NF1 Thesis Analysis

Marshall McQuillen

12/10/2018

Contents

Study Aims

1. To assess whether parental knowledge of NF1 is dependent on the child's mode of inheritance of the condition (familial vs de novo mutations)
2. To determine how exposure to genetic counseling affects knowledge of NF1.

Data Overview & Cleaning

There were a total of 303 responses to the questionare:

- 11 respondents were "Not sure" if they had NF1. These respondents were dropped from the data set since both study aims are contingent upon knowing this information.
- One respondent skipped the question, "Does your child's other parent have NF1?" This respondent was kept since they answered the question, "Do you have NF1?" and their answer to "Does your child's other parent have NF1?" was imputed as "No."
- One respondent skipped the question, "Doe you have NF1?" This respondent was dropped from the data set since both study aims are contingent upon knowing this information.

- 14 respondents do not have NF1 and are "Not sure," if the childs other parent has NF1. These respondents were dropped from the data set since both aims of the thesis are dependent on this information.
- Three additional respondents were dropped from the study due to leaving other questions that will be discussed blank.

After the above adjustments, the final data set contains 274 respondents, with an age ranging from 25 to 82 and a mean age of 42.1135531. The standard deviation of the ages is 9.9638095.

Testing Methodology

Although there are only two separate study aims, the first will require a total of 3 pairwise statistical tests, comparing the average test scores of the three groups below to one another using Welch's t-Test.

- **Group A** - Respondents who do not have NF1 and the child's other parent also does not have NF1.
- **Group B** - Respondents who do not have NF1 and the child's other parent does have NF1.
- **Group C** - Respondents who do have NF1 and the child's other parent does not have NF1.

These three tests, combined with the test for the second study aim, create a total of 4 statistical tests. In order to control the family wise error rate, the original significance level ($\alpha = 0.05$) was adjusted using the Bonferroni Correction, reducing the level of statistical significance to $0.05/4 = 0.0125$.

A table with summar statistcs of the three groups is displayed below.

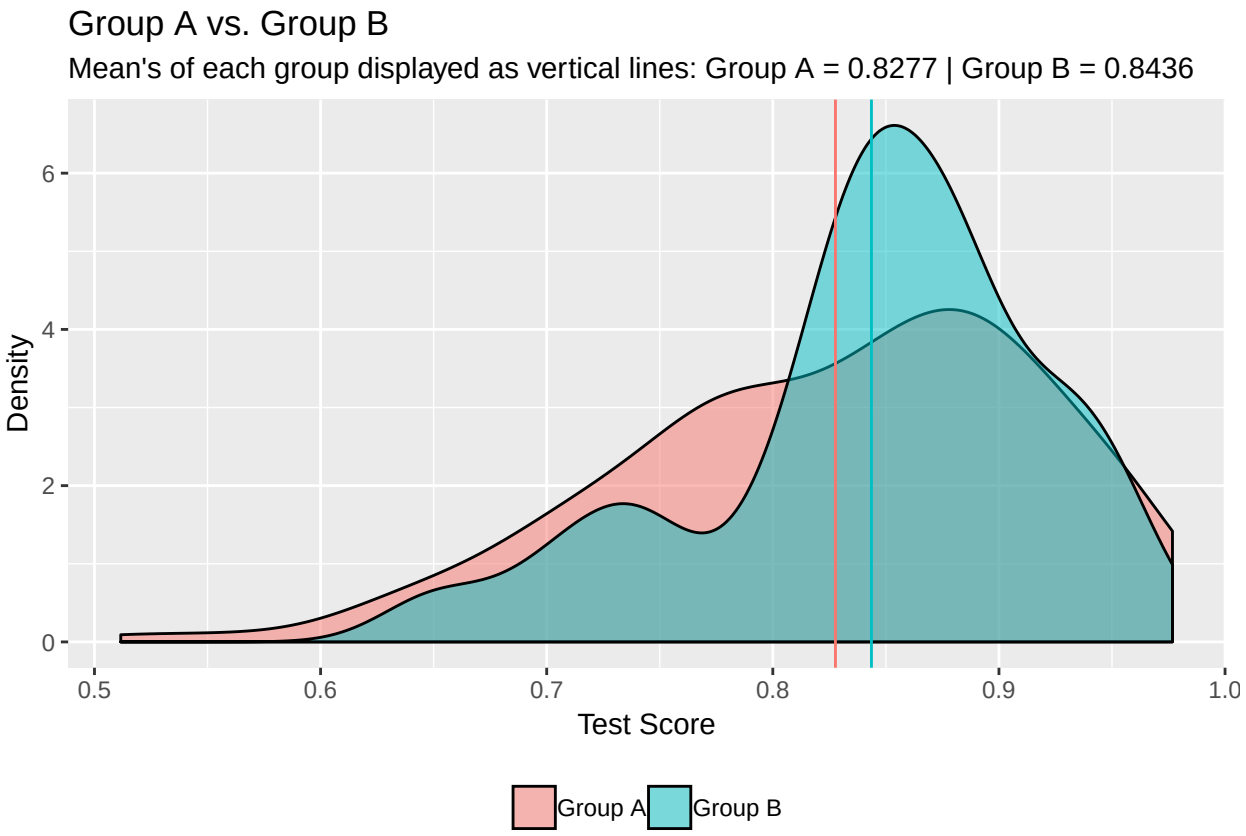
Table 1: Test Score Statistics

	Mean	StandardDeviation
Group A	0.8276901	0.0919368
Group B	0.8436247	0.0750741
Group C	0.7768336	0.0866807

Study Aim 1 - Does the mode of inheritance (familial vs de novo mutations) of the child have an effect on parental knowledge of NF1?

Test 1

- H_0 : Respondents in Group B do not have a different average test score than respondents in Group A.
- H_a : Respondents in Group B have a different average test score than respondents in Group A.

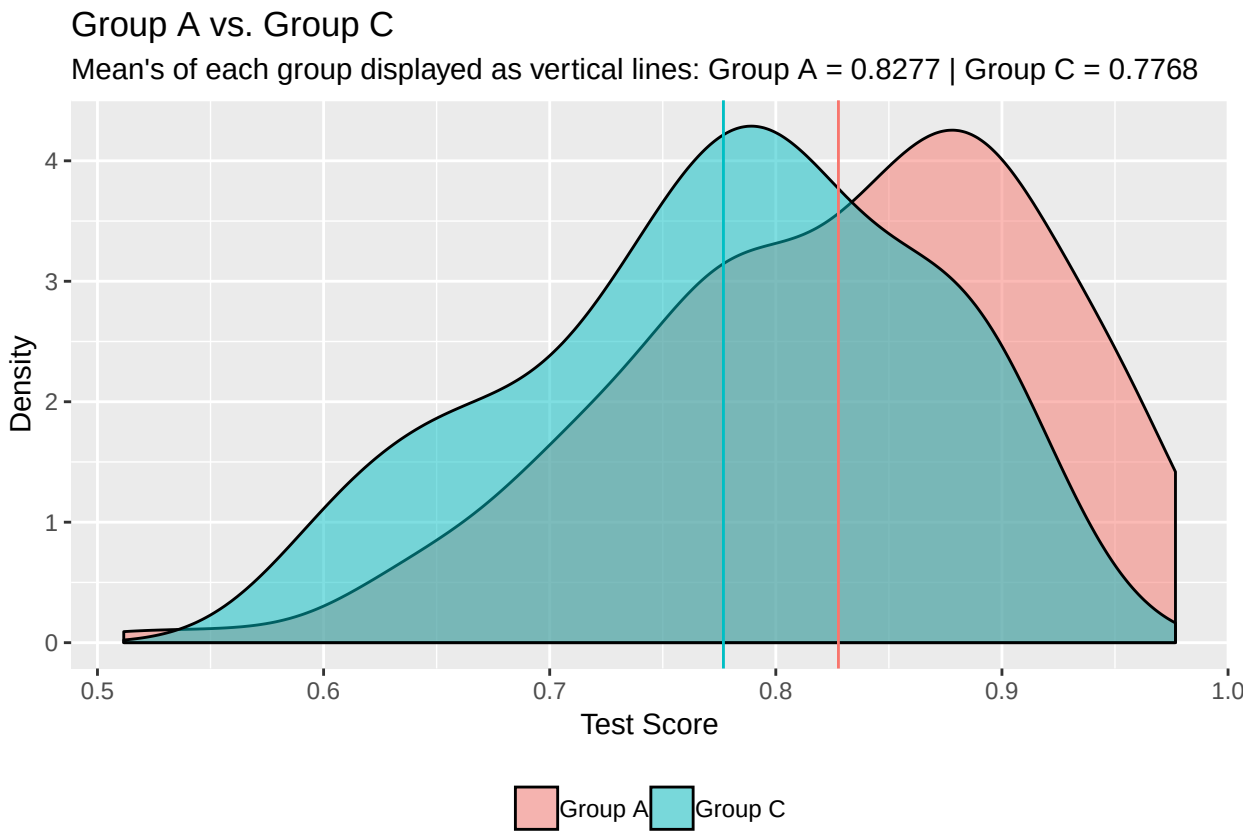



```
## [1] "Test 1 p value = 0.307749444337635"
```

Welch's two-tailed t-Test returns the above p value, with which we fail to reject the null hypothesis.

Test 2

- H_0 : Respondents in Group A do not have a higher average test score than respondents in Group C.
- H_a : Respondents in Group A have a higher average test score than respondents in Group C.

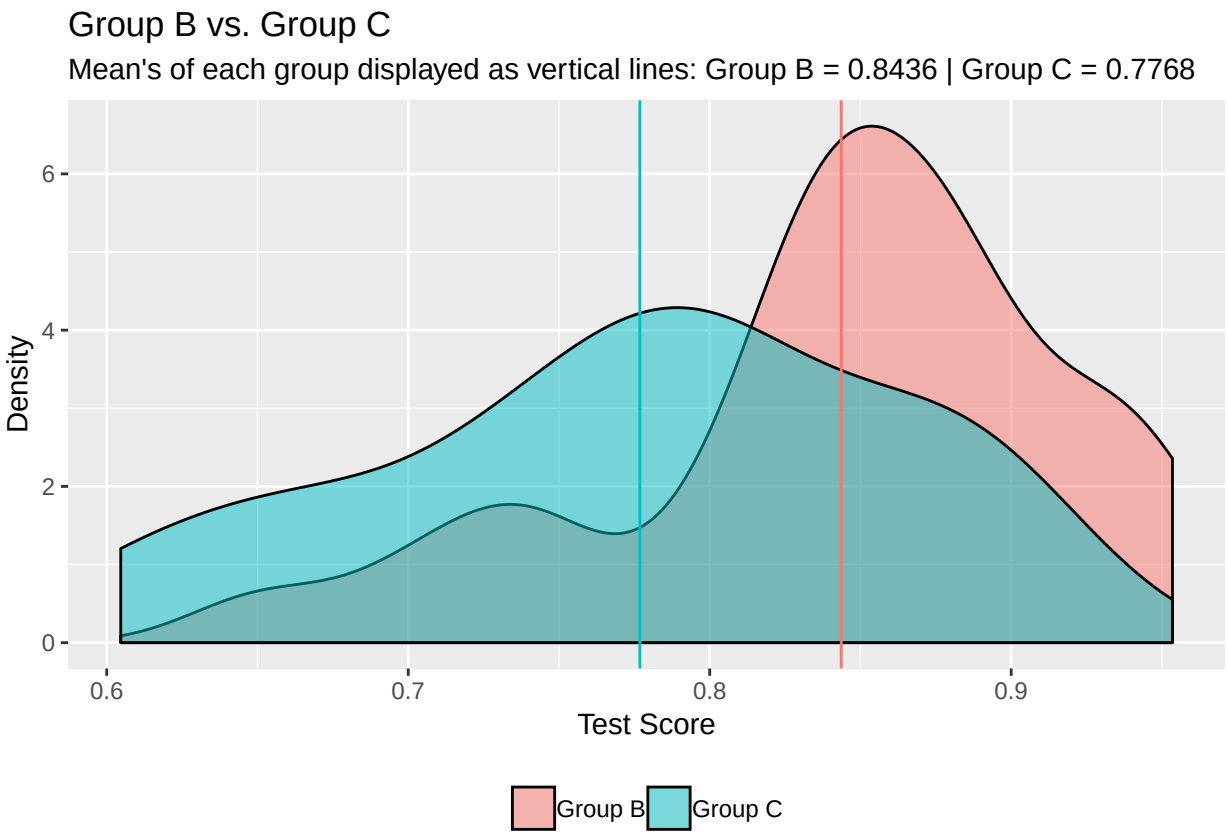



```
## [1] "Test 2 p value = 0.000187363981785434"
```

Welch's one-sided t-Test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, the null hypothesis is rejected in favor of the alternative.

Test 3

- H_0 : Respondents in Group B do not have a higher average test score than respondents in Group C.
- H_a : Respondents in Group B have a higher average test score than respondents in Group C.



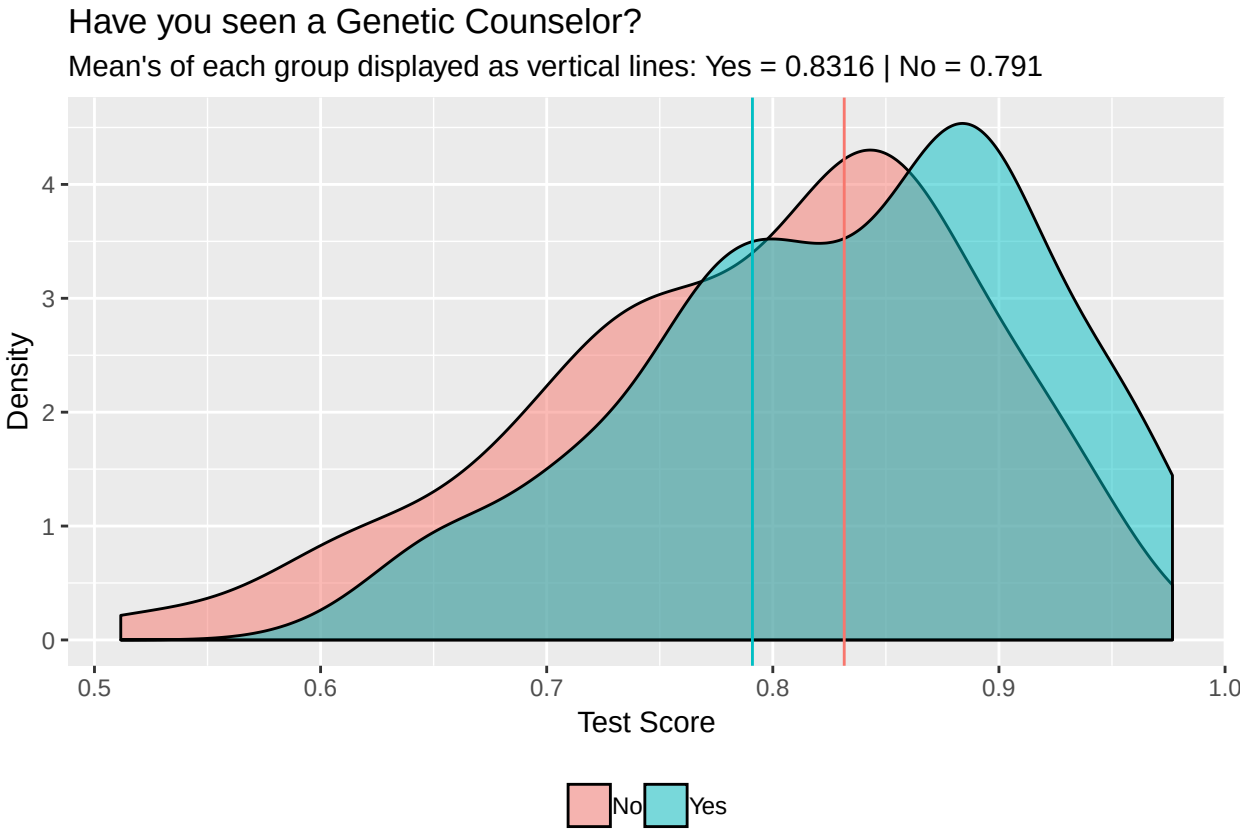

```
## [1] "Test 3 p value = 0.000280154187709805"
```

Welch's one-sided t-Test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, the null hypothesis is rejected in favor of the alternative.

Study Aim 2 - Does exposure to genetic counseling affect knowledge of NF1?

Test 4

- H_0 : Respondents who have seen a genetic counselor do not have a higher average test score than those that have not seen a genetic counselor.
- H_a : Respondents who have seen a genetic counselor have a higher average test score than those that have not seen a genetic counselor




```
## [1] "Test 4 p value = 0.000280154187709805"
```

Welch's one-sided t-Test returns the above p value. Being well below the Bonferroni corrected significance level of 0.0125, the null hypothesis is rejected in favor of the alternative.

Cronbach's Alpha

Using 43 quiz questions returned a Cronbach Alpha value of 0.7010484.

Test Limitations

As with any statistical test, there are limitations imposed by the methods I have chosen to use. Contrary to Student's t-test, Welch's t-test does not assume that the variances of the two samples are equal. However, the assumption that the data is Normally distributed is maintained.

The Shapiro-Wilk test is a statistical test used to determine whether that particular sample came from a population that is normally distributed. The null hypothesis of this test is that the sample came from a Normally distributed population, the alternative therefore being that the sample did not originate from a Normally distributed population. As shown in the table below, all samples have a "low" p value from the Shapiro-Wilk test, groups B and C being the only ones that are above the Bonferroni adjusted significance threshold.

Table 2: Shapiro-Wilk Test P-Values

	PValue
Group A	0.0000616
Group B	0.0623626
Group C	0.0390724
Seen Genetic Counselor	0.0000620
Not Seen Genetic Counselor	0.0034579

This, however, leads to another important topic, the problem of multiple comparisons. When one decides to correct for multiple comparisons, as I have done here using the Bonferroni Correction, the goal is to reduce the probability of a false positive. This practice *necessarily* increases the probability of a false negative. I adjusted the original significance level of 0.05 by accounting for the 4 statistical tests that addressed the study aims of this thesis. However, had I included the above 5 statistical tests that might fall under the heading of *Assumption Testing*, the adjusted significance level would fall to 0.0055556. Clearly, deciding where the bounds of a "family" of statistical tests begin and end has a large impact on the whether a particular test is deemed "significant," and, "there is no firm rule on this."[1]

The problem of multiple comparisons is an area of active research and, seeing as, "There is no universally accepted approach for dealing with the problem of multiple comparisons,"[1] deciding on the Bonferroni Correction was one of many possible solutions to the problem.

Demographic Analysis

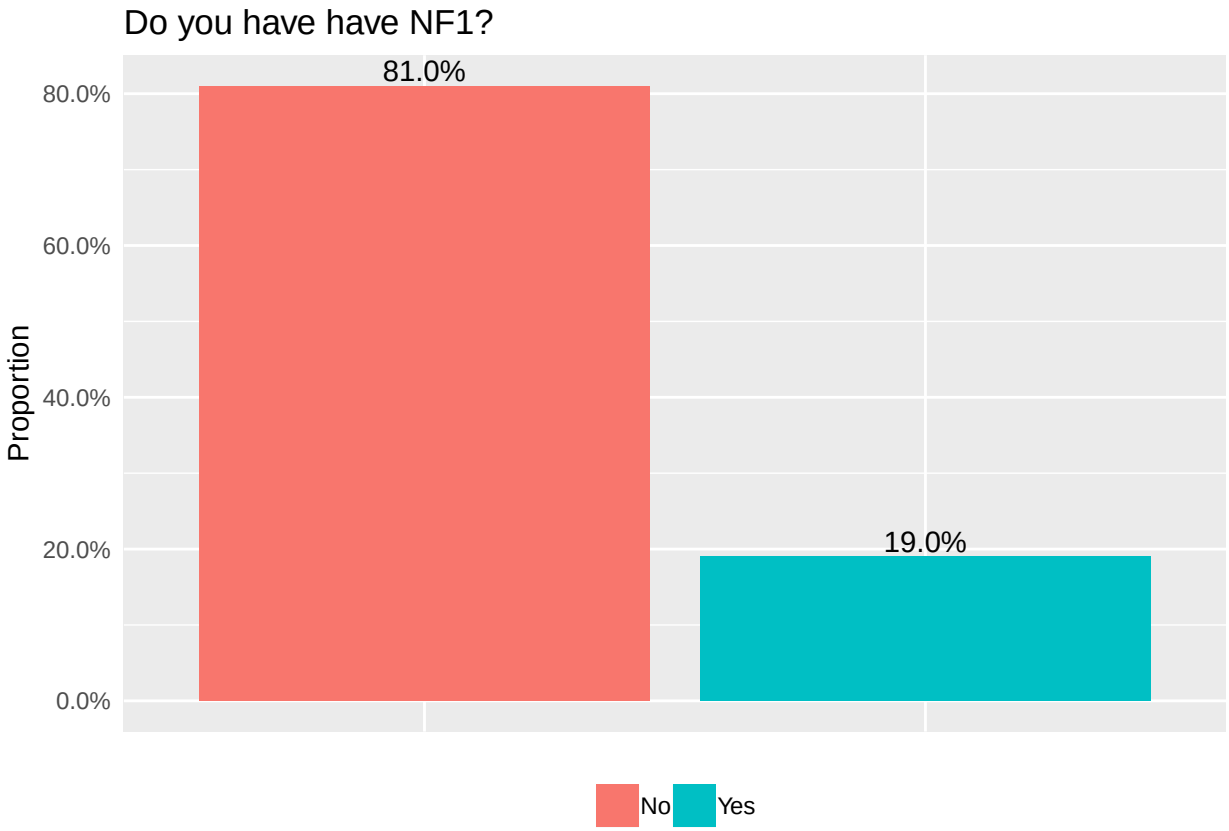


Table 3: Do you have NF1?

Response	Count
No	222
Yes	52

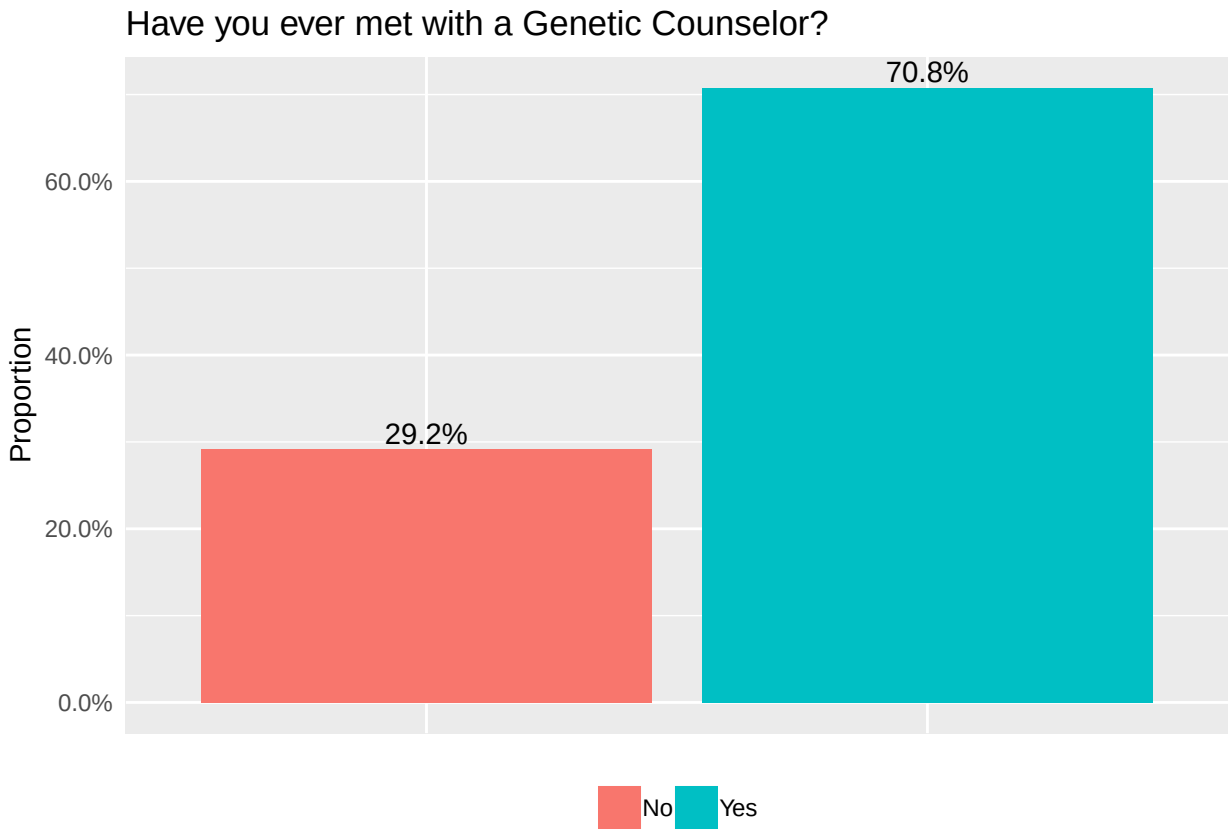


Table 4: Have you ever met with a Genetic Counselor?

Response	Count
No	80
Yes	194

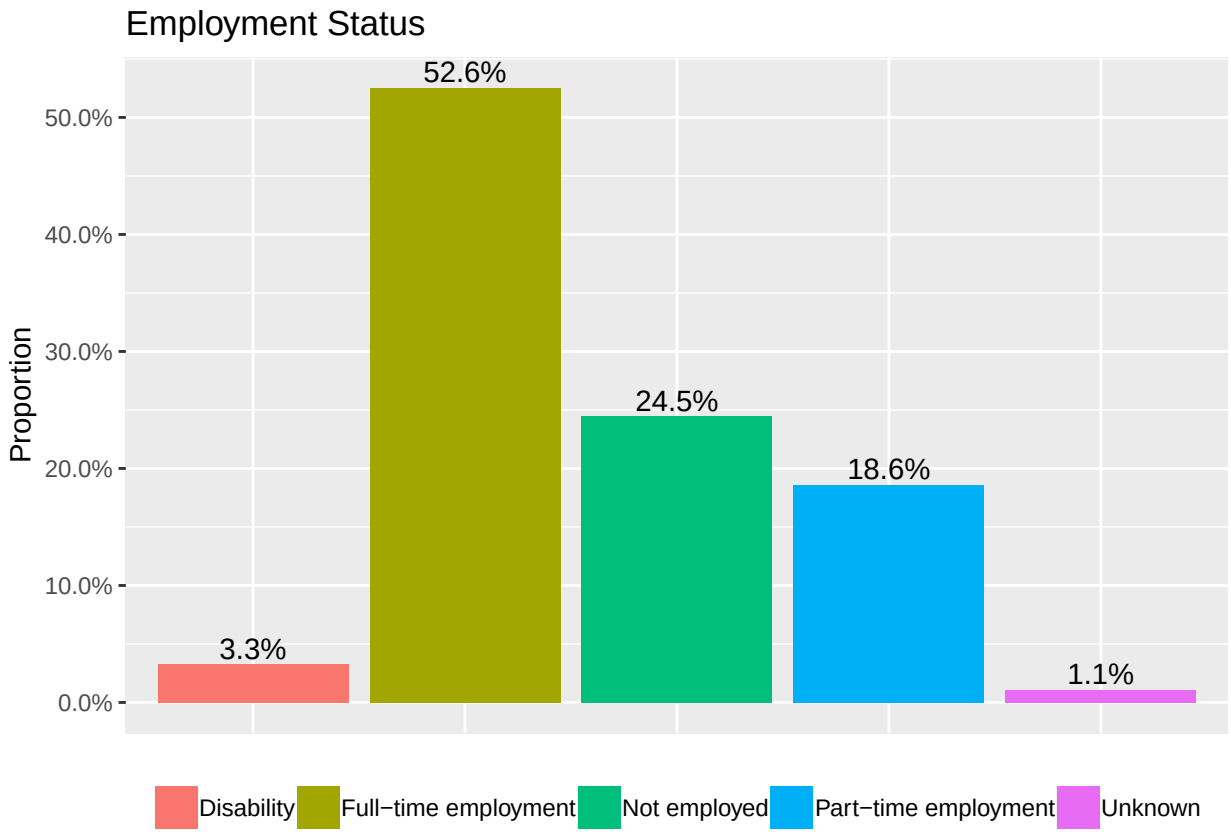


Table 5: What is your employment status?

Response	Count
Disability	9
Full-time employment	144
Not employed	67
Part-time employment	51
Unknown	3

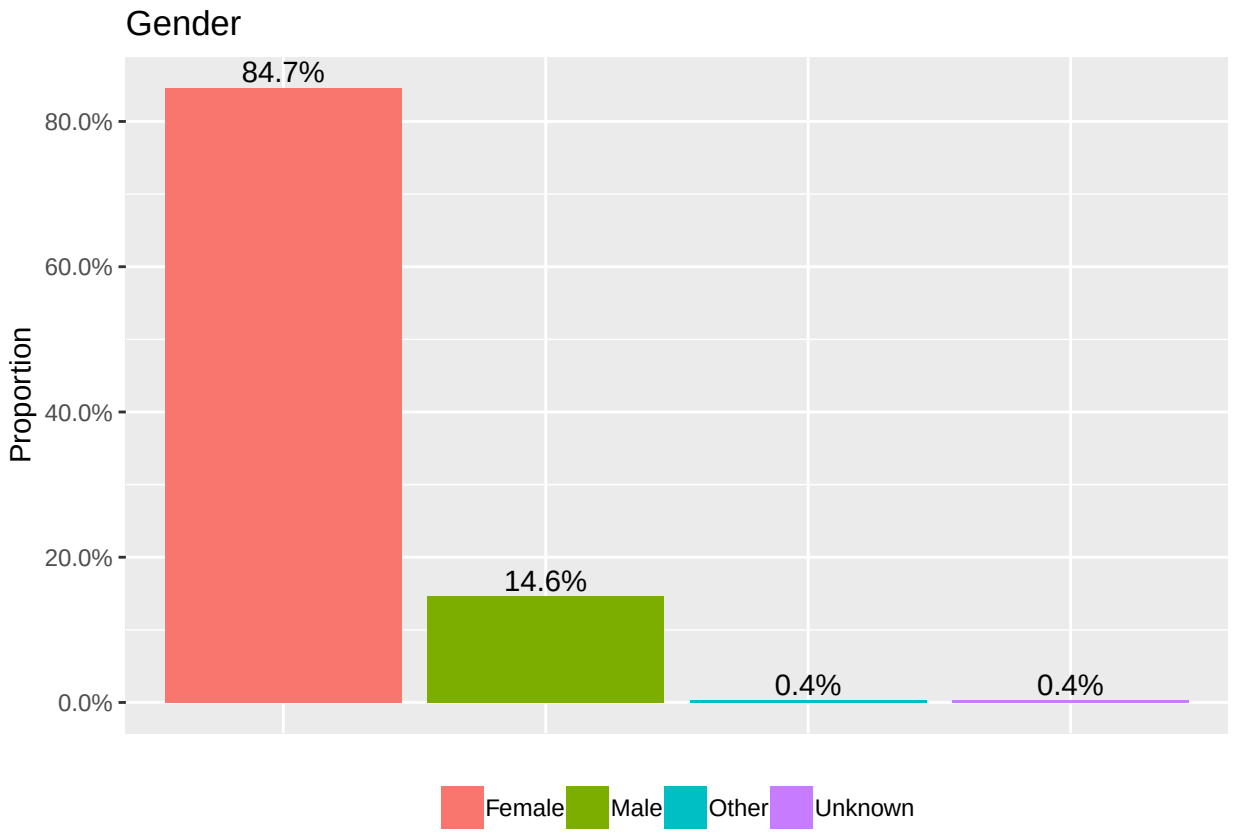


Table 6: Gender

Response	Count
Female	232
Male	40
Other	1
Unknown	1

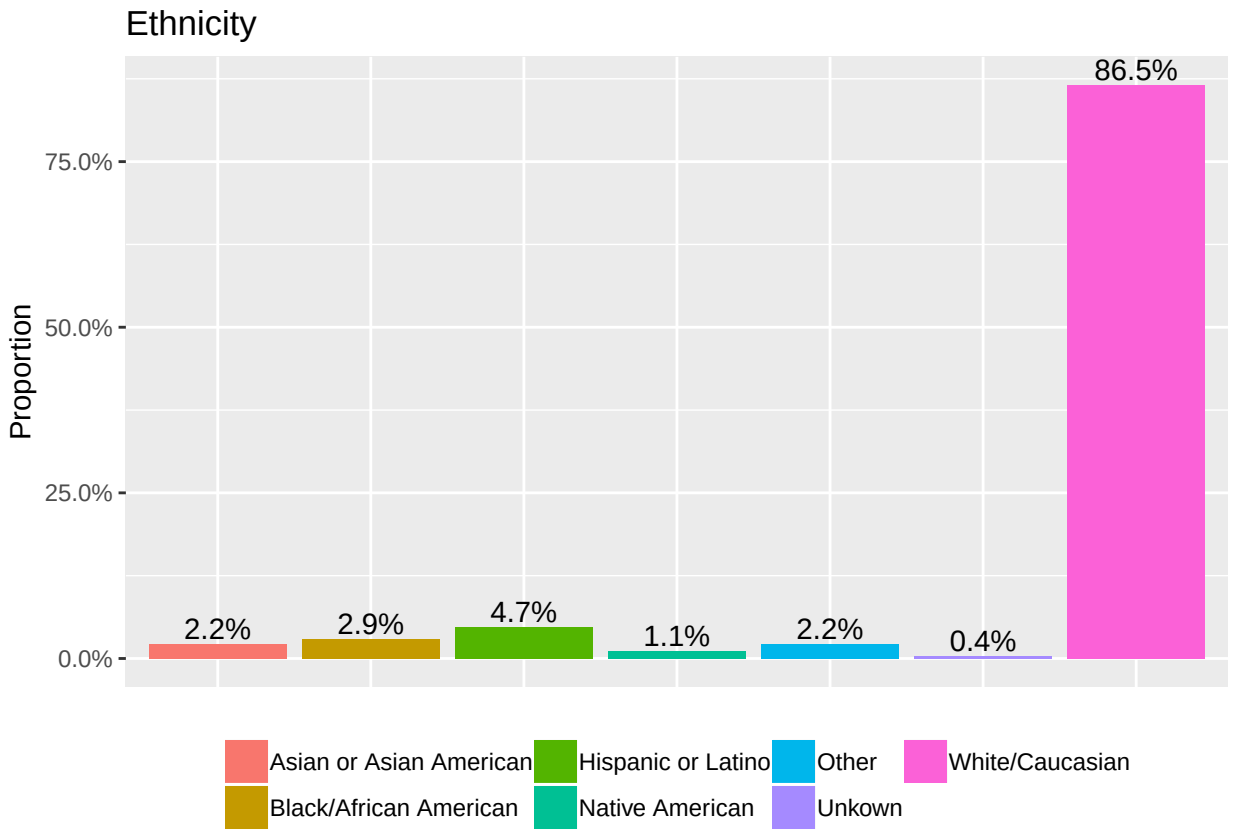


Table 7: Ethnicity

Response	Count
Asian or Asian American	6
Black/African American	8
Hispanic or Latino	13
Native American	3
Other	6
Unkown	1
White/Caucasian	237

One person did not answer the question regarding their educational background, and therefore was excluded from the below plot.

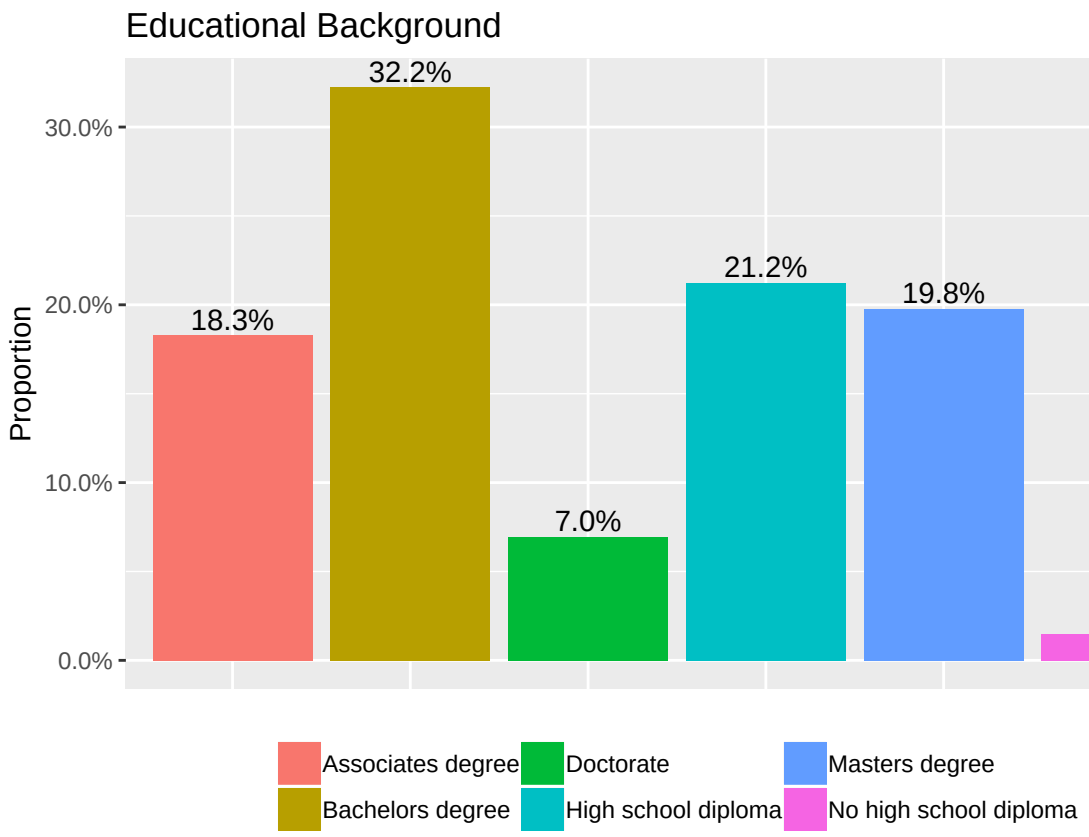
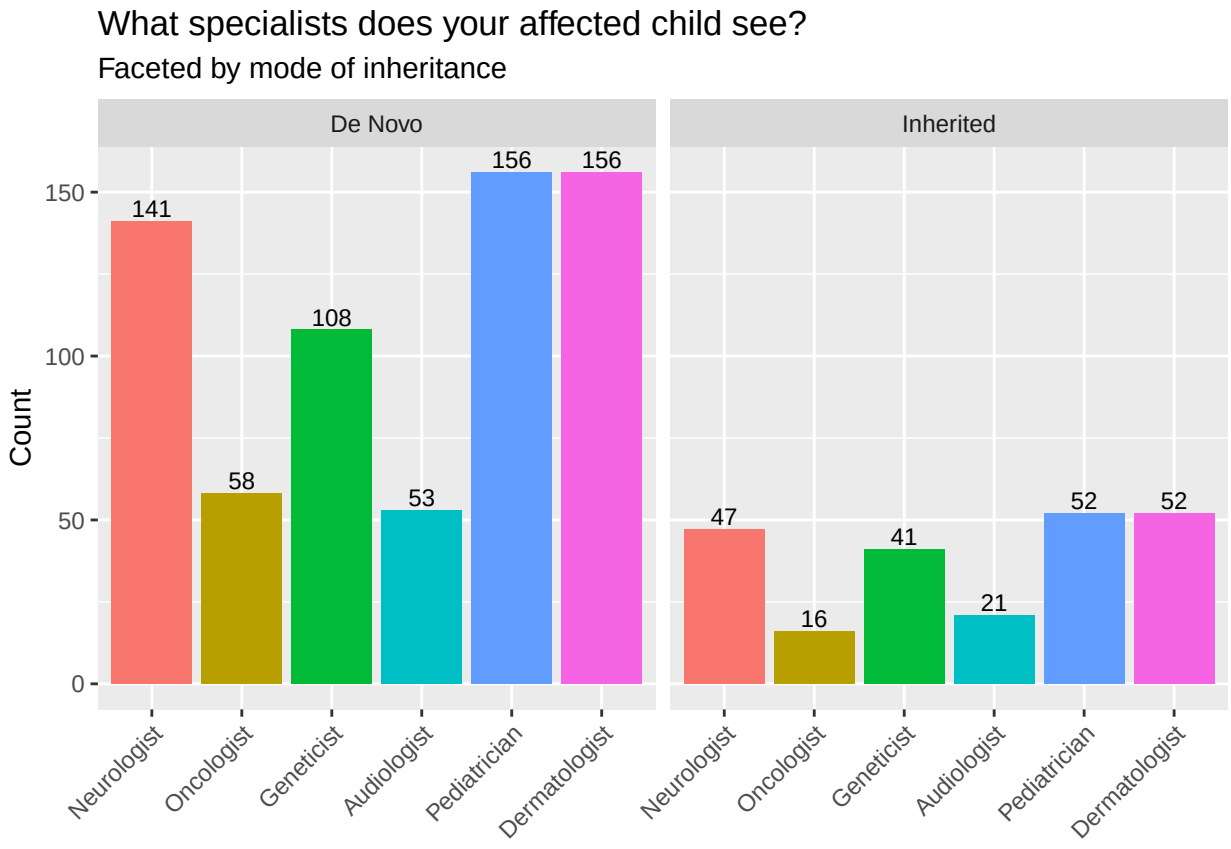


Table 8: Educational Background

Response	Count
Associates degree	50
Bachelors degree	88
Doctorate	19
High school diploma	58
Masters degree	54
No high school diploma	4

It is important to note that the answers to the below question, "What specialists does your affected child see?" are not mutually exclusive - children who see a Neurologist may have seen an Oncologist and a Geneticist as well.



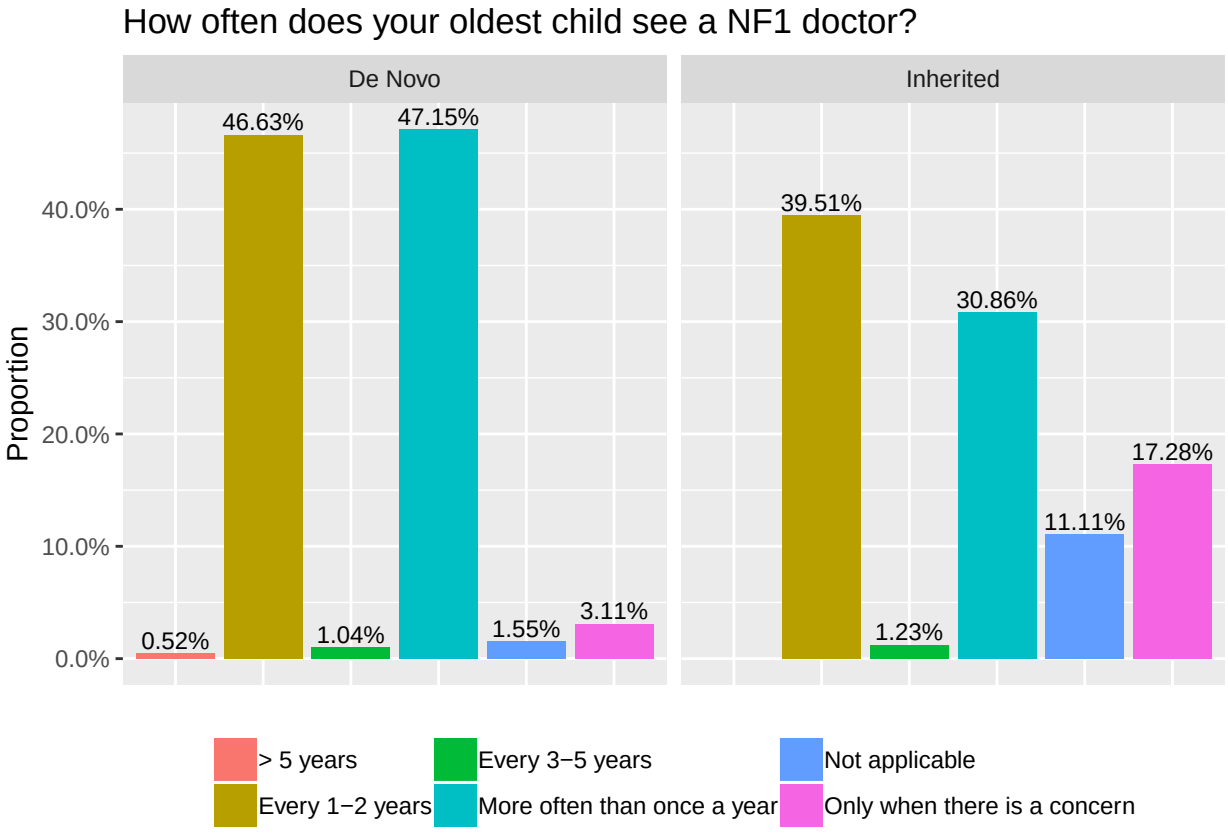
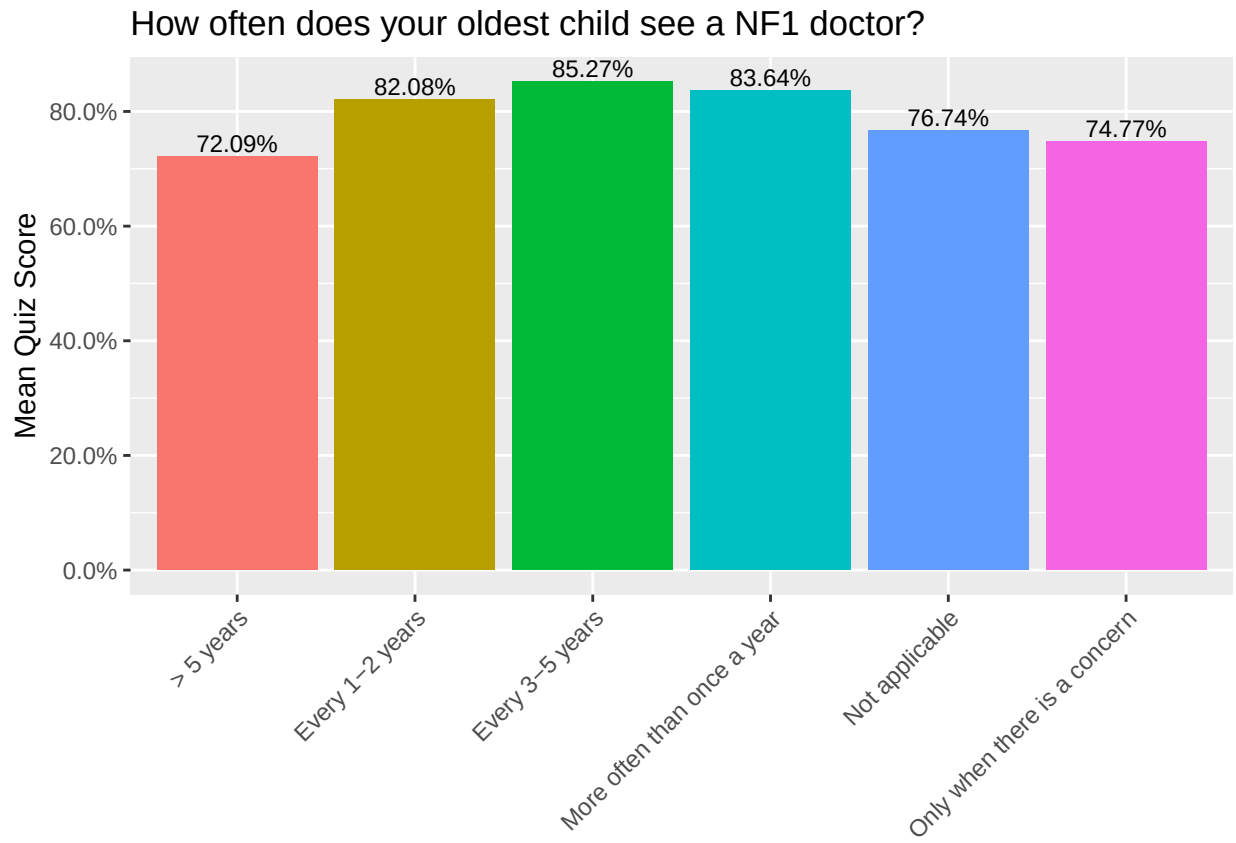
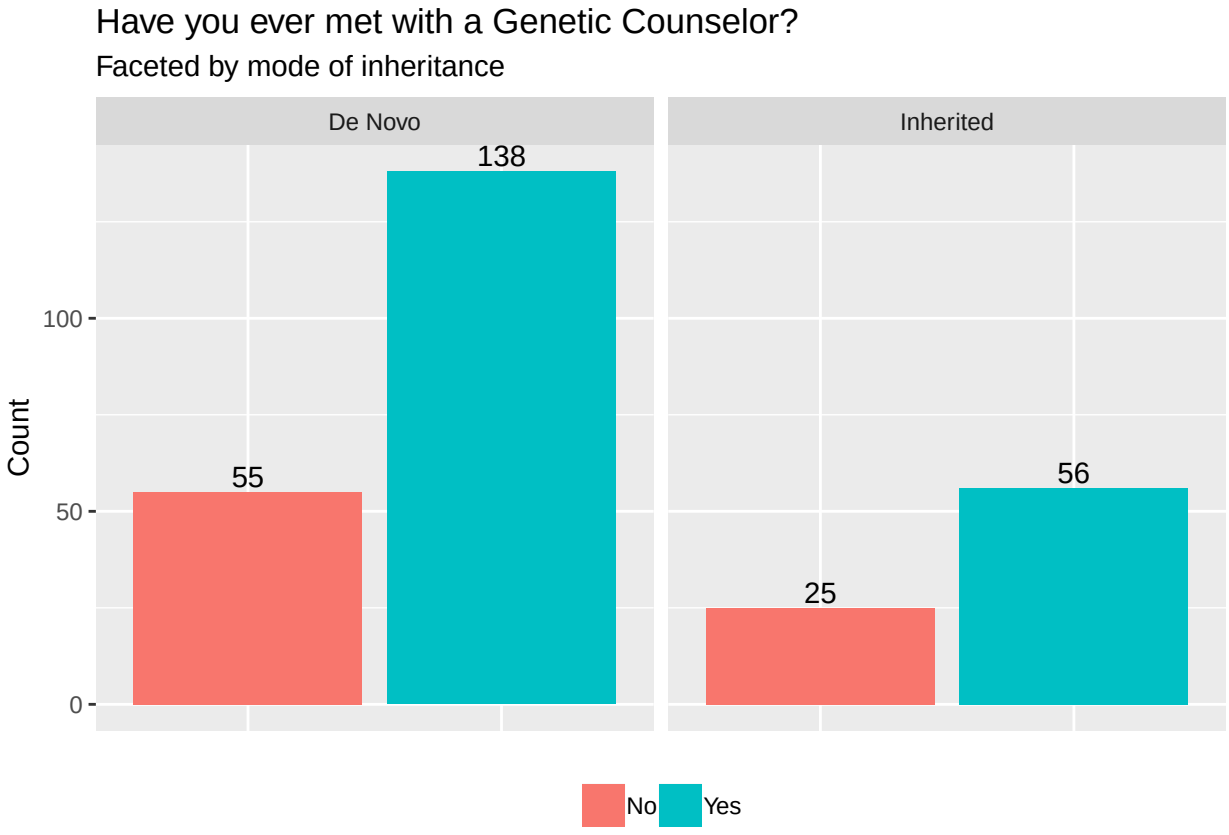


Table 9: How often does your oldest child see a NF1 doctor?

	De Novo	Inherited
> 5 years	1	0
Every 1-2 years	90	32
Every 3-5 years	2	1
More often than once a year	91	25
Not applicable	3	9
Only when there is a concern	6	14


How often does your oldest child see a NF1 doctor?
Mean Quiz Score
80.0%
60.0%
40.0%
20.0%
0.0%
72.09%
82.08%
85.27%
83.64%
76.74%
74.77%
> 5 years
Every 1–2 years
Every 3–5 years
More often than once a year
Not applicable
Only when there is a concern


Have you ever met with a Genetic Counselor?
Faceted by mode of inheritance
De Novo
Inherited
Count
100
50
0
55
138
25
56
No
Yes

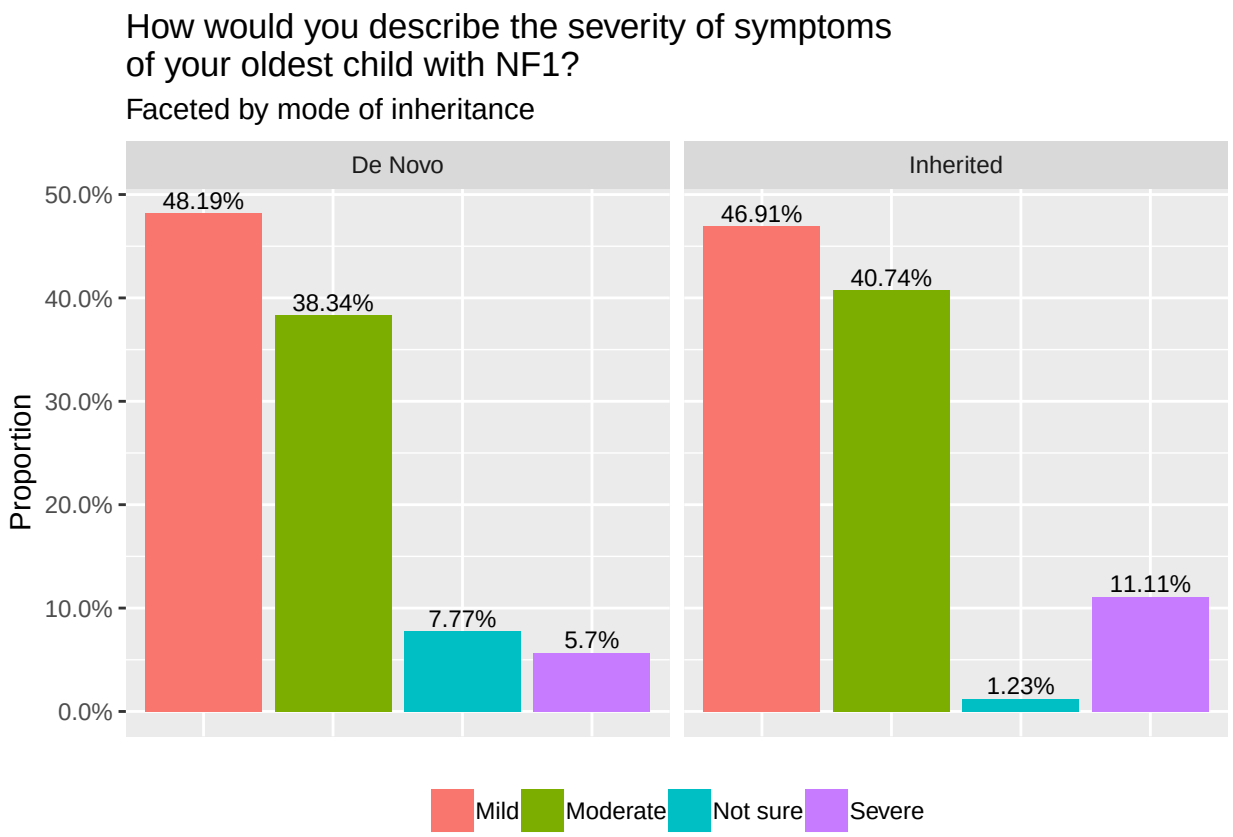
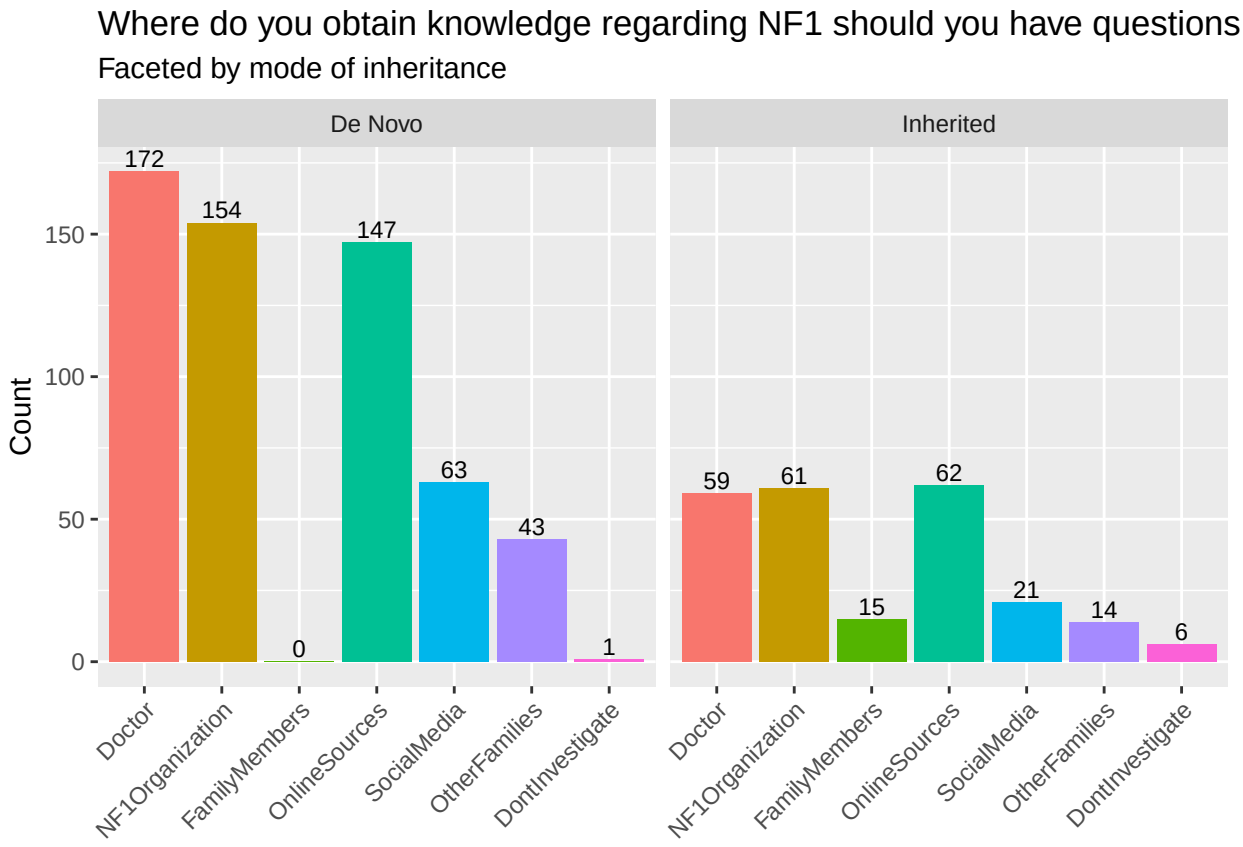
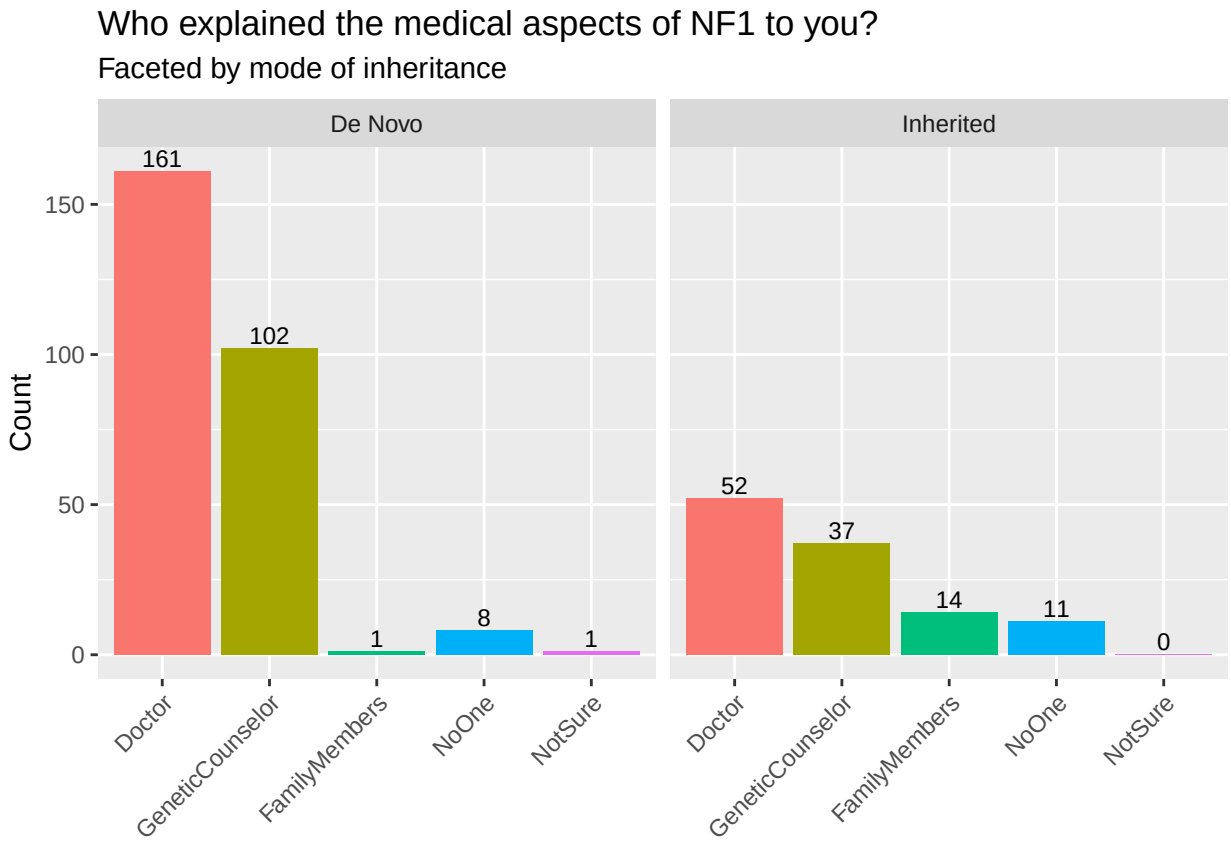


Table 10: How would you describe the severity of symptoms of your oldest child with NF1?

	De Novo	Inherited
Mild	93	38
Moderate	74	33
Not sure	15	1
Severe	11	9

Similar to one of the previous plots, it is important to note that the answers to the below question, "Where do you obtain knowledge regarding NF1 should you have questions?" are not mutually exclusive - parents who ask their doctor might also reference online sources.





References

[1] McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.