|  |  |
| --- | --- |
|  | **Small for Gestational Age (SGA)** |
|  | **Not SGA****(n=650)** | **SGA****(n=105)** | **Total****(n=755)** |
| Mother’s Age, years\* | 24.94 (5.45; 14-43) | 23.85 (4.90; 16-35) | 24.79 (5.39; 14-43) |
| Mother’s Height, cm\* | 157.01 (6.54; 106-176) | 154.56 (5.87; 142-172)b | 156.68 (6.50; 106-176)e |
| Parity, #\* | 1.13 (1.23; 0-6) | 0.90 (1.11; 0-6) | 1.10 (1.21; 0-6) |
| Birthweight, g\* | 3246.21 (402.13; 2510-4730)a | 2231.11 (411.60; 1035-3780)c | 3105.63 (534.46; 1035-4730)f |
| Current Smoker, # (%) | 186 (28.7%)a | 45 (43.3%)c | 231(30.8%)f |
| Male child, % | 339 (52.4%)a |  44 (42.3%)c | 383 (50.9%)f |
| Gestational Age, weeks\* | 39.38 (1.24; 38-44)a | 37.92 (2.20; 30-42)d | 39.18 (1.50; 30-44)g |

\*Provided descriptive statistics are in the format of mean (standard deviation; minimum-maximum)

BMI = body mass index; CVD = cardiovascular disease

a N=647

b N=99

c N=104

d N=103

e N=749

f N=751

g N=750

**Methods:** All analyses were performed in R using the “psych” package and the describe.by function. Descriptive statistics are reported stratified by “small for gestational age (SGA)” status, with SGA defined as < 10th percentile of birthweight for the gestational age at which they were born. For continuous variables (age, mother’s height, parity, birthweight, and gestational age), the mean, standard deviation, and minimum and maximum values are reported. For categorical variables (current smoker and male child), count and percentages are reported.

**Results:** 105 of the 755 (13.9%) participants in this study gave birth to SGA children with a mean birthweight of 2231.11 grams, while 650 participants birthed normally sized infants with a mean birthweight of 3246.21 grams. The presented table summarizes the descriptive statistics within the SGA and non-SGA subsets. Missing observations are noted with a superscript and the corresponding number with data within each subset.

Approximate trends were seen across SGA vs. non-SGA subsets for the variables, age, mother’s height, parity, and gestational age. The mean mother’s age (23.85 years), height (154.56 cm), and parity (0.90 prior births) tended to be lower for mothers with SGA infants as compared to mothers birthing normally sized neonates (24.94 years, 157.01 cm, and 1.13 prior births, respectively). Conversely, mothers birthing SGA children tended to smoke more frequently (43.3% vs. 28.7%). Finally, SGA children themselves tended to be born earlier (mean gestational age 37.92 weeks vs. 39.38 weeks) and were more likely to be female than male (male children % = 42.3 compared to 52.4% in non-SGA children).

2a.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The odds of being born SGA were compared for babies whose mothers smoked and whose mothers did not smoke using a logistic regression model. The Wald statistic was used to perform statistical inference on the regression slope and standard error (SE) parameters (no robust SE option was used due to this being a saturated model). A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned logistic regression parameter estimates.

**Results:** Of the 520 mothers who did not smoke, the odds of giving birth to an SGA child was 0.1279, while for the 231 mothers who did smoke the odds of an SGA child was 0.2419. Based on a 95% confidence interval, the observed odds ratio of 1.8904 would not be unusual if the true population odds ratio comparing SGA births between smoker and non-smoker mothers was between 1.2360 and 2.8912. A two-sided p-value of 0.00336 means we can reject the null hypothesis at the 0.05 level that there is no association between maternal smoking and SGA births.

2b.

|  |  |  |
| --- | --- | --- |
|  | **Not SGA** | **SGA** |
| **Not Maternal Smoker** | 461 | 59 |
| **Maternal Smoker** | 186 | 45 |

**Odds (with regression coefficients):**

Non-smokers: Odds = eBo = e-2.05586 = 0.12798

Smokers: Odds = eBo \* eB1 = e-2.05586 \* e0.63678 = 0.12798 \* 1.89038 = 0.241931

**Odds (simple calculation):**

Non-smokers: Odds = Pi / (1 – Pi) = 59/(461+59) / (1 – (59/(461+59))) = 0.12798

Smokers: Odds = Pi / (1 – Pi) = 45/(186+45) / (1 – 45/(186+45)) = 0.241931

**Probability (with regression coefficients):**

Non-smokers: Pi = eBo / (1+eBo) = 0.12798 / (1+0.12798) = 0.11346

Smokers: Pi = (eBo \* eB1 / ( 1 + eBo \* eB1)) = (0.12798 \* 1.89038 / (1 + 0.12798 \* 1.89038)) = 0.1948

**Probability (simple calculation):**

Non-smokers: Pi = 59/(461+59) = 0.11346

Smokers: Pi = 45/(186+45) = 0.1948

*Because this is a saturated model, we would expect both the simple probability and odds of delivering an SGA infant to be the same as those derived from regression model coefficients. We observe that both the derived (from regression coefficients) and simple calculations of the odds and probabilities are the same in this case of a saturated model.*

2ci.

The estimates for the intercept, slope, and odds ratio for a logistic regression model with the response of SGA and the predictor of maternal non-smoking would have:

-An intercept equal to Bo + B1 in the original model (-2.05 + 0.63 = -1.42). The exponeniated intercept is now 0.2419, or the odds of SGA given maternal smoking = 1.

-The slope is equal to –B1 from the original model (0.6368 = -0.6368).

-The odds ratio is the inverse of the original model OR (1/1.89 = 0.529).

-P-value is the same (p=0.00336)

2cii.

The estimates for the intercept, slope, and odds ratio for a logistic regression model with the response of NOTSGA and the predictor of maternal smoking would have:

 -An intercept of –Bo from the original model (--2.06 = +2.06).

 -The slope is equal to –B1 from the original model (0.6368 = -0.6368).

 - The odds ratio is the inverse of the original model OR (1/1.89 = 0.529).

-P-value is the same (p=0.00336)

2ciii.

The estimates for the intercept, slope, and odds ratio for a logistic regression model with the response of NOTSGA and the predictor of maternal non-smoking would have:

 -An intercept that is –(Bo + B1) from the original model (-(-2.05 + 0.63 = -1.42) = 1.42).

 -The slope is equal to the original model B1 (0.6368 = 0.6368).

-The odds ratio is equal to the original model OR (1.89 = 1.89).

-P-value is the same (p=0.00336)

3a.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The probabilities of being born SGA were compared for babies whose mothers smoked and whose mothers did not smoke using a linear regression model. The Wald statistic was used to perform statistical inference on the regression slope, representing the difference in SGA probabilities for the smoking vs. non-smoking groups. The Huber-White sandwich estimator was used to calculate the standard error (SE) parameters for the slope coefficient. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned linear regression parameter estimates.

**Results:** Of the 520 mothers who did not smoke, the proportion giving birth to an SGA child was 11.35%, while for the 231 mothers who did smoke the proportion of giving birth to a SGA child was 19.48%. Based on a 95% confidence interval, the observed difference in probability of SGA of 8.134% would not be unusual if the true population risk difference comparing SGA births between smoker and non-smoker mothers was between 2.328% and 13.94%. A two-sided p-value of 0.0061 means we can reject the null hypothesis at the 0.05 level that there is no association between maternal smoking and SGA births.

3b.

**Probability (with regression coefficients):**

Non-smokers: Pi = B0 = 0.11346

Smokers: Pi = B0 + B1 = 0.11346 + 0.08134 = 0.1948

**Probability (simple calculation):**

Non-smokers: Pi = 59/(461+59) = 0.11346

Smokers: Pi = 45/(186+45) = 0.1948

**Odds (with regression coefficients):**

*Pi calculated above from regression coefficients and then plugged into this equation*

Non-smokers: Odds = Pi / (1 – Pi) = 0.11346 / (1 – 0.11346) = 0.12798

Smokers: Odds = Pi / (1 – Pi)= 0.1948 / (1 – 0.1948) = 0.241931

**Odds (simple calculation):**

Non-smokers: Odds = Pi / (1 – Pi) = 59/(461+59) / (1 – (59/(461+59))) = 0.12798

Smokers: Odds = Pi / (1 – Pi) = 45/(186+45) / (1 – 45/(186+45)) = 0.241931

*Because this is a saturated model, we would expect both the simple probability and odds of delivering an SGA infant to be the same as those derived from regression model coefficients. We observe that both the derived (from regression coefficients) and simple calculations of the odds and probabilities are the same in this case of a saturated model*.

3ci.

The estimates for the intercept and slope for a linear regression model with the response of SGA and the predictor of maternal non-smoking would have:

 -An intercept equal to B0 + B1 from the original model (0.11346 + 0.08134 = 0.19481)

 -A slope equal to –B1 (-0.08134) from the original model.

 -P-value is the same (p=0.0061)

3cii.

The estimates for the intercept and slope for a linear regression model with the response of NOTSGA and the predictor of maternal smoking would have:

 -An intercept equal to 1 – B0 from the original model (1 – 0.11346 = 0.88654)

 -A slope equal to –B1 (-0.08134) from the original model

 -P-value is the same (p=0.0061)

3ciii.

The estimates for the intercept and slope for a linear regression model with the response of NOTSGA and the predictor of maternal non-smoking would have:

 -An intercept equal to 1 – (B0 + B1) from the original model (1 – (0.11346 + 0.08134) = 0.80519)

 -A slope equal to B1 from the original model

 -P-value is the same (p=0.0061)

4a.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The probabilities of being born SGA during the study period were compared for babies whose mothers smoked and whose mothers did not smoke using a Poisson regression model. The Wald statistic was used to perform statistical inference on the ratio of probabilities of SGA based on the regression slope parameter. The Huber-White sandwich estimator was used to calculate the standard error (SE) parameters for the slope coefficient. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned Poisson regression parameter estimates.

**Results:** Of the 520 mothers who did not smoke, the proportion giving birth to an SGA child was 11.35%, while for the 231 mothers who did smoke the proportion of giving birth to a SGA child was 19.48%. Based on a 95% confidence interval, the observed rate ratio (RR) of SGA probabilities of 1.7169 would not be unusual if the true population rate ratio of SGA births between smoker and non-smoker mothers was between 1.2019 and 2.4527. A two-sided p-value of 0.00302 means we can reject the null hypothesis at the 0.05 level that there is no association between maternal smoking and SGA births.

4b.

**Probability (with regression coefficients):**

Non-smokers: Pi = eB0 = e-2.17629 = 0.11346

Smokers: Pi = eB0+B1 = e-2.17629+0.54054 = 0.1948

**Probability (simple calculation):**

Non-smokers: Pi = 59/(461+59) = 0.11346

Smokers: Pi = 45/(186+45) = 0.1948

**Odds (with regression coefficients):**

*Pi calculated above from regression coefficients and then plugged into this equation*

Non-smokers: Odds = Pi / (1 – Pi) = 0.11346 / (1 – 0.11346) = 0.12798

Smokers: Odds = Pi / (1 – Pi)= 0.1948 / (1 – 0.1948) = 0.241931

**Odds (simple calculation):**

Non-smokers: Odds = Pi / (1 – Pi) = 59/(461+59) / (1 – (59/(461+59))) = 0.12798

Smokers: Odds = Pi / (1 – Pi) = 45/(186+45) / (1 – 45/(186+45)) = 0.241931

*Because this is a saturated model, we would expect both the simple probability and odds of delivering an SGA infant to be the same as those derived from regression model coefficients. We observe that both the derived (from regression coefficients) and simple calculations of the odds and probabilities are the same in this case of a saturated model.*

4ci.

The estimates for the intercept and slope for a linear regression model with the response of SGA and the predictor of maternal non-smoking would have:

 -An intercept equal to B0 + B1 from the original model (-2.17629 + 0.54054 = 1.6358).

 -A slope equal to –B1 (-0.54054) from the original model.

 -A RR equal to 1/RRoriginal (= 1/1.71693 = 0.5824).

 -P-value is the same (p=0.00302).

4cii.

The estimates for the intercept and slope for a linear regression model with the response of NOTSGA and the predictor of maternal smoking would have:

-An intercept and slope are not as easily derived in this comparison. The probability of SGA for a smoker (0.1948) are equal to 1 – exp(Bo + B1) using the coefficients derived from this model, however.

 -The RR and P-value differ, as they are equal to 0.90825 and 0.00769, respectively.

4ciii.

The estimates for the intercept and slope for a linear regression model with the response of NOTSGA and the predictor of maternal non-smoking would have:

-An intercept and slope are not as easily derived in this comparison. The probability of SGA for a smoker (0.1948) are equal to 1 – exp(Bo) using the intercept derived from this model, however.

 -The RR and P-value differ, as they are equal to 1.10102 and 0.00769, respectively.

5.

For the linear regression model in #3, nearly the exact same results can be obtained through using a t-test allowing for unequal variances in mean SGA (probability) between the smoker and non-smoker groups with a two-sided p-value being calculated. In this case, the difference in sample means (-0.08134) is equal to the absolute value of B1 (+0.0813) from the regression model, while B0 is equal to the mean SGA (SGA probability) in the nonsmokers (=0.11346 in both analyses). Both the 95% CIs and p-values differ slightly (-0.0231 to -0.1395 and p=0.0063 in the t-test vs. 0.02328 to 0.13941 and p=0.0061 in the linear regression model).

For the logistic regression model in #2 and the Poisson regression in #4, the exact same OR (1.8904) and RR (1.7169) can be calculated from the raw 2x2/chi-square table since this is a saturated regression model. However, in this case, the coefficients/slope estimates do not correspond to any aspect of the 2x2/chi-square table, as they are each derived from log-links of the odds and rates, respectively. Similarly, the chi-squared derived p-value (p=0.02609) is considerably different than that derived from logistic (p=0.00336) and Poisson (p=0.00302) regressions.

6a.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The difference in probabilities of being born SGA as a function of maternal age (modeled as a linear, continuous variable) was inferred using the Wald statistic to compute the regression slope parameter. The Huber-White sandwich estimator was used to calculate the standard error (SE) parameters for the regression slope coefficient. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned linear regression parameter estimates.

**Results:** Data was available on 755 subjects with a mean age of 24.79 years (SD=5.39; range=14-43). Of these 755 mothers, 105 (13.9% of participants) gave birth to an SGA child. From linear regression estimates of SGA probability by maternal age, we estimate that the probability of giving birth to an SGA child is an absolute 0.451% lower per 1 year increase in maternal age. Based on a 95% CI, this observed difference would not be unusual if the true population difference in SGA proportions suggesting lower rates of SGA with increasing maternal age was between 0.02856% and 0.87445% in one group with prospective mothers 1 year older than another group. With a two-sided p-value of 0.0364 we reject the null hypothesis that there is no association between maternal age and SGA birth risk difference, in favor of a hypothesis that suggests lower SGA with increasing maternal age.

6b.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.A Poisson regression model was used to compare the probabilities of being born SGA during the study period for babies who differed by their maternal age. The Wald statistic was used to perform statistical inference on the ratio of probabilities of SGA based on the regression slope parameter. The Huber-White sandwich estimator was used to calculate the standard error (SE) parameters for the slope coefficient. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned Poisson regression parameter estimates.

**Results:** Data was available on 755 subjects with a mean age of 24.79 years (SD=5.39; range=14-43). Of these 755 mothers, 105 (13.9% of participants) gave birth to an SGA child within the period of the study. From Poisson regression parameters, we estimate that the probability of giving birth to an SGA child was a *relative* 3.384% decrease for each 1 additional year of maternal age (SGA rate ratio = 0.96616). Based on a 95% confidence interval, the observed rate ratio (RR) of SGA probabilities would not be unusual if the true population rate ratio of SGA births between groups of mothers differing by 1 year of age was between 0.93396 and 0.99947. A two-sided p-value of 0.04653 means we can reject the null hypothesis at the 0.05 level that there is no association between maternal age and SGA births probability ratio, in favor of a hypothesis that suggests lower SGA with increasing maternal age.

6c.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The odds of being born SGA were compared for babies based on maternal age using a logistic regression model. The Wald statistic was used to perform statistical inference on the regression slope. Standard error (SE) parameters of the slope were estimated using maximum likelihood. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned logistic regression parameter estimates.

**Results:** Data was available on 755 subjects with a mean age of 24.79 years (SD=5.39; range=14-43). Of these 755 mothers, 105 (13.9% of participants) gave birth to an SGA child within the period of the study. From logistic regression model parameters, we estimate that the odds of a SGA birth were a relative 3.9% lower for each 1 additional year of maternal age (SGA odds ratio = 0.961). Based on a 95% confidence interval, the observed odds ratio of 0.961 would not be unusual if the true population odds ratio comparing SGA births between two groups of mothers differing by 1 year of age was between 0.092416 and 0.99931. A two-sided p-value of 0.0461 means we can reject the null hypothesis at the 0.05 level that there is no association between maternal age and odds of SGA births, in favor of a hypothesis that suggests lower SGA with increasing maternal age.

6d.

Linear regression parameters:

B0 = 0.2509966; B1 = -0.0045152

*Estimate of SGA probability for 20 year old mother:* 0.2509966 + (-0.0045152 \* 20) = 0.1606926

Poisson regression parameters:

B0 = -1.13598; B1 = -0.03442

*Estimate of SGA rate for 20 year old mother:* eB0 \* e(B1\*20) = e-1.13598 \* e(-0.03442\*20) = 0.32111 \* 0.5023792 = 0.161319

Logistic regression parameters:

B0 = -0.85316; B1 = -0.03978

*Estimate of SGA probability for 20 year old mother:* eBo\*eB1 / (1+eBo\*eB1) = e-0.85316\*e-0.03978\*20/(1 + e-0.85316\*e-0.03978\*20)

 = 0.1922882 / (1+0.1922882) = 0.1612766

Estimate from 20-year olds in the data:

40 total 20 year olds in the dataset, 3 with SGA

3/40 = 0.075

All 3 estimates from regression parameters yield relatively similar probabilities of SGA for a 20 year old mother (~0.16). In contrast, the actual data has a proportion of SGA of 0.075 for mothers of age 20. The likely contrast for this data comes from outliers (in this case at the extremely young age, who likely have extremely high rates of SGA relative to the rest of the age groups). When modeling the outliers in addition to the rest of the data, the estimates must accommodate these datapoints and yields parameters that fit the average of the overall rate of SGA for the data.

7.



All of the fitted values from the 3 regression models are fairly similar from age 20 to approximately age 35. However, at the extremes of maternal age in the dataset (<20 years old, >35 years old), the Poisson/logistic regression models diverge from the linear regression model estimates, in that they curve upward. The triangles represent the raw SGA probability within each maternal age group.

8a.

**Methods:** R was used for all analyses using the “regress” function from the package “uwIntroStats”.The odds of being born SGA were compared for babies based on ln(maternal age) using a logistic regression model. As no subjects had maternal age = 0, no transformation was required prior to the natural log transformation. The Wald statistic was used to perform statistical inference on the regression slope. Standard error (SE) parameters of the slope were estimated using maximum likelihood. A two-sided p-value and 95% confidence interval (CI) were calculated based on an approximate standard normal distribution for the aforementioned logistic regression parameter estimates.

**Results:** Data was available on 755 subjects with a mean ln(maternal age) of 3.18 years (SD=0.213; range=2.639-3.761). Of these 755 mothers, 105 (13.9% of participants) gave birth to an SGA child within the period of the study. From logistic regression model parameters, we estimate that the odds of a SGA birth were a relative 66.1% lower for a two-fold increase in maternal age (SGA odds ratio = 0.516). Based on a 95% confidence interval, the observed odds ratio of 0.516 per two-fold increase in maternal age would not be unusual if the true population odds ratio comparing SGA births between two groups of mothers differing by 1 year of age was between 0.3953 and 0.674410. A two-sided p-value of 0.053 means we cannot reject the null hypothesis at the 0.05 level that there is no association between ln(maternal age) and odds of SGA births.

8b. The analysis performed in 6c has the advantages of being much simpler to interpret (two-fold increase in maternal age compares two drastically different groups; e.g., 15 vs 30 or 20 vs 40). As well, there is no biologic rationale to work on the multiplicative scale for maternal age, as a linear/additive trend makes more sense. Finally, there are no real outliers in maternal age that would require natural log transformation.