**Potential outcomes of subsidized medical care: Evaluating premature birth Odds Ratios**

Final Project – Causal Inference, Winter 18-19

Tom Ron

305065658

**Introduction**

The US health care system is unique among advanced industrialized countries. The US does not have a uniform health system nor a universal health care coverage, and only recently passed legislation mandating healthcare coverage for almost everyone. In 2014, 48% of US health care spending came from private funds, with 28% coming from households and 20% coming from private businesses. As a result of this policy, many questions can arise – especially ones regarding the quality and nature of healthcare provided to those who can (and those who cannot) afford themselves an extensive healthcare plan. Because of its unethical and complex nature, a clinical trial trying to prove the effect of having private medical care on a subject's physical condition can never be conducted. Thus, an observational study approach must be taken.

In 1992, The US Department of Health and Human Services published their findings from *ICPSR 2835[1] – National Pregnancy and Health Survey (NPHS): Drug Use Among Women*. The primary objective of the NPHS was to produce annual estimates of the percentages and numbers of mothers of newborns in the US who used selected licit and illicit substances before and during pregnancy. No known attempt to evaluate causal effects from this data was made by the US Department of Health itself; some external publications[2] were produced.

The NPHS data is abundant with internal and external covariates regarding the pregnancy term and pregnancy outcome of many different women. Among its covariates, the type of a subject's healthcare coverage was measured, and several birth outcome parameters was measured as well.

As in many observational-based causal studies, we can carefully leverage old data to answer new questions. The NPHS dataset was aimed at measuring substance abuse behaviors, but

in this work, I will apply causal methodologies to try and answer a different question. My question will focus on finding evidence for causal effect of not having medical insurance on pregnant mothers' newborns health. Many measured covariates (and possible confounders) will allow us to construct a tractable causal model and evaluate *Odds Ratio* estimators of the likeliness of harming one's newborn due to not having a healthcare plan.

**The Data**

The NPHS data consists of 3386 instances of women from 52 hospitals across the US. All women were interviewed and/or questioned shortly after giving birth, recording over 500 covariates regarding their demographic background, economic status, bills and payment for medical services, medical history, substance abuse habits, prenatal care and birth outcome (baby weight, premature birth). The data consists of numeric values (weight, height, pregnancy duration, etc.) as well as categorial values (household income levels, yes/no to substance abuse, etc.). The covariates originate from 3 main sources: data from interview-assisted questionnaires (prefixed 'IAQ'), self-administered questionnaires (prefixed 'SAQ') and other calculated measures (no prefix).

The NPHS documentation has expressed an attempt to correctly sample and stratify the data; women from varied racial, demographic and geographical origins were sampled.

**Challenges**

The first challenge while pre-processing the data was covariate engineering. At over 500 covariates, feature space is very large with little or no data in many segmentations (lack of overlap). Moreover, many covariates measure the same nominal qualities and high feature correlation was prevalent. An effort was made to manually cluster and filter features to reduce redundant or highly correlated ones while keeping those most relevant for causal modeling (i.e., do not harm ignorability assumptions). After strenuous work, the feature space was reduced to roughly 35 covariates, mostly relevant to the research question. Another challenge that was introduced during the preprocessing step was a high rate of missing values in many covariates. Some occurrences of missing values originate to the IAQ/SAQ source itself ("didn't interview" - 22%). These samples were eventually omitted due to missingness of most of the covariates. Another type of missing values were conditioned covariates; values that were missing because of another covariate. E.g., the covariate "last born baby's weight" was (rightfully) marked as missing if this was the subject's first birth. This type of 'Missingness Not at Random' (MNAR) introduced complexity in preprocessing the data and required special attention. Careful imputation was made by assigning the mean/median value to some numeric values and by adding an "Unknown" class to some categorial values.

**Research Question**

Premature birth - background

It is known that premature infants are at greater risk for cerebral palsy, delays in development, hearing problems and sight problems. These risks are greater the earlier a baby is born, and the cause of preterm birth is often not known. Risk factors include diabetes, high blood pressure, being pregnant with more than one baby, being either obese or underweight, vaginal infections, tobacco smoking and psychological stress, among others[7]. Therefore, our research question will deal with another possible cause of premature birth:

Q1: What is the effect of a mother having **private medical insurance** on the chance of undergoing **premature birth**?

In this question, I will denote the covariate *IAQ\_51C* as the **binary treatment** assignment, . *IAQ\_51C* states whether prenatal care bills were covered by a private medical insurance ( or not . To our assumption, "Having private medical insurance" acts as an instrumental to *IAQ\_51C*, but the two are tightly correlated. I will denote the covariate *IAQ\_1CALC* as the **binary outcome** , stating whether a premature birth has occurred (over 2 weeks before a doctor's estimation, ) or not ).

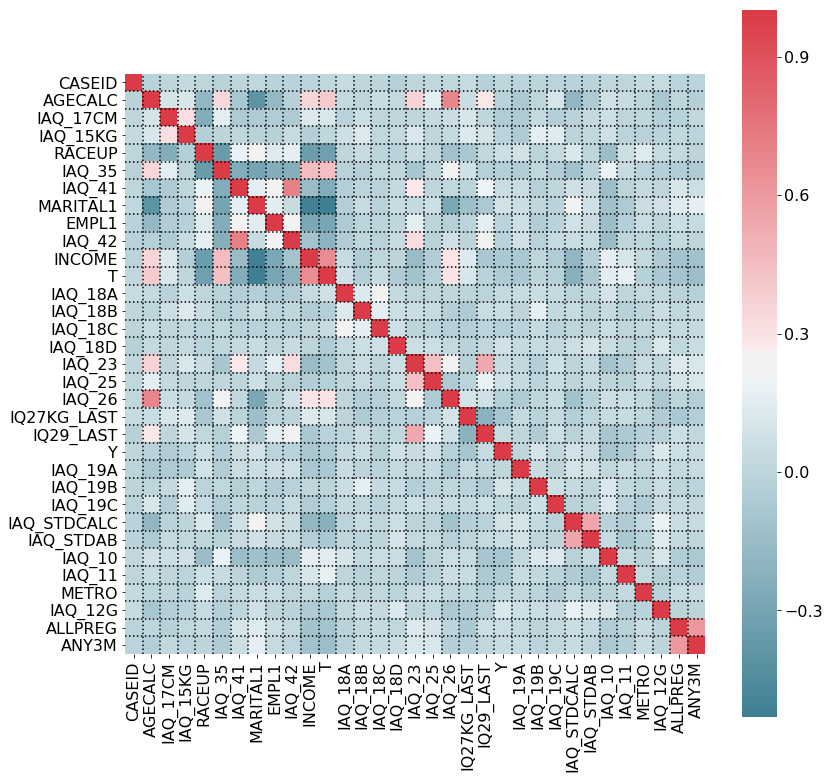


Figure 1 – Q1 Covariate correlation, after feature selection. A full covariate map including feature name, description, datatype, value range, imputation method and comments is available in appendix 2.

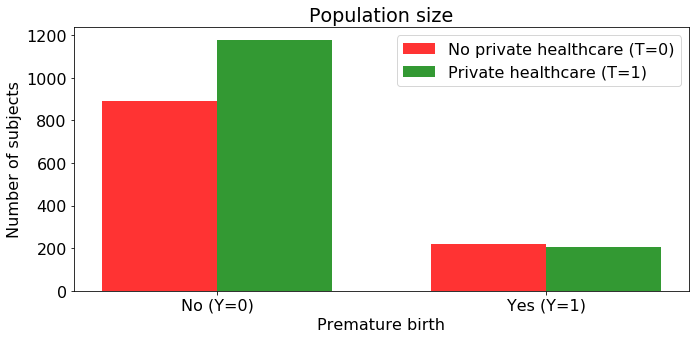


Figure 2 - Q1 population size for treated / control groups with birth outcome

We can notice balanced treated / control group sizes, but unbalanced outcome population. I will address this point later on.

**Towards causal analysis – Assumptions**

Ignorability

The main goal of this work is to conduct a **potential outcome** analysis and calculate Odds Ratio estimators, involving as a binary indicator of a preexisting environmental condition of the mother – having or not having a private medical care. Although it is may be hard to consider as a trivial treatment (i.e. "prescribing a drug") and usually is an outcome of income level and age, I believe that due to the high detail of covariates in the dataset it is possible to isolate its effect. Meaning, ignorability can hold – while conditioning on , the joint distribution of is independent of . In other words, I believe that there are generally no unmeasured confounders.

Overlap

An assessment of overlap was conducted by fitting propensity scores (Figure 5). Although the treated and control populations are inherently different, adequate overlap was found, thus allowing causal analysis.

**Causal Graph**

The first step towards causal analysis was to create a high-level scheme of the causal graph, depicted in Figure 3. Covariate connections and directionality was obtained by exploring NPHS documentation, covariate correlation, temporal precedence constraints and common sense.

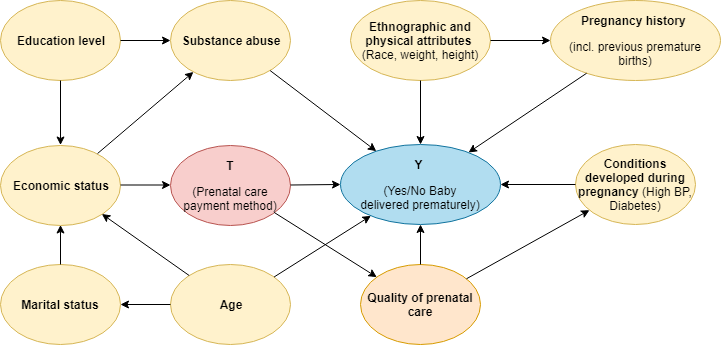


Figure 3 - Q1 Causal Graph rev. 1, A high level represntation.

Many more connections can be made between covariates; this scheme depicts merely the most substantial ones.

An apparent concern from this model arises regarding Post-treatment variables. Treatment assignment can affect "Quality of prenatal care" variables, thus they will explain away some of its effect. A revised approach was to remove these covariates:

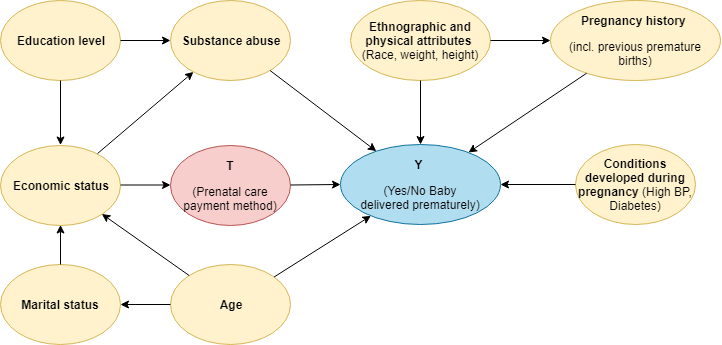


Figure 4 - Q1 Causal Graph rev. 2, A high level represntation.

And this was the final causal scheme I progressed with to Q1 evaluation.

**Methods and Results**

**Propensity Calculation and overlap**

Propensity scores were obtained by a regularized Logistic Regression ( binary classifier. Average scores across 10-fold CV (75-25 train-test split):

|  |  |  |
| --- | --- | --- |
| Average Accuracy | Average F1 Score | Average ROC AUC |
| 0.843 (+/- 0.027 SD) | 0.862 (+/- 0.024 SD) | 0.918 (+/- 0.018 SD) |

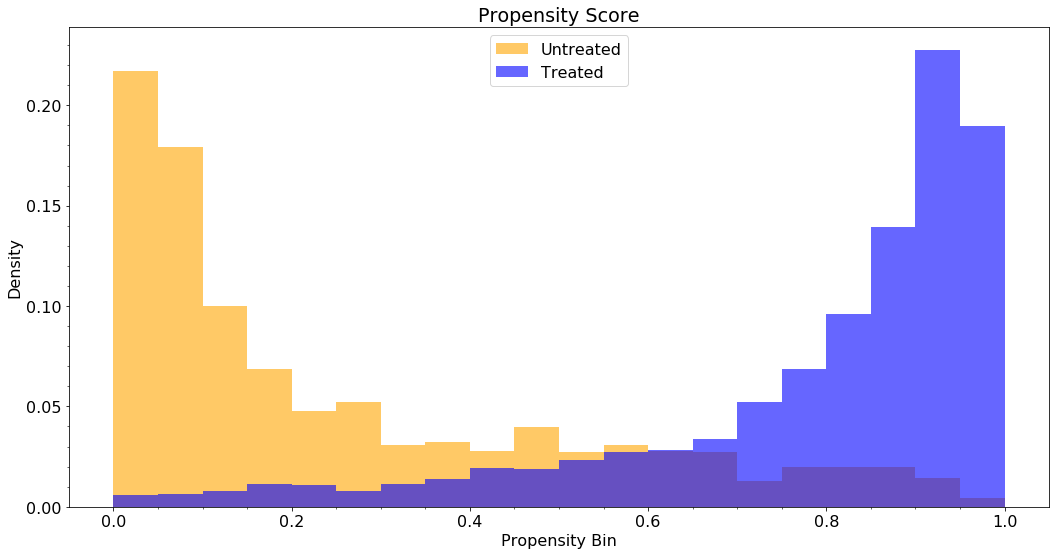


Figure 5 - Propensity score distribution of treated and control units. p(T=1|x) was obtained by a regularized Logistic Regression estimator.

Propensity score distribution demonstrate somewhat of a low overlap and a significant difference between the treated and the control population. Heavy regularization was applied to reduce variance and avoid propensity scores close to 1 or 0.

**Odds ratio approach**

In this question, dealing with a non-linear binary outcome would make estimating ATE or ATT not very meaningful. Instead, we will try to estimate the treatment effect on the outcome by means of .

Contingency tables and counterfactual values

For a potential outcome , we can express the contingency table as:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Table 1 - Y1 Contingency table

And similarly, for :

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Table 2 - Y0 Contingency table

Where denotes the sample size for a specific population. Therefore, we can obtain:

Similarly, for , And where is the treated / control population size.

Finally, the "odds-wise" treatment effect can be expressed by the [3]:

Which we cannot compute due to the counterfactual quantities of

A contingency table for the *observed* data is expressed by:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Table 3 - Observed data contingency table

Under random treatment assignment, an unbiased estimator to is the [3]:

Regarding our data, calculation showed that:

|  |  |
| --- | --- |
|  | 95% CI\* |
| 0.04348 | (0.03529, 0.05358) |

\*CI Computed according to 'Simple' method, Fleiss et al (2003)

Which may imply that women with private health insurance are substantially less prone to premature birth. This estimator obviously does not meet the assumption of random treatment assignment and does not take any other covariates or confounders other than into account. It was made to obtain a sense of directionality or naïve intuition.

**Propensity Score**

IPW estimator

Lunceford and Davidian[3] introduced an inverse propensity weighted estimator for marginal probabilities as follows:

And defined the estimator as follows:

Regarding our data, calculation showed that:

|  |  |
| --- | --- |
|  | 95% CI\* |
| 1.00124 | (0.8597, 1.1659) |

\*CI computed according to weighted 'Simple' method, Fleiss et al (2003)

Which implies **no** significant causal effect (CI contains ) – women with private health insurance are **not** less prone to premature birth than ones with public health insurance.

Stratified Propensity estimator

An arising concern was that although regularization, the propensity scores have high variance and introduce extreme values when using the IPW weighting method. A more conservative approach can by derived based on the common Mantel–Haenszel[3] estimator. MH suggests stratifying the data (for stratas) based on some criterion, following which we can calculate the MH estimator as follows:

When the contingency table for MH notation is expressed as (for strata k):

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table 4 - MH notation for observed data contingency table, strata k

Graf and Schumacher[3] suggested stratifying data based on the **propensity score** and then computing the MH estimator for each strata.

Therefore, I divided our data into 5 equal stratums:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***k=1*** |  |  |  | ***k=2*** |  |  |  |
|  | 4 | 100 | 104 |  | 25 | 76 | 101 |
|  | 16 | 379 | 395 |  | 96 | 302 | 398 |
|  | 20 | 479 | **499** |  | 121 | 378 | **499** |
| ***k=3*** |  |  |  | ***k=4*** |  |  |  |
|  | 53 | 35 | 88 |  | 62 | 7 | 69 |
|  | 265 | 145 | 410 |  | 380 | 50 | 430 |
|  | 318 | 180 | **498** |  | 442 | 57 | **499** |
| ***k=5*** |  |  |  |  |  |  |  |
|  | 63 | 2 | 65 |  |  |  |  |
|  | 421 | 13 | 434 |  |  |  |  |
|  | 484 | 15 | **499** |  |  |  |  |

Propensity quantiles were:

[0.0, 0.1136, 0.4665, 0.7908, 0.9187, 1.0]

And the MH estimator value:

|  |  |
| --- | --- |
|  | 95% CI\* |
| 0.95361 | (0.6392, 1.4224) |

\*CI computed according to MH[4]

Again, we encounter strong evidence that there is **no** significant causal effect (CI contains ) – women with private health insurance are **not** less prone to premature birth than ones with public health insurance.

**T-Learner**

Kunzel et. Al[5] firstly introduced the T-Leaner and the S-Learner algorithms for ATE estimation via counterfactual prediction. In their work, they implied that the T-Learner constellation is better suited for scenarios with balanced size treatment / control groups, a complex treatment and non-continuous outcomes. Therefore, in this section I suggest a simple derivation of the T-Learner for *Marginal Odds Ratio* estimation. The original paper illustrates the ATE calculation as:

1. Fit two separate models on treated and control samples:
2. Calculate as follows:

Instead of calculating , we can directly extrapolate the contingency tables for as follows:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Table 5 - Predicted contingency table for T-Learner approach

And then calculate:

Various models (Random Forrest, SVM and Logistic Regression) were tested as the () classifiers. I applied SMOTE[6] oversampling to the train set during model selection in order to compensate a natural unbalanced ratio of 1:5 vs. (in both treated and control groups). The most successful model was a double Logistic Regression classifier.

Results over 10-fold CV (75-25 split):

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Data Points | Avg. Test Accuracy | Avg. Test ROC AUC |
|  | 1109 | 0.803 (+/- 0.012) | 0.659 (+/- 0.058) |
|  | 1385 | 0.848 (+/- 0.007) | 0.636 (+/- 0.051) |

Accuracy measure can be misleading because of an unbalanced test set (TP Rate of the dominant class). Results show somewhat of a struggle for the models to maximize AUC criteria. Therefore, we should limit our confidence in the contingency table extrapolation and Odds Ratio calculation.

Contingency table extrapolation

Calculation of *Table 5* produced:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | 207 (actual) | 60 (estimated) |
|  | 1178 (actual) | 1009 (estimated) |
|  | 179 (estimated) | 220 (actual) |
|  | 334 (estimated) | 889 (actual) |

And the Marginal Odds Estimator:

|  |  |
| --- | --- |
|  | 95% CI\* |
| 0.3742 | (0.3156, 0.4435) |

\*CI computed according to weighted 'Simple' method, Fleiss et al (2003)

Which might imply an opposite causal directionality compared to previous methods. CI does not account for the variability and only assumes binomial distribution on of the estimated population sizes (which might not be true). Therefore, I believe that it does not represent a true confidence interval and we cannot truly admit a significant result.

Due to the poor performance of the models and the inconsistency of the T-Learner based estimator with previous methods, I tend to disapprove of these results and method (which I suggested myself).

**Possible Drawbacks**

One of the unaddressed issues was the size of the data set. In Figure 2, There is a noticeable skew with as little as ~200 datapoints for in the treated and control group. Combined with a relatively high level of covariate dimensionality, this sparse label makes counterfactual prediction a hard task. We saw evidence for this in the T-Learner method, when models trying to predict performed poorly. This may generally imply that more data is needed in order to gain more confidence in the results, even for propensity-based methods that do not try to predict counterfactuals directly.

Another possible drawback is a wrong modeling of the causal graph. Although I measured correlation and meticulously filtered out covariates, the process was done manually with no professional domain knowledge. I might have omitted important covariates (and confounders) that can affect our results and conclusion. On the other hand, I might have left too many covariates that explain-away the treatment effect and / or impede model's performance in predicting the outcome.

Moreover, the inherent difference between the treated and control groups might have skewed the Marginal Odds Ratio discussion. As it is analogous to ATE, perhaps a revised approach was to estimate a within-treatment Odds Ratio – which is analogous to ATT, specifically considering women *without* private healthcare as the treated group, because one can be inclined in showing a negative effect not having a proper healthcare plan.

Finally, the Odds Ratio approach was somewhat of a new territory for me. Although its notion was easy to understand, it was the first time I tried to evaluate these kinds of estimators. As they are non-linear (on the contrary to ATE), I needed to take extra care when trying to estimate their statistical significance. There might have been some mistakes made when estimating confidence intervals.

**Discussion**

Throughout this work I demonstrated various methods for estimating the Marginal Odds Ratio for the research question at hand. All the propensity-based methods implied no causal effect of having a private medical care on newborn's health. A reasonable explanation to this result can be that women without private medical care are provided with the **same standard** of prenatal care as do women with private healthcare. The main difference between the groups reside on how they *afford* the medical care, but this might not necessarily affect its *quality*. Interestingly enough, the Crude Odds Ratio demonstrated a significant difference between the groups, but the covariate adjustment re-balanced the ratio and eliminated the effect.

I believe that If my results are valid, they are somewhat encouraging. Meaning, with respect to premature births outcome, there is no apparent bias from healthcare providers towards disadvantaged populations. Since 1992, The US healthcare system has progressed greatly – i.e. with the introduction of "Obama Care" that was supposed to alleviate financial strains caused by funding private healthcare.

**Future Work**

When considering the NPHS data, many other questions can arise due to its covariate richness. My first attempt was to try and evaluate a causal effect of substance abuse during pregnancy, but the harsh imbalance of treated vs. control group size made this dissection futile (although reassuring that not many women use drugs during pregnancy). Regarding other datasets, it could be interesting to evaluate the same causal question regarding pregnant women or even other populations with different medical outcomes.

Altogether, the abundance of observational data and the activeness of the causal research community in recent years is surely making causal analysis more and more appealing. I have greatly enjoyed this work and hope to continue practice causal inference research in the future.

**Appendix**

**1. Bibliography**

[1] <https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/2835/summary>

[2] <https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/2835/publications>

[3] <https://journals.sagepub.com/doi/pdf/10.1177/0962280214541995>

[4] <https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Mantel-Haenszel_Test.pdf>

[5] <https://pdfs.semanticscholar.org/cf20/db64bdc22b5303a4dcd10b5f00e8f93befeb.pdf>

[6] <https://www.jair.org/index.php/jair/article/view/10302>

[7] <https://en.wikipedia.org/wiki/Preterm_birth>

**2. Feature Map**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Feature Column** | **Feature Description** | **Datatype** | **LB** | **UB** | **Didn't Interview Code** | **Missing Code(s)** | **Inapplicable Code(s)** | **Imputation strategy** | **Categorial Description** | **Comment** |  |
| **General** | CASEID | Id | int |  |  |  |  |  |  |  |  |  |
| **Physical attributes** | AGECALC | Age (at birth) | int |  |  |  |  |  |  |  |  |  |
|  | IAQ\_17CM | Height (cm) | double | 0 | 250 | 996 |  | 997.00 | Mean |  | Calculated based on IAQ\_17FT and IAQ\_17CM | |
|  | IAQ\_15KG | Mother's weight before pregnancy (kgs) | double | 0 | 250 | 996 |  | 997.00 | Mean |  | Calculated based on IAQ\_15 | |
| **Ethographic Background** | RACEUP | Race | categorial | 1 | 4 | 6 |  |  | Class Zero (N/A) | 1=WHITE, 2=BLACK, 3=HISPANIC, 4=OTHER |  |  |
| **Education** | IAQ\_35 | Last grade school | int | 0 | 17 | 96 | 98, 99 |  | Median | 0 = 7TH GRADE OR LESS, 8 = 8TH GRADE, 9 = HIGH SCHOOL, 10 = HIGH SCHOOL, 11 = HIGH SCHOOL, 12 = HIGH SCHOOL, 13 = COLLEGE, 14 = COLLEGE, 15 = COLLEGE, 16 = COLLEGE, 17 = GRAD/PROFESSIONAL SCHOOL | | |
| **Household and Socio-Economic Status** | IAQ\_41 | People in household incl. subject (12 months prior to delivery) | int | 1 | 50 | 96 | 98, 99 |  | Median |  |  |  |
|  | MARITAL1 | Marital Status | categorial | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=MARRIED, 2=NOT MARRIED |  |  |
|  | EMPL1 | Employment Status | categorial | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=WORKING, 2=NOT WORKING |  |  |
|  | IAQ\_42 | Number of relatives in household | int | 0 | 11 | 96 | 98, 99 | 97 | Median |  |  |  |
|  | INCOME | Household income | categorial | 1 | 16 | 96 |  |  | Median | 1 = NONE, 2 = UNDER $5,000, 3 = $5,000 - $6,300, 4 = $6,301 - $8,450, 5 = $8.451 - $10,600, 6 = $$10,601 - $12,700, 7 = $12,701 - $14,850, 8 = $14,851 - $17,000, 9 = $17,001 - $19,000, 10 = $19,001 - $21,300, 11 = $21,301 - $25,000, 12 = $25,001 - $30,000, 13 = $30,001 - $40,000, 14 = $40,001 - $50,000, 15 = $50,001 - $75,000, 16 = $75,001 + | | |
|  | IAQ\_43B | Public assitance / welfare | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_43C | Food stamps | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
| **Health Insurance and Medical bills** | IAQ\_51C | Prenatal care paid by medical insurance | categorial | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
| **Pre-Pregnancy Medical History** | IAQ\_18A | Pre-pregnancy Diabetes | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_18B | Pre-pregnancy High BP | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_18C | Pre-pregnancy Epilepsy | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_18D | Pre-pregnancy other chronic conditions | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
| **Previous Pregnancy History** | IAQ\_23 | Number of previous pregnancies | int | 0 | 7 | 96 | 98, 99 |  | Mean |  |  |  |
|  | IAQ\_24B | Number of previous miscarriges (spontaneous abortion before week 20) | int | 0 | 7 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IAQ\_24C | Number of previous stillborns (spontaneous abortion after week 20) | int | 0 | 7 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IAQ\_24D | Number of previous ectopic pregnencies | int | 0 | 7 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IAQ\_24E | Number of previous induced abortions | int | 0 | 7 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IAQ\_25 | Outcome of last pregnancy | categorial | 1 | 5 | 6 | 8, 9 | 7 | Class Zero (N/A) | 1 = LIVEBIRTH, 2 = MISCARRIAGE, 3 = STILLBIRTH, 4 = INDUCED ABORTION, 5 = ECTOPIC PREGNANCY | | |
|  | IAQ\_26 | Mother's age at last pregnancy | int | 11 | 42 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IAQ\_CNTR | Number of previous babies born | int | 1 | 10 | 96 |  | 97 | Mean if 96, otherwise 0 |  |  |  |
|  | IQ27KG\_LAST | Baby weight (kg), last pregnancy | int | 0 | 30 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  | Calculated based on IQ27LB\_LAST and IQ27OZ\_LAST | |
|  | IQ28\_LAST | Mother's age at last pregnancy | int | 0 | 30 | 96 | 98, 99 | 97 | Mean if 96, otherwise 0 |  | Calculated based on last 10 pregnancies | |
|  | IQ29\_LAST | Was baby delivered more than 2 weeks early (last pregnancy) | bool | 1 | 2 | 6 | 8, 9 | 97 | Class Zero (N/A) | 1=Yes, 2=No | Calculated based on last 10 pregnancies | |
| **Baby** | IAQ\_1CALC | Was baby delivered more than 2 weeks early (current pregnancy) | bool | 1 | 2 | 96 | 98, 99 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
| **Conditions Developed During Pregnancy** | IAQ\_19A | Anemia | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_19B | High BP (incl. Toxemia or Pre-Eclampsia) | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_19C | Diabetes (or sugar diabetes) | bool | 1 | 2 | 6 | 8, 9 |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_STDCALC | Any STD | bool | 1 | 2 |  |  | 7 | Class Zero (N/A), 0 if missing | 1=Yes, 2=No | Calculated based on OR condition between IAQ\_20A-F | |
|  | IAQ\_STDAB | Any antibiotics taken | bool | 1 | 2 |  |  | 7 | Class Zero (N/A), 0 if missing | 1=Yes, 2=No | Calculated based on OR condition between IAQ\_21A-F | |
|  | IAQ\_10 | Number of prenatal visits | int | 1 | 70 | 96 | 98, 99 | 97 | Mean |  |  |  |
|  | IAQ\_11 | Type of care/clinic | categorial | 1 | 10 | 6 | 8, 9 | 7 | Class Zero (N/A) | 1 = PRIVATE DOCTOR'S OR NURSE MIDWIFE'S OFFICE 2 = COUNTY OR CITY HEALTH DEPARTMENT 3 = COMMUNITY OR NEIGHBORHOOD HEALTH CENTER 4 = HMO/HEALTH MAINTENANCE ORGANIZATION 5 = CLINIC AT WORK OR SCHOOL 6 = CLINIC IN A HOSPITAL 7 = EMERGENCY ROOM IN A HOSPITAL 8 = NAME OF PLACE GIVEN, TYPE NOT SPECIFIED 9 = CLINIC, NOT OTHERWISE SPECIFIED 10 = OTHER PLACE (SPECIFY) - (MAKE PROBLEM CARD) | | |
| **Prenatal Care** | METRO | Metro / Non metro hospital | bool | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | IAQ\_12G | Emergency visits | bool | 1 | 2 | 6 | 8, 9 | 7 | Class Zero (N/A) | 1=Yes, 2=No |  |  |
| **Substance Abuse** | ALLPREG | Any illicit drug use during pregnancy | bool | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |
|  | ANY3M | Any illicit drug use during 3 months prior to pregnancy | bool | 1 | 2 | 6 |  |  | Class Zero (N/A) | 1=Yes, 2=No |  |  |