

Gamete Compatibility

Sahar Mozaffari

September 17, 2015

Contents

Summary Statistics	1
: Allele frequencies and sample size	1
: Number of children by genotype	1
: Birth rate by genotype	4
: Additional figures and plots	4
: Linear regression model & anova	7
Association tests using GEMMA	7
: GEMMA for birth rate	9
: GEMMA for family size	9

Summary Statistics

: Allele frequencies and sample size

I started by looking at the numbers of couples in each genotype group. I focused on rs4844573 in males and rs2075520 in females since we had insufficient numbers of Hutterites with genotypes for rs2906999 —only 506 out of 1415 individuals. Overall, there were 344 couples confidently called for rs4844573 and rs2075520 as shown in **Table 1**. Using the pedigree information I was able to impute genotypes in 17 more couples (**Table 2**); and phased data allowed me to impute one more couple (362 couples total) summarized in **Table 3**.

The minor allele frequency at rs2075520 in Hutterite females is very similar to other reported European populations:

362 Hutterite females: 0.232 A (Val = ancestral)
HapMap CEU: 0.221 A
1000G CEU: 0.27 A
1000G global: 0.42 A

The ancestral (Thr) allele frequency at rs4844573 in Hutterite males is higher than other reported European populations and likely due to the founder effect:

362 Hutterite males: 0.543 C
HapMap CEU: 0.367 C
1000G CEU: 0.37 C
1000G global: 0.467 C

: Number of children by genotype

The following analysis looks at birth rate and number of children for 362 couples. Note that it includes incompleted families (couples who are still in their childbearing years).

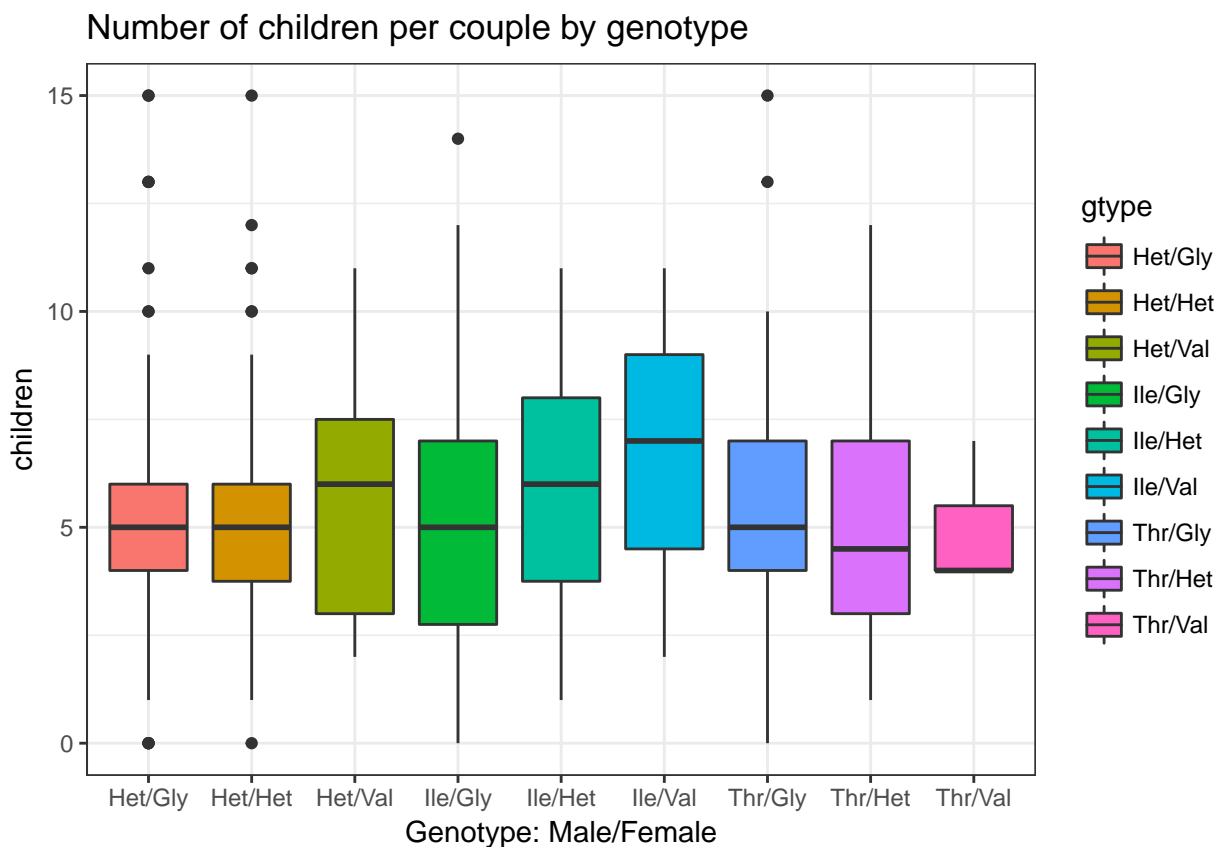


Figure 1: **Figure 1:** Boxplot of the number of children for each couple in each genotype group. The number of couples in each genotype group is labeled below the plot. The black bars are medians, and the boxes are the 25th -75th percentiles. The upper whisker extends from the third quartile to the highest value that is within $1.5 \times \text{IQR}$ of the third quartile, where IQR is the inter-quartile value, or the distance between the first and third quartiles

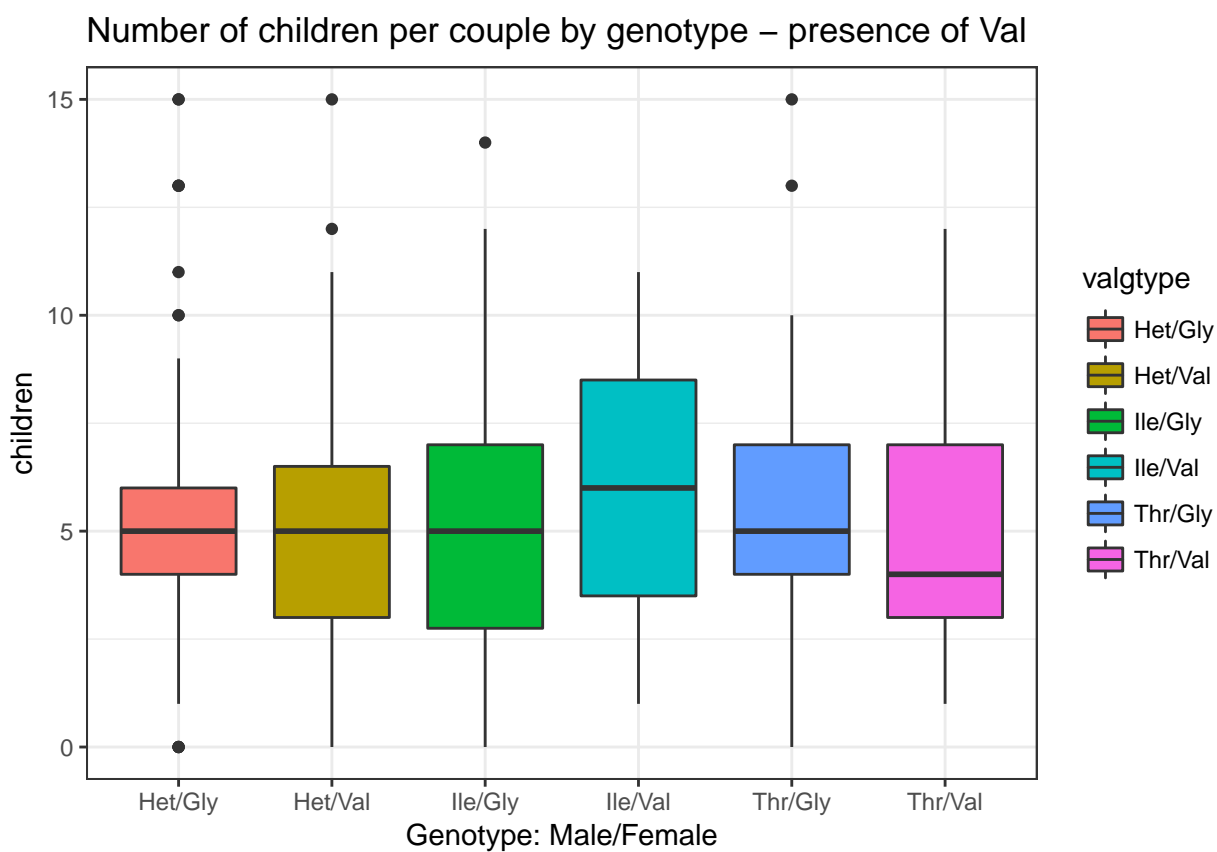


Figure 2: **Figure 2:** Boxplot of the number of children for each couple in each genotype group with the females stratified by presence of the ancestral allele.

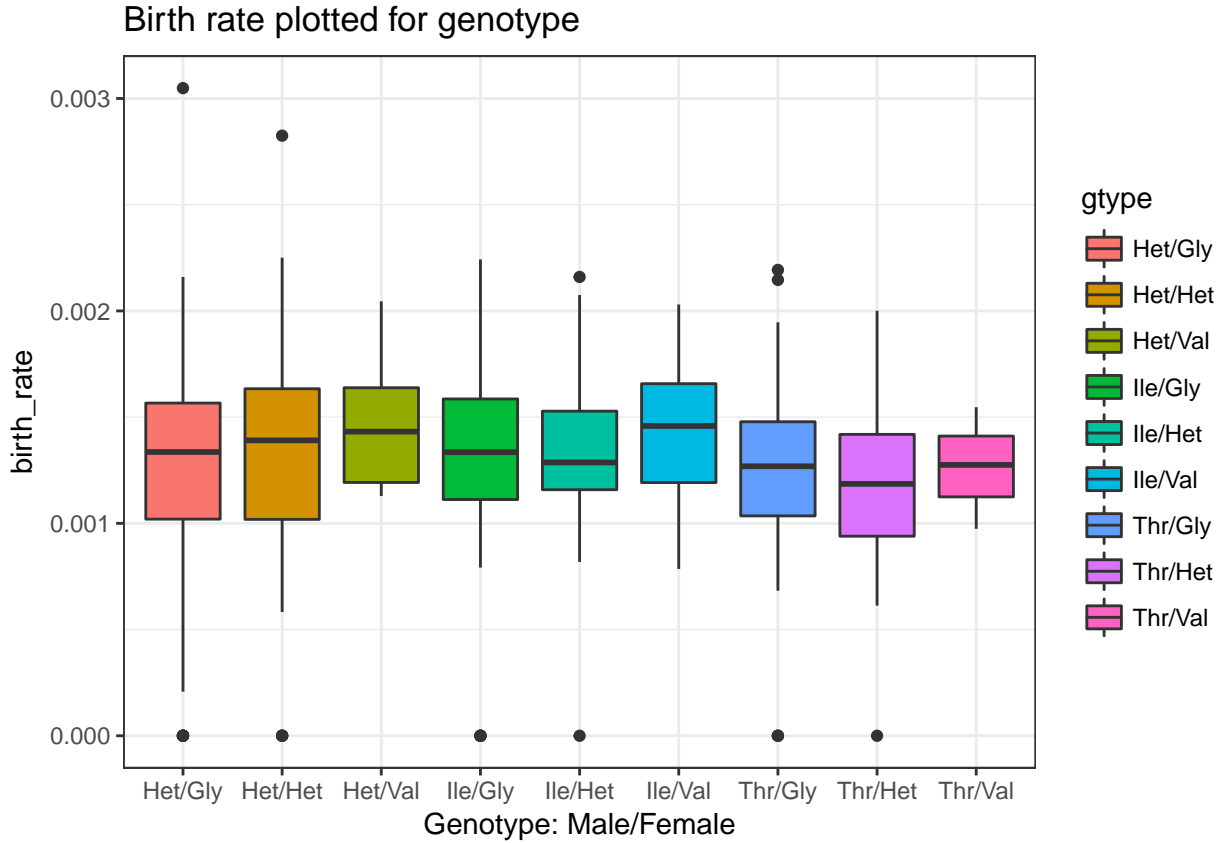


Figure 3: **Figure 3:** Boxplot of the birth rate for each couple in each genotype group.

: Birth rate by genotype

Following Kosova et al. 2010, I calculated birth rate as $[(\#births-1)/(\text{Sum of the interbirth intervals})]$.² As previously noted in studies in 262 of these same couples, the wife's birth year is highly correlated with the husband's birth year ($r^2 = 0.98$ in Kosova, $r^2 = 0.96$ in our study).² Both the wife's birth year ($r = -0.65$, Pearson's p -value = Pearson's p -value $< 2.2e-16$) and time from marriage to last birth ($r = -0.078$, Pearson's p -value = 0.1419) were significant predictors of birth rate. As in Kosova et al. 2010, I included both covariates in a multivariate linear regression model examining the correlation between the birth rate residuals with genotype.

None of the genotype groups differed with respect to number of children or birthrate (either unadjusted or the residual). **Figure 3** shows a boxplot for birthrate for each genotype group. The boxplots in **Figure 4** and **5** below show the birthrate residuals plotted for each genotype group. I did t-tests between the groups in many different combinations for both number of children and birth rate residuals, as well as an anova, and nothing was even close to significant.

: Additional figures and plots

Here are some other plots I made to look at the data similar to Kosova et al. (2010).¹ Similar to the correlation reported in Kosova ($r = -0.62$), the correlation between wife's birth year and number of births in our data set is $r = -0.65$.¹

The number of children and time from marriage to last birth were correlated with $r = 0.875$, similar to Kosova et al. (2010) $r = 0.87$ as shown in Figure 6 below.¹

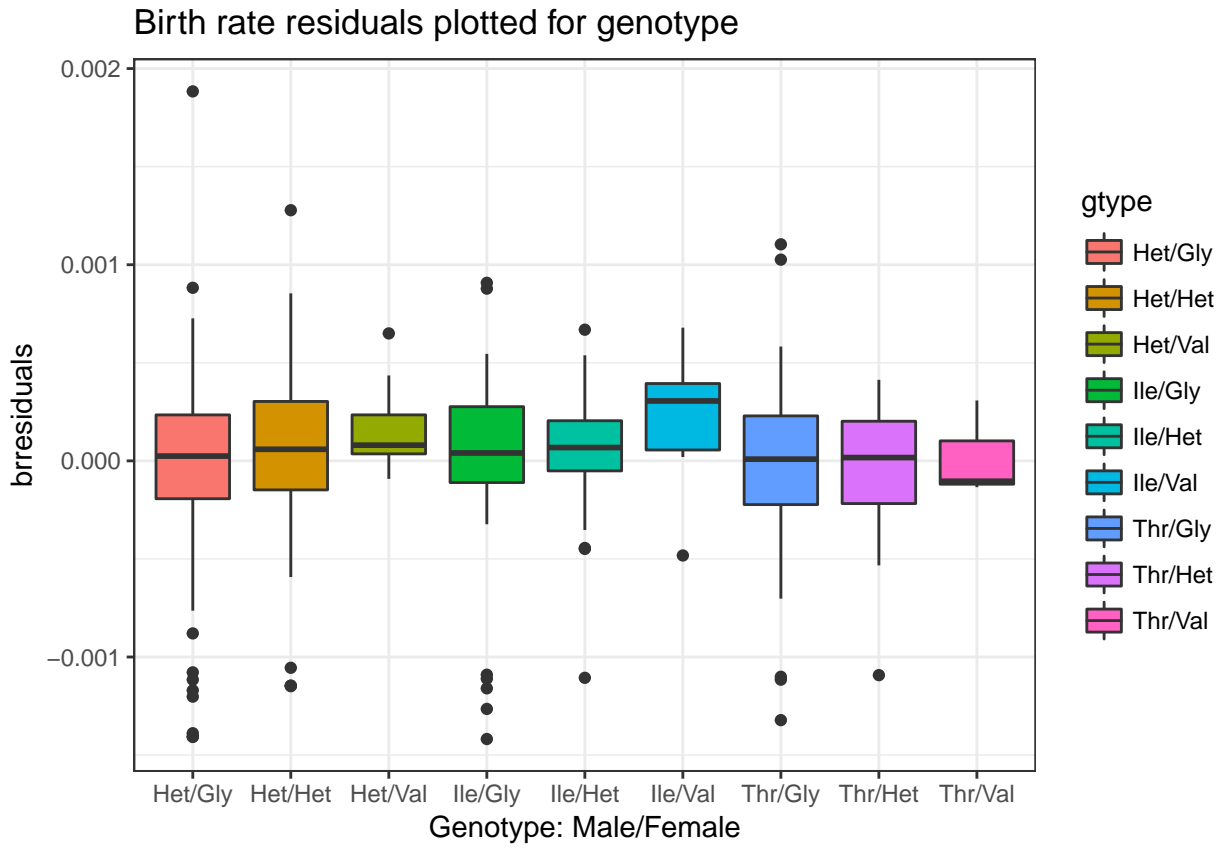


Figure 4: **Figure 4:** Boxplot of the birth rate residuals for each couple in each genotype group. Birth rate residuals were calculated using wife's birth year and time from marriage to last birth as covariates.

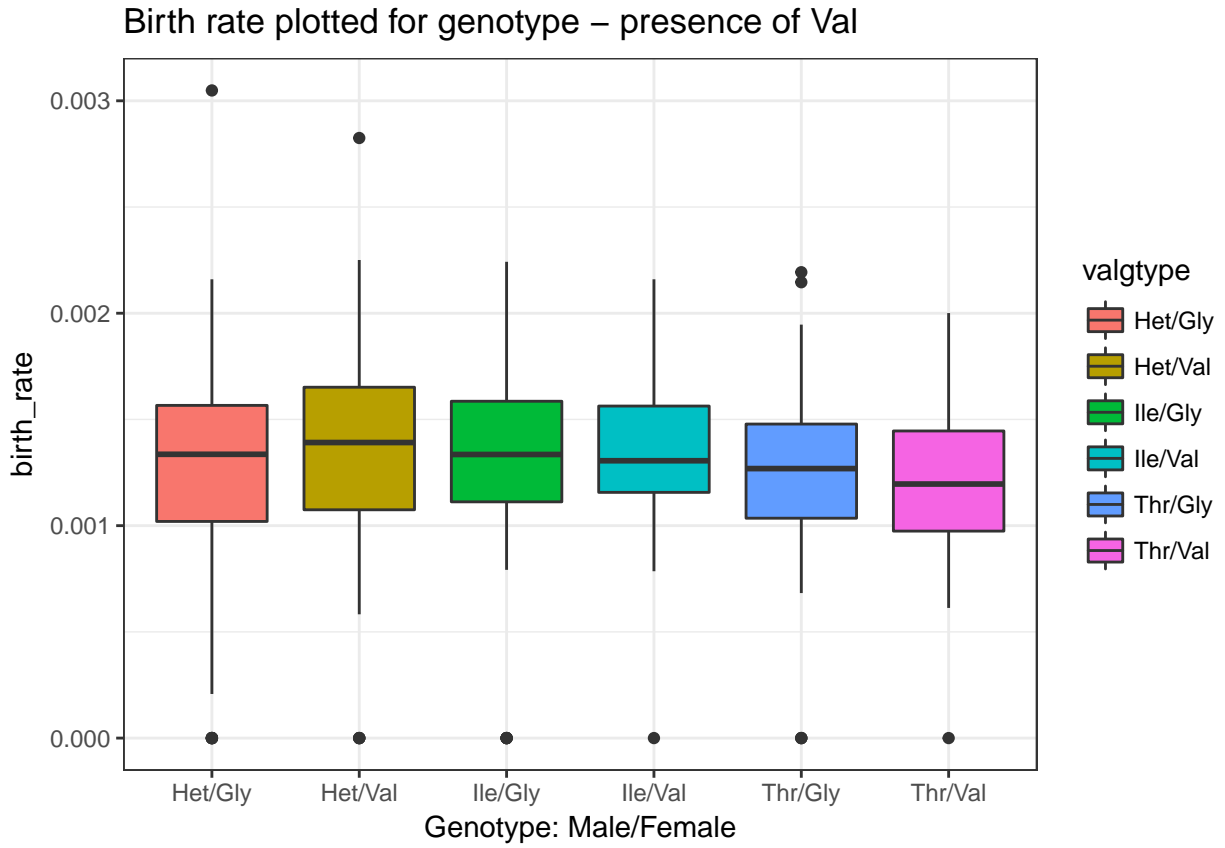


Figure 5: **Figure 5:** Boxplot of the number of children for each couple in each genotype group with the females grouped by presence of the ancestral allele. Birth rate residuals were calculated using wife's birth year and time from marriage to last birth as covariates.

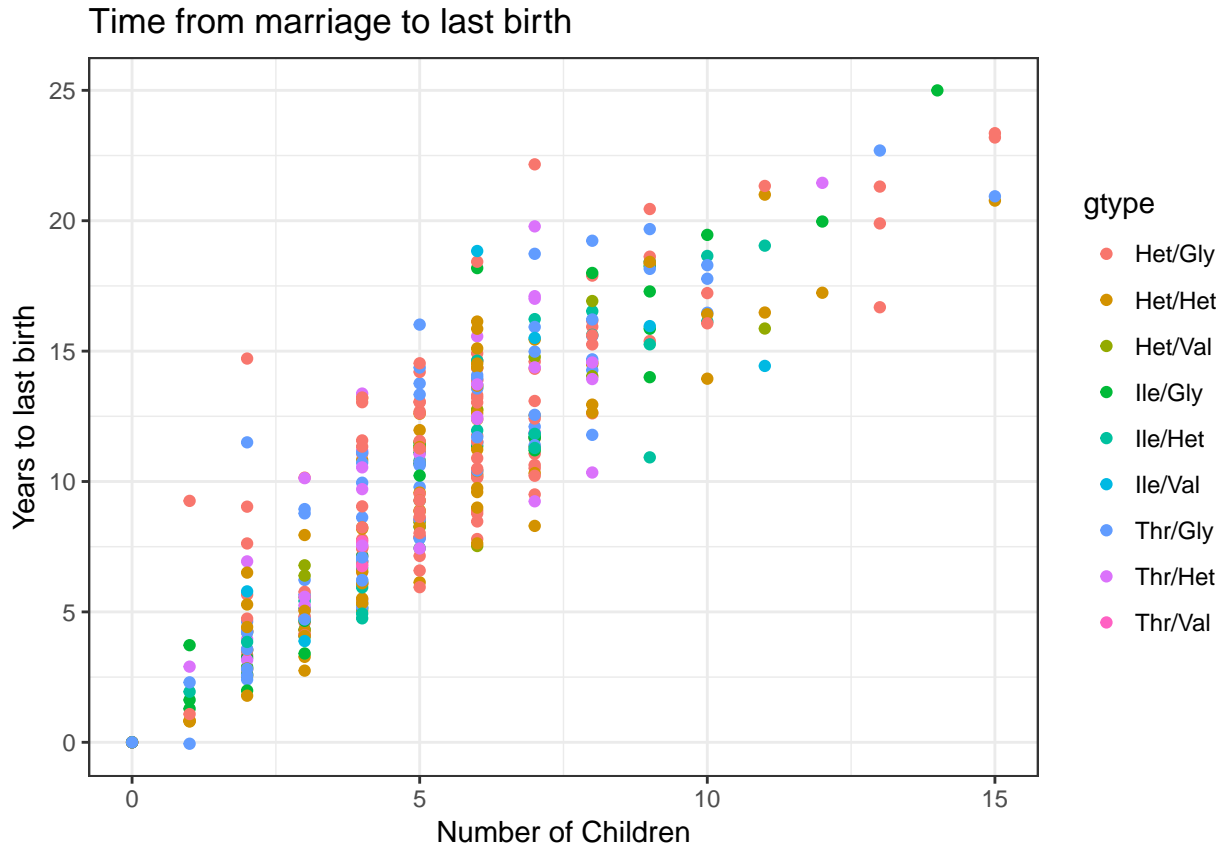


Figure 6: **Figure 6:** Scatterplot of the time from marriage to last (most recent) birth in years on the y-axis plotted against number of children in the family on the x-axis. Each point is a couple, and different colors denote different couple genotype groups.

The remaining figures are histograms looking at the number of children and birth rate colored by genotype groups.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

: Linear regression model & anova

Residuals were calculated with wife's birth year and time from marriage to last birth as covariates, nothing significant

Association tests using GEMMA

I used a linear model as implemented in GEMMA3 to test for association between genotype and birth rate, correcting for the relatedness among individuals. I used an additive model for both rs2075520 (in women) and rs4844573 (in men), where the three genotypes were coded as 0, 1, 2 corresponding to the number of derived alleles. The kinship coefficients used to correct for relatedness between individuals were estimated from pedigree data.

Following Kosova et al. 2010, I calculated birth rate as $[(\#births-1)/(\text{sum of the interbirth intervals})].2$ This analysis includes only couples with at least 2 children (to allow calculation of interbirth intervals). I also

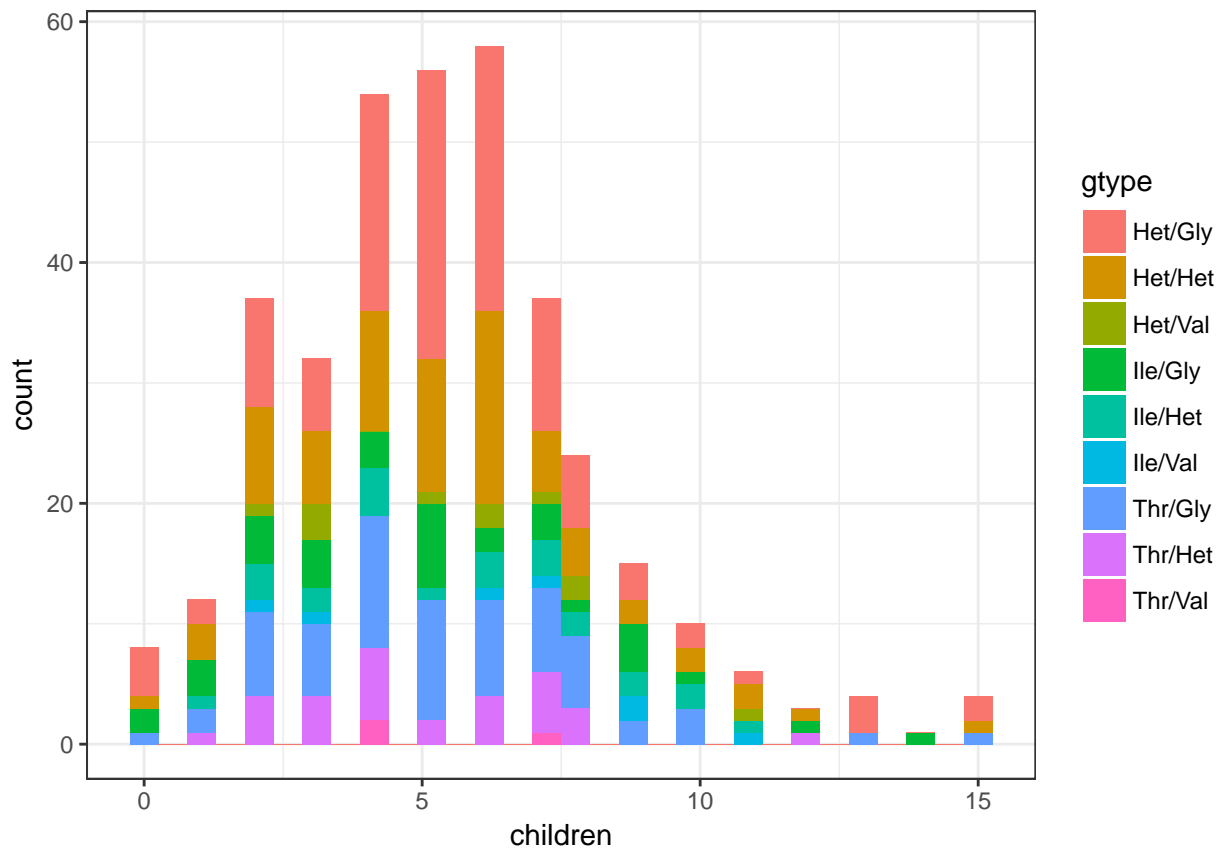


Figure 7: **Figure 7:** Histogram of the number of couples binned by the number of children. Couples are colored by genotype group.

considered family size in couples with at least one child (proven fertility) as well as in all couples (including childless couples).

Following Kosova et al. 2012, I included wife's birth year, wife's age at marriage, and time from marriage to last birth as covariates to adjust for the cohort effects, maternal age effects, and the length of reproductive period and the incompleteness of some families (analysis includes both completed and in-completed families), respectively.⁴ In separate analyses I included the spouse's genotype at the opposite SNP as a covariate (wherever it is noted there are 4 covariates).

: GEMMA for birth rate

: Husband's genotype

I first tested for association between husband's genotype with birth rate for couples with at least 2 children (341 couples). In the analysis without covariates, there was an association between husband's genotype and birth rate with p-value: 0.0054 (Figure 1). The boxplots in Figure 1 do not represent data corrected for relatedness among individuals or for the relevant covariates. Birth rate residuals, calculated by including three covariates (days from marriage to last birth, wife's birth year, and wife's age at marriage), were significantly associated with husband's genotype (p-value = 0.020; Figure 2). Finally, when wife's genotype was included as an additional covariate the p-value increased to 0.034 (Figure 3). These data indicate that there is a correlation between husband's genotype at rs4844573 and birth rate (in an additive model), and that the wife's genotype at rs2075520 does not contribute to the association. In a final model, I included an interaction term for husband's genotype at rs4844573 and wife's genotype at rs2075520 in addition to including the 4 covariates and correcting for relatedness between husbands. This model was not significant (p-value = 0.147). Using R and not correcting for relatedness between individuals, the interaction term had a coefficient estimate of -0.00001.4 (SE 0.00004.3) (p=0.65). The addition of the interaction term did not improve the model (likelihood ratio test, p=0.65).

: Wife's genotype

I also tested for association of wife's genotype at rs2075520 with birth rate for couples with at least two children (341 couples), but it was not significant. Without covariates, the association p-value = 0.33 (Figure 4). Using the same three covariates described above, the p-value = 0.29 (Figure 5) and when the husband's genotype at rs4844573 was included as a covariate the p-value = 0.43 (Figure 6). In a final model, I included an interaction term for husband's genotype at rs4844573 and wife's genotype at rs2075520 in addition to including the 4 covariates and correcting for relatedness between wives. This model was not significant (p-value = 0.852).

: GEMMA for family size

: Couples with at least one child

To test for associations with family size, I included couples with at least one child (353 couples). Without covariates, there was no significant association (p-value = 0.39) between husband's genotype at rs4844573 and family size (**Figure 7**). In analyses including the three covariates described above, there was a trend toward significance with p = 0.066 (**Figure 8**). The association was less significant (p = 0.14) when the wife's genotype at rs2075520 was included as an additional covariate (**Figure 9**). In a final model, I included an interaction term for husband's genotype at rs4844573 and wife's genotype at rs2075520 in addition to including the 4 covariates and correcting for relatedness between husbands. This model was not significant (p-value = 0.307). The addition of the interaction term did not improve the model (likelihood ratio test, p=0.911).

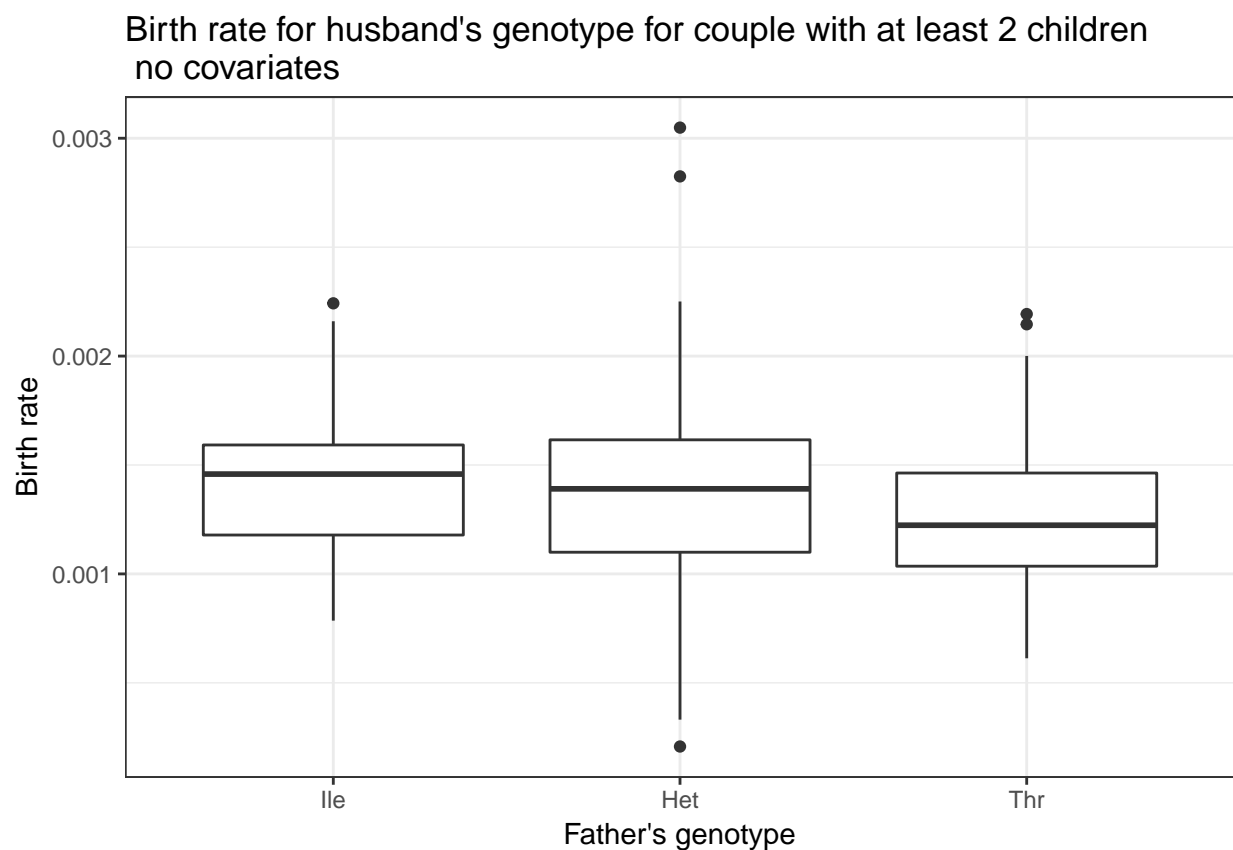


Figure 8: **Figure 1.** Boxplots showing husband's genotype at rs4844573 (x-axis) and birth rate (y-axis) for couples with at least two children (p-value 0.0054). The black bars are medians, and the boxes show the 25th -75th percentiles. The upper whisker extends from the third quartile to the highest value that is within $1.5IQR$ of the third quartile, where IQR is the inter-quartile value, or the distance between the first and third quartiles. The lower whisker extends from the first quartile to the lowest value that is within $1.5IQR$ of the first quartile.

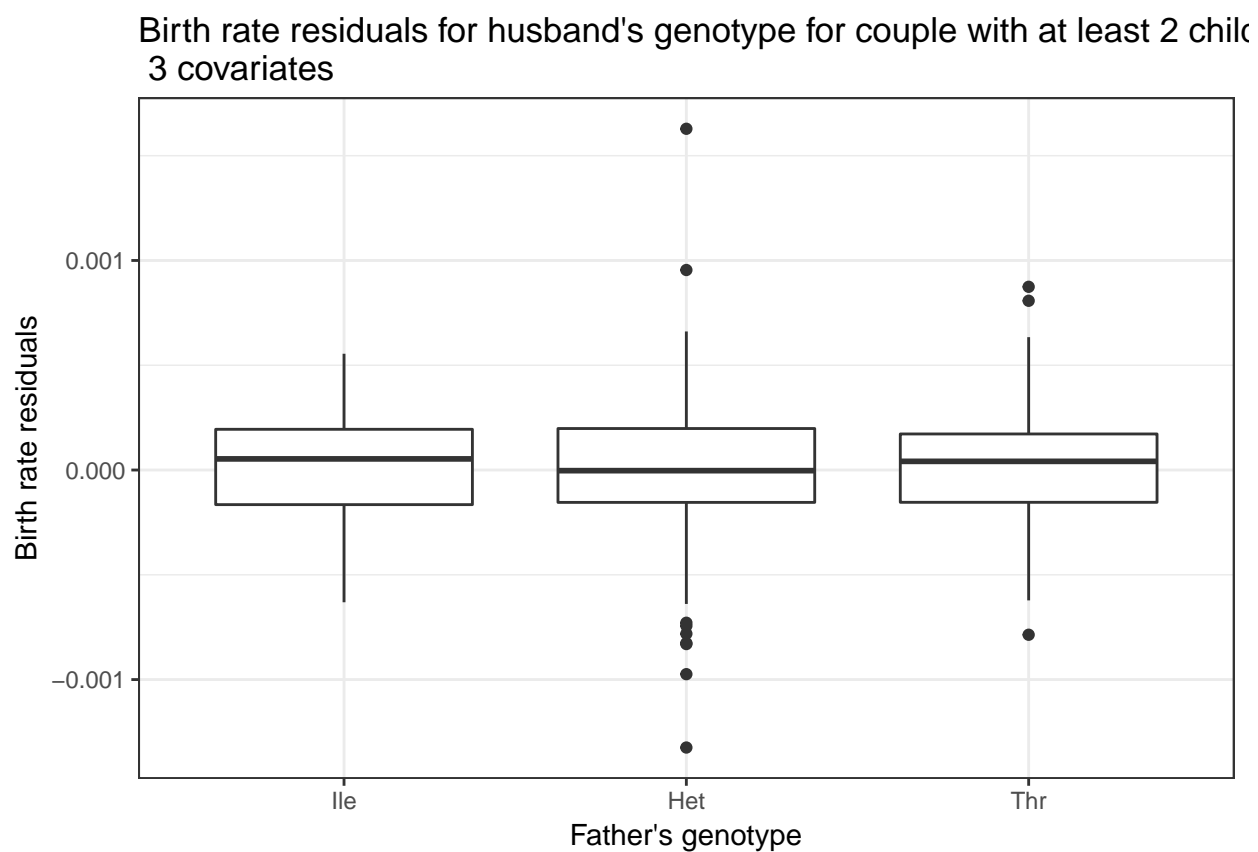


Figure 9: **Figure 2.** Boxplots showing husband's genotype at rs4844573 (x-axis) and birth rate (y-axis) adjusted for wife's age at marriage, wife's birth year, and time from marriage to last birth (p-value 0.020).

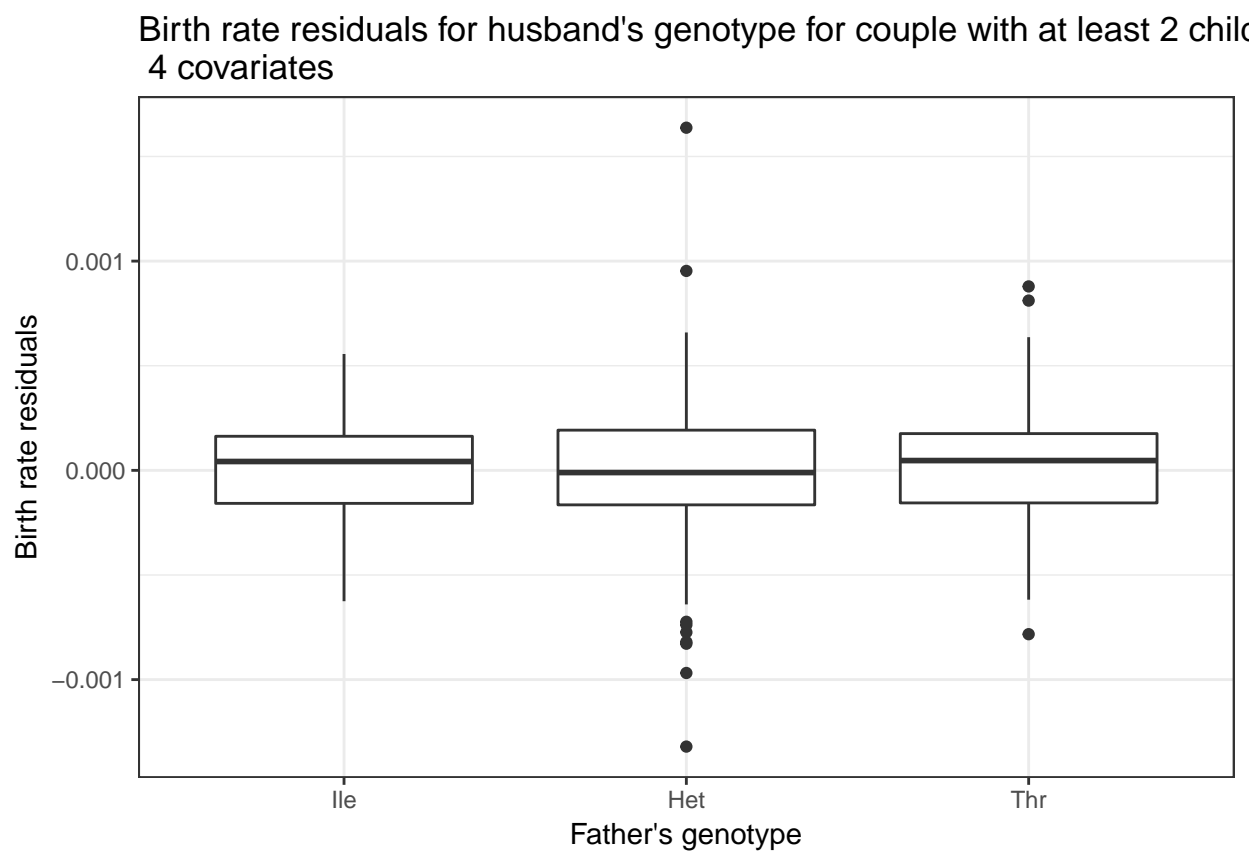


Figure 10: **Figure 3.** Boxplots showing husband's genotype at rs4844573 (x-axis) and birth rate adjusted for wife's age at marriage, wife's birth year, time from marriage to last birth, and wife's genotype at rs2075520 (y-axis) (p-value 0.034).

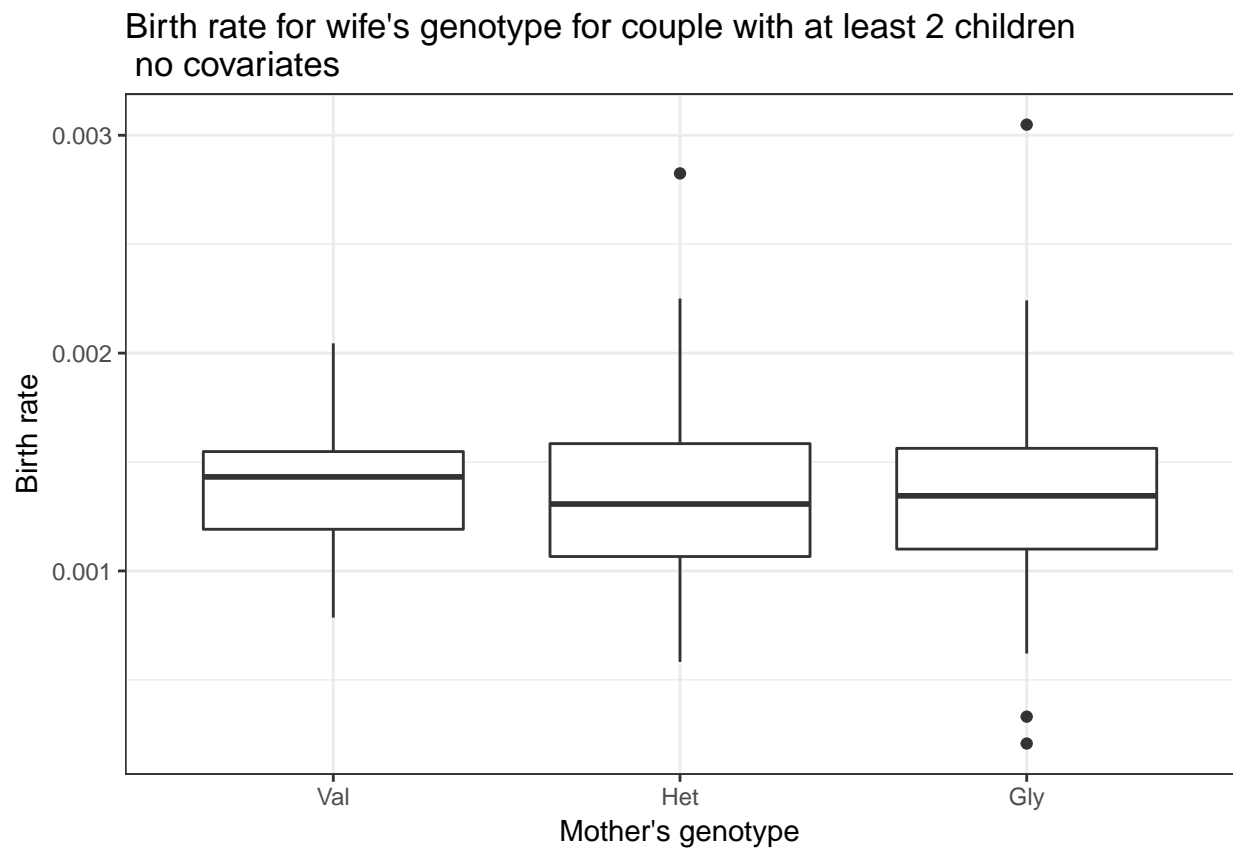


Figure 11: **Figure 4.** Boxplots showing wife's genotype and birth rate for couples with at least two children (p-value 0.033).

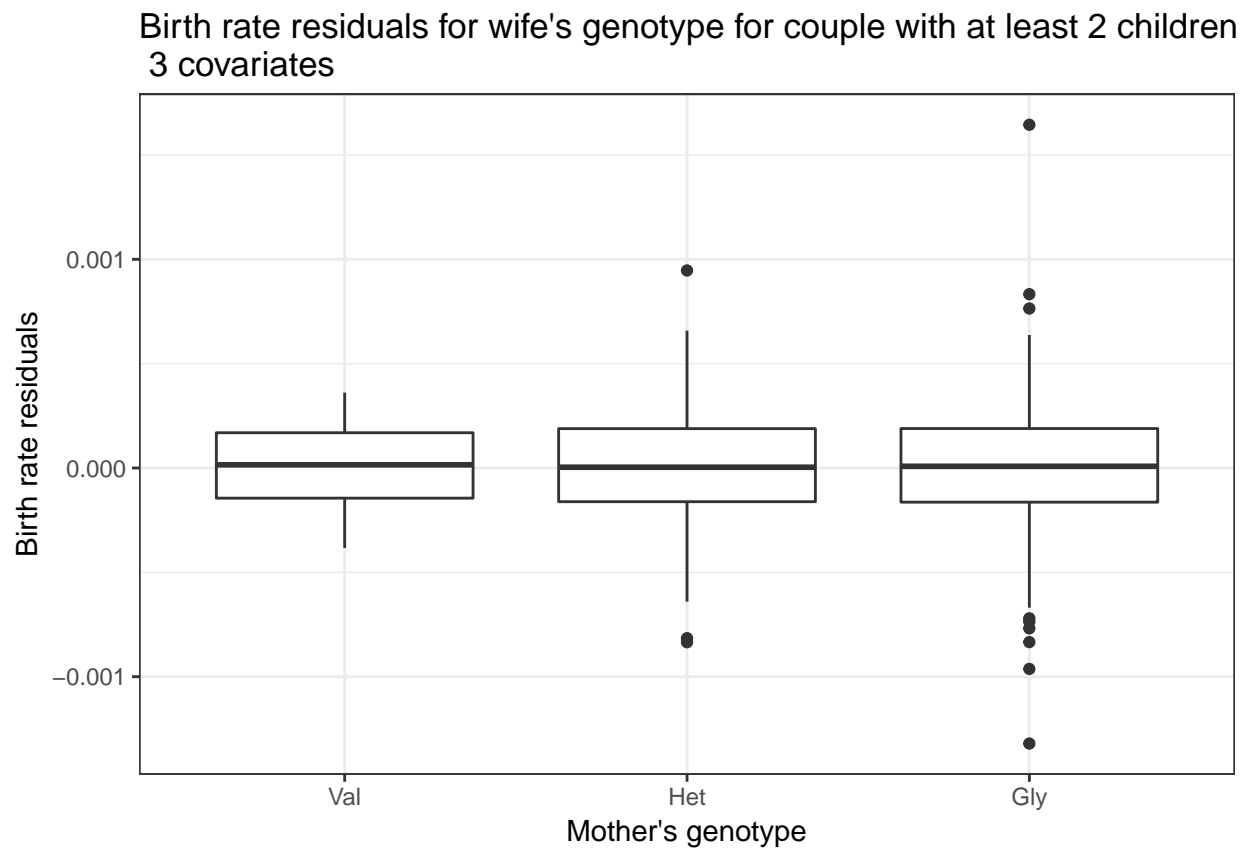


Figure 12: **Figure 5.** Boxplots showing wife's genotype against birth rate residuals calculated using wife's age at marriage, wife's birth year and time from marriage to last birth (p-value 0.29).

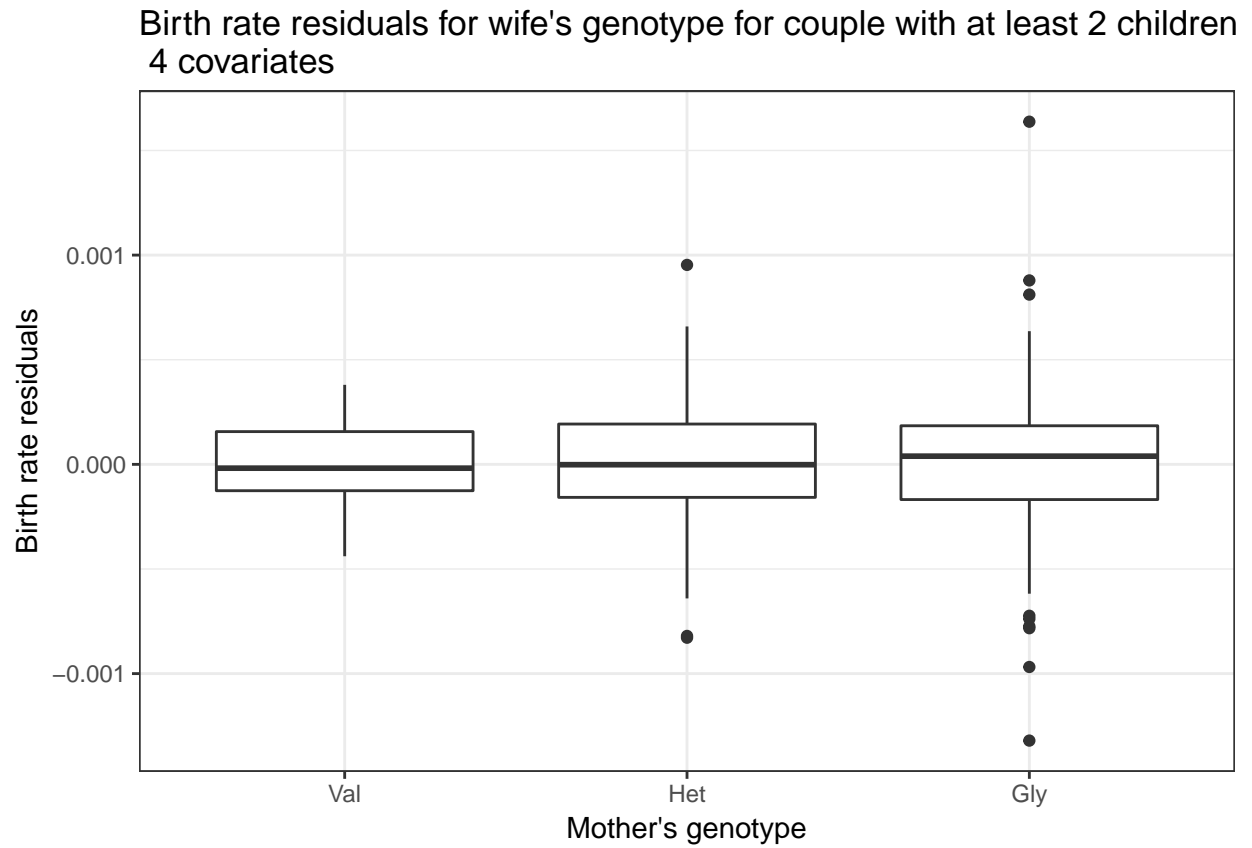


Figure 13: **Figure 6.** Boxplots showing wife's genotype and birth rate residuals calculated using wife's age at marriage, wife's birth year, time from marriage to last birth, and husband's genotype at rs4844573 (p-value 0.43).

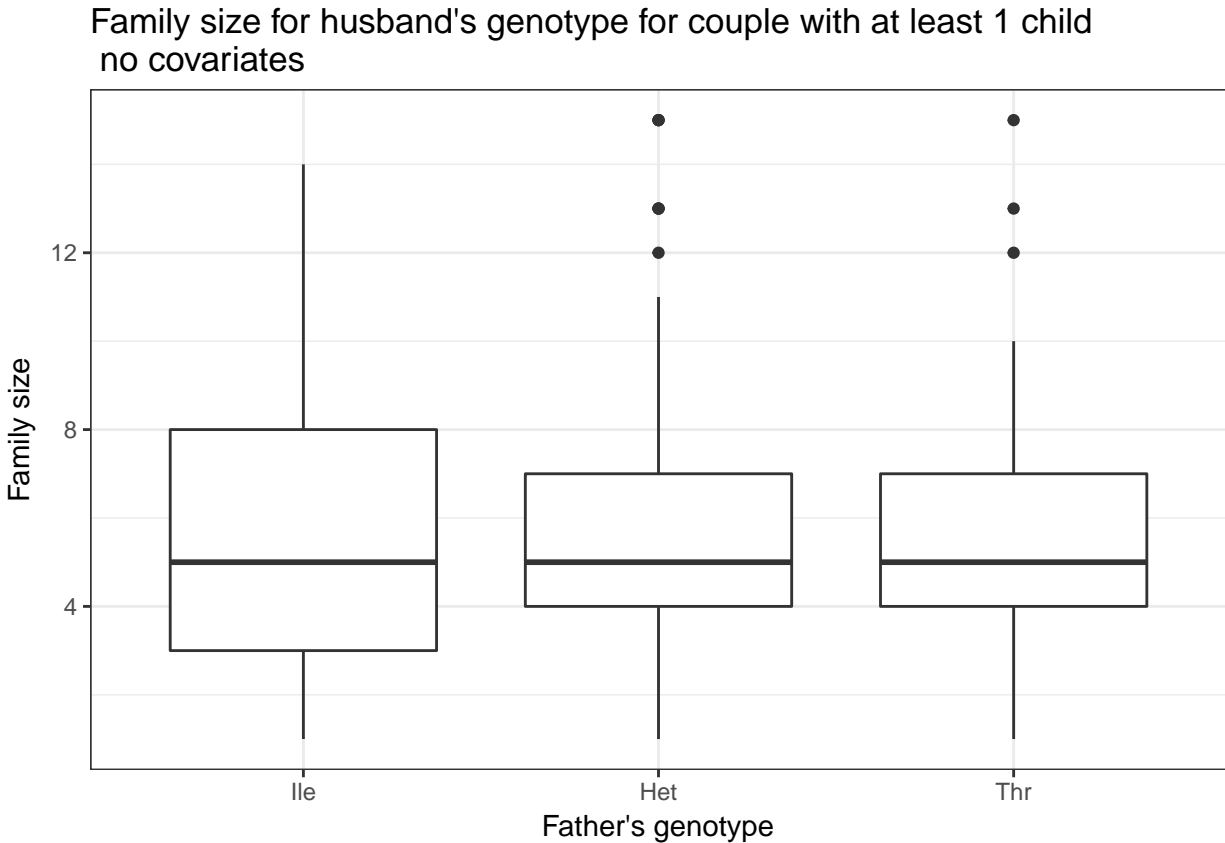


Figure 14: **Figure 7.** Boxplots showing husband's genotype and family size for couples with at least one child (p-value 0.39).

: Husband's genotype

: Wife's genotype

I repeated the same tests but now using the wife's genotype as the main effect. The association between wife's genotype and family size in analyses without covariates were not significant ($p = 0.10$; **Figure 10**). This association becomes marginally more significant when the three covariates are included ($p = 0.075$; **Figure 11**), and less significant when the husband's genotype is included as a covariate ($p = 0.12$; **Figure 12**). In a final model, I included an interaction term for husband's genotype at rs4844573 and wife's genotype at rs2075520 in addition to including the 4 covariates and correcting for relatedness between wives. This model was not significant ($p = 0.419$).

: All Couples

Although we do not know the cause of infertility among the childless couples (and therefore this is likely a very heterogeneous group), I repeated the analyses of family size including seven couples without children (361 couples). There was no association with husband's genotype and family size in the analyses without covariates (Figure 13, p-value: 0.57). This association became more significant in the analysis including the three covariates ($p = 0.089$; **Figure 14**) and less significant when wife's genotype was included as the fourth covariate ($p = 0.19$; **Figure 15**). In a final model, I included an interaction term for husband's genotype at rs4844573 and wife's genotype at rs2075520 in addition to including the 4 covariates and correcting for

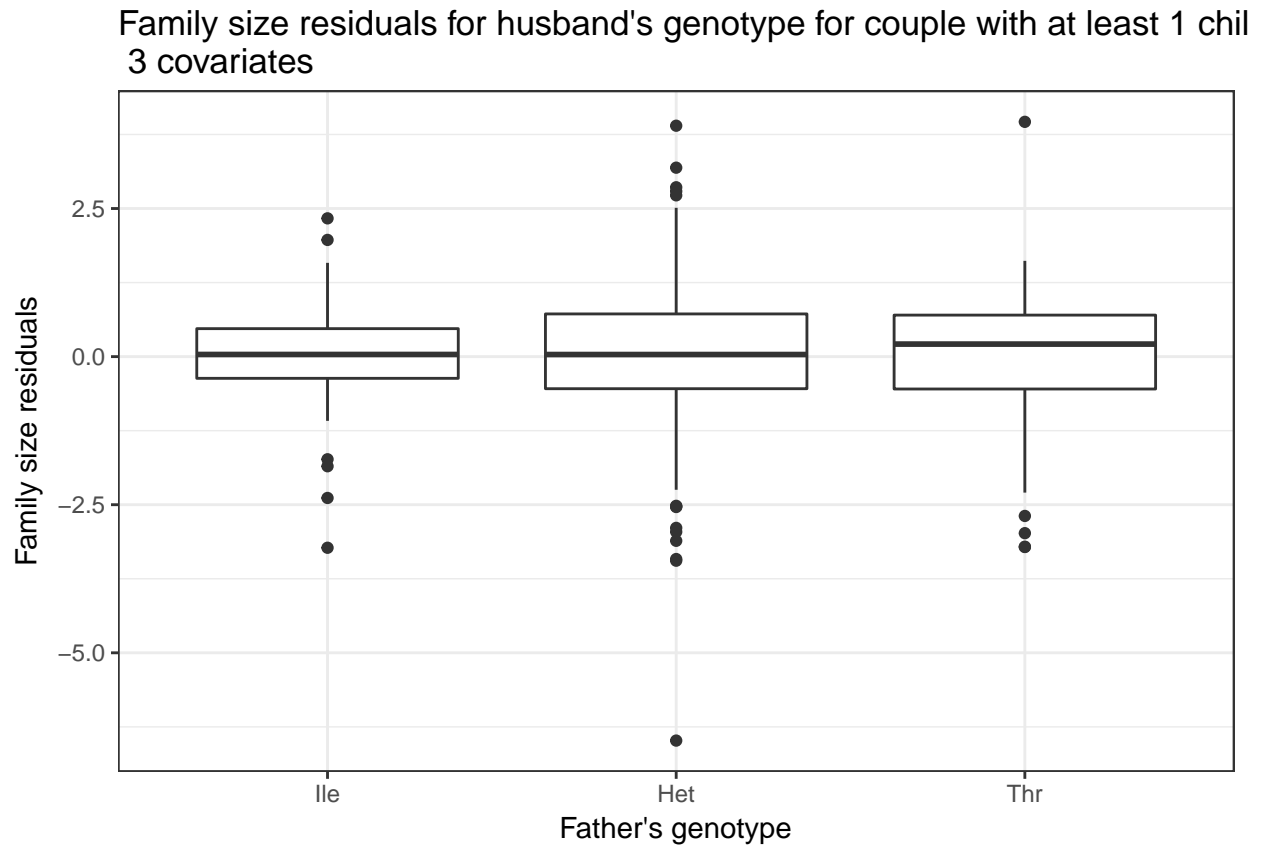


Figure 15: **Figure 8.** Boxplots showing husband's genotype at rs4844573 (x-axis) and birth rate adjusted for wife's age at marriage, wife's birth year, time from marriage to last birth, and wife's genotype at rs2075520 (y-axis) (p-value 0.034).



Figure 16: **Figure 9.** Boxplots for association of husband's genotype with family size residuals calculated using wife's age at marriage, wife's birth year, time from marriage to last birth, and wife's genotype (as an additive model) (p-value 0.14).

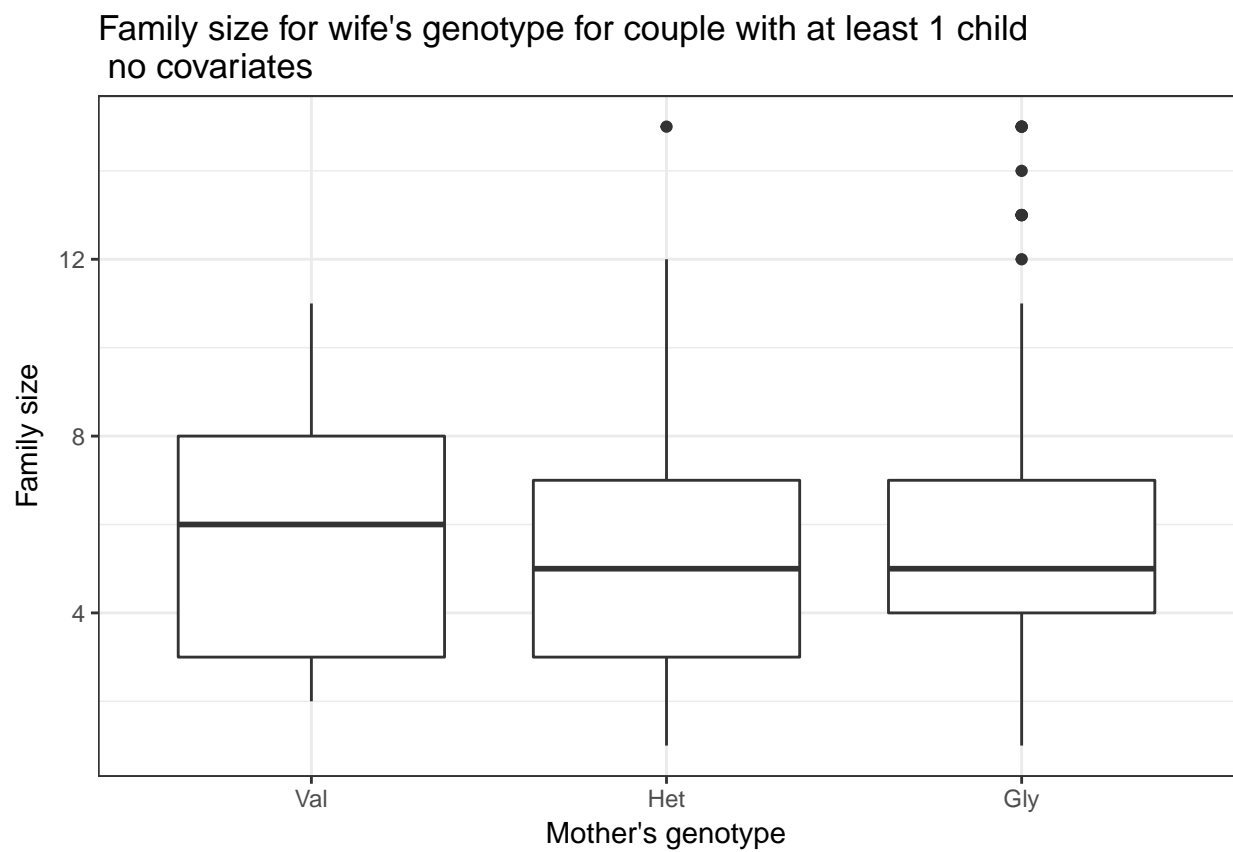


Figure 17: **Figure 10.** Boxplots for association of wife's genotype with family size for couples with at least one child (p-value 0.10).

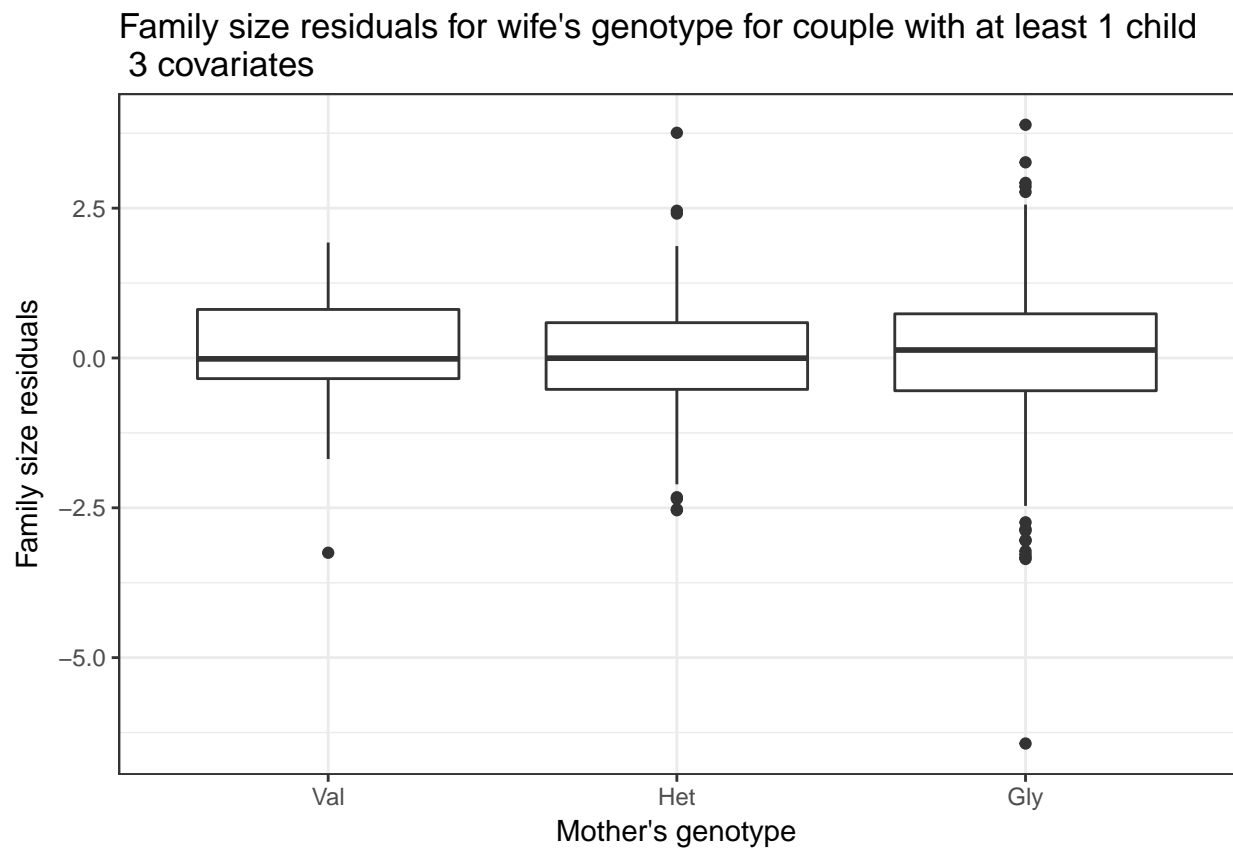


Figure 18: **Figure 11.** Boxplots for association of wife's genotype with family size residuals calculated using wife's age at marriage, wife's birth year and time from marriage to last birth (p-value 0.075).

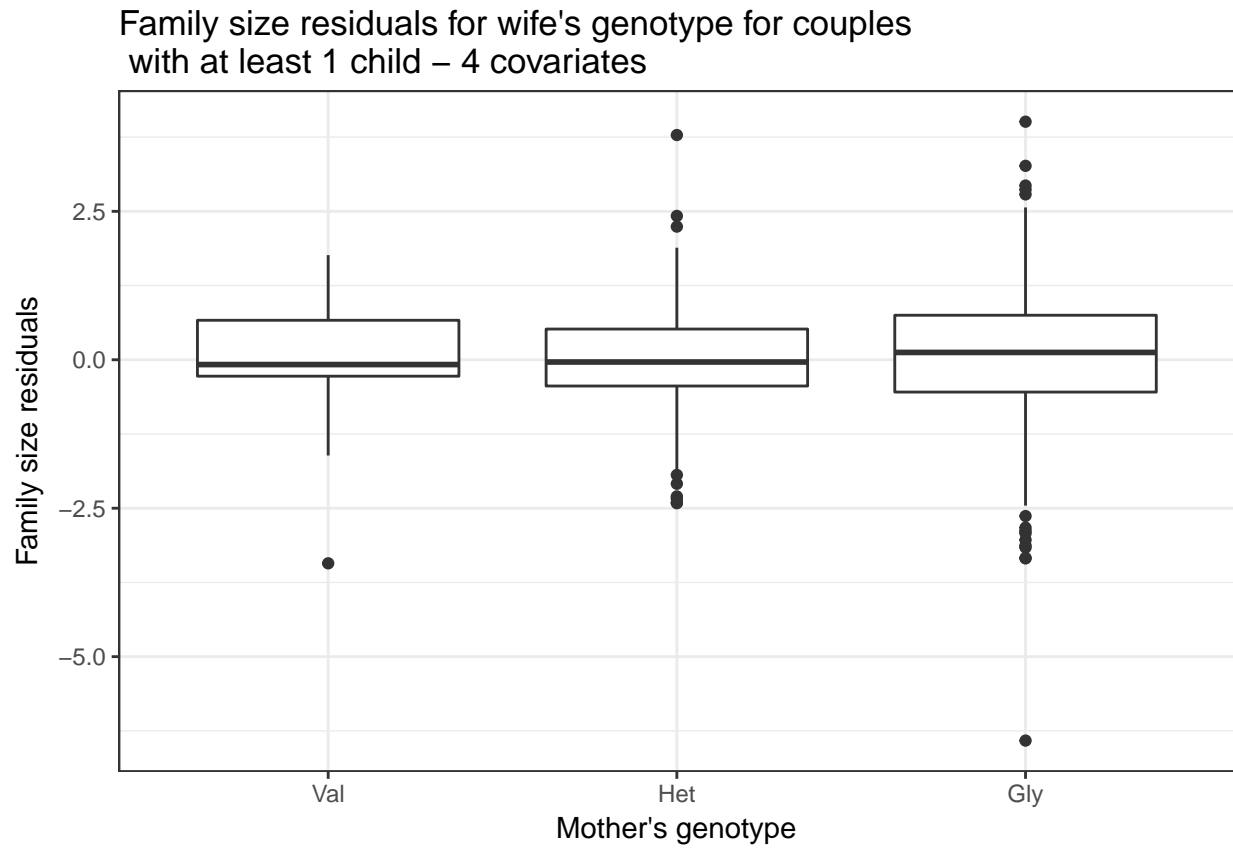


Figure 19: **Figure 12.** Boxplots for association of wife's genotype with family size residuals calculated using wife's age at marriage, wife's birth year, time from marriage to last birth, and husband's genotype (as an additive model) (p-value = 0.12).

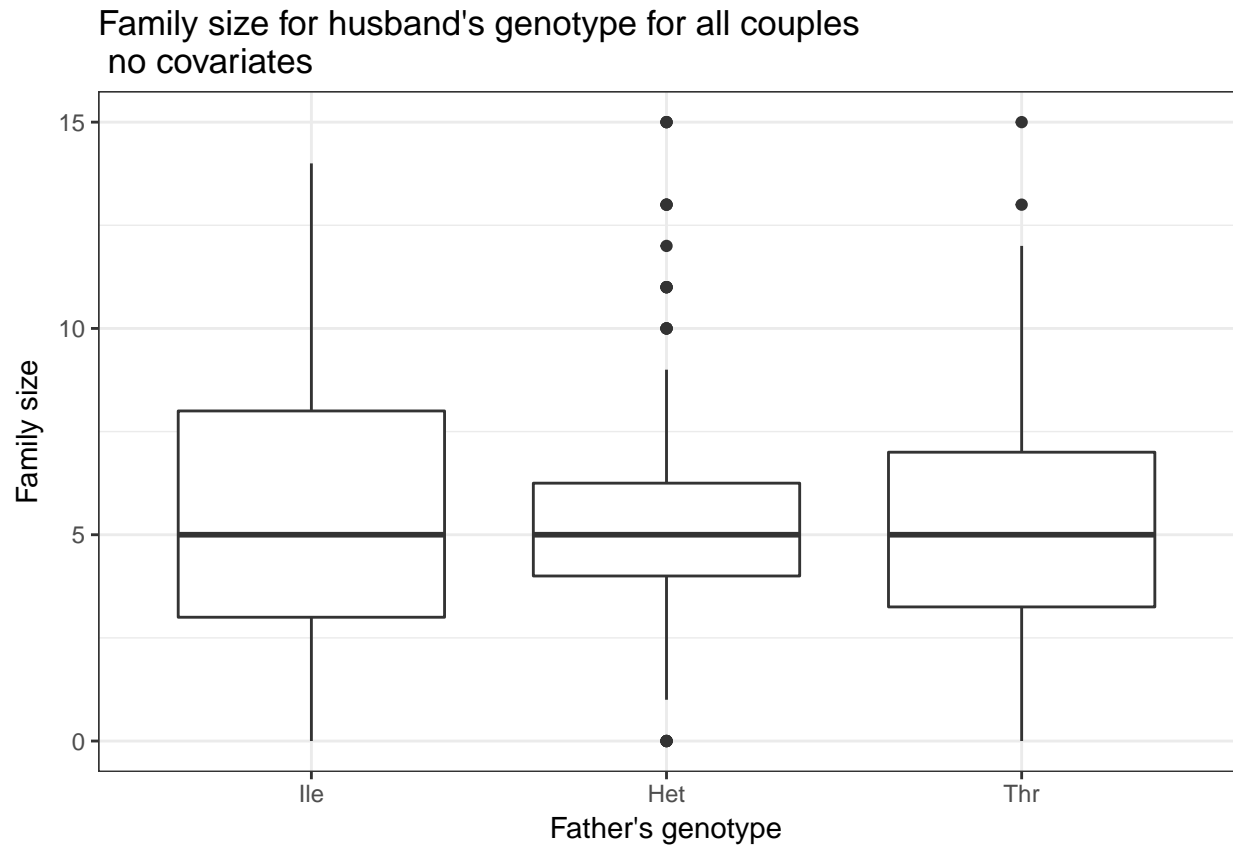


Figure 20: **Figure 13.** Boxplots for association of husband's genotype with family size for all couples (p-value 0.57).

relatedness between husbands. This model was not significant ($p = 0.30$). The addition of an interaction term did not improve the model (likelihood ratio test, $p = 0.837$).

: Husband's genotype

: Wife's genotype

In summary, the husbands' genotype at rs4844573 was associated with birth rate residuals ($p = 0.020$), with the derived (Ile) allele associated with higher birth rates. The wife's genotype at rs2075520 was associated with family size residuals ($p = 0.075$ when considering families with at least one child; $p = 0.040$ when considering all families), with the derived allele associated with smaller families. In all analyses, the addition of the partner's genotype at the alternate SNP reduced the evidence for association suggesting that the effects on fertility are not due to interactions between the husbands' and wives' genotypes. There was no evidence for interaction effects between husbands and wives' genotypes on birth rate or family size after correcting for 3 covariates (wife's age at marriage, wife's birth year, and time from marriage to last birth), and relatedness among individuals.

library("rmarkdown")

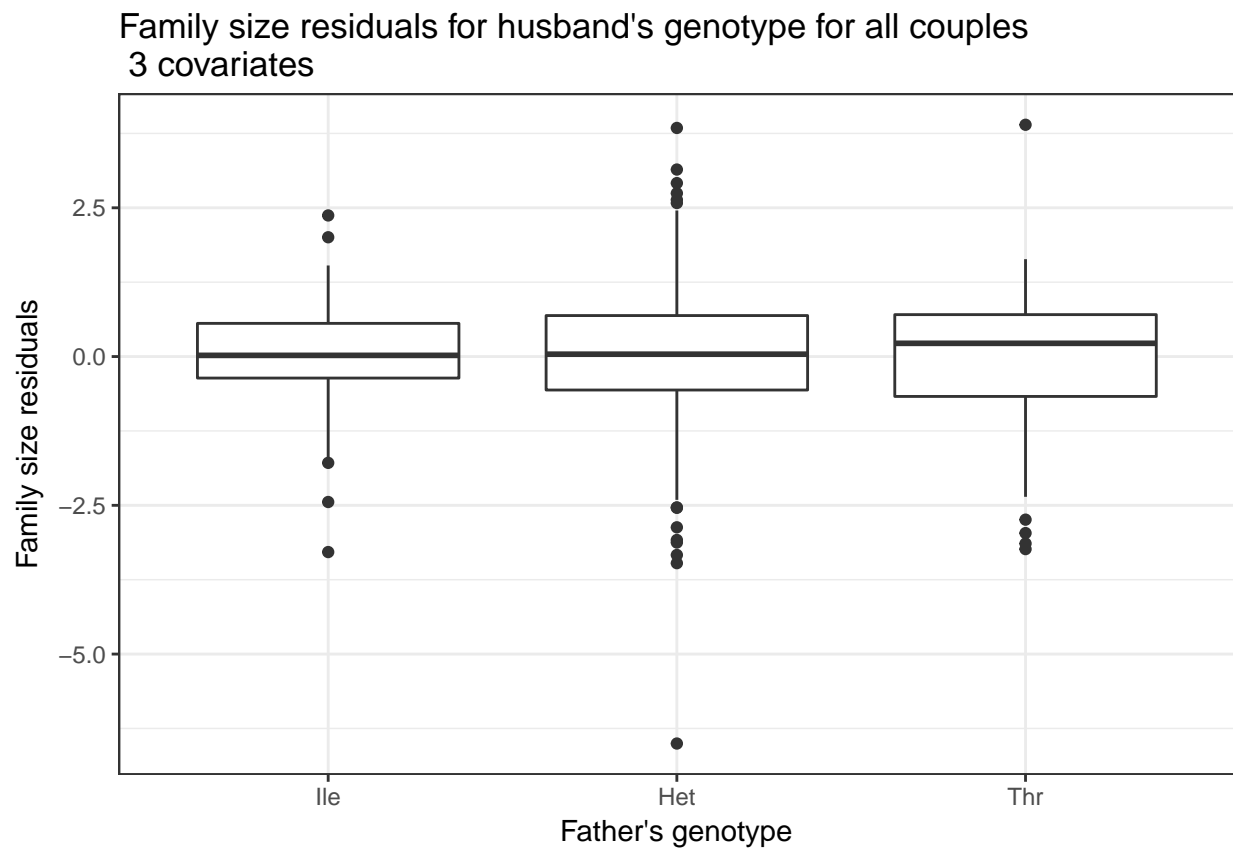


Figure 21: **Figure 14.** Boxplots for association of husband's genotype with family size residuals calculated using wife's age at marriage, wife's birth year and time from marriage to last birth (p-value 0.089).

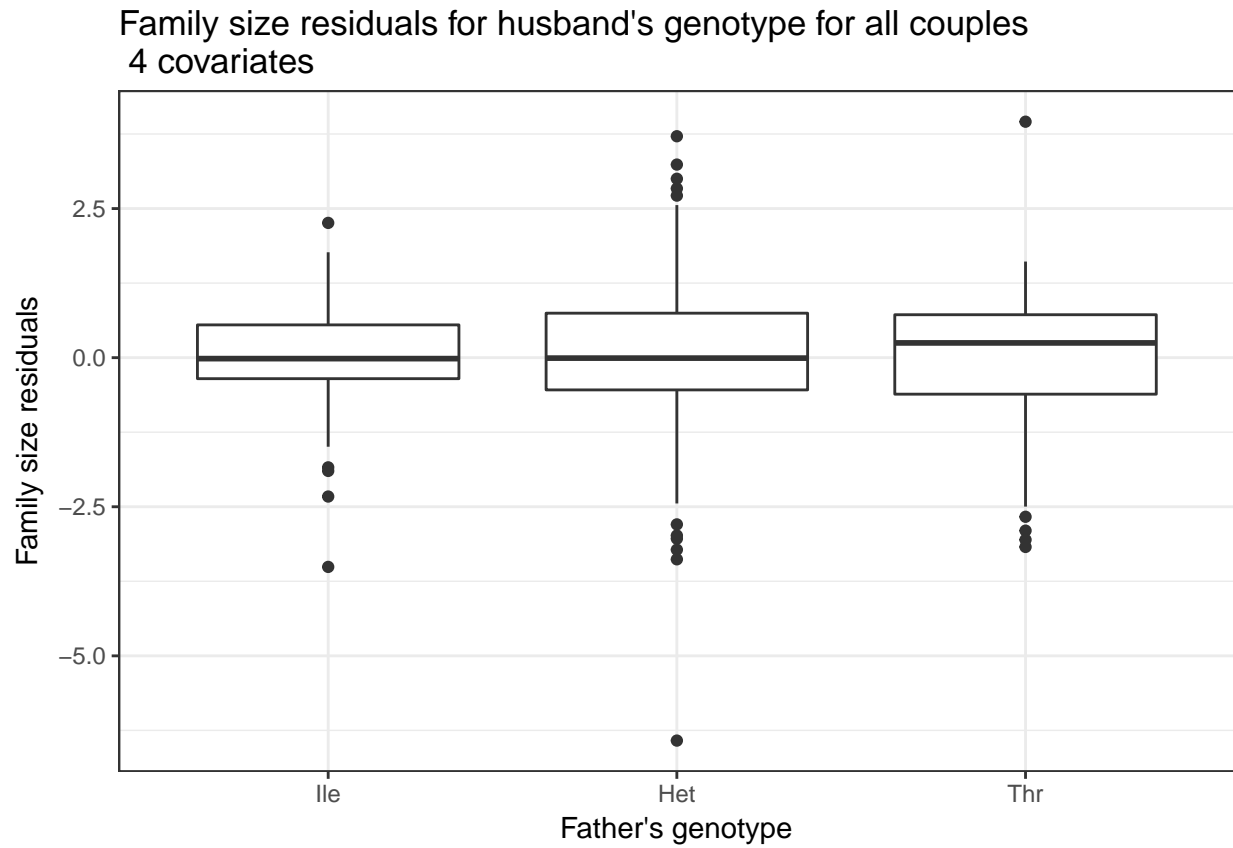


Figure 22: **Figure 15.** Boxplots for association of husband's genotype with family size residuals calculated using wife's age at marriage, wife's birth year, time from marriage to last birth, and wife's genotype (p-value 0.19).

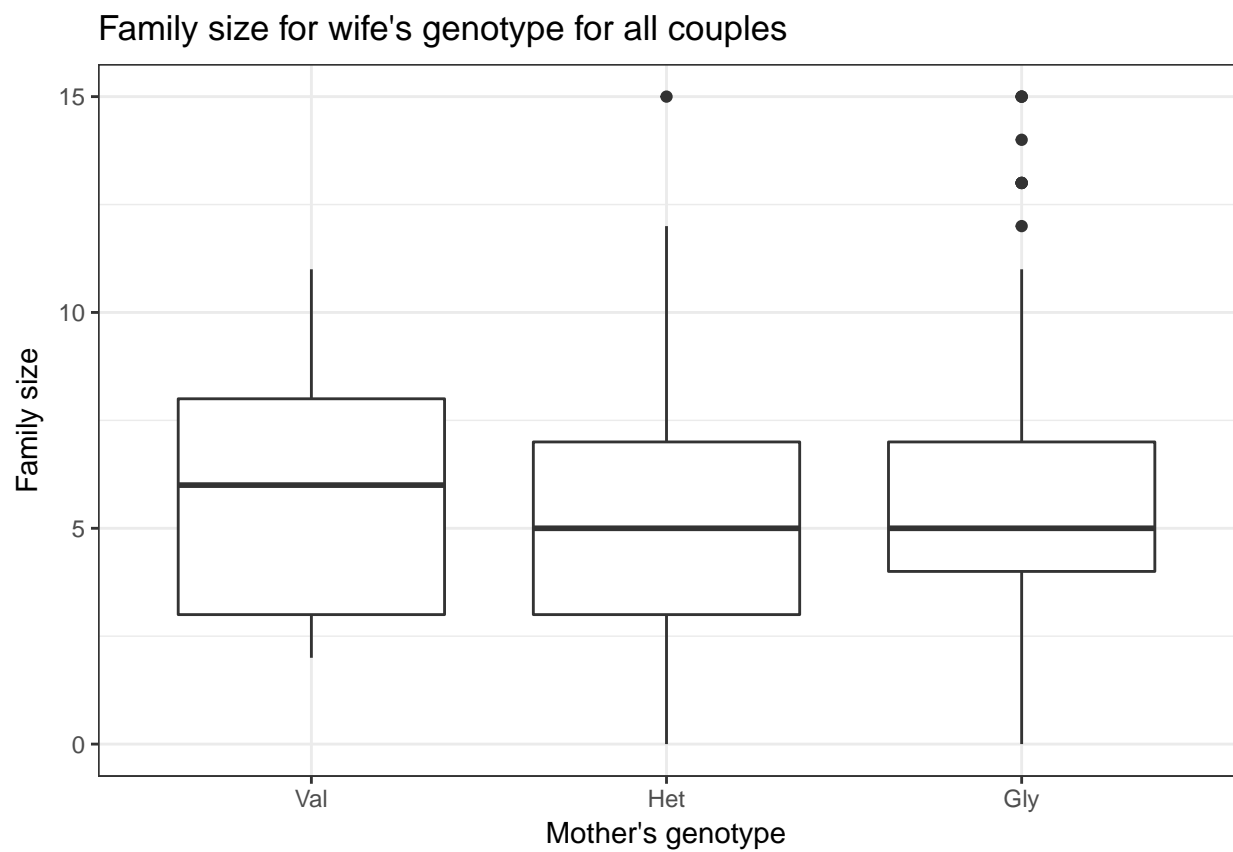


Figure 23: **Figure 16.** Boxplots for association of wife's genotype with family size for all couples (p-value 0.64)

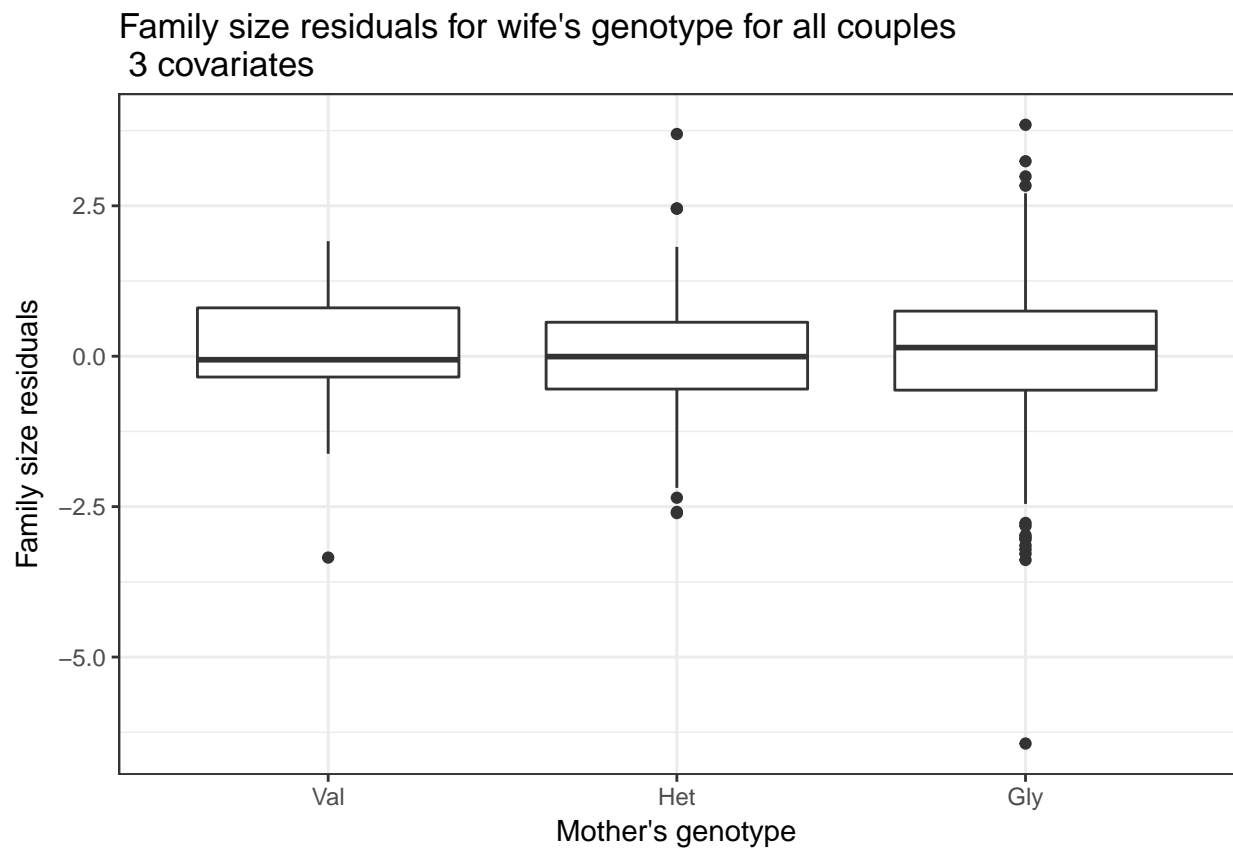


Figure 24: **Figure 17.** Boxplots for association of wife's genotype with family size residuals calculated using wife's age at marriage, wife's birth year and time from marriage to last birth (p-value 0.040).

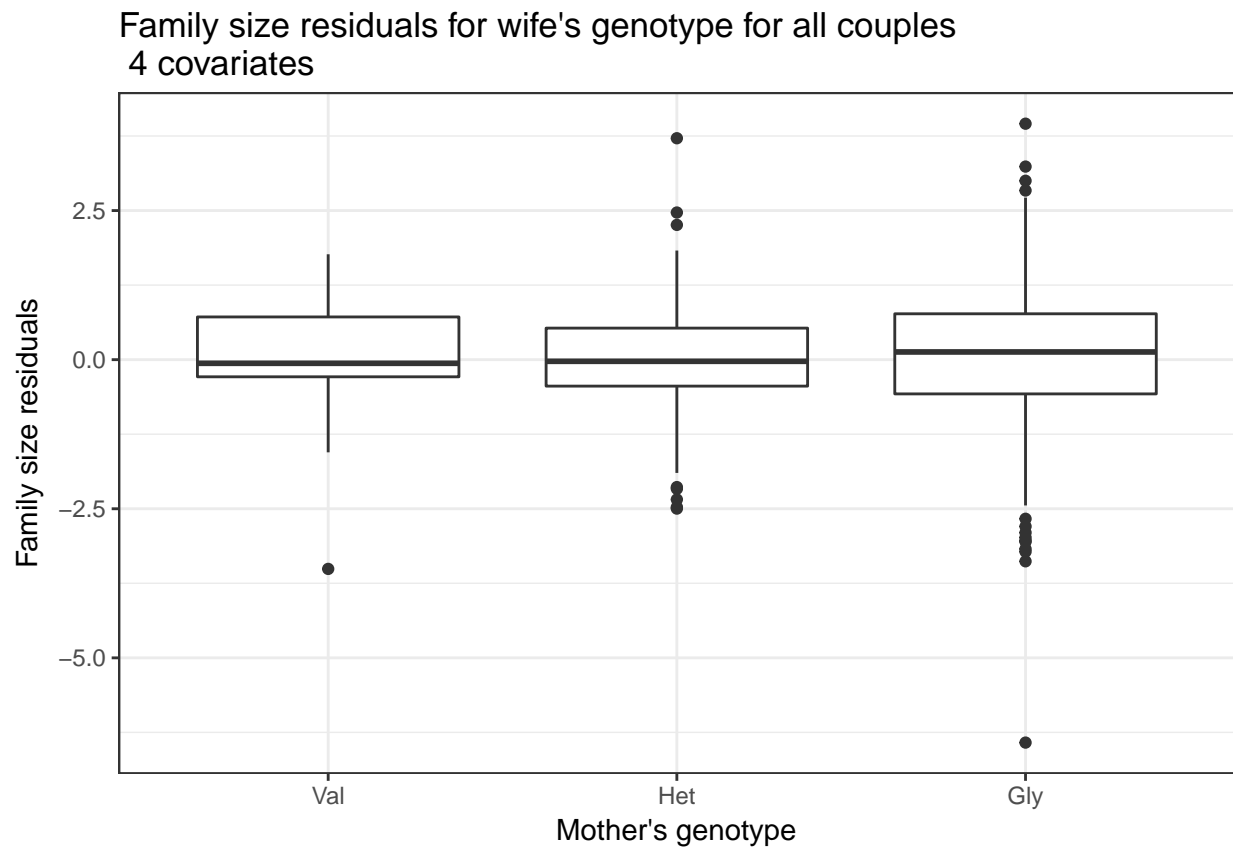


Figure 25: **Figure 18.** Boxplots for association of wife's genotype with family size residuals calculated using wife's age at marriage, wife's birth year, time from marriage to last birth, and husband's genotype (p-value 0.063).