

An Analytical Study of Demographic and Socioeconomic Factors Influencing Contraceptive Method Choice

School of Science & Technology
Department of Computer Science
City St. George's, University of London

I. INTRODUCTION

Access to and effective use of contraceptive methods plays a critical role in improving women's health outcomes, supporting reproductive autonomy, and contributing to broader social and economic development. Global public-health reports indicate that contraceptive uptake varies substantially across populations and is strongly influenced by demographic and socioeconomic factors such as age, education, family size, employment status, and access to information [1]. Evidence from large-scale demographic surveys further suggests that these factors influence not only whether contraception is used, but also the type of method adopted [2].

This study investigates contraceptive method choice using survey-based demographic data collected from married women. Contraceptive behaviour is categorised into three groups: no-use, long-term methods, and short-term methods. Rather than focusing solely on predictive performance, this project adopts an analytical approach to examine how demographic characteristics, such as age, number of children, and education, alongside socioeconomic indicators including standard of living, employment status, media exposure, and husband's occupation, vary across these contraceptive groups. This allows for a clearer and more interpretable understanding of contraceptive decision-making.

Previous empirical studies have shown that education, household conditions, and access to information are among the most important determinants of contraceptive use, although their relative influence differs across contexts and populations [2] [3]. However, much of the existing literature treats contraceptive behaviour as a binary outcome, focusing on use versus non-use. Fewer studies examine differences between short-term and long-term methods. By combining exploratory data analysis with an interpretable machine-learning model, this study seeks to identify which factors most strongly influence contraceptive method choice and how these factors interact.

The findings aim to provide actionable insight for public-health practitioners and policymakers by identifying population groups that may benefit from targeted awareness initiatives, improved access to information, or socioeconomic support, thereby contributing to more effective and equitable family-planning strategies.

II. ANALYTICAL QUESTIONS AND DATA

This study is guided by four analytical questions. First (Q1), how do key demographic characteristics, including wife's age, number of children ever born, and education level, differ across contraceptive method groups? This question is informed by earlier work showing strong links between life stage, family size, and reproductive decisions [2]. Second (Q2), how do socioeconomic factors such as standard of living, husband's occupation, and media exposure relate to contraceptive method choice? Understanding these relationships is important, as access to resources and information may influence both awareness and adoption of contraception [1]. Third (Q3), does a woman's employment status appear to be associated with contraceptive use, potentially reflecting differences in autonomy and planning? Finally (Q4), when all variables are considered together, which factors emerge as the strongest predictors of contraceptive method choice?

The dataset used in this analysis is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey, made publicly available through the UCI Machine Learning Repository [4]. The sample consists of married women who were either not pregnant or uncertain of their pregnancy status at the time of interview. The dataset includes commonly studied demographic and socioeconomic variables, making it appropriate for addressing the analytical questions posed. Importantly, the outcome variable distinguishes between no use, short-term methods, and long-term methods, allowing for a more nuanced examination of contraceptive behaviour than a simple use versus non-use classification.

Some limitations should be noted. The data are cross-sectional and observational, which restricts causal interpretation, and several variables are measured in categorical or ordinal form. In addition, cultural and contextual factors influencing contraceptive decisions are not directly observed. Despite these limitations, the dataset provides a suitable and informative basis for the analytical objectives of this study.

III. ANALYSIS

A. Data Preparation

The analysis begins with a focused data preparation stage designed to support interpretability and direct alignment with the analytical objectives. As the primary goal of the study is analytical rather than predictive, preparation decisions were guided by the need to preserve the original meaning of variables rather than to optimise model performance.

The initial assessment of the data revealed that there were no missing data on the predictors and the dependent variable. Therefore, the task of imputation or deletion was not necessary. Each variable was then checked to ensure consistency with its documented coding scheme. Count-based variables, such as wife's age and number of children ever born, were retained in their original form, while categorical variables were kept in binary or ordinal format where appropriate (e.g., education levels, standard of living index, media exposure). As the dataset was already numerically encoded and analytically interpretable, no additional transformation or encoding steps were applied.

A class distribution check was conducted to understand the balance of the outcome variable (Table 1). This revealed a moderate imbalance across contraceptive categories, which was noted as an important contextual factor for subsequent analysis. However, since the aim is to answer analytical questions rather than to build an optimised predictive system, the imbalance was treated as an interpretive consideration rather than something to be corrected through resampling or weighting.

Contraceptive Method	Count	Percentage
1	629	42.701969
3	511	34.691107
2	333	22.606925

Table 1: Distribution of Contraceptive Method Usage

B. Analytical Design for Q1: Demographic Factor

The first research question examines how key demographic characteristics differ across contraceptive method categories. The relevant factors for analyzing will be the age of the wife, the number of children ever born, and the level of education of the wife because these factors adequately reflect life stage, family formation, and decisional capacity. Demographic information is analyzed first because it essentially reflects background information determining reproduction independent of other sociocultural factors.

The methodology of the analysis for the question will be exploration and comparative. The use of visualization tools enables the comparison of data distributions and proportions between the contraceptive alternatives without requiring the use of summary statistics. For the age of the wife and the number of children ever born, box graphs will be utilized because they enable the visualization of data distributions and the overlap of groups. This is a more complex method than a bar graph of mean data because it emphasizes the "noise" and

indeterminacy of contraceptive decision-making. For the education level of the wife, a count comparison will be utilized because the level of education is an ordinal category and the differences between the data distributions are of major interest.

C. Analytical Design for Q2: Socioeconomic Factors

The second research question examines the relationship between the selection of contraceptive practices and their socioeconomic context. Specifically, this analysis will highlight standard of living, exposure to mass media, and husbands' occupation, all of which measure information and resource access through their position in society. Each of these factors combined allows for the examination of whether differences in family conditions influence contraceptive practices.

For socioeconomic variables, comparative visualizations using count-based distributions enabled proportional comparisons across contraceptive use categories. This method aligns with the analytical objective of identifying group differences in categorical and ordinal variables.

D. Analytical Design for Q3: Employment Status

The third research question is related to whether a woman's employment status is linked with the choice of contraceptive method. The employment status of a woman has been considered independently as a research question, as it may be a proxy for the afore-mentioned dimension of autonomy and allocation of time, and is not entirely captured by the existing indicators of socioeconomic status.

Due to the binary nature of employment status, proportional comparison was used to examine contraceptive type distributions between employed and non-employed women. This approach appropriately compares group representation rather than attempting to capture complex dynamics, while treating employment status as a social rather than strictly economic factor.

E. Analytical Design for Q4: Identify Influential Factors

The fourth research question is an attempt to discover which variables play a crucial role in determining contraceptive method choice, considering all joint demographic and socioeconomic variables. This is accomplished by employing a Random Forest classifier not as a prediction algorithm but as a tool of analysis, by using a Random Forest algorithm because of its ability to process a variety of variables, in addition to offering a stable measure of the importance of variables by averaging through a forest of trees [5].

The type was set up for Gini Importance (Mean Decrease in Impurity). For each node of decision trees, Gini impurity was calculated as follows:

$$G = 1 - \sum_{i=1}^n p_i^2$$

The weight given to a feature, based on how well a given life-stage or socioeconomic component partitions data, is measured as the weighted average of total impurity reduction contributed by all trees. It serves as a robust measure in determining which feature has most greatly contributed to data partitioning, surpassing the knowledge offered by correlation coefficients in measuring 'influence'.

The model was developed on all available predictor variables, and importance scores for those variables were obtained to identify the variables in terms of importance for classification. This method allows the analysis to go beyond simple correlations and examine multiple variables working in concert. While the accuracy of this model was approximately 0.57, this is to be expected given the overlap of classes and the lack of variables related to behavior/culture in the data.

Analysis of the confusion matrix (Figure 1) shows that most misclassification occurs between non-use and short-term methods, indicating that these categories are the most difficult to distinguish using the available predictors. As a result, the primary value of the model lies not in predictive accuracy but in the ordered ranking of variables, which highlights the demographic and socioeconomic factors most influential in differentiating contraceptive method choice.

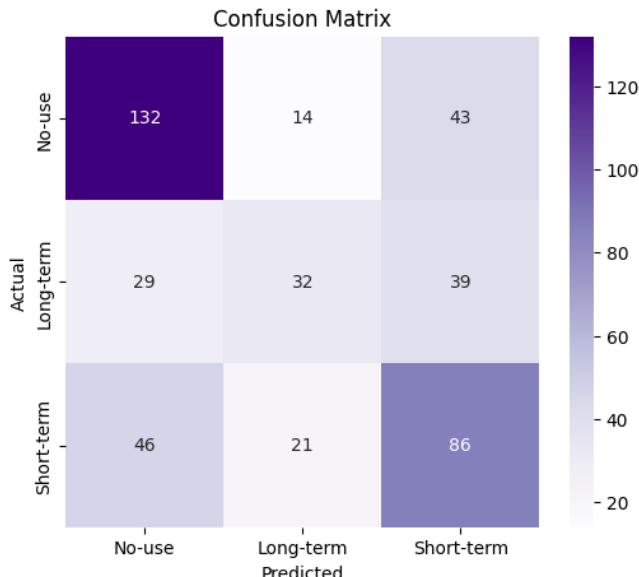


Figure 1: Confusion Matrix

F. Analytical Synthesis

Findings from Q1–Q3 provide evidence of how demographic, socioeconomic, and employment-related factors vary across contraceptive method categories. These insights are consolidated in Q4, where the same variables are evaluated jointly to assess their relative influence. This synthesis links exploratory comparisons with multivariate analysis, allowing interpretations from individual analyses to be considered within a broader analytical context while remaining constrained by the scope of the available data.

IV. Findings, Reflections and Further Work

A. Findings for Q1: Demographic Factors

In figure 2, the difference in age distributions by categories of contraceptive use is highlighted. There appears to be slightly higher median age among women who use long-term contraceptive methods (2) compared to non-users (1) and those who use short-term methods (3) but considerable overlap exists.

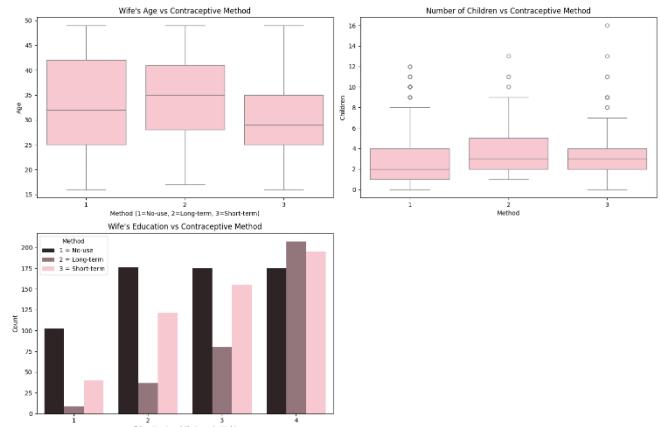


Figure 2: Demographics (Wife's Age, Children, Education) vs. Contraceptive Choice

Non-users generally have smaller family sizes, whereas both short-term and long-term users display larger median values and larger spreads. Long-term users demonstrate only a slight elevation in central tendency compared to non-users, and there is some overlap among non-users and short-term users. Nonetheless, family size is indicated as being linked to family planning practices but is non-determinant.

Education level is also divergent by categories, with higher levels of education among contraceptive users and lower levels among non-users. Taken together, these trends suggest that demographic variables influence the use of contraceptives but are not a complete explanation of an individual's actions.

B. Findings for Q2: Socioeconomic Factors

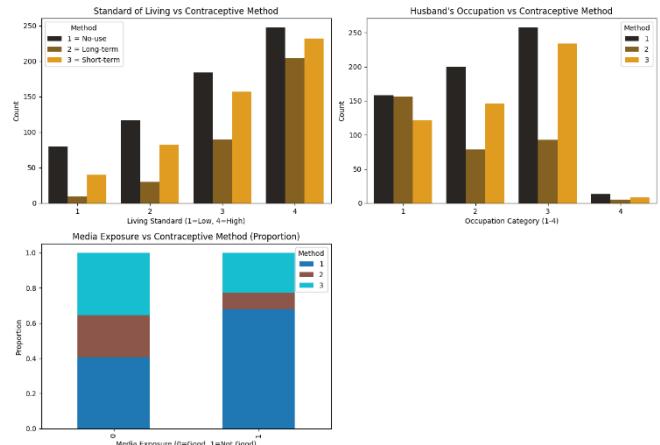


Figure 3: Socioeconomic Factors (Standard of Living, Husband's Occupation, Media Exposure) vs. Contraceptive Choice

From Figure 3 above, at lower levels of living standards, lack of use of contraception is more common, while the use of both short-term and long-term methods of contraception rises with increasing living standards. At the highest level of living standards, use of contraception, especially short-term use, becomes more prominent compared to lack of use.

Variation is also found when categorized by the occupation of the husband. Even though non-use is widespread in all categories, both short-term and long-term methods are practiced to a relatively greater extent in higher occupation categories, especially category 3.

Additional differentiation comes with exposure to the media in contraceptive behavior. Women who are better exposed to the media have higher percentages of contraceptive practice, while non-practice prevails among those who are less exposed to the media information. The above results indicate that the socioeconomic context influences the choice of contraceptive methods, yet the overlapping distributions show that the dimensions of these factors are not mutually exclusive.

C. Findings for Q3: Employment Status

No-use method is the most frequent, followed by short-term methods, while long-term methods are least used, as indicated in Figure 4. It appears that non-working women exhibit a slightly higher level of short-term method use, while long-term method use is marginally lower than those of working women.

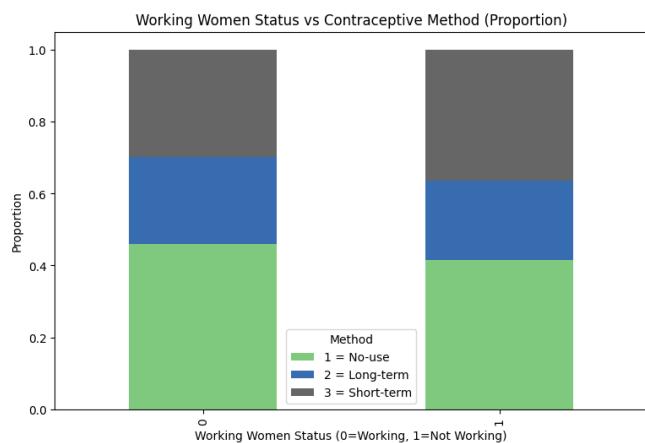


Figure 4: Employment Status vs. Contraceptive Choice

Nonetheless, considering these differences, the level of resemblance in patterns tends to suggest that employment status by itself is not a significantly determining factor of the use of contraceptives.

D. Findings for Q4: Influential Factors

Figure 5 presents feature-importance rankings from the Random Forest model. Wife's age and number of children emerge as the most influential variables, indicating that life-stage characteristics play a central role in distinguishing contraceptive method choice. These variables contribute substantially more than any other factor.

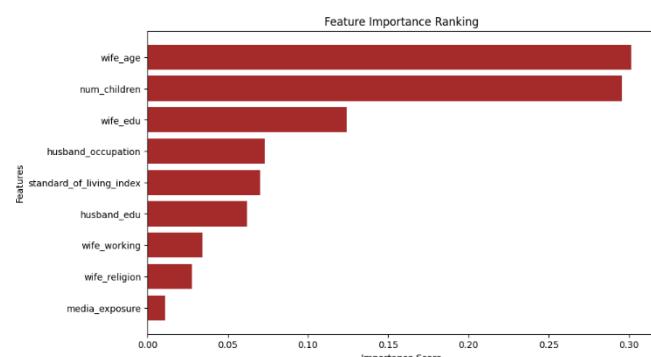


Figure 5: Feature Importance for Contraceptive method choice

Wife's education forms a second tier of influence,

contributing meaningfully but remaining secondary to age and number of children. Socioeconomic indicators such as husband's occupation, standard of living index, and husband's education show moderate importance, suggesting that household context adds explanatory value but does not dominate decision-making.

Employment status, religion, and media exposure exhibit relatively low importance, indicating limited additional contribution once demographic and household characteristics are accounted for. Overall, these results show that demographic life-stage variables are most influential when factors are considered jointly.

The dataset is cross-sectional and limited to demographic and socioeconomic variables, restricting causal inference and excluding behavioural or attitudinal influences. Findings should therefore be interpreted at a population level.

E. Further Work

Future work could use longitudinal data or behavioral data to allow a better examination of patterns of contraceptive use at different life stages. Including qualitative or attitudinal measures and conducting stratified analyses by age, region, or socioeconomic subgroup may reveal more nuanced relationships not visible in the aggregated data.

V. References

- [1] Cleland, John G, Ndigwa, Robert P & Zulu, Eliya M (2011). Family planning in sub-Saharan Africa: progress or stagnation?. *Bulletin of the World Health Organization*, 89(2), 137 - 143. World Health Organization. <https://doi.org/10.2471/BLT.10.077925>
- [2] Haq I, Sakib S, Talukder A.(2017). Sociodemographic Factors on Contraceptive Use among Ever-Married Women of Reproductive Age: Evidence from Three Demographic and Health Surveys in Bangladesh. *Med Sci (Basel)*.doi:10.3390/medsci5040031
- [3] Bongaarts, John and Robert G. Potter. 1983. Fertility, Biology, and Behavior: An Analysis of the Proximate Determinants. Elsevier, <https://doi.org/10.1016/C2009-0-03021-9>.
- [4] T. Lim. "Contraceptive Method Choice," UCI Machine Learning Repository, 1999. [Online]. Available: <https://doi.org/10.24432/C59W2D>.
- [5] Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
- [6] AI Assistance Declaration:

I declare that the project objective, experimental design, and data analysis presented in this notebook are my own work. I utilized AI tools to assist in structuring specific Python functions, debugging logic, and refining technical terminology and precising the sentence for less word counts for documentation and analytical discussions. All final implementation, data preprocessing, and validation of results were performed independently.

VI. Word Counts

- Introduction: 297 words
- Analytical Questions and Data: 283 words
- Analysis: 996 words
- Findings, Reflections and Further Work:
594 words
- Total (excluding references, figures, captions):
2170 words