# Pithy Title

Jared Cummings

November 2021

**Abstract**

Abstract will go here.

## Contents

# Introduction

In the middle of the 20th century there was a global concern about overpopulation. With the benefit of hindsight, the doomsday prophecies about impending mass famine that would blight humanity nearly to extinction were overblown. Neverless, these concerns sparked a non-negligable degree of scientific inquiry into fertility and family planning, especially in the developing world where population growth rates have been highest. Of the antinatalist proponents, there were two camps that conceived of different solutions. The first posited that the birth rate would decline with expanded access to contraceptives and was bolstered by surveys suggesting that there was an unmet demand for birth control in the developing world. The second asserted that socioeconomic development was the ultimate solution. Those in this second camp explained that as a woman's working hours became more lucrative, the extra hands provided by a large family would be increasingly less necessary to make ends meet. Both positions were based off of tenuous research, and the World Fertility Survey (WFS) was launched to increase the scientific standards of the debate. The broader impact of the survey is discussed in Cleland and Verma (1989).

From the survey results, the claim that policy aimed to meet the demand for birth control was the optimal solution would be bolstered by findings suggesting that the women in developing countries generally don't use birth control even when they do not want more children. Alternatively, the claim that socioeconomic development was the underlying driver would be supported by findings suggesting a link between interest in having children and measures of economic advancement, such as education. Using survey results of 1,607 married and fecund Fijian women, this report aims to evaluate the merits of the two claims using the Generalized Linear Model (GLM) framework. Conclusions will be evaluated at the default scientific standard, the 0.05 level of significance. Although in the decades since this data was gathered policy debates have turned away from promoting restrictive family planning to promoting higher fertility, this study still provides an interesting window into the applicability of data in decision making.

# Data

Little and Pullum (1979) reported some of the summarized responses of the Fijian stage of the WFS, the Fiji Fertility Survey, run in 1974. Because there are nontrivial behavioral disparities between ethnic groups in Fiji, these responses are specific to Indian women living in the country. As such, conclusions drawn from the data should be understood in this specific context. Of interest to this analysis is the part of the survey where the women reported ever having used any efficient contraceptive method. These results are summarized by current age of the woman, her desire for more children, and her level of education. The data, *which is linked in the appendix*, is summarized in Figure 1.

# Methodology

When not summarized into factor-level proportions, this data is reported as binary indicators for each individual for variables representing education, contraceptive use, and desire for more children. The age is summarized into four distinct categories. The data in its totality is then summarized into a 2x2x2x4 table of counts. Using the contraceptive use status of the respondents as the dependent variable, the contingency table can best be understood as a 2x2x4 table of binomial random variables. A regression analysis on this data should respect this structure. Using the GLM framework, the binomial success probability can appropriately be modeled using a link function, $g(p_{ijk})$. Formally, the model can be defined

$$\begin{aligned} \texttt{contraceptives\_used}_{ijk} &\overset{ind.}{\sim} \mathcal{B}(\texttt{contraceptives\_used}_{ijk} + \texttt{contraceptives\_not\_used}_{ijk},\ p_{ijk}) \quad (1) \\ g(p_{ijk}) &= X\beta \end{aligned}$$

where $\mathcal{B}$ denotes the binomial distribution and $X\beta$ is a linear regression function of $\texttt{education}_i$, $\texttt{desire}_j$, and $\texttt{age}_k$ to be defined later.

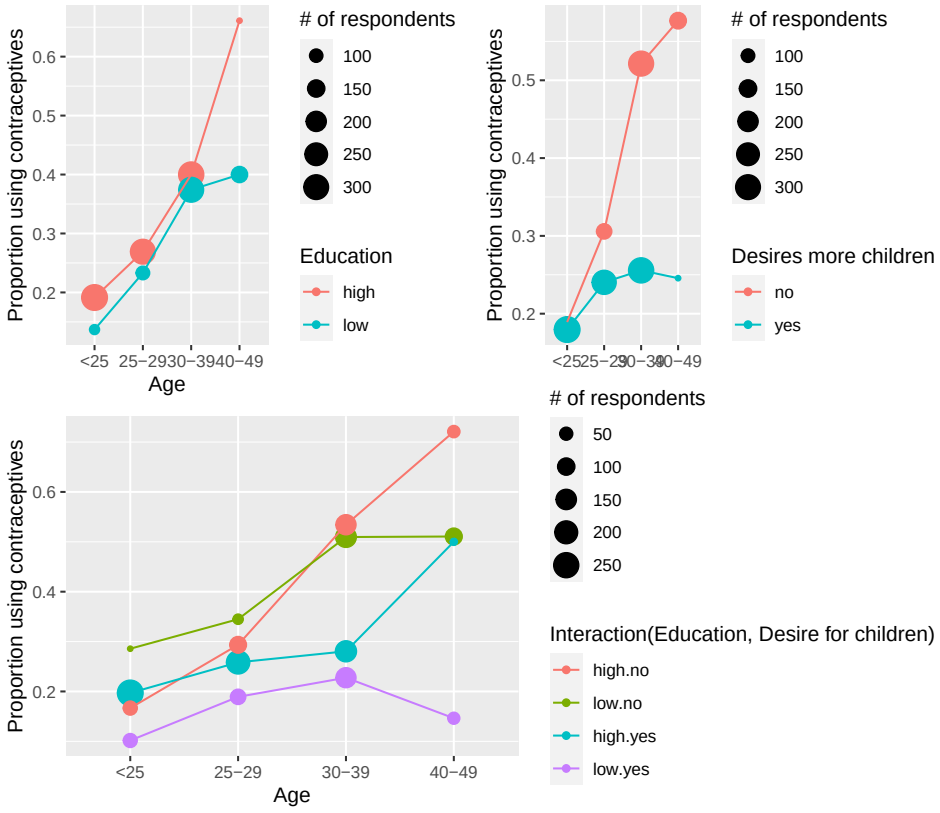


Figure 1: Marginal and joint visualizations of partial results of the Fiji Fertility Survey. (Top left) Proportion who have used an efficient contraceptive increases with age for women in both eduacation levels. The increasing trends noted are expected, older women having had more life in which to use an efficient contraceptive method. Sample size disparities are evident, which reflects the development of the Fijian educational system, which was not fully structured in the 1930s during the childhoods of the oldest women in the survey. (Top right) Proportion who have used an efficient contraceptives strongly increases with age among women who do not want more children. While an increase is aparent among women who do want more children, it noticably less and possibly explained by an age effect. Small sample sizes can again be seen. (Bottom) Propotion who have used an efficient contraceptive appears to increase with age in all joint categorizations of education and desire for children. There is a vissible change in behavior in the oldest age bucket.

This logistic regression model defined in (1) makes fairly minimal assumptions about the data. Firstly, the data is assumed to follow the stated binomial distribution. Secondly, the observed counts are assumed to be independent. Thirdly, monotonicity is assumed between the binomial probabilities and the regression equation. Some would include a fourth assumption here, that the cell counts in the contingency table being modeled are sufficiently large, though this is assumption only concerns itself with assessing statistical significance and not the propriety of fitting the model.

# Results

Through the process of manual backward selection, beginning at a saturated model and walking backward, the linear regression component of the binomial GLM chosen to model this data relationship was

$$\begin{aligned} \log \frac{p_{ijk}}{1-p_{ijk}} &= \beta_{1,1} \cdot \mathbf{1}(\texttt{age}_k =' < 25') + \beta_{1,2} \cdot \mathbf{1}(\texttt{age}_k =' 25-29') + \beta_{1,3} \cdot \mathbf{1}(\texttt{age}_k =' 30-39') + \\ &\quad \beta_{1,4} \cdot \mathbf{1}(\texttt{age}_k =' 40-49') + \beta_2 \cdot \mathbf{1}(\texttt{desire}_j = \text{yes}) + \beta_3 \cdot \mathbf{1}(\texttt{education}_i = \text{low}) + \\ &\quad \alpha_1 \cdot \mathbf{1}(\texttt{age}_k =' 25-29' \wedge \texttt{desire}_j = \text{yes}) + \alpha_2 \cdot \mathbf{1}(\texttt{age}_k =' 30-39' \wedge \texttt{desire}_j = \text{yes}) + \\ &\quad \alpha_3 \cdot \mathbf{1}(\texttt{age}_k =' 40-49' \wedge \texttt{desire}_j = \text{yes}) \end{aligned} \tag{2}$$

where $\beta$ coefficients are first-order effects of the variables and the $\alpha$ coefficients are first order interactions. The logistic function was chosen over a Probit link for its odds ratio interpretation. The data as seen in Figure 1 did not suggest evidence of skewness that might be better modeled with the log-log or complementary log-log links. Table 1 reports the estimated coefficients, their standard errors, and the associated z scores and p-values for the model defined in (2). Significance is determined at the 0.05 level. `R` code needed to reproduce the model is linked in the appendix.

Table 1: Parameter estimates, standard errors, z scores and p-values from the model defined in (2). Significance ('Sig.') is determined at a 0.05 cutoff. The increasing contraceptive usage with increasing age noted in Figure 1 is seen in the estimates of the beta 1,* coefficients. A desire for more children and lower education are both associated with lower contraceptive usage, though the effect of desire for more children is most strongly felt in conjuction with the higher two age buckets.

| | Coeficient | Estimate | Std. Error | z score | Pr(>\|z\|) | Sig. |
|---|---|---|---|---|---|---|
| beta 1,1 | Age < 25 | -1.396 | 0.298 | -4.688 | 0.000 | Y |
| beta 1,2 | Age 25-29 | -0.743 | 0.200 | -3.721 | 0.000 | Y |
| beta 1,3 | Age 30-39 | 0.263 | 0.133 | 1.981 | 0.048 | Y |
| beta 1,4 | Age 40-49 | 0.545 | 0.194 | 2.802 | 0.005 | Y |
| beta 2 | Wants more kids | -0.066 | 0.331 | -0.200 | 0.841 | N |
| beta 3 | Low education | -0.341 | 0.126 | -2.709 | 0.007 | Y |
| alpha 1 | Age 25-29 : Wants more kids | -0.259 | 0.410 | -0.633 | 0.527 | N |
| alpha 2 | Age 30-39 : Wants more kids | -1.113 | 0.374 | -2.975 | 0.003 | Y |
| alpha 3 | Age 40-49 : Wants more kids | -1.362 | 0.484 | -2.811 | 0.005 | Y |

## Model fit

The model is a strong fit to the data. The Pearson $\chi^2$ statistic was found to be 8.94 and the Neyman $\chi^2$ statistic was found to be 7.83, both with 7 degrees of freedom. Both statistics support the conclusion that the model in (2) is not significantly different than the saturated model with p-values 0.257 and 0.348 respectively. Since both statistics support the same conclusion and are relatively similar in value, the cell

counts in the contingency table being modeled are sufficiently large for statistical significance of the model and parameter estimates to be considered valid.

The assumptions of the model all appear to hold. The data being modeled is indeed binomial counts. The survey design and execution does not prompt doubt about the independence assumption. The model in (2) does not contain any estimates of slope, only intercepts for various combinations of the factors in question, so the mononicity assumption is also met by default.

### Discussion

Recall from the introduction that there were two camps interested in using the results of this survey to suggest policy for the governments of developing countries. The goal of both ideologies was to reduce family sizes and was motivated by a now defunct fear of overpopulation. One group claimed that policy aimed to meet the demand for birth control was the optimal solution. This group's stance would be bolstered by findings suggesting that the women in developing countries generally don't use birth control more readily even when they do not want more children. The second group claimed that socioeconomic development was the underlying driver of lower fertility rates. This perspective would be supported by findings suggesting a link between interest in having children and measures of economic advancement, such as education.

The first group will not be totally disappointed. Younger women do not have a statistically lower history of contraceptive use when partitioned by desire for more children, controlling for education level. A woman younger than 25 who wants more children is just as likely to have a history of contraceptive use as a woman younger than 25 with the same level of education who does not want more children (p-value = 0.841). The same is true of women in the 25-29 age group (p-value = 0.527). However, this behavior does reverse itself among women in the 30-39 and 40-49 age groups. Women in the 30-39 age group who do not want more children are three times more likely to have a history of contraceptive use than an equally educated comparison (p-value = 0.003). The multiplier for women in the 40-49 age group is an even greater value of 3.9 (p-value = 0.005). This together suggests that younger women have children at the same rate regardless of preference, and that the increased access to contraceptives that corresponds to older age likely drives down the fertility rate. Policy recommending increased access to contraceptives earlier in life could be pitched as a way to decrease the number of children being born to each woman.

This is not the only way to interpret these findings. Could it be that younger women who do not want more children are not as financially established as older women who do not want more children? Could the relatively lower socioeconomic status be the causal component of the different contraceptive usage behaviors? The data does not allow for insight into this question, and the second group could contest the first groups interpretation of the results on this basis. Something that the second group can assert that the first group cannot reject is that education, which proxies to some degree as a measure of socioeconomic comfort, is a significant differentiator of contraceptive use among the survey respondents. Controlling for age and desire for children, a woman with more education is 40.6% more likely to have a history of contraceptive use (p-value = 0.007). The decline in birth rates that has been witnessed over the decades since the WFS was conducted in countries that have a larger educated professional class of women has an analog that was measurable 1974 in Fiji. The second group has history and statistics on their side.

## Conclusion

The aim of this analysis was to determine the weight of evidence, if any, that the WFS provided to the arguments of the two camps of policy promoters previously described. This has been achieved. The insights in support of the "increase access" policy recommendations have been shown to be arguable, but tenuous in the absence of follow-up work. Additional research should be conducted into the potential socioeconomic shift that occurs as the women age. The model fitted on the survey responses demonstrably supports the "strengthen socioeconomic status" argument.

As with all research, insights do need to be interpreted in context. The data available, and hence the applicability of the fitted model and ensuing discussion, is strictly restricted to ethnically Indian women living in Fiji who are physically capable of childbearing. To generalize the findings, additional segments of the WFS can be analyzed and summarized. Until doing so, it is not advisable to use the findings contained in this analysis to promote specific action over broader demographics.

# Bibliography

Cleland, J, and Vijay Verma. 1989. "The World Fertility Survey: An Appraisal of Methodology." *Journal of the American Statistical Association* 84 (407): 756–67.

Little, Roderick JA, and Thomas W Pullum. 1979. "The General Linear Model and Direct Standardization: A Comparison." *Sociological Methods & Research* 7 (4): 475–501.

# Appendix