Modeling the Effect of ASIP Gene on Melanoma Grade in Horses
Xinling Xu
Advisor: Aaron K. Rendahl

Abstract

About 70—80% of Gray horses older than 15 years old have melanomas. The gene ASIP is thought to affect the severity of melanoma. Previously, the effect of ASIP on melanoma grade was studied by Teixeira and McCue, et al. (2012), using a linear model with age as a cubic term. In this paper, their data was reanalyzed, and several new models were compared, including simpler linear models, non-linear models and quasi-binomial models. The quasi-binomial model was found to be preferred, as it does not assume a specific distribution, meanwhile, it has appropriate mean and variance for melanoma grade data.

Background

Melanoma is a skin disease of developing dark pigment in skin. It affects many species. In this paper, the main focus is to find out the effect of ASIP gene on the severity of melanoma disease for Gray horses. A Gray horse is of color gray, carrying the Gray allele. Pielberg and Golovko, et al. (2008), suggest once a gray horse develops melanoma, the dark spots can occur under the tail root, in the perianal, lips and eyelids region. This condition becomes more severe as age increases. The metastasized melanoma will invade internal organs and eventually reduce longevity.

Several genes have been linked to melanoma. Pielberg and Golovko, et al. (2008), found that a mutation in ASIP (agouti-signaling protein) gene, which increases MC1R (melanocorin-1-receptor) pathway signaling, affects the severity of melanoma in Gray horses. MC1R and ASIP
are antagonized with each other. *MC1R* signaling would increase melanoma severity. When *ASIP* gene is not mutated, that is when it is being *ASIP*\(^{A/A}\), it would antagonize *MC1R* signaling. So the melanoma would not become severe. When *ASIP* gene is mutated, that is when it becomes *ASIP*\(^{a/a}\), it can no longer antagonize *MC1R*. Thus *MC1R* signaling would increase and lead to severe melanoma.

According to Teixeira and McCue, et al. (2012), the effect of *ASIP* and *MC1R* are not statistically significant, meanwhile, the age effect is significant as a cubic term in their models. The focus of this paper will be on the effect of *ASIP* genotype on melanoma grade, as well as the effect of age. Since the age effect is statistically significant, it will be included in all models. The effect of other genes, such as *STX17* and *MC1R* on melanoma, will not be discussed here.

*ASIP* genotypes are recorded as *ASIP*\(^{A/A}\) (homozygous wild-type), *ASIP*\(^{A/a}\) (heterozygous) and *ASIP*\(^{a/a}\) (homozygous), among which *ASIP*\(^{a}\) is the mutated gene that can lead to severe melanoma. Pielberg and Golovko, et al. (2008), found that Gray horses, who were older than 6 years carrying *ASIP*\(^{a}\), had a higher incidence in getting severe melanoma (p=0.0006). Horses with *ASIP*\(^{A/A}\) genotype had overall lower melanoma grades, while horses with *ASIP*\(^{A/a}\) stayed between these two groups.

Teixeira and McCue, et al. (2012), also performed similar study. Both linear regression and logistic regression were used with age as a cubic term. The response for the linear model was the grade of melanoma. The response for the logistic model was case/control status. The result of their study was that no statistically significant difference was found between *ASIP* genotype groups. Age had a strong effect on melanoma grade. As will be seen, this analysis had some
concerns; the current study is to improve this analysis as well as compare the result with Pielberg and Golovko, et al. (2008).

Our data set

To further approach the effect of ASIP on melanoma, the data from Teixeira and McCue, et al. (2012), College of Veterinary Medicine, University of Minnesota, is used through the whole study. In this data set, there are 335 Gray Quarter Horses (QH) with and without dermal melanomas. The ages of the horses are between 1 and 33 with mean age 9.20. The range of melanoma grade is 0 to 4, with 4 being the severest case. The mean melanoma grade of this cohort is 0.35. Grades are recorded at 0, 1, 1.5, 2, 2.5, 3 and 4. ASIP is analyzed assuming an additive effect, where the number of ASIP \(a\) is used in all models. That is, the effect of ASIP is 0 if ASIP gene is ASIP \(A/A\), the effect is 1 if ASIP gene is ASIP \(A/a\), the effect is 2 if ASIP gene is ASIP \(a/a\). This is because of the biology mechanism behind genes. Other descriptive statistics are also recorded for each horse, such as age, gender and breed farms.

Since this study is focused on the effect of ASIP and age on melanoma grade, only horses with complete ASIP, age and melanoma grade records were kept. Thus, all the following study was based on 325 horses.

Exploratory Statistics and Previous Models

The raw data scatterplot is shown in Figure 1. There are 272 out of 325 horses (83.69%) having melanoma grade at 0. More zeros clustered at an early age than after 15 years old, which agrees with Teixeira and McCue, et al. (2012), as they suggest that the prevalence of melanoma becomes severe after a certain age.
In this cohort, 30 horses (9%) have ASIP \( a/a \), with mean melanoma grade 0.28. 151 horses (46%) have ASIP \( A/a \) with mean melanoma grade 0.39. 144 horses (44%) have ASIP \( A/A \) with mean melanoma grade 0.32. The wild-type allele ASIP \( A \) is the major allele with frequency 0.68 and 439 counts. Allele ASIP \( a \) has frequency 0.32 with 211 counts. The number of horses in each group is very different. In Figure 2c, only three non-zero points are in ASIP \( a/a \) group, this might lead to some problems in finding the proper model for ASIP effect. Besides, for each ASIP group, many points are clustering at grade zero.
Pielberg and Golovko, et al. (2008), suggested to use linear models with *age* as a cubic term. Teixeira and McCue, et al. (2012), used these suggested models with different interaction terms between *age* and *ASIP*, which lead to non-significant result of genotype *ASIP* effect. The models from Teixeira and McCue, et al. (2012), will be explored first. The drawbacks of their models will be shown and further improvement can then be made based on these drawbacks.

Teixeira and McCue, et al. (2012), used two models, both with a cubic *age* term. The first one is model *mc1*, which includes *age* as a cubic term with different intercepts for each *ASIP* group. There is no interaction terms between *age* and *ASIP*. The second model *mc2* includes *age* (cubic) and *ASIP*, as well as the interaction term of *age* and *ASIP*. To complete the cubic models, model *mc3*, which includes the interaction term between *ASIP* and *age* (*quadratic*) as well as model *mc4*, which includes all interaction terms between *age* and *ASIP* are also fitted.

In R, the models are written as below. The name of the dataset used in this paper is *casip*.

\[ mc1 = \text{lm}(\text{MELANOMA.GRADE} \sim \text{poly(AGE, 3)} + \text{ASIP.add}, \text{data} = \text{casip}) \]
\begin{verbatim}
mc2 = lm(MELANOMA.GRADE ~ poly(AGE, 3) + AGE*ASIP.add, data = casip)
mc3 = lm(MELANOMA.GRADE ~ poly(AGE, 3) + poly(AGE, 2) * ASIP.add, data = casip)
mc4 = lm(MELANOMA.GRADE ~ poly(AGE, 3) * ASIP.add, data = casip)
\end{verbatim}

For model \textit{mc1}, \textit{age (cubic)} term is statistically significant with \textit{p}<0.00001. The effect of \textit{ASIP} is not statistically significant with \textit{p}=0.38. For model \textit{mc2}, both the \textit{ASIP} (\textit{p}=0.52) and its interaction term (\textit{p}=0.16) with \textit{age} are not statistically significant. For model \textit{mc3}, all terms with \textit{ASIP} are not statistically significant at the 0.05 level. In model \textit{mc4}, when the interaction term of \textit{age (cubic)} and \textit{ASIP} is included, the main effect of \textit{age (cubic)} is not significant with \textit{p}=0.15. And the effect of \textit{ASIP} is not statistically significant either with \textit{p}=0.27.

Since these four models are nested, the model selection can be performed based on likelihood ratio test. Calculations were done in R, using the \textit{lmtest} package. The smallest model \textit{mc1} is preferred with \textit{p}=0.16 compared to model \textit{mc2}. 

In Figure 3, all fitted lines for ASIP<sup>AA</sup> group decrease after horses are around 30 years old. It is reasonable to believe that as age increases, the melanoma disease should always get more severe. For model mc1, the fitted curves of all three groups share the same shape with only a difference in the intercept. This might not be true, since ASIP<sup>a</sup> gene may affect melanoma grade in a different way as the mean melanoma grade changes and age increases.
Many issues about the cubic models need to be addressed. First, the decreasing trend of the fitted lines for all three ASIP groups in models mc1 and mc2, also for ASIP A/A group in models mc3 and mc4. Second, for all four models, a melanoma grade decreases before 6. This does not seem to be reasonable. When horses are young, they might not develop melanoma for all three groups, but a true decreasing trend seems to be unlikely. Third, the fitted lines in models mc2, mc3 and mc4 cross each other. The effect of the ASIP gene, if any, is expected to stay at the same effect level as age increases. That is, if a horse with ASIP a tends to get severe melanoma, it should stay true even for young horses. Besides, for models mc4, the increasing rate of melanoma grade in ASIP a/a group seems to be too steep and the grade would eventually go beyond 4. The melanoma grade is between 0 and 4, thus a model within this range is more ideal. Therefore, finding a non-decreasing model with no crossing issue and stays between 0 and 4 is the goal of this paper.

**New model 1: Quadratic model**

As a first attempt, the cubic model was simplified. Quadratic models with age as a quadratic term were fitted. If there was no special reason for a cubic model, a simpler model that could explain the data well might be preferred.

The following models were fitted in R.

\[
\begin{align*}
mq1 &= \text{lm}(\text{MELANOMA.GRADE} \sim \text{poly(AGE,2)} + \text{ASIP.add}, data=\text{casip}) \\
mq2 &= \text{lm}(\text{MELANOMA.GRADE} \sim \text{poly(AGE,2)} + \text{AGE:ASIP.add}, data=\text{casip}) \\
mq3 &= \text{lm}(\text{MELANOMA.GRADE} \sim \text{poly(AGE,2)} + \text{AGE*ASIP.add}, data=\text{casip}) \\
mq4 &= \text{lm}(\text{MELANOMA.GRADE} \sim \text{poly(AGE,2)} * \text{ASIP.add}, data=\text{casip})
\end{align*}
\]
The first model $mq1$ is assuming there is no interaction effect between ASIP and age. In the second model $mq2$, ASIP affects melanoma grade only in a linear way with age, without different intercepts for ASIP groups. In the third model $mq3$, the interaction term of ASIP and age is included as well as a different intercept for each ASIP group. Model $mq4$ is the full quadratic model including all interaction terms.

Figure 4. Quadratic models with fitted lines for each model.
One advantage of these quadratic models is that they happen to be non-decreasing except for model $mq4$, which has a decreasing trend when horses are younger. Since horses do not develop melanoma until a later age, this is not the main issue. By only looking at age after 6, all four models show the trend of increasing grade as age increases. The “crossing” issue still exists. For model $mq3$, this problem has even been amplified, which leads to the negative grades of horses from $ASIP^{a/a}$ group at early ages. All models are increasing steeply with age and some curves go beyond grade 4, such as the fitted line of $ASIP^{a/a}$ group from model $mq3$.

Models $mq2$, $mq3$ and $mq4$ are nested and likelihood ratio test are used to make model selection. Models $mq4$ and $mq3$ are not statistically significant different from each other, with $p=0.13$. Models $mq3$ and $mq2$ are also not statistically significant different at the 0.05 level. The p-value for comparing these two models is 0.33. However, models $mq2$ and $mq1$ are not nested in each other. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used to compare these two models. Model $mq2$ has both smaller AIC and BIC. Hence, model $mq2$ is selected.

The final quadratic model is $mq2$. After substituting the estimates into the model, it is,

$Melanoma\ Grade=0.28+6.49*age+0.01*(age*ASIP)+2.03*age(\text{quadratic})$

The p-value of the estimated coefficient of $age*ASIP$ term is 0.05.
Figure 5 shows the 95% confidence bands for $ASIP^{A/A}$ and $ASIP^{a/a}$ groups. The overlapping of the confidence bands indicates the increasing uncertainty of the effect of $ASIP^{a}$ on melanoma grade as the mean of melanoma grade increases. These two confidence bands above are based on the final quadratic model $mq2$. Before age 5, the confidence bands include negative values. And some fitted grade values for young horses are below zero too. Even though this model has some drawbacks, for age greater than 5, it can reasonably explain the data.

After selecting this model, Quantile-Quantile plot and residual plot were used for model diagnostics. The assumptions of this linear regression model are: the melanoma grades are normally distributed and are independent between horses, with a constant variance. The biggest violation of these assumptions is the normality assumption. As it is quite clear that, the melanoma grade is between 0 and 4, and is only recorded as discrete values with many zeros,
which makes it very non-normal. Moreover, linear regression works better for continuous and symmetric data, which is not true in this case.

If melanoma grade is normally distributed, the data should fall on a straight line in the Quantile-Quantile plot. Figure 6a is far from a straight line. In Figure 6b, a plot of the residuals vs. fitted values was used to check the fit of a model. However, in this case, grade is recorded only at discrete points, which makes the response appear to be like count data. It is hard to diagnose the fit of this quadratic model. For the independence assumption, the parent/offspring pair information is not available, so this assumption might be loosely established.

In conclusion, even though this quadratic model is not ideal, one advantage of it is that the interpretation for ASIP effect is quite straightforward. For a fixed age, say 15, the mean melanoma grade for horses in ASIP $A/a$ group is 0.15 higher than the mean melanoma grade for horses in ASIP $A/A$ group. The mean melanoma grade for horses in ASIP $a/a$ group is 0.30 higher
than the mean melanoma grade for horses in ASIP \(^{A/A}\) group. Besides, within the range, the fitted curves appear to be reasonable for this data.

**Nonlinear and Quasibinomial Regression**

Since the fitted curve of an ideal model has a shape, which levels off when it reaches 4, it is reasonable to consider a nonlinear regression model, which has the curvature that is desirable.

The nonlinear regression models are also often used in biological field. A nonlinear regression model has the form

\[ Y_i = f(x_i, \beta) + \varepsilon_i, i = 1, \ldots, n, \]

where \( f \) is a known function of \( x \) and \( \beta \), and \( \varepsilon \) are random errors with mean zero and constant variance. Notice that, the constant variance assumption might not hold. The variability might increase as the mean of melanoma grade increases. The constant variance assumption should be loosened.

For this data set, several nonlinear models were tried, including,

**Nonlinear model 1:** *Melanoma Grade* = \( \frac{age}{K + \exp(a) + ASIP + age} \times 4 \)

**Nonlinear model 2:** *Melanoma Grade* = \( \frac{\left[ \frac{a \exp(a)}{1 + \exp(a)} \right] + ASIP + \left[ \frac{a \exp(b)}{1 + \exp(b)} \right] (2 - ASIP)}{2} \times \frac{age}{K + age} \)
For nonlinear model 1,

\[ \text{Melanoma Grade} = \frac{\text{age}}{K + \text{age}} \times 4, \text{ when ASIP}=0. \]

\[ \text{Melanoma Grade} = \frac{\text{age}}{K + \exp(a) + \text{age}} \times 4, \text{ when ASIP}=1. \]

\[ \text{Melanoma Grade} = \frac{\text{age}}{K + \exp(a) \times 2 + \text{age}} \times 4, \text{ when ASIP}=2. \]

By choosing the appropriate values for \( K \) and \( a \), the values of \( K + \exp(a) \) and \( K + \exp(a) \times 2 \) are always greater than 0, so the fraction part is always between 0 and 1. This model allows the shape of fitted lines for ASIP groups to be different. Figure 7 shows different curves for ASIP groups when \( K=5, a=0.5 \) and \( K=10, a=0.9 \). The steep slopes at an early age indicates this combination of parameters might be more suitable for a breakout type of disease. In order to change the shape of these lines, the appropriate start values for \( K \) and \( a \) need to be selected.
For nonlinear model 2, the values of both \( \frac{\exp(a)}{1+\exp(a)} \) and \( \frac{\exp(b)}{1+\exp(b)} \) are between 0 and 4. Since the value of ASIP can be 0, 1 or 2, the first term in the equation is between 0 and 4. The maximum melanoma grade for different ASIP groups is different, as the choice of \( a \) and \( b \) decides how much difference there can be. The shape of their lines is the same since the slope of \( age \) is the same across three groups, which is 1.

When dealing with nonlinear regression, the estimation of parameters is often done by iterative computation. However, after trying many starting points for nonlinear models, all iterative processes failed to converge. One reason might be there are many zeros in the dataset. The appropriate starting parameter values are hard to find.

So far all work has been based on ordinary least square regression. Models were fitted using the original form of the data. Transforming the original data is one direction to continue the study. By dividing melanoma grade by 4, which is the highest melanoma grade, the response is restricted between 0 and 1, including 0 and 1. Now a new response \( \text{MELANOMA.GRADE/4} \) has been created, which can be interpreted as the severity of melanoma in a 0 to 1 scale. This new response will be used in the following models.

It seems to be no known distribution for melanoma grade. Normal distribution is far from accurate, while Poisson distribution is for count data, assuming mean equals variance. One approach of dealing with unknown distribution is Quasi-Likelihood. To use Quasi-Likelihood, one only needs to specify the link and variance functions.

Now a new thought is to use Quasi-Likelihood on the new response \( \text{MELANOMA.GRADE/4} \).

**Quasi-binomial Background**
The normality assumption of the error term in linear regression models is very crucial. Most inferences of the model are based on this assumption, such as the predicted confidence intervals for the mean and the confidence intervals for parameters. If the sample size is large enough, by law of large numbers, the distribution of the sample mean asymptotically follows a normal distribution. The inference can still be based on the normal distribution.

For Generalized linear models, the link and variance functions are also very important. The link function presents the form of the model connecting the non-normal distribution to a normal distribution, while inferences are based on the variance function. When using GLMs, one has an understanding of what the distributions are. For example, if the response is count data, a Poisson distribution with the canonical link might be considered.

When the distribution of the data is unknown, but one might have an idea about the form of the variance function as well as the link function, quasi-likelihood can be used in dealing with this situation. Since when using quasi-likelihood there will be no assumption of the distribution, the estimates of parameters are unable to be obtained by maximum likelihood. Faraway (2006), suggests that a substitute for the likelihood is necessary. The theoretical part of quasi-likelihood will not be discussed here, instead, the focus will be on the choice of link and variance functions for this melanoma dataset.

In this dataset, quasi-binomial generalized linear models will be fitted. The first reason is that the response (MELANOMA.GRADE/4) is now between 0 and 1, so a logit link function seems to be natural. Secondly, the variance of the response (melanoma grade) is expected to change as the mean of it changes. When the mean of melanoma grade is small, say 0.5, the variability of it should not be large. That is, there is really not much of melanoma grade difference within ASIP
groups. The age effect has already been shown to be important. Thus, when horses are getting older, their melanoma grades are expected to be increasing. As the mean of melanoma grade increases, the variance of it might increase as well. The melanoma grade can be really low for horses in ASIP $A/A$ group, meanwhile, horses can get high melanoma grade for horses in ASIP $a/a$ group. When the mean of melanoma grade reaches really high, maybe near 4, the grade becomes stable and levels off eventually without much variability. Therefore, a variance function of form $\mu(1 - \mu)$ is appropriate. The quasi-binomial satisfied these two requirements. It has the canonical link logit function and has the variance function in the form of $\phi \mu(1 - \mu)$.

**Quasi-binomial Modeling**

Building on the previous quadratic models, the same four quadratic models are fitted with the response MELANOMA.GRADE/4 instead, using quasi-binomial. The four fitted models in R are,

\[
\begin{align*}
\text{quasi1} &= \text{glm}(\text{MELANOMA.GRADE}/4 \sim \text{poly}(\text{AGE}, 2) + \text{ASIP}, \text{family}=\text{quasibinomial}, \text{data}=\text{casip}) \\
\text{quasi2} &= \text{glm}(\text{MELANOMA.GRADE}/4 \sim \text{poly}(\text{AGE}, 2) + \text{AGE:ASIP}, \text{family}=\text{quasibinomial}, \text{data}=\text{casip}) \\
\text{quasi3} &= \text{glm}(\text{MELANOMA.GRADE}/4 \sim \text{poly}(\text{AGE}, 2) + \text{AGE*ASIP}, \text{family}=\text{quasibinomial}, \text{data}=\text{casip}) \\
\text{quasi4} &= \text{glm}(\text{MELANOMA.GRADE}/4 \sim \text{poly}(\text{AGE}, 2) * \text{ASIP}, \text{family}=\text{quasibinomial}, \text{data}=\text{casip})
\end{align*}
\]
To be consistent with all earlier models, even though models were fitted using 0 to 1 scale, the response in the plots was transformed back to the original 0 to 4 scale. When the fitted lines are on the log odds scale, the mean and the median are equal. However, after taking the inverse logit transformation, the mean does not stay as the mean on the original scale. Therefore, all fitted lines in Figure 7 are for the median of melanoma grade. Similarly, all responses that are transformed back into the original scale are the estimated median of melanoma grade.
For quasi-likelihood models, model selection is unable to be done in the usual way, such as using likelihood ratio test or Akaike’s information criterion (AIC). In R, \textit{drop1} function specifying F-test can be used for model selection. Since the estimated effect of \textit{ASIP} is not significant at the 0.05 level, using \textit{drop1} function deletes the two interaction terms of \textit{ASIP} and \textit{age}, \textit{ASIP} and \textit{age} (quadratic). This leads to model \textit{quasi1}. By using the function again, the model eventually contains no \textit{ASIP} term. However, the effect of \textit{ASIP} is of main interest. It will be kept in the final model as an interaction term with \textit{age}.

In Figure 7a, the fitted lines show that all three \textit{ASIP} groups have the same increasing trend of melanoma grade as age increases. The only difference is that they start at the different melanoma grade baseline. For \textit{ASIP} \textit{A/A} group, the median of melanoma grade levels off at a lower level compared with the other two groups.

The p-values for \textit{age} (quadratic) term in models \textit{quasi1}, \textit{quasi2}, \textit{quasi3} and \textit{quasi4} are 0.104, 0.166, 0.199 and 0.241 separately. In model \textit{quasi4}, the p-value for the interaction term between age (quadratic) and \textit{ASIP} is 0.852. Thus, \textit{age} (quadratic) term can be deleted, as well as the interaction term of it and \textit{ASIP}. Besides, to make the grades for all three groups start at the same value, the main effect of \textit{ASIP} is dropped.

The final model is,

$$final\text{\textit{quasi}}=\text{\textit{glm}}(\text{\textit{MELANOMA.GRADE}/4}\sim \text{\textit{AGE}}:\text{\textit{ASIP}}+\text{\textit{AGE}},\text{\textit{family}}=\text{\textbf{quasibinomial}},\text{\textit{data}}=\text{\textit{casip}})$$

The final model includes an \textit{age} term and an interaction of \textit{age} and \textit{ASIP}. All groups share the same intercept, because it is assumed that all horses have the same melanoma grade at age 0.
To make a simpler expression in terms of the original scale, the "ilogit" function is defined as the inverse-logit function. Using the "ilogit" expression, the final fitted model is,

\[
Melanoma\ Grade = 4 \times \text{ilogit}(\ -4.31 + 0.15 \times \text{age} + 0.03 \times \text{age} \times \text{ASIP})
\]

Figure 9. The final quasi-binomial model with fitted lines. Horses with ASIP genotype a/a have overall the highest melanoma grade.

The interpretation of this model is on the log odds scale. The slope and intercept mentioned below are not the usual slope and intercept as in linear regression models. In this final model, the estimated coefficient of the interaction term of age and ASIP is 0.03, with p=0.08 and 95% confidence interval (-0.00, 0.06), which is the additional slope of age for each level of ASIP. For example, when fixing age at 6, the slope of age for ASIP a/a group is 0.18 greater than it is for ASIP A/a group and 0.36 greater than it is for ASIP A/A group. The estimated coefficient of age is 0.15 with p<0.00001 and 95% confidence interval (0.11, 0.19). That is, for a certain ASIP group,
one unit increase in age increases the log odds of melanoma grade by $0.15 + 0.03 \times \text{ASIP}$. In here, the ‘odds’ does not have the traditional meaning. It is defined as $\text{melanoma.grade}/(4-\text{melanoma.grade})$. This new definition for odds provides a more intuitive way to interpret the parameters in this model. Using this new definition, the effect of ASIP is not as statistically significant at the 0.05 level, in contrast to Gerli, et al. To interpret the results in terms of odds ratio, the fitted model for each group is,

$$
\log(\text{odds}_{A/A}) = 0.15 \times \text{age} - 4.31 \quad \text{when ASIP}=0
$$

$$
\log(\text{odds}_{A/a}) = 0.18 \times \text{age} - 4.31 \quad \text{when ASIP}=1
$$

$$
\log(\text{odds}_{a/a}) = 0.21 \times \text{age} - 4.31 \quad \text{when ASIP}=2
$$

The odds ratio of getting severe melanoma disease between $\text{ASIP}\; A/a$ and $\text{ASIP}\; A/A$ groups is $\exp(0.03 \times \text{age})$. That is, for every unit increase in age, the odds ratio between these two groups is multiplied by $\exp(0.03) = 1.03$. The odds ratio between $\text{ASIP}\; a/a$ and $\text{ASIP}\; A/A$ groups is $\exp(0.06 \times \text{age})$. That is, for every unit increase in age, the odds ratio between these two groups is multiplied by $\exp(0.06) = 1.06$.

Table 1. Estimated 95% confidence intervals of median melanoma grade for certain ages.

<table>
<thead>
<tr>
<th></th>
<th>Age=6</th>
<th>Age=15</th>
<th>Age=25</th>
<th>Age=35</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ASIP}; A/A$</td>
<td>0.13 (0.08, 0.20)</td>
<td>0.43 (0.30, 0.62)</td>
<td>1.38 (0.88, 1.99)</td>
<td>2.78 (1.85, 3.43)</td>
</tr>
<tr>
<td>$\text{ASIP}; A/a$</td>
<td>0.15 (0.09, 0.22)</td>
<td>0.61 (0.46, 0.80)</td>
<td>2.01 (1.41, 2.61)</td>
<td>3.40 (2.69, 3.75)</td>
</tr>
<tr>
<td>$\text{ASIP}; a/a$</td>
<td>0.17 (0.12, 0.27)</td>
<td>0.84 (0.48, 1.36)</td>
<td>2.64 (1.47, 3.46)</td>
<td>3.74 (2.80, 3.9)</td>
</tr>
</tbody>
</table>
The confidence intervals of median melanoma grade at age 6, 15, 25 and 35 are shown in Table 1. Even though the estimated grades appear to be very different, the confidence intervals for corresponding estimates are overlapped. These confidence intervals include the variability of the age effect. To better understanding ASIP effect alone, age 6 and 25 are selected, because they represent the different stages of horses. After accounting for the age effect, the difference of melanoma grade between groups are computed below.

The estimated median of melanoma grade for horses in ASIP $^{A/A}$ group at age 6 is 0.13. Assuming 0.13 is the true value, the estimated median of melanoma grade for horses in ASIP $^{A/a}$ group at the same age is between 0.13 and 0.18. The estimated median of melanoma grade for horses in ASIP $^{a/a}$ group at the same age is between 0.12 and 0.24. Similarly, the estimated median of melanoma grade for horses in ASIP $^{A/A}$ group at age 25 is 1.38. Assuming 1.38 is the true melanoma grade, the estimated median of melanoma grade for horses in ASIP $^{A/a}$ group is between 1.38 and 2.88. It is between 1.23 and 3.60 for horses in ASIP $^{a/a}$. Therefore, the effect of ASIP alone is not obviously different for three groups.

The computation above was done in R first using the predict function with se.fit=TRUE. This estimations are on the log odds scale. After transforming it back to the MELANOMA.GRADE/4 scale, multiply it by 4, then the estimated median can be obtained. The 95% confidence intervals were calculated in the same fashion.
Cook’s distance was used to check for influential horses. Horse 42 is an outlier case as it is shown in Figure 10. This horse is 33 years old with no melanoma disease. After excluding this horse from the dataset, the same model was fitted again. The fitted model becomes:

\[ \text{Melanoma Grade} = 4\text{ilogit}(-4.41 + 0.02\text{age}\times ASIP + 0.16\text{age}) \]

The estimated intercept decreases by 0.1 and the other two coefficients have changed by 0.01. Now the estimated effect of ASIP is even less significant with \(p=0.17\).

The 95% confidence interval for the estimated coefficient of the interaction term of age and ASIP is \((-0.01, 0.05)\). This changes of estimated coefficients and confidence intervals might due to the fact that horse 42 makes the age effect less significant. In total, there are three horses at age 30 and one horse at age 33, all carry ASIP \(A/A\). These three 30 years old horses have melanoma
grades 4, 2.5 and 1. Since many studies have already shown the age effect is significant, horse 42 is clearly an outlier and will be excluded from the study for ASIP effect.

In Figure 10, the two 95% confidence bands for ASIP$^{A/A}$ and ASIP$^{a/a}$ groups are shown. Again, they are overlapped. The band of ASIP$^{a/a}$ even contains the estimated median melanoma grade of ASIP$^{a/a}$ group.

![Confidence Bands for A/A and a/a](image)

**Figure 11. Confidence bands of median for ASIP$^{a/a}$ and ASIP$^{A/A}$ groups.**

Again, to better understanding the effect of ASIP alone, age 6 and 25 are selected. After accounting for the age effect, the estimated median of melanoma grade for horses in ASIP$^{A/A}$ group at age 6 is 0.12. Assuming 0.12 is the true value, the estimated median of melanoma grade for horses in ASIP$^{A/a}$ group at the same age is between 0.12 and 0.16. The estimated median of melanoma grade for horses in ASIP$^{a/a}$ group at the same age is between 0.11 and 0.22. The estimated mean of melanoma grade for horses in ASIP$^{A/A}$ group at age 25 is 1.60. Assuming 1.60
is the true melanoma grade, the estimated median of melanoma grade for horses in ASIP \(A/a\) group is between 1.36 and 2.79. It is between 1.15 and 3.56 for horses in ASIP \(a/a\). Now the log odds of getting severe melanoma for different groups are,

\[
\log(\text{odds}_{A/A}) = 0.16\text{age} - 4.41 \quad \text{when } \text{ASIP}=0
\]

\[
\log(\text{odds}_{A/a}) = 0.18\text{age} - 4.41 \quad \text{when } \text{ASIP}=1
\]

\[
\log(\text{odds}_{a/a}) = 0.20\text{age} - 4.41 \quad \text{when } \text{ASIP}=2
\]

The odds ratio between ASIP \(a/a\) and ASIP \(A/A\) groups is \(\exp(0.04\text{age})\). It is \(\exp(0.02\text{age})\) between ASIP \(A/a\) and ASIP \(A/A\) groups.

The Pearson residuals plot was used to check the final model, which was after deleting horse 42. There are still some issues with the final model. Because that the response has only seven different values, the residuals fall on seven curves. Recall that the response values are 0, 0.5, 1, 1.5, 2, 2.5, 3 and 4. Comparing with the residuals plot of the quadratic model, it seems that this fitted model does not improve much. However, this model does have the non-decreasing, non-negative and non-crossing properties. Most importantly, the normality assumption is no longer necessary.
**Conclusion and Discussion**

Under the quasi-binomial model, the effect of *ASIP* is not statistically significant at the 0.05 level. Age effect is statistically significant with p<0.00001.

As Teixeira and McCue, et al. (2012), mentioned, even though the estimated mean of melanoma grade is increasing as the number of *ASIP* increases, the effect of *ASIP* mutation is not significant at the 0.05 level after adjusting for the age effect.

For further study, other genes such as MC1R might affect melanoma grade. Together with MC1R, *ASIP* might be influential on melanoma grade.
Teixeira and McCue, et al. (2012), mentions that gene ASIP has effect on melanoma grade for human. But maybe for horses, the way ASIP works is different. So within the range of this dataset and both quadratic models and quasi-binomial models, the effect of ASIP is not statistically significant.

References

