

Do People from Big Families Love Big Families?

A study on the relationship between sibling count and conceptions of ideal family size

Nura Hossainzadeh, Ayushi Goel, Ross Vrbanac, Richard Lumpi

Contents

1	Introduction	2
2	Data and Methodology	2
3	Model Specifications and Results	3
4	Discussion	5
5	Appendix	6
6	Data Exploration with Exploration Set - Not included in PDF Hand In	7
6.1	Playing around with Visualisation	7
6.2	Models	10
6.3	Selected Tests	19

1 Introduction

Modernity has often been described as an era of increased social isolation; small and intimate local communities have been unraveled by exodus to industrial cities; globalization and increased facility of travel have brought individuals to live their lives in unfamiliar lands; most recently, social media has often replaced real-life interaction even between people living in the same locality. Perhaps the most intimate of social ties, the family, too, has been affected by the degradation of social ties in the modern world; the nuclear family has become measurably smaller, and fertility rates are declining both in the US and on a global level. According to a study by The Lancet, by the year 2100, 97% of countries will have a fertility rate that is not adequate to sustain their population¹.

Much has been written on the relationship between family size and individual aptitudes and achievements; on the one hand, a focus on the career successes and academic aptitudes of individual children, and on the other, a focus on relationship success (in marriage, for example) and social aptitudes². However, little has been written on family size and the attitudes and perceptions of its individual members. In this study we explore the relationship between the number of siblings an individual has and that individual's perception of an ideal number of children for a family. While our study does not seek to trace a cause-and-effect explanatory relationship between family size and perceptions of an ideal family, we are interested in exploring the relationship between these two factors as a preliminary step to potential future studies. Our research investigates the following question:

How do people with different numbers of siblings perceive the ideal number of children for a family?

2 Data and Methodology

In our study, we conceptualize “number of siblings” as the number of genetically or non-genetically related individuals with whom a respondent shared a family environment. This includes step-siblings and adopted siblings, capturing the full social experience of growing up in a multi-sibling household. We conceptualize “ideal number of children” as the number of children a person believes a family should have, reflecting their views on overall family dynamics.

Our data was sourced from the General Social Survey (GSS), a nationally representative survey administered to adult Americans by NORC (formerly the National Opinion Research Center). NORC is an independent research institution funded by both private and public sponsors, with over half its staff being University of Chicago faculty members and administrators. The GSS, launched in 1972, continues to survey adult Americans annually on a broad range of demographic, behavioral, and attitudinal questions. To operationalize our concepts, we downloaded the 2022 survey and compared the responses given by each person on two chosen questions. Number of siblings was operationalized by examining responses to the question, “How many brothers and sisters did you have? Please count those born alive, but no longer living, as well as those alive now. Also include stepbrothers and stepsisters, and children adopted by your parents”. We believe this phrasing adequately captures the social experience of a multi-sibling household. Ideal number of children was operationalized by examining responses to the question, “What do you think is the ideal number of children for a family to have?” We believe this phrasing adequately captures the idea of emphasizing the family unit rather than an individual perspective.

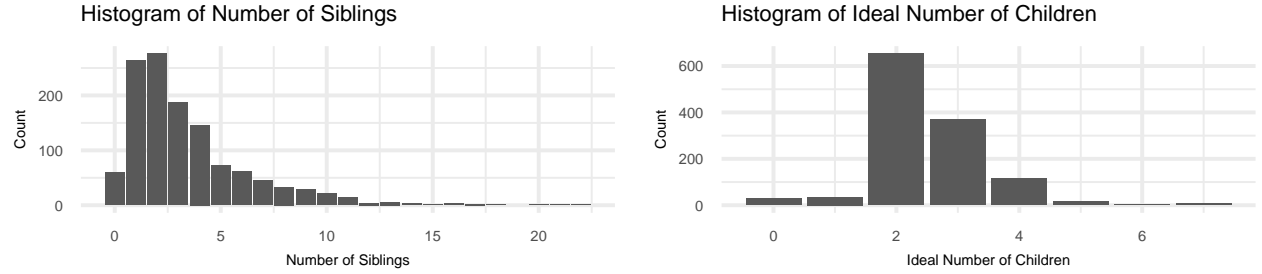
We filtered the data by removing rows where questions were inapplicable. Notably, the question about the ideal number of children was only posed in two of the three ballots (Ballot A and B) in the 2022 survey (2376 of 3544 respondents). We excluded responses indicated “as many as you want,” as we seek to measure the content of an ideal, rather than inquire whether the ideal exists (555 of the 2376 respondents). Responses

¹“The Lancet: Dramatic declines in global fertility rates set to transform global population patterns by 2100,” Accessed 15 July 2024. <https://www.healthdata.org/news-events/newsroom/news-releases/lancet-dramatic-declines-global-fertility-rates-set-transform>

²See, for example, Judith Blake and Douglas Downey on the “research dilution hypothesis,” which says that intellectual and educational outcomes are significantly worse for children in larger families because the family’s limited resources are spread thinly between children (Blake, Judith. Family size and achievement. Vol. 3. Univ of California Press, 2022.; Downey, Douglas B. “When bigger is not better: Family size, parental resources, and children’s educational performance.” American sociological review (1995): 746-761.).

with no or invalid answers were removed, leaving us with 1758 valid responses. The response for “7 or more” children was transformed to 7 to maintain the metric scale, though this may introduce a slight bias.

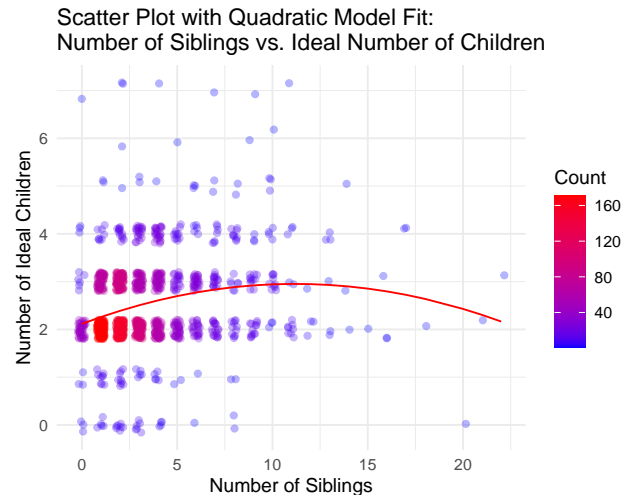
To analyze the relationship between the number of siblings and the ideal number of children, we used ordinary least squares (OLS) regression. The dataset was split into an exploration set (30%, $n = 527$) and a validation set (70%, $n = 1231$). The study’s graphs and analysis are based on the validation set. To run a linear regression model for both the exploration and the validation set, relying on large-sample properties, rests on two assumptions, namely that the data is independent and identically distributed (IID) and that a best linear predictor (BLP) exists and is unique. **IID:** GSS is based on probability multi-stage cluster sampling for the US. Large sample and sampling design guarantees that the data can be reasonably considered IID for the purpose of the study. **BLP:** A best linear predictor exists if the covariance between the output variable and each feature is finite. Histograms show that the number of siblings has a pronounced right skew (skewness = 1.9) with an indication of a long tail (maximum number of siblings in the dataset was 22), and the ideal number of children has a slight right skew (skewness = 0.87). The analysis assumes the population distribution of siblings is reasonably bounded to ensure that the variance-covariance matrix between the number of siblings and the ideal number of children is finite and a BLP exists. Univariate analysis and dropping defaults in categorical transformations of the independent variable ensures no perfect colinearity, supporting the assumption that the BLP is unique.



3 Model Specifications and Results

The visualization of the sample joint distribution in the scatterplot, together with the histograms, indicates that most respondents see two or three children as the ideal number of children (mean = 2.49). Visual inspection does not show a clear pattern, or structural break in the data, even though a slight indication of a parabolic shape emerges. Furthermore, most observations in the sample are drawn in a narrow space, introducing limited variability in the number of siblings (mean = 3.44, sample standard deviation = 2.94), making the analysis susceptible to noise from the outlier values.

Model Estimates for Number of Ideal Children		
	Dependent variable:	
	Base (1)	Quadratic (2)
Number of Siblings	0.066*** (0.012)	0.150*** (0.024)
Number of Siblings Squared		-0.007*** (0.002)
Constant	2.266*** (0.042)	2.115*** (0.052)
Observations	1,231	1,231
R ²	0.044	0.058
Adjusted R ²	0.043	0.056
Residual Std. Error	0.914 (df = 1229)	0.908 (df = 1228)
F Statistic	56.023*** (df = 1; 1229)	37.550*** (df = 2; 1228)
Note:	*p<0.1; **p<0.05; ***p<0.01 Model coefficients with robust standard errors.	



To begin the investigation into whether a statistically significant relationship exists between the number of siblings and number of ideal children, we estimate the “Base” model:

Base Model: $NumberOfIdealChildren = \beta_0 + \beta_1 NumberOfSiblings + e$

Visual inspection of the error terms (see Appendix) further suggests a parabolic relationship. We add a quadratic term to better capture the data generating process (susceptible to sample variability driven by outliers), referred to as the “Quadratic” model (plotted in the scatterplot):

Quadratic Model: $NumberOfIdealChildren = \delta_0 + \delta_1 NumberOfSiblings + \delta_2 NumberOfSiblings^2 + e$

The regression table presents the coefficient estimates for both models. The Base Model exhibits a positive, statistically significant relationship. A t-test on the coefficient β_1 rejects the null hypothesis ($\beta_1 = 0$) with a p-value of less than 0.01. A heteroskedasticity robust Wald test rejects the null hypothesis of the Base model in favor of the Quadratic model (F-statistic: 12.52 with a p-value < 0.01). For the quadratic model, the ideal number of children increases up to around eleven siblings (inflection point at 10.71), after which the relationship declines. Using, the Quadratic model, we found that adjusting the number of siblings—for example, from one to two—increases the number of ideal children from 2.41 to 2.54. The model estimates an ideal number of closer to three children until its inflection point at eleven siblings, after which it begins to decline. Both models poorly explain the variability, indicated by low R^2 values. To further investigate the descriptive potential of a linear model, we estimated a more flexible model by converting each number of siblings into a dummy variable, referred to as the “Indicator” model.

Indicator Model: $NumberOfIdealChildren = \lambda_0 + \sum_{siblings} \lambda_i NumberOfSiblings + e$

The explanatory ability of the Indicator model only slightly increases, ($R^2 = 0.08$). Again, we can observe an upward trend, with each indicator for more than 3 siblings up to the 10th sibling significant with a p-value < 0.01. Beyond the 10th sibling, we observe a mix of significant and insignificant dummy coefficients, both positive and negative, supporting our reservations to draw inference from small numbers of observations. The results are consistent across both the exploration and validation set, apart from the F-test, which in the exploration set did not reject the null. A challenge in achieving strong model fit is likely a result of the ideal number of children being a discrete variable with large relative gaps between values.



4 Discussion

Our analysis indicates that there is a statistically significant, positive relationship between siblings and the ideal number of children. Both the Base and Quadratic model capture a range between roughly 2 and 3 for respondents growing up with zero to ten children, beyond which can reasonably be considered outliers in contemporary US families ($n = 37$). The concentration of observations makes it challenging to draw inference for people from large families. In practical terms, where having children is a discrete rather than a continuous matter, both models indicate a positive relationship with rather limited practical significance, indicating more of a leaning towards 3 rather than 2 children the more siblings a respondent grew up with. Even the highly flexible specification with dummy variables for each number of siblings explains very little of the variation in the outcome variable despite a large number of coefficients.

Ultimately, our study suggests that there is no practical pattern between siblings and the ideal number of children, and that the ideal family size in the US centers around two to three children. The actual fertility rate in the US sits at an estimated 1.62 per woman³, which suggests that whether it be cultural or monetary reasons, US families may be having fewer children than they ideally prefer. A 2023 Gallup study on Americans' ideal family size drew a similar inference, noting that while participants preferred an average family size of 2.7 children, declining birth rates indicate that "Americans' views of the ideal may not be their personal reality"⁴. Future research should focus on other variables describing the variability in the ideal number of children, rather than solely the number of siblings a person has.

³Hamilton, Brady, et al. Vital Statistics Rapid Release Births: Provisional Data for 2023. 2024.

⁴Inc, Gallup. "Americans' Preference for Larger Families Highest since 1971." Gallup.com, 25 Sept. 2023, news.gallup.com/poll/511238/americans-preference-larger-families-highest-1971.aspx.

5 Appendix

Link to Data Source

- GSS Data Explorer: <https://gssdataexplorer.norc.umd.edu/variables/vfilter>
- Variables of interest: chldidel and sibs

Model Specifications

We have tried the following specifications which were discussed in the main report

- Mean Model: Not straight forward from the plots if there is a relationship at all - first just look at the mean.
- Base Model: Investigate whether there is evidence for a linear relationship, we found a parabola in the error terms, motivated the next model.
- Quadratic Model: Better capture sample pattern with a quadratic relationship, errors are now centered around zero - see plot below.
- Indicator Model: Find potential structural breaks in the data, findings were used to set up the large family model. First significant term was sibling number 4 in the evaluation set.

To investigate how well simpler models with structural breaks and trend interactions could approach the Indicator Model we tried the following:

- Siblings vs No Siblings Model: A binary dummy variable indicating if a respondent had a sibling or not. Poor explanatory performance.
- Siblings vs No Siblings Model with a trend for Siblings: Adding a trend term to the model made the coefficient on the dummy insignificant.
- Large Family Model: A binary dummy variable for having more than three siblings, indication that bigger families lean more towards 3 rather than 2.
- Large Family Model with Trend interaction Term: Trend for siblings with an interaction term to have structural break and change in slope, all coefficients individually insignificant.

We also tried higher polynomial models to capture more variability. They only marginally but not meaningfully increased R^2 ($R^2 = 0.053$ for polynomial of degree 5), and not useful in reasoning about the relationship. Logarithmic transformations have been omitted as both the dependent and independent variables have meaningful 0 values.

Residual Plots

