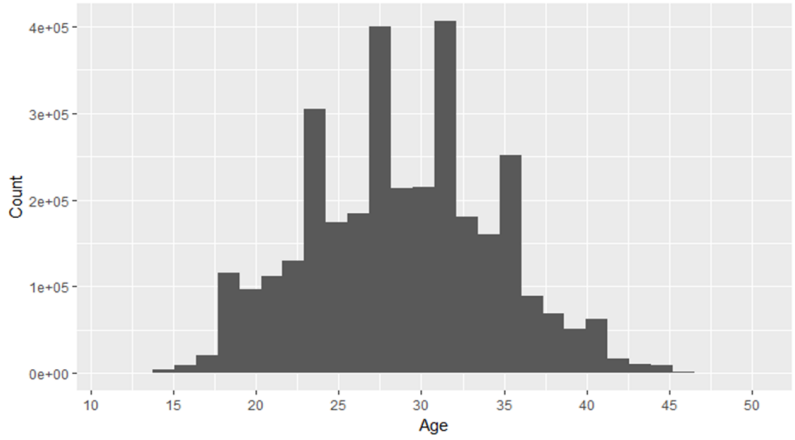
** For the APGAR score, from 0 to 10, the higher the better health condition of the newborn.

On average, male parent is at age 31.35, 2.44 years older than the female parent at 28.91. Somewhat violates common intuition, the male parent has a lower education level than the female parent on the 1 to

8 scale applied here[3]. In 2015, 32% of the newborns are delivered by unmarried couples. Regarding newborn health conditions, illness occurrence percentages are generally low, all below 0.1% of all newborn children.

Compared to females in nearby age groups, significantly more females get pregnant at age 24, 26, 32, and 35 (Figure 1). 24 is the age people just finishing college degrees, and 26 the Master's. They represent transformative points of an individual's life, which could lead to people's decision of having a baby. The other ages could signify physiological pressures coming from the risk of having newborn illness. For instance, the occurrence rate of Down syndrome increases rapidly after age 30 and further spikes after 35.

Figure 1.
Mother's Age Distribution (bin = 30)



The correlation between variables of parent age, education level, and APGAR score is provided in Table 3. In general, parent ages and educations positively correlate with each other. Especially between ages of male and female parents and between education levels of male and corresponding female parents—both pairs have correlation coefficients over 0.7—conforms to the fact that people tend to seek partners at a similar age and a comparable educational level as they have. Surprisingly, associations are not obvious between these parent demographic factors and newborn's health condition, measured by the

[3] The scales are coded as following: 1 = 8th grade or less; 2 = 9th to 12th grade without diploma; 3 = high school graduate; 4 = college credit without degree; 5 = Associate degree; 6 = Bachelor's degree; 7 = Master's degree; 8 = Doctorate or professional degree

APGAR score—all coefficients are near to 0. The low correlation implicates a non-linear relationship between the demographic elements and the health conditions. The relationship can be observed and will be discussed further later.

Table 3. Pearson's Correlation Coefficients between Parents' Demographic Background and Newborn's Health Condition

	Mother's Age	Mother's Education	Father's Age	Father's Education	5-min APGAR Score
Mother's Age	1.0000	0.3881	0.7394	0.3414	0.0122
Mother's Education	0.3881	1.0000	0.2747	0.7043	0.0135
Father's Age	0.7394	0.2747	1.0000	0.2751	0.0123
Father's Education	0.3414	0.7043	0.2751	1.0000	0.0156
5-min APGAR Score	0.0122	0.0135	0.0123	0.0156	1.0000

In Figure 2 and Figure 3, I compared the APGAR scores between different racial groups. Regardless parent's sex, a similar pattern is revealed, because people tend to seek partners of the same racial group as shown in Table 4 and Table 5, from perspectives of female parent and male parent respectively. In general, Asian parents have babies with the best health condition (with average APGAR score 8.85), followed by Hispanics and Whites; Black, native American (AIAN), and native Pacific Islander (NHOPI) have the worst health reports (with average APGAR scores slightly higher than 8.7).

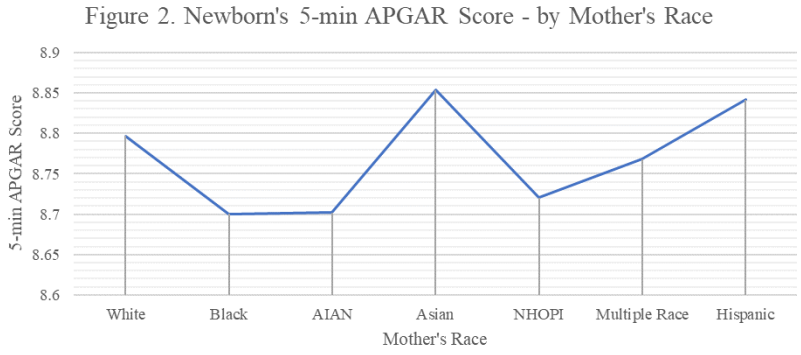


Figure 3. Newborn's 5-min APGAR Score - by Father's Race

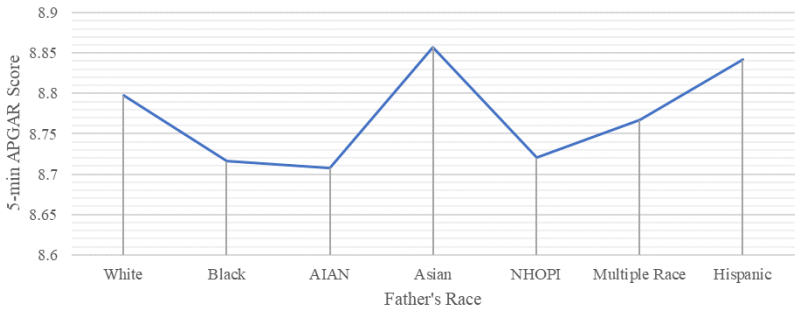


Table 4. Partner's Race Distribution by Mother's Races

Mother's Race	White	Black	AIAN	Asian	NHOPI	Multiple Race	Hispanic
	Partner's Race						
White	88.22%	2.99%	0.39%	0.85%	0.08%	1.52%	5.96%
Black	4.46%	90.39%	0.11%	0.26%	0.04%	1.52%	3.22%
AIAN	28.21%	4.65%	50.83%	0.62%	0.31%	3.12%	12.26%
Asian	14.55%	1.70%	0.07%	78.31%	0.14%	1.76%	3.46%
NHOPI	13.27%	7.79%	0.43%	2.67%	62.00%	5.34%	8.50%
Multiple Race	40.81%	20.64%	1.16%	3.65%	0.99%	19.25%	13.49%
Hispanic	12.26%	3.78%	0.23%	0.69%	0.11%	1.19%	81.76%

Table 5. Partner's Race Distribution by Father's Races

Father's Race	White	Black	AIAN	Asian	NHOPI	Multiple Race	Hispanic
	Partner's Race						
White	90.22%	0.90%	0.36%	1.81%	0.05%	1.41%	5.26%
Black	12.82%	76.13%	0.25%	0.89%	0.13%	2.98%	6.80%
AIAN	32.87%	1.77%	53.21%	0.74%	0.14%	3.30%	7.98%
Asian	7.81%	0.47%	0.07%	87.77%	0.10%	1.13%	2.66%
NHOPI	17.64%	1.88%	0.92%	3.98%	57.19%	7.85%	10.53%
Multiple Race	46.90%	9.21%	1.20%	6.62%	0.65%	20.04%	15.38%
Hispanic	14.21%	1.51%	0.36%	1.00%	0.08%	1.08%	81.75%

In Figure 4 and Figure 5, the occurrences of specific illnesses are examined by mother and father's racial group, respectively. In general, they confirm my observation with the APGAR scores, with the data line representing Asian parents at the bottom. According to the figures, some illnesses are more common to certain racial groups. For example, White parents suffer from higher probability to have babies with hypospadias. On the contrary, some diseases are connected more strongly with one side of the parents than another. For instance, the relations between AIAN fathers and gastroschisis and cleft lip babies are

more obvious than between AIAN mothers and the same ailments; on the other hand, Down syndrome and chromosomal disorder are less related with AIAN fathers than with AIAN mothers. These associations could be attributed to demographic groups' special DNA sequences or unique life styles, and require further investigations to determine.

Figure 4. Newborn's Health Condition Occurrence - by Mother's Race

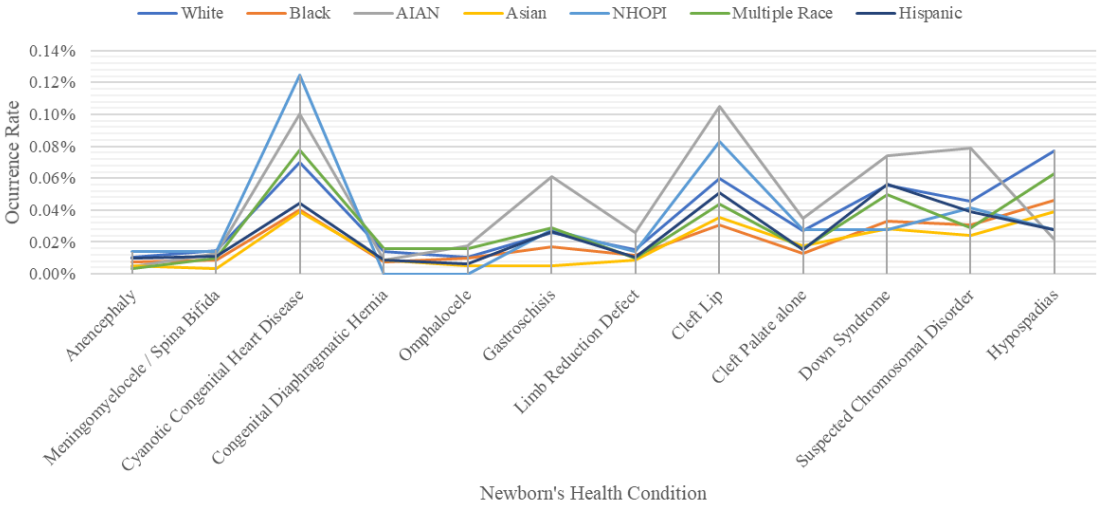
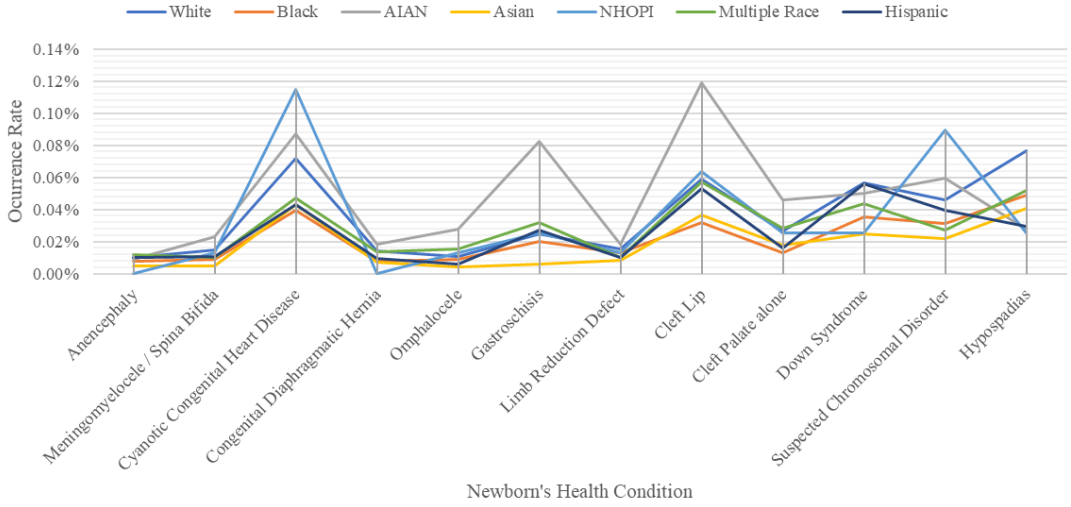


Figure 5. Newborn's Health Condition Occurrence - by Father's Race



Generally, illness occurrence rates decrease as parents' education levels increase (Figure 6, 7). The pattern is more obvious to certain diseases, such as Down syndrome, chromosomal disorder, gastroschisis, and cleft lip, but not the others.

Figure 6. Newborn's Health Condition Occurrence - by Mother's Education Level

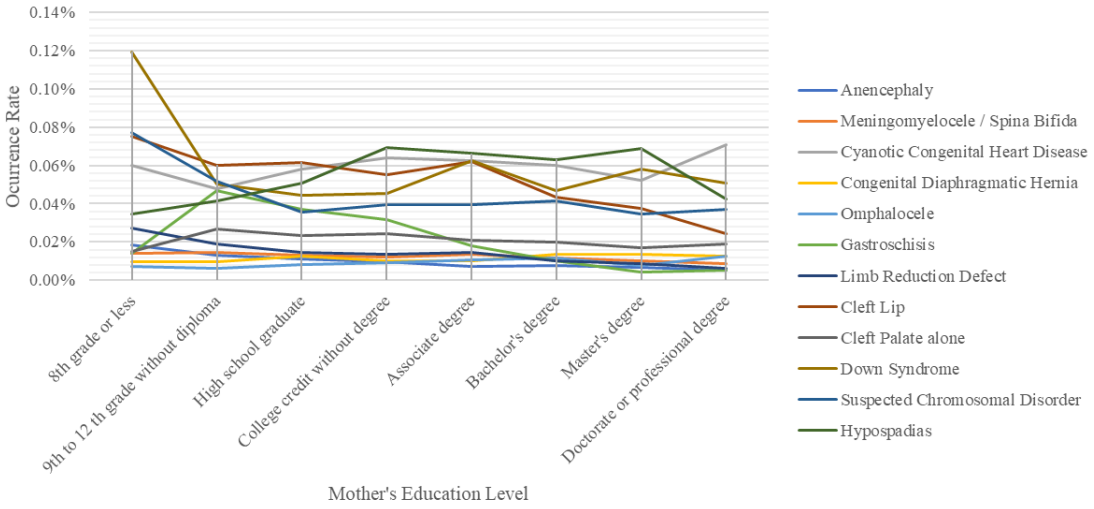
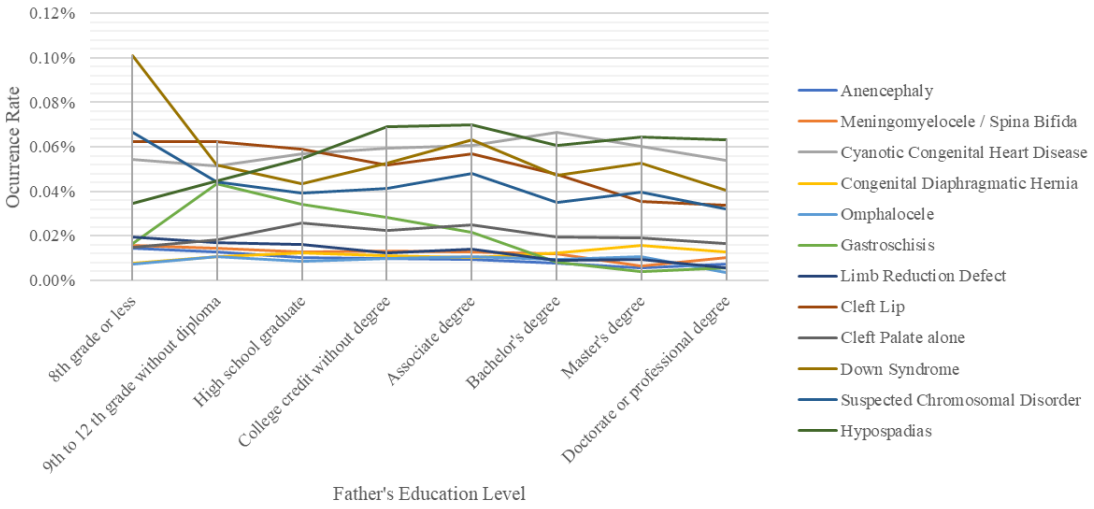


Figure 7. Newborn's Health Condition Occurrence - by Father's Education Level



However, a reversely U-shaped relationship is shown between parents' education levels and general baby health condition, the APGAR score (Figure 8, 9)—newborns tend to be generally healthier from parents with both high and low level educations, but not from the middle ones. This relationship is also

illustrated earlier in Table 3 by the small coefficient between education level and APGAR score. This seemingly unintuitive correlation can, indeed, be explained by simultaneously considering parent education's and age's effect. Assume parent's age, in fact, negatively correlates with the babies' health condition (I have to assume here, while, in the correlation table, the relation between the two is unseen); meanwhile, parent's education positively correlates with the babies' health condition. These are the relationships conforming to traditional beliefs, as aging increases the risk of babies having a newborn illness, and, coming from education, knowledge ensures pregnant women are provided appropriate care. While age and education level themselves have a positive relationship (Table 3), the negative effect of age on babies' health can be counterbalanced by the positive effect of education on the health. These two opposite effects can interfere with each other and result in low correlations between age and APGAR, and education and APGAR in Table 3. When young age's positive effect outweighs low education's negative one, and high education's positive effect outweighs elder age's negative one, the reversed U-shape relationship is therefore fasioned.

Figure 8. Newborn's 5-min APGAR Score - by Mother's Education Level

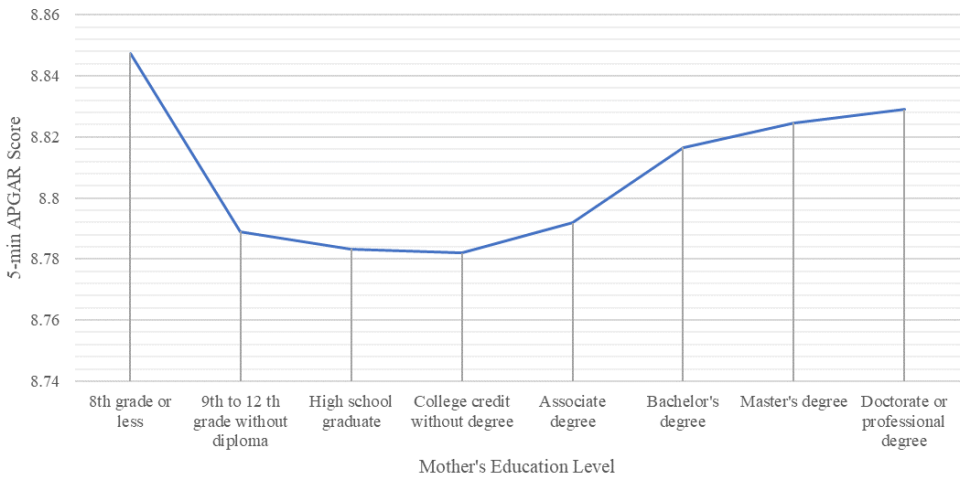
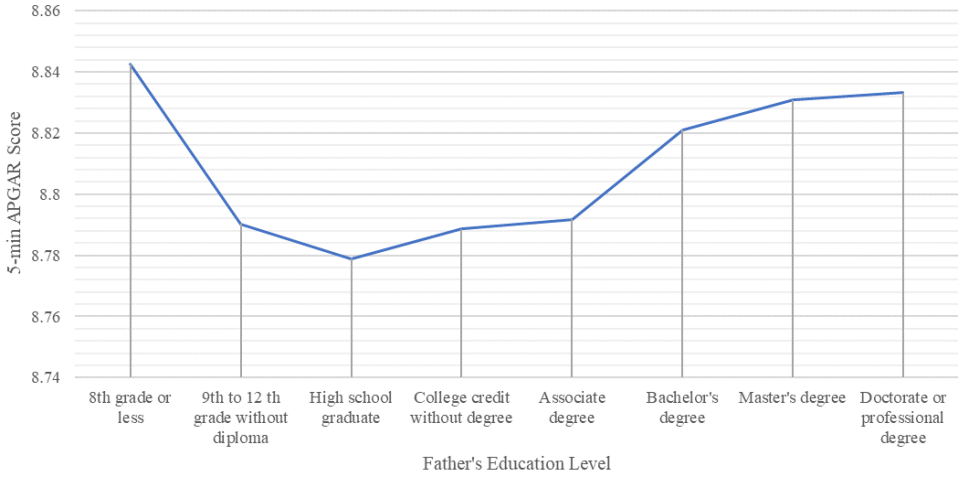
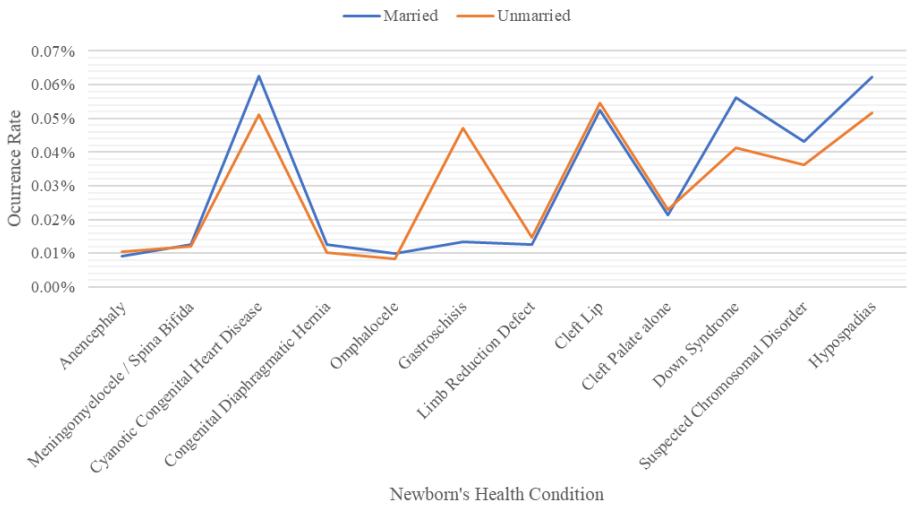


Figure 9. Newborn's 5-min APGAR Score - by Father's Education Level



Lastly, married couples tend to deliver healthier newborns. On average, APGAR score of married couples is 8.81, slightly higher than 8.78 of unmarried couples. However, the relationship is generally rebutted in the corresponding figure (Figure 10), where married couples have babies with higher diagnosis occurrences in several illness categories including cyanotic congenital heart disease, Down syndrome, chromosomal disorder, and hypospadias. Gastroschisis is the only category where babies of married couples perform significantly better than of unmarried couples. The reason of these relationships is not obvious and requires further study, perhaps combined with observations of couples' behaviors.

Figure 10. Newborn's Health Condition Occurrence - by Marital Status



Reference

Alexander, G. R., Himes, J. H., Kaufman, R. B., Mor, J., & Kogan, M. (1996). A United States national reference for fetal growth. Obstetrics & Gynecology, 87(2), 163-168.

Bettegowda, V. R., Dias, T., Davidoff, M. J., Damus, K., Callaghan, W. M., & Petrini, J. R. (2008). The relationship between cesarean delivery and gestational age among US singleton births. Clinics in perinatology, 35(2), 309-323.

Boulet, S. L., Alexander, G. R., Salihu, H. M., Kirby, R. S., & Carlo, W. A. (2006). Fetal growth risk curves: defining levels of fetal growth restriction by neonatal death risk. American journal of obstetrics and gynecology, 195(6), 1571-1577.

Davidoff, M. J., Dias, T., Damus, K., Russell, R., Bettegowda, V. R., Dolan, S., ... & Petrini, J. (2006, February). Changes in the gestational age distribution among US singleton births: impact on rates of late preterm birth, 1992 to 2002. In Seminars in perinatology (Vol. 30, No. 1, pp. 8-15). WB Saunders.

Dickert-Conlin, S., & Chandra, A. (1999). Taxes and the Timing of Births. Journal of political Economy, 107(1), 161-177.

Finer, L. B., & Henshaw, S. K. (2006). Disparities in rates of unintended pregnancy in the United States, 1994 and 2001. Perspectives on sexual and reproductive health, 38(2), 90-96.

Finer, L. B., & Zolna, M. R. (2011). Unintended pregnancy in the United States: incidence and disparities, 2006. Contraception, 84(5), 478-485.

Finer, L. B., & Zolna, M. R. (2016). Declines in unintended pregnancy in the United States, 2008–2011. New England Journal of Medicine, 374(9), 843-852.

Holman, R. C., Curns, A. T., Belay, E. D., Steiner, C. A., & Schonberger, L. B. (2003). Kawasaki syndrome hospitalizations in the United States, 1997 and 2000. Pediatrics, 112(3), 495-501.

Ventura, S. J., Hamilton, B. E., & Matthews, T. J. (2014). National and state patterns of teen births in the United States, 1940-2013. National vital statistics reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 63(4), 1-34.

MacDorman, Marian F., Fay Menacker, and Eugene Declercq. "Cesarean birth in the United States: epidemiology, trends, and outcomes." Clinics in perinatology 35.2 (2008): 293-307.

MacDorman, M. F., Matthews, T. J., Mohangoo, A. D., & Zeitlin, J. (2014). International comparisons of infant mortality and related factors: United States and Europe, 2010. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 63(5), 1-6.

Menacker, F., & Hamilton, B. E. (2010). Recent trends in cesarean delivery in the United States (pp. 1-8). US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Oken, E., Kleinman, K. P., Rich-Edwards, J., & Gillman, M. W. (2003). A nearly continuous measure of birth weight for gestational age using a United States national reference. BMC pediatrics, 3(1), 6.

Zhang, J., & Bowes Jr, W. A. (1995). Birth-weight-for-gestational-age patterns by race, sex, and parity in the United States population. Obstetrics & Gynecology, 86(2), 200-208.