

# TDLU Counts and PIH

Suzanne Dufault

May 29, 2019

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Examining the Data</b>                               | <b>4</b>  |
| <b>2</b> | <b>Zero-Inflated versus Hurdle Models</b>               | <b>8</b>  |
| 2.1      | Zero-Inflated . . . . .                                 | 8         |
| 2.2      | Hurdle . . . . .  | 8         |
| 2.3      | Follow-Up Questions . . . . .                           | 8         |
| <b>3</b> | <b>Testing for Zero-Inflation</b>                       | <b>9</b>  |
| 3.1      | Goodness-of-Fit Comparisons . . . . .                   | 9         |
| <b>4</b> | <b>Unadjusted Negative Binomial Model Results</b>       | <b>10</b> |
| 4.1      | Adjusted Negative Binomial Model Results . . . . .      | 10        |
| 4.1.1    | Using Delta Method for Covariate Combinations . . . . . | 11        |
| 4.2      | Adjusted Stratified Models . . . . .                    | 12        |
| <b>5</b> | <b>Tables and Figures for Paper</b>                     | <b>16</b> |
| <b>6</b> | <b>Supplemental</b>                                     | <b>22</b> |

## Abstract

Just as a brief background, the outcome, TDLU counts (terminal ductal lobular units), are counted by a specially trained pathologist, and represent the locations where breast cancer originates. The number of these units correlates with breast cancer risk, and our hypothesis is that women with PIH and the TT genotype will have lower TDLU counts, as this would support a possible mechanism for the association seen in prior studies. I am thinking you would do a similar type of genetic model as you did in the UK Biobank study to see if interaction (and possibly trend?) is present.

## List of Tables

|    |  |    |
|----|--|----|
| 1  | Summary statistics for whole cohort and broken down by PIH status. . . . .   | 5  |
| 2  | [AMONG PIH POSITIVE WOMEN] Summary statistics by genotype. . . . .   | 6  |
| 3  | [AMONG PIH NEGATIVE WOMEN] Summary statistics by genotype. . . . .   | 7  |
| 4  | Number of non-zero TDLU counts. . . . .  | 7  |
| 5  | Testing for zero-inflation via model comparisons. The first three comparisons are nested models (with the latter nested in the former) and are compared via a likelihood ratio test. The last comparison is between non-nested models and instead used the Vuong test for non-nested models. . . . .   | 9  |
| 6  | Results of unadjusted negative binomial model. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor with reference level of no t.alleles (t.alleles = 0). It appears that women with PIH are associated (not at 0.05 significance) with a lower incidence rate of TDLUs. Women with at least one T allele experience lower incidence rates of TDLUs (closer to a 0.05 significance rate). . . . .  | 10 |
| 7  | Results of unadjusted negative binomial model. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a binary variable with the reference of no T alleles (T alleles = 0). T alleles appear to have a protective effect (significant at 0.05 level) and PIH appears protective (not significant at 0.05). . . . .  | 10 |
| 8  | Results of adjusted negative binomial model without interaction term. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study, family history, and parity all seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates). Increasing age has a protective effect. Some family history has a harmful effect. Increasing parity has a harmful effect. . . . . | 10 |
| 9  | Results of adjusted negative binomial model with interaction terms. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study and parity are both significantly associated with TDLU count (when adjusting for the rest of the model covariates). . . . .  | 11 |
| 10 | Results of adjusted negative binomial model without interaction term. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study, family history, and parity all seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates). Increasing age has a protective effect. Some family history has a harmful effect. Increasing parity has a harmful effect. . . . . | 12 |
| 11 | Results of adjusted negative binomial model with interaction terms. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study and parity both seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates). . . . .   | 12 |
| 12 | [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH positive women, having two T alleles versus no T alleles appears protective (significant at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level). . . . .   | 13 |
| 13 | [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH positive women, having two T alleles versus no T alleles appears protective (approaching significance at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level). . . . .  | 13 |
| 14 | [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated linearly (e.g. as a linear trend). The reference for family history is no family history. Among PIH positive women, increasing the number of T alleles appears to have a protective association (significant at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level). . . . .   | 13 |

|    |   |    |
|----|---|----|
| 15 | [AMONG PIH POSITIVE WOMEN ONLY] Goodness of fit. As seen, the most flexible model (Model 1) has no significant improvement in goodness of fit over the less flexible models (binary categorization and linear trend). Further, the non-nested Vuong test between the binary and trend models does not reject the null hypothesis that the models are indistinguishable. This provides evidence that either the binary or the linear model provide a sufficient fit to the data relative to the other models and both require the same degrees of freedom. . . . . | 14 |
| 16 | [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a factor. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful). . . . .   | 14 |
| 17 | [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as binary. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful). . . . .   | 15 |
| 18 | [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a linear variable. The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful). . . . .   | 15 |
| 19 | [AMONG PIH NEGATIVE WOMEN ONLY] Goodness of fit. As seen, the most flexible model (Model 1) has no significant improvement in goodness of fit over the less flexible models (binary categorization and linear trend). Further, the non-nested Vuong test between the binary and trend models does not reject the null hypothesis that the models are indistinguishable. This provides evidence that either the binary or the linear model provide a sufficient fit to the data relative to the other models and both require the same degrees of freedom. . . . . | 15 |
| 1  | Summary statistics for whole cohort and broken down by PIH status. . . . .  | 17 |
| 2  | [AMONG PIH POSITIVE WOMEN] Summary statistics by genotype. . . . .  | 18 |
| 3  | [AMONG PIH NEGATIVE WOMEN] Summary statistics by genotype. . . . .  | 19 |
| 2  | Results of adjusted negative binomial model with interaction terms. . . . .   | 20 |
| 3  | Summary table of the adjusted T allele CRs from the models stratified on HDP-status. . . . .  | 21 |
| S1 | Goodness of fit test results . . . . .  | 22 |
| S2 | Complete adjusted negative binomial model with interaction terms . . . . .  | 22 |
| S3 | Complete adjusted negative binomial models stratified by HDP status . . . . .   | 23 |
| S4 | Complete adjusted negative binomial models stratified by HDP status . . . . .   | 23 |
| S5 | Complete adjusted negative binomial models stratified by HDP status with alleles treated linearly for trend . . . . .   | 23 |
| S6 | Complete adjusted negative binomial models stratified by HDP status with alleles treated linearly for trend . . . . .   | 24 |

## List of Figures

# 1 Examining the Data

TABLE 1: Summary statistics for whole cohort and broken down by PIH status.

|  | Total (N = 191) | PIH: 0 (N = 115) | PIH: 1 (N = 76) |
|--|-----------------|------------------|-----------------|
| <b>T Alleles</b>                           |                 |                  |                 |
| 0 T Alleles                                | 42 (21.99%)     | 24 (20.87%)      | 18 (23.68%)     |
| 1 T Alleles                                | 98 (51.31%)     | 61 (53.04%)      | 37 (48.68%)     |
| 2 T Alleles                                | 51 (26.70%)     | 30 (26.09%)      | 21 (27.63%)     |
| <b>Age at Entry to Study (years)</b>       |                 |                  |                 |
| Mean                                       | 45.9            | 46.2             | 45.45           |
| SD   | 10.61           | 10.45            | 10.89           |
| Min  | 27              | 27               | 27              |
| Max  | 66              | 66               | 66              |
| <b>BMI</b>                                 |                 |                  |                 |
| Mean                                       | 30.02           | 28.43            | 32.43           |
| SD   | 8.01            | 7.38             | 8.36            |
| Min  | 14.63           | 14.63            | 20.05           |
| Max  | 70.6            | 58.8             | 70.6            |
| <b>Parity</b>                              |                 |                  |                 |
| 1 Child                                    | 43 (22.51%)     | 26 (22.61%)      | 17 (22.37%)     |
| 2 Children                                 | 100 (52.36%)    | 64 (55.65%)      | 36 (47.37%)     |
| 3+ Children                                | 48 (25.13%)     | 25 (21.74%)      | 23 (30.26%)     |
| <b>Age at First Birth</b>                  |                 |                  |                 |
| Mean                                       | 27.02           | 27.32            | 26.57           |
| SD   | 5.07            | 4.9              | 5.33            |
| Min  | 15              | 19               | 15              |
| Max  | 43              | 43               | 39              |
| <b>Age at Menarche (years)</b>             |                 |                  |                 |
| Mean                                       | 12.52           | 12.6             | 12.41           |
| SD   | 1.51            | 1.47             | 1.58            |
| Min  | 8               | 9                | 8               |
| Max  | 17              | 17               | 16              |
| <b>Family History: 1st Degree Relative</b> |                 |                  |                 |
| No   | 140 (73.30%)    | 85 (73.91%)      | 55 (72.37%)     |
| Yes  | 51 (26.70%)     | 30 (26.09%)      | 21 (27.63%)     |
| <b>TDLUs</b>                               |                 |                  |                 |
| Mean                                       | 11.01           | 11.75            | 9.88            |
| SD   | 14.27           | 15.09            | 12.96           |
| Min  | 0               | 0                | 0               |
| Max  | 99              | 99               | 70              |

TABLE 2: [AMONG PIH POSITIVE WOMEN] Summary statistics by genotype.

|  | T alleles: 0 (N = 18) | T alleles: 1 (N = 37) | T alleles: 2 (N = 21) |
|--|-----------------------|-----------------------|-----------------------|
| <b>Age at Entry to Study (years)</b>       |                       |                       |                       |
| Mean                                       | 44                    | 44.73                 | 47.95                 |
| SD   | 10.81                 | 10.45                 | 11.8                  |
| Min  | 27                    | 28                    | 29                    |
| Max  | 61                    | 63                    | 66                    |
| <b>BMI</b>                                 |                       |                       |                       |
| Mean                                       | 29.76                 | 32.27                 | 35.02                 |
| SD   | 6.36                  | 9.16                  | 7.94                  |
| Min  | 20.05                 | 20.05                 | 22.15                 |
| Max  | 47.81                 | 70.6                  | 53.17                 |
| <b>Parity</b>                              |                       |                       |                       |
| 1 Child                                    | 5 (27.78%)            | 9 (24.32%)            | 3 (14.29%)            |
| 2 Children                                 | 9 (50.00%)            | 18 (48.65%)           | 9 (42.86%)            |
| 3+ Children                                | 4 (22.22%)            | 10 (27.03%)           | 9 (42.86%)            |
| <b>Age at First Birth</b>                  |                       |                       |                       |
| Mean                                       | 26.89                 | 26.57                 | 26.29                 |
| SD   | 5.89                  | 5.18                  | 5.35                  |
| Min  | 17                    | 15                    | 18                    |
| Max  | 38                    | 39                    | 36                    |
| <b>Age at Menarche (years)</b>             |                       |                       |                       |
| Mean                                       | 12.67                 | 12.14                 | 12.67                 |
| SD   | 1.57                  | 1.64                  | 1.46                  |
| Min  | 9                     | 8                     | 11                    |
| Max  | 15                    | 16                    | 16                    |
| <b>Family History: 1st Degree Relative</b> |                       |                       |                       |
| No   | 11 (61.11%)           | 29 (78.38%)           | 15 (71.43%)           |
| Yes  | 7 (38.89%)            | 8 (21.62%)            | 6 (28.57%)            |
| <b>TDLUs</b>                               |                       |                       |                       |
| Mean                                       | 16.28                 | 9.14                  | 5.71                  |
| SD   | 17.85                 | 12.22                 | 5.81                  |
| Min  | 0                     | 0                     | 0                     |
| Max  | 70                    | 47                    | 22                    |

TABLE 3: [AMONG PIH NEGATIVE WOMEN] Summary statistics by genotype.

|  | T alleles: 0 (N = 24) | T alleles: 1 (N = 61) | T alleles: 2 (N = 30) |
|--|-----------------------|-----------------------|-----------------------|
| <b>Age at Entry to Study (years)</b>       |                       |                       |                       |
| Mean                                       | 48.58                 | 46.16                 | 44.37                 |
| SD   | 10.45                 | 10.04                 | 11.25                 |
| Min  | 30                    | 27                    | 29                    |
| Max  | 63                    | 65                    | 66                    |
| <b>BMI</b>                                 |                       |                       |                       |
| Mean                                       | 29.31                 | 28.7                  | 27.15                 |
| SD   | 9.03                  | 6.62                  | 7.48                  |
| Min  | 19.41                 | 17.96                 | 14.63                 |
| Max  | 51.98                 | 47.53                 | 58.8                  |
| <b>Parity</b>                              |                       |                       |                       |
| 1 Child                                    | 1 (4.17%)             | 17 (27.87%)           | 8 (26.67%)            |
| 2 Children                                 | 18 (75.00%)           | 31 (50.82%)           | 15 (50.00%)           |
| 3+ Children                                | 5 (20.83%)            | 13 (21.31%)           | 7 (23.33%)            |
| <b>Age at First Birth</b>                  |                       |                       |                       |
| Mean                                       | 27.08                 | 27.56                 | 27.03                 |
| SD   | 5.06                  | 5.14                  | 4.37                  |
| Min  | 20                    | 19                    | 19                    |
| Max  | 38                    | 43                    | 40                    |
| <b>Age at Menarche (years)</b>             |                       |                       |                       |
| Mean                                       | 12.5                  | 12.25                 | 13.4                  |
| SD   | 1.5                   | 1.31                  | 1.48                  |
| Min  | 9                     | 9                     | 11                    |
| Max  | 17                    | 16                    | 16                    |
| <b>Family History: 1st Degree Relative</b> |                       |                       |                       |
| No   | 18 (75.00%)           | 43 (70.49%)           | 24 (80.00%)           |
| Yes  | 6 (25.00%)            | 18 (29.51%)           | 6 (20.00%)            |
| <b>TDLUs</b>                               |                       |                       |                       |
| Mean                                       | 13.46                 | 10.77                 | 12.37                 |
| SD   | 22.71                 | 12.83                 | 11.85                 |
| Min  | 0                     | 0                     | 0                     |
| Max  | 99                    | 59                    | 45                    |

| TDLU_cat | n   | percent |
|----------|-----|---------|
| non-zero | 169 | 88.48   |
| zero     | 22  | 11.52   |

TABLE 4: Number of non-zero TDLU counts.



## 2 Zero-Inflated versus Hurdle Models

Given that 10% of our data is comprised of zeros, we want to consider a zero-inflated model alongside the standard Poisson and Negative Binomial regression fits.

### 2.1 Zero-Inflated

Zero-inflated models assume that the zeros arise from two sources: *structural* and *sampling*. The *structural* component is those not at risk (e.g. non-smokers smoke 0 cigarettes, non-fishers catch 0 fish, etc.). The *sampling* component is those who are at risk, but by the count process, sample 0 anyway (e.g. fishers can catch 0 fish, sexually active individual may have 0 risky encounters, etc.)

The zero-inflated Poisson distribution for participant  $i$  can be defined as: [Source: Rose et al.]

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i} & y_i = 0 \\ (1 - p_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & y_i > 0 \end{cases} \quad (1)$$

where  $p_i$  is the probability of being an excess zero and is often modeled using logistic regression:

$$\text{logit}(p_i) = X\beta \quad (2)$$

“Interpretation of the ZIP model depends upon what is being modeled. For medical studies the zero-inflated portion can be thought of as the odds of moving from the non-risk to the at-risk group. Once in the at-risk group we can determine the expected number of events or the risk of an event for one group versus another group.”

A zero-inflated negative binomial model can be parameterized similarly.

### 2.2 Hurdle

A hurdle model, on the other hand, assumes all excess zeros are structural (e.g. only non-smokers smoke 0 cigarettes.) As such, it has two parts: 1) a binary response model, 2) a truncated-at-zero count model.

A hurdle model can be expressed as:

$$P(Y = 0) = f_1(0) = p \quad (3)$$

$$P(Y = i) = (1 - p)\frac{f_2(i)}{1 - f_2(0)} = (1 - p)f_2'(i) \quad i > 0 \quad (4)$$

for any probability density functions  $f_1$  and  $f_2$  corresponding to non-negative integers.  $f_1$  is the hurdle and  $f_2$  represents the count process.

“This allows us to interpret the positive outcomes ( $>0$ ) that result from passing the zero hurdle (threshold). The hurdle portion of the two-part model estimates the probability that the threshold is crossed. Theoretically the threshold could be any value, but its usually taken as zero because this is most often meaningful in the context of the study objectives.”

### 2.3 Follow-Up Questions

1. For an individual who is screened, are zero counts structural? In other words, at the time of screening do the zero counts arise strictly from those who do not have breast cancer? Or are we allowing the possibility that a woman with breast cancer could also have a TDLU count of 0?

### 3 Testing for Zero-Inflation

We'll do comparative model fitting of the following models:

- Poisson model
- Negative binomial model (NB)
- Zero-inflated Poisson (ZIP)
- Zero-inflated negative binomial (ZINB)
- Hurdle Model with Poisson counting process (HPois)
- Hurdle Model with negative binomial counting process (HNB)

Nested models directly compared via likelihood ratio test:

- Poisson is nested within the NB model.
- Poisson is nested within ZIP.
- NB is nested with ZINB.

The Vuong test for non-nested models will be used to distinguish between the previous set of models and the hurdle models.

#### 3.1 Goodness-of-Fit Comparisons

Models: TDLUs  $\sim$  PIH + I(T Alleles = 1) + I(T Alleles = 2)

| alternative.hypothesis | p.value |
|------------------------|---------|
| NB > Poisson           | 0.000   |
| ZIP > Poisson          | 0.000   |
| ZINB > NB              | 0.660   |
| NB > HNB               | 0.046   |

TABLE 5: Testing for zero-inflation via model comparisons. The first three comparisons are nested models (with the latter nested in the former) and are compared via a likelihood ratio test. The last comparison is between non-nested models and instead used the Vuong test for non-nested models.

According to this test, the standard negative binomial model fits better or just as well as any other model. We'll proceed with the negative binomial model.

## 4 Unadjusted Negative Binomial Model Results

|             | Coefficient | Standard Error | Z Value | p-value | IRR    | CI.lb  | CI.ub  |
|-------------|-------------|----------------|---------|---------|--------|--------|--------|
| (Intercept) | 2.803       | 0.200          | 14.014  | 0.000   | 16.487 | 11.141 | 24.400 |
| PIH         | -0.234      | 0.179          | -1.306  | 0.191   | 0.791  | 0.557  | 1.124  |
| t.alleles1  | -0.399      | 0.222          | -1.800  | 0.072   | 0.671  | 0.434  | 1.036  |
| t.alleles2  | -0.473      | 0.251          | -1.884  | 0.060   | 0.623  | 0.381  | 1.019  |

TABLE 6: Results of unadjusted negative binomial model. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor with reference level of no t.alleles (t.alleles = 0). It appears that women with PIH are associated (not at 0.05 significance) with a lower incidence rate of TDLUs. Women with at least one T allele experience lower incidence rates of TDLUs (closer to a 0.05 significance rate).

|                        | Coefficient | Standard Error | Z Value | p-value | IRR   | CI.lb | CI.ub |
|------------------------|-------------|----------------|---------|---------|-------|-------|-------|
| (Intercept)            | 2.80        | 0.20           | 14.00   | 0.00    | 16.44 | 11.11 | 24.34 |
| PIH                    | -0.23       | 0.18           | -1.28   | 0.20    | 0.80  | 0.56  | 1.13  |
| ‘T alleles‘one or more | -0.42       | 0.21           | -2.01   | 0.04    | 0.65  | 0.43  | 0.99  |

TABLE 7: Results of unadjusted negative binomial model. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a binary variable with the reference of no T alleles (T alleles = 0). T alleles appear to have a protective effect (significant at 0.05 level) and PIH appears protective (not significant at 0.05).

A likelihood ratio test between these two unadjusted models shows that the binary treatment of the T alleles variable is no worse of a fit than the model comparing 0, 1, and 2 T alleles (p-value: 0.725).

### 4.1 Adjusted Negative Binomial Model Results

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR     | CI.lb  | CI.ub    |
|----------------------------------|-------------|----------------|---------|---------|---------|--------|----------|
| (Intercept)                      | 5.351       | 1.121          | 4.775   | 0.000   | 210.924 | 23.448 | 1897.319 |
| PIH                              | -0.134      | 0.171          | -0.787  | 0.432   | 0.874   | 0.626  | 1.222    |
| ‘T alleles‘1                     | -0.178      | 0.205          | -0.865  | 0.387   | 0.837   | 0.560  | 1.252    |
| ‘T alleles‘2                     | -0.200      | 0.234          | -0.854  | 0.393   | 0.819   | 0.517  | 1.296    |
| AFB                              | 0.006       | 0.017          | 0.345   | 0.730   | 1.006   | 0.973  | 1.040    |
| Age                              | -0.044      | 0.008          | -5.602  | 0.000   | 0.957   | 0.943  | 0.972    |
| ‘Family Hist‘some family history | 0.378       | 0.185          | 2.043   | 0.041   | 1.460   | 1.016  | 2.099    |
| BMI                              | -0.022      | 0.011          | -1.987  | 0.047   | 0.978   | 0.957  | 1.000    |
| Parity                           | 0.305       | 0.106          | 2.870   | 0.004   | 1.357   | 1.102  | 1.671    |
| Menarche                         | -0.091      | 0.058          | -1.570  | 0.116   | 0.913   | 0.815  | 1.023    |

TABLE 8: Results of adjusted negative binomial model without interaction term. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study, family history, and parity all seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates). Increasing age has a protective effect. Some family history has a harmful effect. Increasing parity has a harmful effect.

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR     | CI.lb  | CI.ub    |
|----------------------------------|-------------|----------------|---------|---------|---------|--------|----------|
| (Intercept)                      | 5.056       | 1.135          | 4.453   | 0.000   | 156.895 | 16.954 | 1451.892 |
| PIH                              | 0.208       | 0.345          | 0.604   | 0.546   | 1.231   | 0.627  | 2.419    |
| 'T alleles'1                     | -0.027      | 0.268          | -0.102  | 0.918   | 0.973   | 0.575  | 1.645    |
| 'T alleles'2                     | 0.099       | 0.307          | 0.322   | 0.747   | 1.104   | 0.605  | 2.016    |
| AFB                              | 0.008       | 0.017          | 0.498   | 0.619   | 1.008   | 0.975  | 1.043    |
| Age                              | -0.041      | 0.008          | -5.232  | 0.000   | 0.960   | 0.945  | 0.975    |
| 'Family Hist'some family history | 0.372       | 0.185          | 2.008   | 0.045   | 1.451   | 1.009  | 2.086    |
| BMI                              | -0.022      | 0.011          | -1.918  | 0.055   | 0.979   | 0.957  | 1.000    |
| Parity                           | 0.334       | 0.107          | 3.132   | 0.002   | 1.396   | 1.133  | 1.720    |
| Menarche                         | -0.102      | 0.058          | -1.763  | 0.078   | 0.903   | 0.806  | 1.011    |
| PIH:'T alleles'1                 | -0.310      | 0.417          | -0.743  | 0.457   | 0.734   | 0.324  | 1.661    |
| PIH:'T alleles'2                 | -0.740      | 0.482          | -1.536  | 0.124   | 0.477   | 0.185  | 1.227    |

TABLE 9: Results of adjusted negative binomial model with interaction terms. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study and parity are both significantly associated with TDLU count (when adjusting for the rest of the model covariates).

#### 4.1.1 Using Delta Method for Covariate Combinations

```

broom::tidy(m_nb_adj_int_centered)

## # A tibble: 12 x 5
##   term                                estimate std.error statistic    p.value
##   <chr>                                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                        1.85      0.494       3.74  1.83e-4
## 2 PIH                                0.208     0.345       0.604  5.46e-1
## 3 `T alleles`1                       -0.0275    0.268     -0.102  9.18e-1
## 4 `T alleles`2                        0.0990    0.307       0.322  7.47e-1
## 5 AFB                                0.00845    0.0170     0.498  6.19e-1
## 6 Age                                -0.0408    0.00779    -5.23  1.68e-7
## 7 `Family Hist`some family history  0.372     0.185       2.01  4.47e-2
## 8 BMI                                -0.0215    0.0112     -1.92  5.51e-2
## 9 Parity                             0.334     0.107       3.13  1.73e-3
## 10 Menarche                         -0.102     0.0578     -1.76  7.79e-2
## 11 PIH:`T alleles`1                 -0.310     0.417     -0.743  4.57e-1
## 12 PIH:`T alleles`2                 -0.740     0.482     -1.54  1.24e-1

new_names <- str_replace(names(m_nb_adj_int_centered$coef), ":", "_")
new_names <- str_replace_all(new_names, "`", "")
new_names <- str_replace_all(new_names, " ", "")

# Comparing TT to GG for HDP+
comp_TT_GG_HDPplus <- deltaMethod(m_nb_adj_int_centered,
  "Talleles2 + PIH_Talleles2",
  parameterNames = new_names)

2*pnorm(comp_TT_GG_HDPplus$Estimate/comp_TT_GG_HDPplus$SE, mean = 0, sd = 1)
## [1] 0.07988333

# Comparing GT to GG for HDP+
comp_GT_GG_HDPplus <- deltaMethod(m_nb_adj_int_centered,
  "Talleles1 + PIH_Talleles1",
  parameterNames = new_names)

2*pnorm(comp_GT_GG_HDPplus$Estimate/comp_GT_GG_HDPplus$SE, mean = 0, sd = 1)
## [1] 0.2884442

# Comparing TT to GG for HDP-
broom::tidy(m_nb_adj_int_centered) %>%
  filter(term == "`T alleles`2")

```

```
## # A tibble: 1 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 `T alleles`2  0.0990    0.307    0.322    0.747

# Comparing GT to GG for HDP-
broom::tidy(m_nb_adj_int_centered) %>%
  filter(term == "`T alleles`1")

## # A tibble: 1 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 `T alleles`1 -0.0275    0.268   -0.102    0.918
```

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR     | CI.lb  | CI.ub    |
|----------------------------------|-------------|----------------|---------|---------|---------|--------|----------|
| (Intercept)                      | 5.381       | 1.091          | 4.933   | 0.000   | 217.278 | 25.615 | 1843.072 |
| PIH                              | -0.134      | 0.171          | -0.783  | 0.434   | 0.875   | 0.626  | 1.222    |
| `T alleles`one or more           | -0.185      | 0.194          | -0.954  | 0.340   | 0.831   | 0.569  | 1.215    |
| AFB                              | 0.006       | 0.017          | 0.348   | 0.728   | 1.006   | 0.973  | 1.040    |
| Age                              | -0.044      | 0.008          | -5.606  | 0.000   | 0.957   | 0.943  | 0.972    |
| `Family Hist`some family history | 0.381       | 0.185          | 2.056   | 0.040   | 1.463   | 1.018  | 2.103    |
| BMI                              | -0.022      | 0.011          | -2.009  | 0.045   | 0.978   | 0.957  | 0.999    |
| Parity                           | 0.303       | 0.106          | 2.871   | 0.004   | 1.354   | 1.101  | 1.665    |
| Menarche                         | -0.093      | 0.056          | -1.662  | 0.097   | 0.911   | 0.817  | 1.017    |

TABLE 10: Results of adjusted negative binomial model without interaction term. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated a a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study, family history, and parity all seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates). Increasing age has a protective effect. Some family history has a harmful effect. Increasing parity has a harmful effect.

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR     | CI.lb  | CI.ub    |
|----------------------------------|-------------|----------------|---------|---------|---------|--------|----------|
| (Intercept)                      | 5.090       | 1.108          | 4.594   | 0.000   | 162.368 | 18.508 | 1424.405 |
| PIH                              | 0.197       | 0.346          | 0.570   | 0.569   | 1.218   | 0.619  | 2.397    |
| `T alleles`one or more           | 0.005       | 0.256          | 0.021   | 0.983   | 1.005   | 0.609  | 1.660    |
| AFB                              | 0.009       | 0.017          | 0.523   | 0.601   | 1.009   | 0.976  | 1.043    |
| Age                              | -0.043      | 0.008          | -5.464  | 0.000   | 0.958   | 0.944  | 0.973    |
| `Family Hist`some family history | 0.349       | 0.185          | 1.883   | 0.060   | 1.418   | 0.986  | 2.040    |
| BMI                              | -0.021      | 0.011          | -1.863  | 0.062   | 0.979   | 0.958  | 1.001    |
| Parity                           | 0.322       | 0.106          | 3.040   | 0.002   | 1.379   | 1.121  | 1.697    |
| Menarche                         | -0.098      | 0.056          | -1.752  | 0.080   | 0.907   | 0.813  | 1.012    |
| PIH:`T alleles`one or more       | -0.440      | 0.399          | -1.103  | 0.270   | 0.644   | 0.295  | 1.407    |

TABLE 11: Results of adjusted negative binomial model with interaction terms. PIH compares non-PIH women (PIH = 0) to PIH women (PIH = 1). T alleles are treated as a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Age at time of study and parity both seem to be significantly associated with TDLU count (when adjusting for the rest of the model covariates).

Using a likelihood ratio test to compare the models (without interaction terms) with 0, 1, and 2 T alleles versus the model with binary classification of one or more T alleles, we obtain a p-value of 0.913, indicating that the model with more levels is not a better fit than the one with fewer. We can use the model with the binary treatment of T alleles.

## 4.2 Adjusted Stratified Models

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR      | CI.lb   | CI.ub      |
|----------------------------------|-------------|----------------|---------|---------|----------|---------|------------|
| (Intercept)                      | 8.545       | 1.887          | 4.528   | 0.000   | 5142.603 | 127.282 | 207778.285 |
| 'T alleles'1                     | -0.500      | 0.326          | -1.536  | 0.124   | 0.606    | 0.320   | 1.148      |
| 'T alleles'2                     | -0.760      | 0.386          | -1.967  | 0.049   | 0.468    | 0.219   | 0.997      |
| AFB                              | -0.045      | 0.028          | -1.628  | 0.104   | 0.956    | 0.905   | 1.009      |
| Age                              | -0.023      | 0.013          | -1.793  | 0.073   | 0.978    | 0.954   | 1.002      |
| 'Family Hist'some family history | 0.071       | 0.303          | 0.235   | 0.814   | 1.074    | 0.593   | 1.946      |
| BMI                              | -0.027      | 0.018          | -1.522  | 0.128   | 0.973    | 0.940   | 1.008      |
| Parity                           | -0.048      | 0.179          | -0.267  | 0.790   | 0.953    | 0.671   | 1.354      |
| Menarche                         | -0.224      | 0.095          | -2.360  | 0.018   | 0.799    | 0.663   | 0.963      |

TABLE 12: [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH positive women, having two T alleles versus no T alleles appears protective (significant at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level).

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR      | CI.lb   | CI.ub      |
|----------------------------------|-------------|----------------|---------|---------|----------|---------|------------|
| (Intercept)                      | 8.805       | 1.791          | 4.917   | 0.000   | 6665.206 | 199.293 | 222912.492 |
| 'T alleles'one or more           | -0.576      | 0.310          | -1.857  | 0.063   | 0.562    | 0.306   | 1.033      |
| AFB                              | -0.044      | 0.028          | -1.585  | 0.113   | 0.957    | 0.906   | 1.011      |
| Age                              | -0.025      | 0.013          | -1.970  | 0.049   | 0.975    | 0.952   | 1.000      |
| 'Family Hist'some family history | 0.062       | 0.304          | 0.204   | 0.839   | 1.064    | 0.587   | 1.929      |
| BMI                              | -0.028      | 0.017          | -1.609  | 0.108   | 0.972    | 0.940   | 1.006      |
| Parity                           | -0.078      | 0.175          | -0.444  | 0.657   | 0.925    | 0.657   | 1.303      |
| Menarche                         | -0.233      | 0.092          | -2.526  | 0.012   | 0.792    | 0.661   | 0.949      |

TABLE 13: [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a binary variable. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH positive women, having two T alleles versus no T alleles appears protective (approaching significance at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level).

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR      | CI.lb   | CI.ub      |
|----------------------------------|-------------|----------------|---------|---------|----------|---------|------------|
| (Intercept)                      | 8.415       | 1.812          | 4.645   | 0.000   | 4516.286 | 129.566 | 157424.573 |
| 'T alleles'                      | -0.380      | 0.194          | -1.960  | 0.050   | 0.684    | 0.468   | 1.000      |
| AFB                              | -0.046      | 0.028          | -1.645  | 0.100   | 0.955    | 0.904   | 1.009      |
| Age                              | -0.022      | 0.013          | -1.767  | 0.077   | 0.978    | 0.954   | 1.002      |
| 'Family Hist'some family history | 0.103       | 0.300          | 0.342   | 0.733   | 1.108    | 0.615   | 1.996      |
| BMI                              | -0.028      | 0.018          | -1.560  | 0.119   | 0.973    | 0.939   | 1.007      |
| Parity                           | -0.032      | 0.178          | -0.182  | 0.856   | 0.968    | 0.684   | 1.371      |
| Menarche                         | -0.220      | 0.093          | -2.375  | 0.018   | 0.802    | 0.669   | 0.962      |

TABLE 14: [AMONG PIH POSITIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated linearly (e.g. as a linear trend). The reference for family history is no family history. Among PIH positive women, increasing the number of T alleles appears to have a protective association (significant at 0.05 level). Age at first birth is approaching significance (protective). Increasing age at study entry appears to be protective (significant at 0.05 level).

Tests can be used to compare these three models with respect to goodness of fit. Specifically, we have three models:

**Model 1** Treating T alleles as a factor. This is the most flexible model and it allows the effect of one T allele to differ from the effect of two T alleles.

**Model 2** Treating T alleles as a binary variable. This essentially assumes a plateau effect where having one T allele is the same as having two T alleles.

**Model 3** Treating T alleles as a linear variable. This enforces a linear dose-response relationship between the number of T alleles and the log count of TDLUs.

| Model Comparison                          | $\chi^2$ p-value |
|---|------------------|
| Model 1 v Model 2 (nested, LRT)           | 0.451            |
| Model 1 v Model 3 (nested, LRT)           | 0.664            |
| Model 2 v Model 3 (non-nested Vuong Test) | 0.337            |

TABLE 15: [AMONG PIH POSITIVE WOMEN ONLY] Goodness of fit. As seen, the most flexible model (Model 1) has no significant improvement in goodness of fit over the less flexible models (binary categorization and linear trend). Further, the non-nested Vuong test between the binary and trend models does not reject the null hypothesis that the models are indistinguishable. This provides evidence that either the binary or the linear model provide a sufficient fit to the data relative to the other models and both require the same degrees of freedom.

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR    | CI.lb | CI.ub   |
|----------------------------------|-------------|----------------|---------|---------|--------|-------|---------|
| (Intercept)                      | 3.789       | 1.393          | 2.721   | 0.007   | 44.219 | 2.885 | 677.757 |
| ‘T alleles’1                     | 0.013       | 0.260          | 0.052   | 0.959   | 1.014  | 0.608 | 1.689   |
| ‘T alleles’2                     | 0.080       | 0.302          | 0.264   | 0.792   | 1.083  | 0.600 | 1.955   |
| AFB                              | 0.022       | 0.021          | 1.022   | 0.307   | 1.022  | 0.980 | 1.065   |
| Age                              | -0.043      | 0.010          | -4.302  | 0.000   | 0.958  | 0.940 | 0.977   |
| ‘Family Hist’some family history | 0.441       | 0.229          | 1.924   | 0.054   | 1.554  | 0.992 | 2.435   |
| BMI                              | -0.020      | 0.014          | -1.369  | 0.171   | 0.980  | 0.953 | 1.009   |
| Parity                           | 0.398       | 0.131          | 3.028   | 0.002   | 1.489  | 1.151 | 1.927   |
| Menarche                         | -0.041      | 0.073          | -0.556  | 0.578   | 0.960  | 0.831 | 1.109   |

TABLE 16: [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a factor. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful).

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR    | CI.lb | CI.ub   |
|----------------------------------|-------------|----------------|---------|---------|--------|-------|---------|
| (Intercept)                      | 3.699       | 1.371          | 2.699   | 0.007   | 40.419 | 2.752 | 593.541 |
| ‘T alleles’one or more           | 0.035       | 0.248          | 0.142   | 0.887   | 1.036  | 0.637 | 1.684   |
| AFB                              | 0.022       | 0.021          | 1.031   | 0.303   | 1.022  | 0.981 | 1.065   |
| Age                              | -0.043      | 0.010          | -4.322  | 0.000   | 0.958  | 0.940 | 0.977   |
| ‘Family Hist’some family history | 0.427       | 0.228          | 1.872   | 0.061   | 1.533  | 0.980 | 2.397   |
| BMI                              | -0.019      | 0.014          | -1.332  | 0.183   | 0.981  | 0.953 | 1.009   |
| Parity                           | 0.401       | 0.132          | 3.052   | 0.002   | 1.494  | 1.154 | 1.933   |
| Menarche                         | -0.035      | 0.070          | -0.500  | 0.617   | 0.966  | 0.842 | 1.107   |

TABLE 17: [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as binary. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful).

|                                  | Coefficient | Standard Error | Z Value | p-value | IRR    | CI.lb | CI.ub   |
|----------------------------------|-------------|----------------|---------|---------|--------|-------|---------|
| (Intercept)                      | 3.737       | 1.346          | 2.776   | 0.006   | 41.951 | 2.999 | 586.751 |
| ‘T alleles’                      | 0.041       | 0.151          | 0.270   | 0.787   | 1.042  | 0.775 | 1.400   |
| AFB                              | 0.022       | 0.021          | 1.031   | 0.303   | 1.022  | 0.981 | 1.065   |
| Age                              | -0.043      | 0.010          | -4.308  | 0.000   | 0.958  | 0.940 | 0.977   |
| ‘Family Hist’some family history | 0.436       | 0.228          | 1.911   | 0.056   | 1.547  | 0.989 | 2.419   |
| BMI                              | -0.020      | 0.014          | -1.348  | 0.178   | 0.981  | 0.953 | 1.009   |
| Parity                           | 0.402       | 0.131          | 3.071   | 0.002   | 1.495  | 1.157 | 1.933   |
| Menarche                         | -0.039      | 0.072          | -0.550  | 0.582   | 0.961  | 0.836 | 1.106   |

TABLE 18: [AMONG PIH NEGATIVE WOMEN ONLY] Results of adjusted negative binomial model. T alleles are treated as a linear variable. The reference for family history is no family history. Among PIH negative women, T alleles appear to have no association. Increasing age at study entry appears to be protective (significant at 0.05 level). Increasing parity appears to be harmful (significant at 0.05 level). Family history approaches significance (harmful).

Again, tests can be used to compare these three models with respect to goodness of fit. Specifically, we have three models:

**Model 1** Treating T alleles as a factor. This is the most flexible model and it allows the effect of one T allele to differ from the effect of two T alleles.

**Model 2** Treating T alleles as a binary variable. This essentially assumes a plateau effect where having one T allele is the same as having two T alleles.

**Model 3** Treating T alleles as a linear variable. This enforces a linear dose-response relationship between the number of T alleles and the log count of TDLUs.

| Model Comparison                          | $\chi^2$ p-value |
|---|------------------|
| Model 1 v Model 2 (nested, LRT)           | 0.796            |
| Model 1 v Model 3 (nested, LRT)           | 0.902            |
| Model 2 v Model 3 (non-nested Vuong Test) | 0.466            |

TABLE 19: [AMONG PIH NEGATIVE WOMEN ONLY] Goodness of fit. As seen, the most flexible model (Model 1) has no significant improvement in goodness of fit over the less flexible models (binary categorization and linear trend). Further, the non-nested Vuong test between the binary and trend models does not reject the null hypothesis that the models are indistinguishable. This provides evidence that either the binary or the linear model provide a sufficient fit to the data relative to the other models and both require the same degrees of freedom.



## 5 Tables and Figures for Paper

TABLE 1: Summary statistics for whole cohort and broken down by PIH status.

|  | Total (N = 191) | PIH: 0 (N = 115) | PIH: 1 (N = 76) |
|--|-----------------|------------------|-----------------|
| <b>T Alleles</b>                           |                 |                  |                 |
| 0 T Alleles                                | 42 (21.99%)     | 24 (20.87%)      | 18 (23.68%)     |
| 1 T Alleles                                | 98 (51.31%)     | 61 (53.04%)      | 37 (48.68%)     |
| 2 T Alleles                                | 51 (26.70%)     | 30 (26.09%)      | 21 (27.63%)     |
| <b>Age at Entry to Study (years)</b>       |                 |                  |                 |
| Mean                                       | 45.9            | 46.2             | 45.45           |
| SD   | 10.61           | 10.45            | 10.89           |
| Min  | 27              | 27               | 27              |
| Max  | 66              | 66               | 66              |
| <b>BMI</b>                                 |                 |                  |                 |
| Mean                                       | 30.02           | 28.43            | 32.43           |
| SD   | 8.01            | 7.38             | 8.36            |
| Min  | 14.63           | 14.63            | 20.05           |
| Max  | 70.6            | 58.8             | 70.6            |
| <b>Parity</b>                              |                 |                  |                 |
| 1 Child                                    | 43 (22.51%)     | 26 (22.61%)      | 17 (22.37%)     |
| 2 Children                                 | 100 (52.36%)    | 64 (55.65%)      | 36 (47.37%)     |
| 3+ Children                                | 48 (25.13%)     | 25 (21.74%)      | 23 (30.26%)     |
| <b>Age at First Birth</b>                  |                 |                  |                 |
| Mean                                       | 27.02           | 27.32            | 26.57           |
| SD   | 5.07            | 4.9              | 5.33            |
| Min  | 15              | 19               | 15              |
| Max  | 43              | 43               | 39              |
| <b>Age at Menarche (years)</b>             |                 |                  |                 |
| Mean                                       | 12.52           | 12.6             | 12.41           |
| SD   | 1.51            | 1.47             | 1.58            |
| Min  | 8               | 9                | 8               |
| Max  | 17              | 17               | 16              |
| <b>Family History: 1st Degree Relative</b> |                 |                  |                 |
| No   | 140 (73.30%)    | 85 (73.91%)      | 55 (72.37%)     |
| Yes  | 51 (26.70%)     | 30 (26.09%)      | 21 (27.63%)     |
| <b>TDLUs</b>                               |                 |                  |                 |
| Mean                                       | 11.01           | 11.75            | 9.88            |
| SD   | 14.27           | 15.09            | 12.96           |
| Min  | 0               | 0                | 0               |
| Max  | 99              | 99               | 70              |

TABLE 2: [AMONG PIH POSITIVE WOMEN] Summary statistics by genotype.

|  | T alleles: 0 (N = 18) | T alleles: 1 (N = 37) | T alleles: 2 (N = 21) |
|--|-----------------------|-----------------------|-----------------------|
| <b>Age at Entry to Study (years)</b>       |                       |                       |                       |
| Mean                                       | 44                    | 44.73                 | 47.95                 |
| SD   | 10.81                 | 10.45                 | 11.8                  |
| Min  | 27                    | 28                    | 29                    |
| Max  | 61                    | 63                    | 66                    |
| <b>BMI</b>                                 |                       |                       |                       |
| Mean                                       | 29.76                 | 32.27                 | 35.02                 |
| SD   | 6.36                  | 9.16                  | 7.94                  |
| Min  | 20.05                 | 20.05                 | 22.15                 |
| Max  | 47.81                 | 70.6                  | 53.17                 |
| <b>Parity</b>                              |                       |                       |                       |
| 1 Child                                    | 5 (27.78%)            | 9 (24.32%)            | 3 (14.29%)            |
| 2 Children                                 | 9 (50.00%)            | 18 (48.65%)           | 9 (42.86%)            |
| 3+ Children                                | 4 (22.22%)            | 10 (27.03%)           | 9 (42.86%)            |
| <b>Age at First Birth</b>                  |                       |                       |                       |
| Mean                                       | 26.89                 | 26.57                 | 26.29                 |
| SD   | 5.89                  | 5.18                  | 5.35                  |
| Min  | 17                    | 15                    | 18                    |
| Max  | 38                    | 39                    | 36                    |
| <b>Age at Menarche (years)</b>             |                       |                       |                       |
| Mean                                       | 12.67                 | 12.14                 | 12.67                 |
| SD   | 1.57                  | 1.64                  | 1.46                  |
| Min  | 9                     | 8                     | 11                    |
| Max  | 15                    | 16                    | 16                    |
| <b>Family History: 1st Degree Relative</b> |                       |                       |                       |
| No   | 11 (61.11%)           | 29 (78.38%)           | 15 (71.43%)           |
| Yes  | 7 (38.89%)            | 8 (21.62%)            | 6 (28.57%)            |
| <b>TDLUs</b>                               |                       |                       |                       |
| Mean                                       | 16.28                 | 9.14                  | 5.71                  |
| SD   | 17.85                 | 12.22                 | 5.81                  |
| Min  | 0                     | 0                     | 0                     |
| Max  | 70                    | 47                    | 22                    |

TABLE 3: [AMONG PIH NEGATIVE WOMEN] Summary statistics by genotype.

|  | T alleles: 0 (N = 24) | T alleles: 1 (N = 61) | T alleles: 2 (N = 30) |
|--|-----------------------|-----------------------|-----------------------|
| <b>Age at Entry to Study (years)</b>       |                       |                       |                       |
| Mean                                       | 48.58                 | 46.16                 | 44.37                 |
| SD   | 10.45                 | 10.04                 | 11.25                 |
| Min  | 30                    | 27                    | 29                    |
| Max  | 63                    | 65                    | 66                    |
| <b>BMI</b>                                 |                       |                       |                       |
| Mean                                       | 29.31                 | 28.7                  | 27.15                 |
| SD   | 9.03                  | 6.62                  | 7.48                  |
| Min  | 19.41                 | 17.96                 | 14.63                 |
| Max  | 51.98                 | 47.53                 | 58.8                  |
| <b>Parity</b>                              |                       |                       |                       |
| 1 Child                                    | 1 (4.17%)             | 17 (27.87%)           | 8 (26.67%)            |
| 2 Children                                 | 18 (75.00%)           | 31 (50.82%)           | 15 (50.00%)           |
| 3+ Children                                | 5 (20.83%)            | 13 (21.31%)           | 7 (23.33%)            |
| <b>Age at First Birth</b>                  |                       |                       |                       |
| Mean                                       | 27.08                 | 27.56                 | 27.03                 |
| SD   | 5.06                  | 5.14                  | 4.37                  |
| Min  | 20                    | 19                    | 19                    |
| Max  | 38                    | 43                    | 40                    |
| <b>Age at Menarche (years)</b>             |                       |                       |                       |
| Mean                                       | 12.5                  | 12.25                 | 13.4                  |
| SD   | 1.5                   | 1.31                  | 1.48                  |
| Min  | 9                     | 9                     | 11                    |
| Max  | 17                    | 16                    | 16                    |
| <b>Family History: 1st Degree Relative</b> |                       |                       |                       |
| No   | 18 (75.00%)           | 43 (70.49%)           | 24 (80.00%)           |
| Yes  | 6 (25.00%)            | 18 (29.51%)           | 6 (20.00%)            |
| <b>TDLUs</b>                               |                       |                       |                       |
| Mean                                       | 13.46                 | 10.77                 | 12.37                 |
| SD   | 22.71                 | 12.83                 | 11.85                 |
| Min  | 0                     | 0                     | 0                     |
| Max  | 99                    | 59                    | 45                    |

TABLE 2: Results of adjusted negative binomial model with interaction terms.

|                            | Coefficient | Standard Error | Z Value | p-value | CR (95% CI)          |
|----------------------------|-------------|----------------|---------|---------|----------------------|
| HDP                        | 0.208       | 0.345          | 0.604   | 0.546   | 1.231 (0.627, 2.419) |
| T alleles = 1              | -0.027      | 0.268          | -0.102  | 0.918   | 0.973 (0.575, 1.645) |
| T alleles = 2              | 0.099       | 0.307          | 0.322   | 0.747   | 1.104 (0.605, 2.016) |
| HDP $\times$ T alleles = 1 | -0.310      | 0.417          | -0.743  | 0.457   | 0.734 (0.324, 1.661) |
| HDP $\times$ T alleles = 2 | -0.740      | 0.482          | -1.536  | 0.124   | 0.477 (0.185, 1.227) |

HDP compares HDP-positive women to HDP-negative women. T alleles are treated as a factor variable. The reference for T alleles is no T alleles (T alleles = 0). These results are adjusted for family history, age at biopsy, parity, age at menarche, age at first birth, and BMI. Full model covariates can be found in the supplemental material, Table S2.

TABLE 3: Summary table of the adjusted T allele CRs from the models stratified on HDP-status.

|                    | <b>HDP-Negative</b>  |         | <b>HDP-Positive</b>  |         |
|--------------------|----------------------|---------|----------------------|---------|
|                    | CR (95% CI)          | p-value | CR (95% CI)          | p-value |
| Factor Model       |                      |         |                      |         |
| T Alleles = 0      | 1 (ref)              |         | 1 (ref)              |         |
| T Alleles = 1      | 1.014 (0.608, 1.689) | 0.959   | 0.606 (0.320, 1.148) | 0.124   |
| T Alleles = 2      | 1.083 (0.600, 1.955) | 0.792   | 0.468 (0.219, 0.997) | 0.049   |
| Linear Trend Model |                      |         |                      |         |
| $\Delta$ T Alleles | 1.042 (0.775, 1.400) | 0.787   | 0.684 (0.468, 1.000) | 0.050   |

These results are adjusted for family history, age at biopsy, parity, age at menarche, age at first birth, and BMI. Full model covariates can be found in the supplemental material, Table S3 and Table S4.

```
## PIH
## 0 1
## 115 76
## T alleles
## 0 1 2
## 42 98 51
## Parity
## 1 2 3 4 5
## 43 100 38 9 1
## Family Hist
## 0 1 2 3
## 140 40 10 1
## # A tibble: 1 x 8
##   afb.mean age.mean men.mean bmi.mean t.alleles.mode pih.mode parity.mode
##   <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
## 1 27.0 45.9 12.5 30.0 1 0 2
## # ... with 1 more variable: family.mode <chr>
```

## 6 Supplemental

TABLE S1: Goodness of fit test results

| Alternative Hypothesis | p-value |
|------------------------|---------|
| GLM-NB > Poisson       | 0.000   |
| ZIP > Poisson          | 0.000   |
| ZINB > GLM-NB          | 0.660   |
| GLM-NB > HNB           | 0.046   |

GLM-NB = generalized linear model with negative binomial distribution

Poisson = generalized linear model with Poisson distribution

ZIP = zero-inflated poisson model

ZINB = zero-inflated negative binomial model

HNB = negative binomial hurdle model.

TABLE S2: Complete adjusted negative binomial model with interaction terms

|                   | Coefficient | Standard Error | Z Value | p-value | CR (95% CI)          |
|-------------------|-------------|----------------|---------|---------|----------------------|
| HDP               | 0.208       | 0.345          | 0.604   | 0.546   | 1.231 (0.627, 2.419) |
| T Alleles = 1     | -0.027      | 0.268          | -0.102  | 0.918   | 0.973 (0.575, 1.645) |
| T Alleles = 2     | 0.099       | 0.307          | 0.322   | 0.747   | 1.104 (0.605, 2.016) |
| AFB               | 0.008       | 0.017          | 0.498   | 0.619   | 1.008 (0.975, 1.043) |
| Age               | -0.041      | 0.008          | -5.232  | 0.000   | 0.96 (0.945, 0.975)  |
| Family History    | 0.372       | 0.185          | 2.008   | 0.045   | 1.451 (1.009, 2.086) |
| BMI               | -0.022      | 0.011          | -1.918  | 0.055   | 0.979 (0.957, 1.000) |
| Parity            | 0.334       | 0.107          | 3.132   | 0.002   | 1.396 (1.133, 1.720) |
| Menarche          | -0.102      | 0.058          | -1.763  | 0.078   | 0.903 (0.806, 1.011) |
| HDP×T alleles = 1 | -0.310      | 0.417          | -0.743  | 0.457   | 0.734 (0.324, 1.661) |
| HDP×T alleles = 2 | -0.740      | 0.482          | -1.536  | 0.124   | 0.477 (0.185, 1.227) |

T alleles are treated as factor variables. HDP compares HDP+ to HDP- women. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history.

TABLE S3: Complete adjusted negative binomial models stratified by HDP status

|                | Coefficient | Standard Error | Z Value | p-value | CR (95% CI)              |
|----------------|-------------|----------------|---------|---------|--------------------------|
| (Intercept)    | 4.157       | 0.803          | 5.177   | 0.000   | 63.899 (13.242, 308.331) |
| T Alleles = 1  | -0.500      | 0.326          | -1.536  | 0.124   | 0.606 (0.320, 1.148)     |
| T Alleles = 2  | -0.760      | 0.386          | -1.967  | 0.049   | 0.468 (0.219, 0.997)     |
| AFB            | -0.045      | 0.028          | -1.628  | 0.104   | 0.956 (0.905, 1.009)     |
| Age            | -0.023      | 0.013          | -1.793  | 0.073   | 0.978 (0.954, 1.002)     |
| Family History | 0.071       | 0.303          | 0.235   | 0.814   | 1.074 (0.593, 1.946)     |
| BMI            | -0.027      | 0.018          | -1.522  | 0.128   | 0.973 (0.940, 1.008)     |
| Parity         | -0.048      | 0.179          | -0.267  | 0.790   | 0.953 (0.671, 1.354)     |
| Menarche       | -0.224      | 0.095          | -2.360  | 0.018   | 0.799 (0.663, 0.963)     |

For HDP+ women only. T alleles are treated as factor variables. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history.

TABLE S4: Complete adjusted negative binomial models stratified by HDP status

|                | Coefficient | Standard Error | Z Value | p-value | IRR (95% CI)         |
|----------------|-------------|----------------|---------|---------|----------------------|
| (Intercept)    | 1.296       | 0.493          | 2.629   | 0.009   | 3.653 (1.390, 9.598) |
| T Alleles = 1  | 0.013       | 0.260          | 0.052   | 0.959   | 1.014 (0.608, 1.689) |
| T Alleles = 2  | 0.080       | 0.302          | 0.264   | 0.792   | 1.083 (0.600, 1.955) |
| AFB            | 0.022       | 0.021          | 1.022   | 0.307   | 1.022 (0.980, 1.065) |
| Age            | -0.043      | 0.010          | -4.302  | 0.000   | 0.958 (0.940, 0.977) |
| Family History | 0.441       | 0.229          | 1.924   | 0.054   | 1.554 (0.992, 2.435) |
| BMI            | -0.020      | 0.014          | -1.369  | 0.171   | 0.98 (0.953, 1.009)  |
| Parity         | 0.398       | 0.131          | 3.028   | 0.002   | 1.489 (1.151, 1.927) |
| Menarche       | -0.041      | 0.073          | -0.556  | 0.578   | 0.96 (0.831, 1.109)  |

For HDP- women only. T alleles are treated as factor variables. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history.

TABLE S5: Complete adjusted negative binomial models stratified by HDP status with alleles treated linearly for trend

|                | Coefficient | Standard Error | Z Value | p-value | IRR (95% CI)             |
|----------------|-------------|----------------|---------|---------|--------------------------|
| (Intercept)    | 4.047       | 0.749          | 5.401   | 0.000   | 57.204 (13.174, 248.392) |
| T Alleles      | -0.380      | 0.194          | -1.960  | 0.050   | 0.684 (0.468, 1.000)     |
| AFB            | -0.046      | 0.028          | -1.645  | 0.100   | 0.955 (0.904, 1.009)     |
| Age            | -0.022      | 0.013          | -1.767  | 0.077   | 0.978 (0.954, 1.002)     |
| Family History | 0.103       | 0.300          | 0.342   | 0.733   | 1.108 (0.615, 1.996)     |
| BMI            | -0.028      | 0.018          | -1.560  | 0.119   | 0.973 (0.939, 1.007)     |
| Parity         | -0.032      | 0.178          | -0.182  | 0.856   | 0.968 (0.684, 1.371)     |
| Menarche       | -0.220      | 0.093          | -2.375  | 0.018   | 0.802 (0.669, 0.962)     |

For HDP+ women only. T alleles are treated linearly. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history.



TABLE S6: Complete adjusted negative binomial models stratified by HDP status with alleles treated linearly for trend

|                | Coefficient | Standard Error | Z Value | p-value | IRR (95% CI)         |
|----------------|-------------|----------------|---------|---------|----------------------|
| (Intercept)    | 1.266       | 0.455          | 2.786   | 0.005   | 3.548 (1.455, 8.648) |
| T Alleles      | 0.041       | 0.151          | 0.270   | 0.787   | 1.042 (0.775, 1.400) |
| AFB            | 0.022       | 0.021          | 1.031   | 0.303   | 1.022 (0.981, 1.065) |
| Age            | -0.043      | 0.010          | -4.308  | 0.000   | 0.958 (0.940, 0.977) |
| Family History | 0.436       | 0.228          | 1.911   | 0.056   | 1.547 (0.989, 2.419) |
| BMI            | -0.020      | 0.014          | -1.348  | 0.178   | 0.981 (0.953, 1.009) |
| Parity         | 0.402       | 0.131          | 3.071   | 0.002   | 1.495 (1.157, 1.933) |
| Menarche       | -0.039      | 0.072          | -0.550  | 0.582   | 0.961 (0.836, 1.106) |

For HDP- women only. T alleles are treated linearly. The reference for T alleles is no T alleles (T alleles = 0). The reference for family history is no family history.