Heterogeneity in Intergenerational Transmission of Health: A Recursive Partitioning Approach from Machine Learning

Dweepobotee Brahma[1] Souvik Ghosh[2]

September 15, 2023

Abstract

Intergenerational transmission of health at birth is particularly affected by maternal circumstances at birth as well as exposure to timely antenatal policies. This paper uses a large nationally representative survey data to explicitly estimate heterogeneity in intergenerational transmission of health by maternal circumstance and policy exposure. Using a novel model-based recursive partitioning algorithm from the Machine Learning literature that uses econometric tests for parameter instability, this study identifies different circumstance profiles characterized by varying coefficients of intergenerational health transmission. We also estimate heterogeneity in health transmission by policy exposure within a given circumstance profile. Results exhibit considerable heterogeneity by both short-run and long run markers of maternal health and reveal that a global model for investigating intergenerational transmission is inadequate. Worse-off circumstances have stronger transmissions of maternal health to infants and only specific antenatal policies have any effect. The results have implications for more targeted policy-making and improve our understanding of how to break intergenerational cycles of ill-health.

Keywords: Intergenerational Transmission of Health, Machine Learning, Heterogeneity

JEL Classification: I15, C5, O12

[1] Centre for Mathematical and Computational Economics, School of Artificial Intelligence and Data Science, IIT Jodhpur; dweepobotee@iitj.ac.in
[2] Centre for Mathematical and Computational Economics, School of Artificial Intelligence and Data Science, IIT Jodhpur; souvikghosh@iitj.ac.in

I. Introduction

Intergenerational transmission of health is a key channel through which health inequalities are perpetuated over time (Bhalotra and Rawlings (2011, 2013). Birthweight – a marker of health stock inherited from parents at birth – has lingering consequences on several adult life outcomes. However, the transmission of maternal health to infant birthweight itself is affected by maternal circumstances. Children are more likely to bear the effects of poor maternal health if they are born in adverse socio-economic conditions (Currie and Moretti 2007; Bhalotra and Rawlings 2011). On the other hand, the transmission of poor maternal health to their newborn's birthweight can be offset at least partially by timely and quality antenatal care. Yet exposure to antenatal care is itself affected by maternal circumstances.

In this study we ask two specific questions that are somewhat sequential in nature. First, we ask how different maternal 'circumstances' – characterized by the combination of environmental, socio-economic, demographic factors – affect this coefficient of transmission of health from mothers to infants. Second, we ask how within these circumstance profiles, this coefficient of intergenerational transmission of health may be affected by receiving antenatal healthcare. We identify several circumstance profiles using a nationally representative household survey from India and document heterogeneity in the coefficient of intergenerational transmission of maternal health to infants across these circumstance profiles.

The study of intergenerational transmission of health has its roots in the idea of intergenerational mobility and distributive justice which benchmarks the amount of inequality a society may be willing to tolerate. A growing body of literature documents the effects of different channels of intergenerational transmission of health through several socio-economic characteristics. Cross-country studies establish the persistent effects of mothers' health on child mortality and anthropometric measures and the importance of socio-economic conditions in underpinning this transmission (Bhalotra and Rawlings (2011, 2013). Country specific studies establish this result in greater depth through the use of hospital linked birth records in the United States (Currie and Moretti (2007)), survey data in Vietnam (Venkataraman (2011)), Philippines (Bevis and Villa (2022)), China (Eriksson, Pan and Qin (2014)), Germany (Coneus and Spiess (2012)) and India (Subramanian et al. (2009) and Kumar and Nahlen (2023)). Coefficient of intergenerational transmission is influenced heavily by household income and wealth with considerable non-linearities in this relationship (Bhalotra and Rawlings (2013)) and by residential area particularly through ZIP codes (Currie and Moretti (2007)). Intergenerational transmission of health is found to be stronger for boys than for girls (Venkataraman (2011)). Intergenerational transmission of health has been found to persist well into adolescence (Bevis and Villa (2022)), and through both maternal and paternal characteristics although the channel of transmission is stronger for maternal than for paternal characteristics (Giuntella, Mattina and Quintana-Domeque (2023)). Intergenerational transmission is explored through a variety of health outcomes for both parents and children with height being the most commonly used indicators for mothers and anthropometric Z scores are used as age-appropriate standardized indicators of child health (Bevis & Villa (2022), Bhalotra and Rawlings (2011), Bhalotra and Rawlings (2013), Giuntella, Mattina and Quintana-Domeque (2023), Venkataraman (2011)) as it can be thought of as a stock measure of health. Studies using linked hospital records are able to use maternal birth weight which captures the stock of health inherited by the mother (Currie and Moretti (2007). Kumar and Nahlen (2023) focus on anaemia for both mothers and children to capture intergenerational transmission of micronutrient deficiency. Self-reported measures of health status are also common as in Halliday et al.(2021).

This paper picks up where the existing literature leaves off. Given the documented importance of other socio-economic factors in affecting maternal transmission of health particularly at birth, this paper identifies and characterizes sets of socio-economic, demographic attributes that constitute different circumstance profiles with varying coefficients of intergenerational transmission. We use a novel inference-based recursive partitioning technique from the Machine Learning literature to identify the circumstance profiles in a data-driven manner. Model-based recursive partitioning is an improvement over the typical regression and classification trees. Model-based recursive partitioning allows a middle ground between estimating parameters through a fully parametric model and a fully non-parametric model. It combines the data driven approach to partitioning the covariate space from the Machine Learning literature with a model-based approach of testing for parameter stability from the econometrics literature. Instead of fitting an overall model to the entire dataset, model-based recursive partitioning estimates local models on subsets of the data that are "learned" through repeated partitioning of the data. The goal here is uncovering heterogeneity in the estimates of intergenerational health transmission using all available information on Xs, rather that testing whether pre-specified Xs are associated with significantly different intergenerational transmissions. This makes Machine Learning methods which adopt a data-driven approach to uncovering heterogeneity a natural choice.

This paper contributes to the existing literature in several ways. First, it makes a methodological contribution by demonstrating how an inference-based Machine Learning algorithm can be used to measure heterogeneity in intergenerational transmission of health at birth across several circumstance profiles and exposure to healthcare. Second, it provides new evidence on heterogeneity in intergenerational transmission of maternal health with regard to both maternal circumstances and exposure to policy. Results illustrate considerable heterogeneity across the identified subgroups indicating that a global model on the full sample is not appropriate to model intergenerational health transmission. We find that heterogeneity is primarily driven by religion, maternal education, maternal age, household wealth index and infant's gender. We identify distinctly different circumstance profiles defined by these socio-economic characteristics. Only a few antenatal healthcare policies have any effect in tempering the strength of intergenerational transmission. The totality of results indicate that worse-off maternal circumstances are associated with a coefficient of intergenerational transmission. This indicates that the persistence of ill-health across generations can be hard to break.

The structure of the paper is as follows. Section II explains the novel model-based recursive partitioning approach. Section III discusses the dataset and variables. Section IV lays out the full set of results with the circumstance profiles. The figures for decision trees and subtrees are illustrated in the Appendix. Section V concludes with a discussion on limitations and policy implications.

II. Empirical Strategy

We define a simplistic framework where an infant's health outcome (H_i) depends on their mother's health status (H_m), maternal circumstances (C_m) and exposure to antenatal care policies (P).

$$Y_i = f(H_m, C_m, P) \tag{1}$$

Out of these, circumstances are the most important set of factors since both mother's health status and exposure to policy depend on maternal circumstances. We can go so far as to rewrite the above equation as

$$Y_i = f\big(H_m, C_m, P(C_m)\big) \tag{2}$$

We model this is a two-step process. First, we identify different circumstance profiles within the population and estimate the intergenerational transmission within each circumstance profile.

$$Y_i = f(H_m, C_m) \tag{2.1}$$

Next, holding a circumstance profile constant, we model heterogeneity in intergenerational transmission of health by policy exposure within a circumstance profile.

$$Y_i = g(H_m, P; \overline{C_m}) \tag{2.2}$$

Empirically, we can model (2.1) through a linear regression of the following general form.

$$y_i = \beta_0 + \beta_1 h_i + \beta_2 X + u \tag{3}$$

where y_i is the health outcome of the infant, h_i is the health outcome of the mother and X contains the set of *circumstances*. β_1 is the key parameter of interest capturing intergenerational transmission of health and X is the set of *circumstances*. Under traditional econometric methods, we have to assume that the relation between circumstances and infant health is linear and separable from the mother's health. We can relax the separability assumption somewhat by incorporating interaction terms with mother's health h_i and any of the circumstances in X to capture heterogeneity in intergenerational transmission by specific circumstance (e.g. mother's age). But such an approach causes us to get trapped in the pool of what we already know. One can check for heterogeneity with respect to specific circumstances particularly those with which we expect there to be heterogeneity. This is where modern data-driven Machine Learning techniques is useful. We use a model-based recursive partitioning approach to consider all possible combinations of circumstances and divide the population into specific circumstance profiles. We fit our empirical equation(s) for all the individual mother-infant dyads defined by a given circumstance profile.

We use the model-based recursive partitioning framework from Zeileis et al. (2008). It starts by fitting a parametric model to the entire dataset. It then tests for parameter instability over the set of covariates. If there exists some overall parameter instability, the data is split into two subgroups at the point of parameter instability. This process continues with successive recursive partitioning with each partitioning creating two subsets of the data until some stopping rule is reached.

The model-based recursive partitioning approach uses the theory of parameter instability using a class of estimators called M – estimators. These are estimators that do not have explicit algebraic expressions and need to be calculated by numerical methods. A common type of such estimator is used in threshold models which form the basis of regression trees and random forests. Zeileis et al. (2008) develop a unified approach for a wide range of statistical models called generalized M – fluctuation tests. The approach has similarities with tests for structural changes widely used in time series econometrics.

The generalized M – fluctuation tests used for splitting decision relies on the estimating the parameter θ which minimizes an objective function ψ.

$$\hat{\theta} = arg \max \sum_{i=1}^{n} \psi(y_i, x_i, \theta) \tag{4}$$

where y_i and x_i , $i = 1, \dots, n$ denote the vector of independent and dependent variables and θ represents the parameter. The estimation process involves calculating the individual contributions of each observation i to the score function

$$\psi(y_i, x_i, \theta) = \frac{\delta \Psi(y_i, x_i, \theta)}{\delta \theta} \tag{5}$$

Next, the pre-specified model is fit on the entire dataset to obtain the initial estimate $\hat{\theta}$. To detect a change in the parameter estimate, over a range of a covariate say X_j, the observations are rearranged and ordered according to their value of X_j. The null hypothesis states that no systematic change in the parameter is present. The null hypothesis is tested against the alternative that one or more parameters of the specified model changes significantly over the ordering induced by X_j.

The test statistic uses the partial derivative of the objective function. The contributions of each observation i to the derivative of the objective function at the current value of the parameter estimate $\hat{\theta}$ is calculated and ordered according to the value of the covariate X_j. Under the null hypothesis of no systemic change, these contributions should be distributed randomly around mean zero. To detect a clear structural break, Zeileis and Hornik (2007) use the cumulative sums of the individual contributions to construct a test statistic for detecting structural breaks and its adjoining asymptotic properties and p – values. If parameter instability is detected, the variable with the lowest p – value is chosen.

Once the covariate with which a split is to be made split is chosen, the next step is to choose the value s at which the split will be made. The cut point s is chosen by a criterion that maximizes the objective function in the two potential subsamples: $L(s) = \{ i | X_{ij} \leq s\}$ and $R(s) = \{i \mid X_{ij} > s \}$. The optimal cut point $s*$ is chosen through the process of maximizing

$$\sum_{i \in L(s*)} \Psi\left(y_i, x_i, \widehat{\theta^{(L)}}\right) + \sum_{i \in R(s*)} \Psi\left(y_i, x_i, \widehat{\theta^{(R)}}\right) \tag{6}$$

over all potential cut point s. This process splits each region into two subregions. This process continues recursively until no further parameter instability is detected.

This process leads to the formation of several non-overlapping regions in the covariate space (known as terminal leaves). In each of these regions the pre-specified parametric model is fit to estimate the parameters of interest. This contrasts with a traditional decision tree where the mean of the dependent variable is calculated to generate a predicted value at that leaf. In model-based recursive partitioning we fit a regression model and generate a parameter estimate in each leaf. The parametric relations we are interested in is

$$log(Birthweight_i) = \beta_0 + \beta_1 log(BMI_i) + u_i \quad (7)$$

$$log(Birthweight_i) = \gamma_0 + \gamma_1 log(Height_i) + \epsilon_i \quad (8)$$

where $i = 1,2, \dots, n$ is the index for observations. Our parameters of interest are β_1 and γ_1 which capture the coefficients of intergenerational transmission of mother's health status to the infant. We estimate equations (7) and (8) in ***each terminal leaf*** to get the coefficient of intergenerational transmission of health for each subgroup.

We use this model based partitioning algorithm twice. First, to explore heterogeneity in the coefficient of intergenerational transmission by other socio-economic factors to identify and characterize several types within the population. Next, model-based partitioning is applied in each subsample (from each terminal node) identified in the previous tree to explore heterogeneity in the coefficient of intergenerational transmission by different prenatal healthcare policies. Put differently, we first explore heterogeneity in the coefficient of intergenerational transmission by socio-economic characteristics to identify several 'types'. Then we explore heterogeneity in the coefficient of transmission within these types by antenatal healthcare policy.

III. Data

We use unit-level data from India from the National Family Health Survey – V conducted during 2019-21. The National Family Health Survey is a nationally representative survey data gathering information on health indicators from 636,699 households and 1,274,250 birth history records. Since we are interested in intergenerational transmission of mother's health on birthweight, we restrict the sample for this study to those infants who were surveyed within 1 year of birth. This ensures two things. One, we attempt to minimize recall error particularly in case of infant's birthweight in cases where the mother is recalling this from memory. Second, we want to restrict the sample to capture mother's BMI around the time of birth to model short-run transmission of health.

We divide the relevant covariates into two non-overlapping categories: (i) socioeconomic and demographic characteristics, and (ii) prenatal healthcare policies. Socio-economic and demographic variables include indicators for different broad religion and caste categories, household wealth index, mother's age, mother's education (years of schooling), age of household age, years of schooling of household head, gender of the infant, birth-order, household size, indicator for urban. The list of prenatal care variables includes indicators for the type of healthcare worker who provided care during antenatal care visits, awareness regarding potential complications during pregnancy, receipt of supplementary nutrition from ICDS/Anganwadi centres, tetanus toxoid injection and number of antenatal care visits. We use a discrete variable to capture whether the mother did not get any antenatal care, got atleast four antenatal care visits[3] or more than four antenatal care visits. A detailed list and explanation of the covariates used is in Tables 2 and 3 in the Appendix.

Indicator of Infant Health

Birthweight[4] can be considered as a measure of the first stock of health inherited from parents. It is a leading indicator of childhood health and is known to have a persistent documented influence on long-term health and other market outcomes in adults.

Indicator of Maternal Health

As indicators of mother's health, we consider two measures – BMI and height. BMI (weight/height^2) serves as a flow measure of health and considering the passthrough of mother's BMI to infant birthweight allows us to measure the transmission of short-run health across generations. In contrast, maternal height is a long-run measure of maternal health.

IV. Results

A. Heterogeneity by Socio-Economic Characteristics.

A.1. Mother's Height

Applying model-based recursive partitioning, we identify several subgroups in the population who have different coefficients of intergenerational transmission of health. The main inference tree fitted through model based recursive partitioning (MOB) is illustrated in Figure 3 in the Appendix. Boys born to mothers with eight or less years of schooling in Hindu or Muslim households ***(Type A)*** who make up 21.1% of the entire sample have a coefficient of intergenerational transmission of 0.466. In contrast, for girls born to mothers with similar characteristics (with eight or less years of schooling in Hindu or Muslim households) the coefficient of transmission is influenced further by the wealth of the household. In poorer households (wealth index in the second or lower quintile of the population) who make up 12.2% of our sample – ***Type B*** – the coefficient of transmission is 0.36, whereas in relatively richer households (wealth index in the third or higher quintile of the population) comprising 8% of the sample – ***Type C*** – the coefficient of transmission is 0.485.

[3] This is following WHO's recommendation regarding at leat four antenatal visits during pregnancy. "Guidelines for antenatal care and skilled attendance at birth by ANMs/LHVs/SNs", Maternal Health Division, Ministry of Health and Family Welfare, Government of India, April 2010. URL: https://nhm.gov.in/images/pdf/programmes/maternal-health/guidelines/sba_guidelines_for_skilled_attendance_at_birth.pdf
"WHO Antenatal Care Randomized Trial: Manual for the Implementation of the New Model" URL: https://apps.who.int/iris/bitstream/handle/10665/42513/WHO_RHR_01.30.pdf?sequence=1&isAllowed=y

[4] Birthweight has typically had missing observations in household survey data like NFHS. However, with improvements in tracking indicators in child health and rise in institutional births the extent of missingness has steadily declined across different rounds of NHFS. NFHS-V which is used here has a response rate of more than 90%.

For mothers with more than 8 years of education in Hindu or Muslim households, intergenerational transmission of health to infants is further influenced by their age. For mothers less than 24 years old ***(Type D)*** the coefficient of intergenerational transmission is 0.54. They make up around 23.18% of the entire sample. In case of infants born to mothers more than 24 years of age with more than 8 years of schooling in Hindu or Muslim households, boys consisting of 12.3% of the sample ***(Type E)*** have a coefficient of intergenerational transmission of health of 0.600, while girls consisting of 11.2% of the sample ***(Type F)*** have a coefficient of 0.393. Lastly the coefficient of intergenerational transmission of health through mother's height to infants in Christian and other religion households ***(Type G)*** is 0.306. This group makes up the remaining 11.7% of the overall sample. The tree built with model-based recursive partitioning to estimate this heterogeneity across the population is illustrated in Figure 3 in the Appendix.

Comparison of Subgroup Estimates with the Full Sample Estimates

The MOB based algorithm provides estimates of intergenerational transmission of health by different circumstance profiles. Since ultimately a linear regression is estimated in each leaf, we also compute the interval estimates and compare with the full sample estimates. Figure 1 below illustrates the interval estimates of the coefficients of intergenerational transmission of health through mother's height for each of the circumstance type identified through model-based recursive partitioning. The overall coefficient of intergenerational transmission of health through mother's height estimated through a parametric regression over the entire sample with the same socio-economic variables as controls in 0.421. The regression results are present in Table 5 in the appendix. The full sample estimate is illustrated through the dotted red line in Figure 1 below.

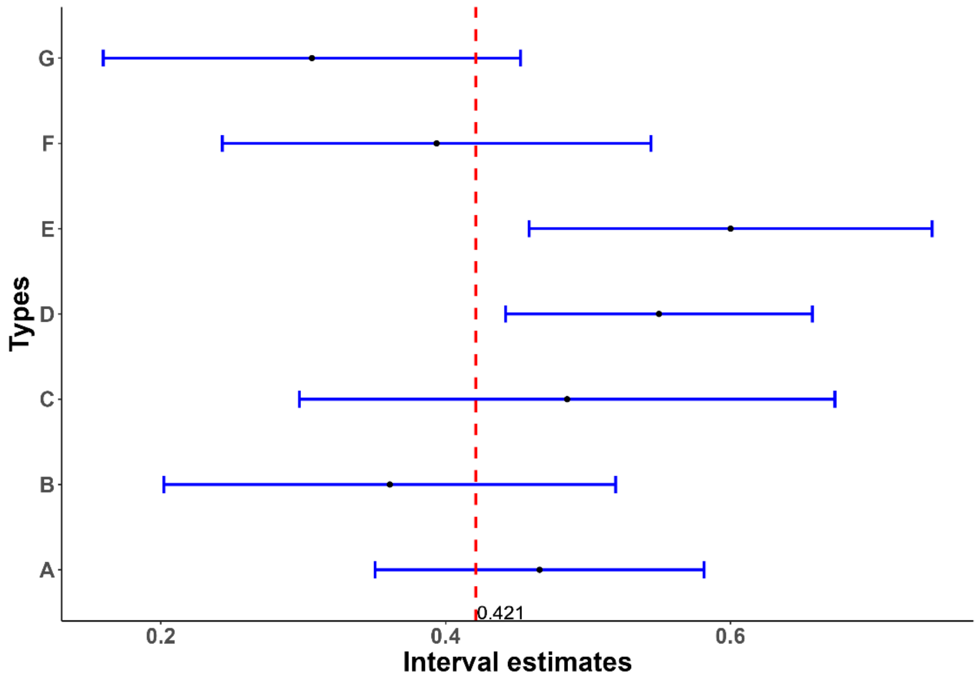

Figure 1: The bars illustrate the interval estimates of intergenerational transmission of mother's height for every circumstance type uncovered through MOB. The red dotted line is the estimated coefficient of intergenerational transmission of health through mother's height. As evident from the x-axis, these interval estimates are statistically significant at the 5% level.

Both the point estimates and interval estimates reveal considerable heterogeneity in the estimated coefficients of intergenerational transmission. While the overall estimate lies within the interval estimates of most groups, it lies outside the interval of circumstance type E – boys born to mothers more than 24 years old with more than 8 years of education in Hindu or Muslim families, and circumstance profile D – infants born to mothers with ≤ 24 years of age and > 8 years of schooling.

A.2. Mother's BMI

While mother's height is considered a stock measure of health, BMI is considered a flow measure of mother's health (Bhalotra and Rawlings (2011)). We also run model based recursive partitioning (MOB) on mother's BMI to identify circumstance profiles that characterize heterogeneity in intergenerational transmission of mother's BMI on infant's birthweight. Since we restrict our data to mother-child pairs whose health outcomes are measured within 12 months of birth, we can use mother's BMI as a reasonably good flow measure of mother's health around pregnancy. The main inference tree for heterogeneity in the intergenerational transmission of health through mother's BMI to infant's birthweight fitted through model based recursive partitioning (MOB) is illustrated in Figure 10 in the Appendix.

The circumstance profiles that we identify in this case are very similar to those identified for mother's health with different estimated coefficients of intergenerational transmission of health. For boys born to mothers with eight or less years of education in Hindu or Muslim families ***(Type A)*** the coefficient of intergenerational transmission is 0.144. In case of girls born to mothers with 8 or less years of education in Hindu or Muslim households, the wealth index of the family further influences the coefficient of transmission. In relatively poorer households with wealth index in the second or lower quintile of the population ***(Type B)***, the coefficient of intergenerational transmission is 0.15. This group consists of 12.2% of our sample. In relatively affluent households with wealth index in the third or higher quintile of the population ***(Type C)***, the coefficient of intergenerational transmission is 0.116. This group comprises about 8% of our sample.

Boys born to mothers with 8 or more years of education in Hindu or Muslim households ***(Type D)***, the coefficient of intergenerational transmission of health is 0.137. This circumstance profile has the highest – about 24% of our sample. For girls in this circumstance profile, the household wealth further influences intergenerational transmission. Girls born to mothers with eight or more years of education in Hindu or Muslim households in relatively less affluent families whose wealth index is in the third or lower quintile of the population ***(Type E)*** who comprise 10.6% of our sample have an intergenerational transmission coefficient of 0.178. In contrast, girls born to mothers with eight or more years of education in Hindu or Muslim households in relatively more affluent families whose wealth index is in the fourth or fifth quintile of the population ***(Type F)*** have a coefficient of intergenerational transmission of only 0.09. This group has 12% of the sample in this study. Infants in households of other religions (Christians, Jains, Buddhists and others) have a coefficient of intergenerational transmission of 0.15. These individuals make up the remaining

While the circumstance profiles identified for mother's height on infant birthweight and for mother's BMI and infant birthweight are quite similar, they differ in one major way. In case of mothers' height, for infants born to mothers with more than eight years of education in Hindu or Muslim families, the age of the mother is chosen as a splitting variable followed by the gender of the infant. In contrast, for mothers' BMI, instead of mother's age the gender of the infant is chosen followed by the wealth index of households as the splitting variable.

Comparison of Subgroup Estimates with the Full Sample Estimates

Figure 2 below compares the interval estimates of coefficients of intergenerational transmission of health in each terminal leaf with the overall estimate generated through a global model fitted on the entire sample. Similar to the case of mother's height, there is considerable heterogeneity in the interval estimates across the different circumstance types.

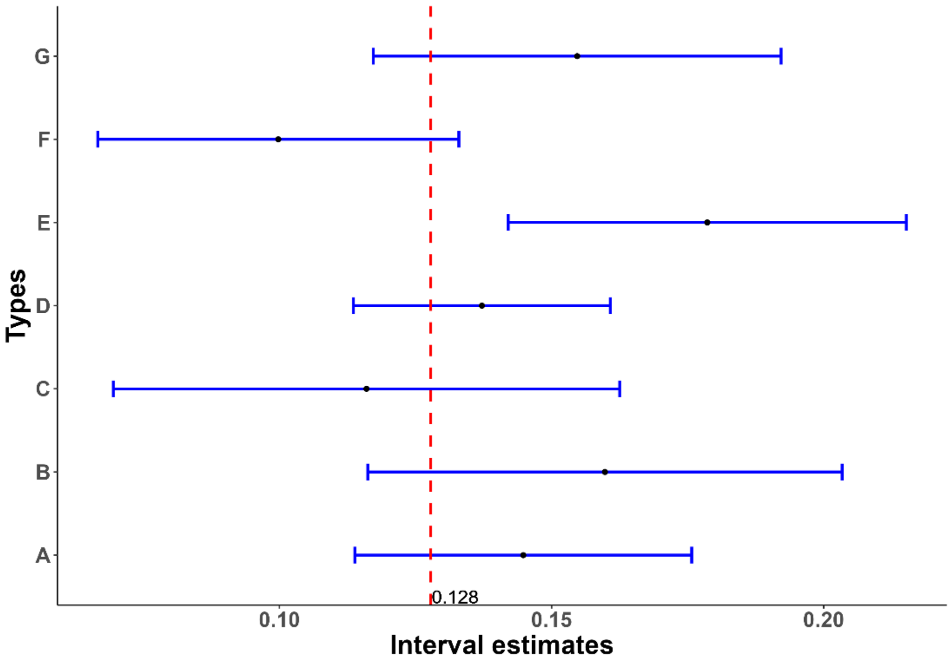

Figure 2: The bars illustrate the interval estimates of intergenerational transmission of mother's BMI for every circumstance type uncovered through MOB. The red dotted line is the estimated coefficient of intergenerational transmission of health through mother's height. As evident from the x-axis, these estimates for each type (in the leaves) are statistically significant at the 5% level.

Statistical Significance

Model based recursive partitioning enables us to uncover heterogeneity across different circumstance profiles in a data-driven manner. A natural question that follows would be whether these differences in intergenerational transmission are statistically different. Since ultimately we fit a simple linear regression in each leaf, we can reverse-engineer these estimates using dummies to define leaf placement of observations. To do so, we estimate the following equation on the full sample of data.

$$Y = \sum_{l} L_l + \sum_{l} \beta_l L_l X + u \tag{9}$$

Here, L_l refers to a dummy signifying whether a given observation lands in leaf l. β_l captures the coefficient on the interaction term between the leaf dummy and X which recovers the estimates of intergenerational transmission in each leaf that we estimated earlier through model-based recursive partitioning. We estimate this above equation twice – once for the specification with maternal height and once for the specification with maternal BMI.

We first test for overall significance using a standard econometric F-test. The null hypothesis of no overall significance is rejected at the 1% level for both maternal height and maternal BMI. We then proceed to do pairwise t-tests for differences in the estimates for a pair of leaves (signifying circumstance profiles). The results of these pairwise t-tests are presented in the tables below.

Table 1: Statistical significance from pairwise t-tests across difference leaves for maternal height

Types	Type A	Type B	Type C	Type D	Type E	Type F	Type G
Type A	-	-	-	-	-	-	-
Type B							
Type C							
Type D		**					
Type E		***					
Type F					**		
Type G	**			***	***		

*p<0.1; **p<0.05; ***p<0.01

Table 2: Statistical significance from pairwise t-tests across difference leaves for maternal BMI

Types	Type A	Type B	Type C	Type D	Type E	Type F	Type G
Type A	-	-	-	-	-	-	-
Type B							
Type C							
Type D							
Type E			***	**			
Type F	***	***		**	***		
Type G						***	

*p<0.1; **p<0.05; ***p<0.01

Across both specifications of maternal height and maternal BMI, several circumstances profiles show statistically significant differences. In the case of maternal height, Type G comprising individuals belonging to households of other religions are statistically different from Type A, Type D and Type E. An interesting case is the fact that Type E and Type F here are statistically different. Both groups consist of infants born to mothers with more than 8 years of schooling and more than 24 years of age. The only difference between them is in the gender of the infant.

In case of maternal BMI, Type F exhibits statistical difference across most groups including Type E. Both Type E and Type F consist of girls born to mothers with more than 8 years of education in Hindu or Muslim families. The difference between Type E and Type F is in terms of the wealth of the families.

B. Heterogeneity by Healthcare Policy

B.1. Maternal Height

Having estimated intergenerational transmission of health from mother to infant across different circumstance profiles, we are interested in estimating the change in this coefficient associated with exposure to several prenatal healthcare interventions. We take the subsamples in each terminal leaf from the previous trees fitted through model-based recursive portioning and fit trees using model-based recursive partitioning using only prenatal healthcare variables. The list of prenatal healthcare variables is given in Table 3. This exercise allows us to look at heterogeneity in intergenerational transmission by exposure to policy within a circumstance profile. Put differently, we use model-based recursive partitioning to uncover heterogeneity in intergenerational health transmission by policy exposure holding circumstance profile fixed illustrated by equation 2.2 earlier.

For the previously identified ***type A*** – boys born to mothers with less than eight years of education in Hindu or Muslim households, the type of healthcare worker providing antenatal care has the strongest signal and leads to most parameter instability within this subsample. A small subsample of boys from Type A whose mothers did not receive any prenatal care ***(type Aa)*** have a coefficient of intergenerational transmission of 0.134. Boys whose mothers received antenatal care from community healthcare workers ***(type Ab)*** have a coefficient of intergenerational transmission of 0.34. On the other hand, the subsample of infants whose mothers who received antenatal care from both doctors as well as at least one of the different types of community health workers such as ASHA, Anganwadi or ANM had a coefficient of intergenerational transmission as 0.57. This is illustrated in Figure 4 in the Appendix.

For girls born to mothers with less than eight years of education in Hindu or Muslim households with wealth index in the second or lower quintile of the population ***(type B),*** the number of antenatal visits is associated with heterogeneity in intergenerational transmission of health. Girls whose mothers did not visit any health facility even once for antenatal care ***(type Ba)*** have an intergenerational transmission of 0.32. Girls whose mothers had at most 4 antenatal visits ***(type Bb)*** have an intergenerational transmission coefficient of 0.26, while those who visited more than 4 times ***(type Bc)*** have a coefficient of intergenerational transmission of 0.50. This is illustrated in Figure 5 in the Appendix.

Model-based recursive partitioning algorithm does not split the subset of observations belonging to ***Type C*** – girls with mothers having eight or less years of education in Hindu or Muslim households with wealth index in the third or higher quintile of the population – into further subgroups based on antenatal policy. This implies that this group is homogenous enough that there are no structural breaks in the coefficient of intergenerational transmission of health associated with exposure to healthcare policies.

Infants born to mothers with more than eight years of education but with age less than 24 years in Hindu or Muslim households ***(type D)*** exhibit heterogeneity in intergenerational transmission of health associated with mother's awareness about potential pregnancy complications and number of visits for antenatal care. Infants whose mothers were unaware of either potential pregnancy complications or where to go in case potential complications ***(type Da)*** have a coefficient of intergenerational transmission of 0.52. Of infants whose mothers were made aware of complications or where to go for complications, those who avail up to 4 antenatal care visits ***(type Db)*** had an intergenerational transmission of 0.51, while those who availed more than 4 antenatal care visits ***(type Dc)*** have a coefficient of intergenerational transmission of 0.57. This is illustrated in Figure 6 in the Appendix.

Within the subgroup of boys born to mothers over 24 years of age with more than 8 years of education in Hindu and Muslim families ***(type E)***, heterogeneity is associated with number of visits for antenatal care. Those who had at least four antenatal care visits ***(type Ea)*** have an intergenerational transmission of 0.48 while those who had more than 4 visits ***(type Eb)*** have an intergenerational transmission of 0.67. This is illustrated in Figure 7 in the Appendix.

In case of ***Type F*** – girls born to mothers with > 24 years of age and > 8 years of schooling in Hindu/Muslim households – the intergenerational transmission of health is further mediated by receipt of supplementary food through the ICDS/Anganwadi scheme and awareness of complications surrounding pregnancy. Girls whose mothers did not receive any supplementary food during pregnancy have an intergenerational transmission of 0.50. Of those who do receive supplementary nutrition, those who are unaware of pregnancy complications have an intergenerational transmission of 0.07 while those who are aware have an intergenerational transmission of 0.376. This is illustrated in Figure 8 in the Appendix.

Finally, ***Type G*** consisting of infants in households belonging to religions other than Hindus and Muslims exhibit heterogeneity in intergenerational transmission associated with type of healthcare worker from whom they receive healthcare and supplementary food. Those who either did not receive any prenatal care or received care only from community healthcare workers and did not receive any supplementary food (***type Ga)*** have an intergenerational transmission of health of 0.1529. Those who either did not receive any prenatal care or received care only from community healthcare workers but did receive supplementary food (***type Gb)*** have an intergenerational transmission of health of 0.29. In contrast, infants who received antenatal care from a combination of community healthcare works and doctors have intergenerational transmission of 0.25 while those who received care from only doctors have an intergenerational transmission of 0.50. This is illustrated in Figure 9 in the Appendix.

B.2. Maternal BMI

A similar analysis is conducted on the heterogenous groups identified by employing model-based recursive partitioning on the relationship between infant birthweight and mother's BMI. Boys born to mothers with less than 8 years of education in Hindu and Muslim families previously characterised as ***type A*** exhibit further heterogeneity associated with the type of healthcare worker who provides antenatal care and the number of antenatal care visits. Infants whose mothers did not receive any form of antenatal care ***(type Aa)*** comprising 5.5% of type A have an intergeneration transmission of 0.105. Infants whose mothers received antenatal care from only traditional and community healthcare workers such as ASHA, Dai or Anganwadi workers ***(type Ab)*** consisting of 41.5% of sample of type A have an intergeneration transmission of health of 0.09. Out of the infants whose mothers received antenatal care from a combination of community healthcare workers and doctors, there is further heterogeneity associated with intensity of antenatal visits. Infants with up to 4 antenatal care visits ***(type Ac)*** which consist of 28.4% of type sample have an intergenerational coefficient of transmission of

0.168. The remaining 24.5% comprising infants with more than 4 antenatal care visits ***(type Ad)*** have an intergenerational transmission of 0.185. This is illustrated in Figure 11 in the Appendix.

In the subsample of **type B** – girls born to mothers with less than 8 years of education in Hindu and Muslim households whose wealth lies in the second or lower quintile of the population, heterogeneity is associated with awareness regarding pregnancy complications, supplementary food and adequate antenatal care visits. Girls whose mothers were not aware of potential pregnancy complications nor where to go in case of complications and did not receive any supplementary food through the ICDS/Anganwadi schemes during pregnancy ***(type Ba)*** have an intergenerational transmission of 0.26. They comprise close to 9% of the sample of type B. On the other hand, girls whose mother were wholly unaware about complications during pregnancy but received supplementary food during pregnancy ***(type Bb)*** comprise 9.6% of type B have an intergenerational transmission of 0.23. Of the girls whose mothers were aware of either pregnancy complications or where to treat complications, there is further heterogeneity associated with number of antenatal visits. Those with up to 4 antenatal visits ***(type Bc)*** comprising 51.4% of the sample of type B have an intergenerational transmission of 0.12, while those with more than 4 antenatal visits ***(type Bd)*** comprising about 31% of type B have an intergenerational transmission of 0.17. Figure 12 in the Appendix illustrates this tree.

Consistent with the results from the case with mother's height, the sample of ***type C*** consisting of girls with mother's with less than 8 years of education in Hindu or Muslim households with wealth in the third of higher quintile of the population are homogeneous with respect to exposure to policy. Model-based recursive partitioning algorithm does not detect sufficient heterogeneity in the coefficient of intergenerational transmission related to exposure to policy in this group.

For infants in the group identified as type D – boys with mothers having more than 8 years of education in Hindu or Muslim households, heterogeneity in intergenerational transmission is associated with awareness about complications during pregnancy and number of antenatal care visits. Boys whose mothers were unaware of potential complications and where to go to treat complications ***(type Da)*** make up 11.6% of the type D subsample and have an intergenerational transmission of 0.17. Of those who are aware of pregnancy complications, those who availed up to 4 antenatal care visits ***(type Db)*** comprise 40.6% of type D and have an intergenerational transmission of 0.14. Those who availed more than 4 antenatal care visits ***(type Dc)*** make up 47.7% of type D and exhibit an intergenerational transmission of 0.11. This is illustrated in Figure 13 in the Appendix.

For girls born to mothers with more than 8 years of education in Hindu or Muslim households with wealth index in the third or lower quintile of the population ***(type E)*** intergenerational transmission of health is affected by number of antenatal care visits. Those who did not have any antenatal care visits have an intergeneration transmission of 0.19 ***(type Ea).*** Those who availed up to 4 antenatal care visits ***(type Eb)*** have intergenerational transmission of 0.19 while those who availed more than 4 visits have intergenerational transmission of 0.16 ***(type Ec).*** This is illustrated in Figure 14 in the Appendix.

Girls who are born to mothers with more than 8 years of education in Hindu or Muslim households with wealth in the fourth or fifth quintile of the population (type F) exhibit heterogeneity in intergenerational transmission with respect to awareness about pregnancy complications. Infants whose mother are unaware of complication related to pregnancy and where to get treated for complications ***(type Fa)*** make up 10.3% of subsample type F and have an intergenerational transmission of 0.12. Infants whose mothers are aware of either potential complications or where to go for treatment in case of complications ***(type Fb)*** make up the remaining 89.7% and have an intergenerational transmission of 0.09. This is illustrated in Figure 15 in the Appendix.

Infants born in households other than Hindus or Muslims (type G) have differential coefficients of intergenerational transmission of health associated with type of facility where they receive antenatal care and receipt of supplementary food during pregnancy. Infants who received absolutely no antenatal care or some antenatal care from community health workers, and no supplementary food ***(type Ga)*** comprise only 7.4% of the sample of type G and have an intergenerational transmission of 0.16. Infants who received absolutely no antenatal care or some antenatal care from community health workers, but do receive supplementary food ***(type Gb)*** make up about 75% of type G and have an intergenerational transmission of 0.18. Infants who received antenatal care from a combination of community healthcare workers and doctors ***(type Gc)*** comprise 8.9% of type G and have an intergenerational transmission of 0.14 while infants who received care from only doctors ***(type Gd)*** are 8.9% of type G with an intergenerational transmission of 0.12. Figure 16 in the Appendix illustrates this tree.

Concluding Discussion

This paper estimates heterogeneity in intergenerational transmission in health by maternal circumstance and exposure to policy. Identifying these circumstances allows us to identify different profiles of socio-economic characteristics that contribute to heterogeneity in birthweight across the population. We focus on birthweight as it is often the start of health inequalities and also the time of life when circumstances matter the most. Additionally, we focus on infants, i.e. children aged one year or less to focus on contemporaneous intergenerational transmission of maternal health. Focusing on contemporaneous transmission also allows us to explore the effects of socioeconomic characteristics and prenatal health care policies on intergenerational transmission separately. Results illustrate considerable heterogeneity across the identified subgroups indicating that a global model on the full sample is not appropriate to model intergenerational health transmission. We find that heterogeneity is primarily driven by religion, maternal education, maternal age, household wealth index and infant’s gender.

The contributions of this paper are two-fold. First, it makes a methodological contribution in demonstrating how a Machine Learning algorithm to measure intergenerational transmission of nutritional status across several circumstance profiles and exposure to healthcare. Second, it provides new evidence on the heterogeneity in the intergenerational transmission of health. We identify and characterize sets of socio-economic, demographic attributes that constitute different circumstance profiles with varying coefficients of intergenerational transmission. Results demonstrate sufficient heterogeneity in the coefficient of transmission with worse-off circumstances exhibiting a higher magnitude of transmission. The totality of the results establishes some stylized facts and contributes to our understanding of the heterogeneity in intergenerational transmission of nutritional status and the role robust antenatal care plays in tempering this transmission.

References

1. Bevis, Leah EM, and Kira Villa. 2022. "Intergenerational Transmission of Maternal Health Evidence from Cebu, the Philippines." Journal of Human Resources 57 (5): 1425–65.
2. Bhalotra, Sonia, and Samantha Rawlings. 2013. "Gradients of the Intergenerational Transmission of Health in Developing Countries." Review of Economics and Statistics 95 (2): 660–72.
3. Bhalotra, Sonia, and Samantha B. Rawlings. 2011. "Intergenerational Persistence in Health in Developing Countries: The Penalty of Gender Inequality?" Journal of Public Economics 95 (3–4): 286–99.
4. Coneus, Katja, and C. Katharina Spiess. 2012. "The Intergenerational Transmission of Health in Early Childhood—Evidence from the German Socio-Economic Panel Study." Economics & Human Biology 10 (1): 89–97.
5. Currie, Janet, and Enrico Moretti. 2007. "Biology as Destiny? Short-and Long-Run Determinants of Intergenerational Transmission of Birth Weight." Journal of Labor Economics 25 (2): 231–64.
6. Eriksson, Tor, Jay Pan, and Xuezheng Qin. 2014. "The Intergenerational Inequality of Health in China." China Economic Review 31: 392–409.
7. Giuntella, Osea, Giulia La Mattina, and Climent Quintana-Domeque. 2022. "Intergenerational Transmission of Health at Birth: Fathers Matter Too!"
8. Halliday, Timothy, Bhashkar Mazumder, and Ashley Wong. 2021. "Intergenerational Mobility in Self-Reported Health Status in the US." Journal of Public Economics 193: 104307.
9. Kumar, Santosh, and Bernard Nahlen. 2023. "Intergenerational Persistence of Health: Evidence from India." Economics Letters 224: 111023.
10. Subramanian, S. V., Leland K. Ackerson, George Davey Smith, and Neetu A. John. 2009. "Association of Maternal Height with Child Mortality, Anthropometric Failure, and Anemia in India." Jama 301 (16): 1691–1701.
11. Venkataramani, Atheendar S. 2011. "The Intergenerational Transmission of Height: Evidence from Rural Vietnam." Health Economics 20 (12): 1448–67.
12. Zeileis, Achim, Torsten Hothorn, and Kurt Hornik. 2008. "Model-Based Recursive Partitioning." Journal of Computational and Graphical Statistics 17 (2): 492–514. https://doi.org/10.1198/106186008X319331.

Appendix

Table 3: Summary of socioeconomic variables

	N	Percent	Mean	SD	Min	Max
Rural						
Yes	29853	80.34				
No	7306	19.66				
Household head is a female						
Yes	5,775	15.54				
No	31,384	84.46				
Wealth index						
poorest	8034	21.62				
poorer	7773	20.92				
middle	7340	19.75				
richer	7138	19.21				
richest	6874	18.50				
Child is a female						
Yes	18101	48.71				
No	19058	51.29				
Religion of household head						
Hindu	28466	76.61				
Muslim	4344	11.69				
Christian	2808	7.56				
Others	1541	4.15				
Caste of household head						
Others	6171	16.61				
SC	8256	22.22				
ST	8078	21.74				

	N	Percent	Mean	SD	Min	Max
OBC	14654	39.44				
Mother's age			25.65	4.76	15	49
Age of household head			46.42	15.33	15	95
Household Size			6.28	2.55	2	34
Order of birth			2.08	1.28	1	14
Mother's education (years)			8.21	5.03	0	20
Household head's education (years)			6.28	2.55	0	20
Total	37159	100				

Table 4: Summary of policy variables

	N	Percent	Mean	SD	Min	Max
Prenatal care received from						
No prenatal care	1647	4.43				
ASHA/Dai/ANM/Midwife/Nurse/Others	12404	33.38				
Doctor and ASHA/Dai/ANM/Midwife/Nurse/Other	13671	36.79				
Only Doctor	9437	25.4				
Aware of pregnancy complications						
Yes	31722	85.37				
No	5437	14.63				
Received supplementary food from Anganwadi/ICDS centre during pregnancy						
Yes	26186	70.47				
No	10973	29.53				
Antenatal visits						
No visits	1647	4.43				
<= 4 visits	19111	51.43				
> 4 visits	16401	44.14				

# Tetanus injections			1.91	0.69	0	7
Total	37159	100				

Table 5: Estimates of Intergenerational Transmission over the full sample

Full Sample Regression results				
	log(birthweight)			
	(1)	(2)	(3)	(4)
log(height_mother)	0.421*** (0.027)	0.419*** (0.027)		
log(BMI_mother)			0.128*** (0.007)	0.127*** (0.007)
Socio-economic Controls	✓	✓	✓	✓
Policy Controls		✓		✓
Observations	37,159	37,159	37,159	37,159
Adjusted R^2	0.035	0.036	0.038	0.038

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6: Description of types and transmission coefficients (Mother's BMI)		
Types	**Description**	**Coefficient**
A	Male children and less educated mothers in HHs with Hindu or Muslim head	0.144
Aa	Received no prenatal care	0.105
Ab	Received prenatal care from CHWs	0.092
Ac	Received prenatal care from a combination of doctors and CHWs and <=4 antenatal visits	0.168
Ad	Received prenatal care from a combination of doctors and CHWs and >4 antenatal visits	0.185
B	Girls born to mothers with ≤ 8 years of schooling in poor (wealth index ≤2nd quintile) HHs with Hindu or Muslim head	0.159
Ba	Unaware of pregnancy complications and not received supplementary food during pregnancy	0.265
Bb	Unaware of pregnancy complications but received supplementary food from during pregnancy	0.232
Bc	Aware of pregnancy complications and <=4 antenatal visits	0.120
Bd	Aware of pregnancy complications and >4 antenatal visits	0.170
C	Girls born to mothers with ≤ 8 years of schooling in richer (wealth index > 2nd quintile) HHs with Hindu or Muslim head	0.116
D	Male children and more educated mothers in HHs with Hindu or Muslim head	0.137
Da	Unaware of pregnancy complications	0.176
Db	Aware of pregnancy complications and <=4 antenatal visits	0.148
Dc	Aware of pregnancy complications and >4 antenatal visits	0.115
E	Girls born to more educated mothers with ≥ 8 years of schooling in poor and middle income (wealth index ≤ 3rd quintile) HHs with Hindu or Muslim head	0.178
Ea	No antenatal care visits	0.194
Eb	At least 4 antenatal care visits	0.197
Ec	More than 4 antenatal care visits	0.162
F	Girls born to more educated mothers in with ≥ 8 years of schooling in richer (wealth index > 3rd quintile) HHs with Hindu or Muslim head	0.099
Fa	Unaware of pregnancy complications	0.124
Fb	Aware of pregnancy complications	0.096
G	HHs with non-Hindu or non-Muslim head	0.154
Ga	Received no antenatal care or only from CHWs and no supplementary nutrition	0.162
Gb	Received no antenatal care or only from CHWs and received supplementary nutrition	0.180
Gc	Received antenatal care from a combination of CHWs and doctors	0.149
Gd	Received antenatal care from only doctors	0.128

Table 7: Description of types and transmission coefficients (Mother's height)		
Types	**Description**	**Coefficient**
A	Boys born to mothers with <8 years of schooling in HHs with Hindu or Muslim head	0.466
Aa	Received no prenatal care	0.134
Ab	Received prenatal care from Community Healthcare Workers	0.349
Ac	Received prenatal care from a combination of doctor or doctor CHWs	0.573
B	Girls born to mothers with <8 years of schooling in poor HHs (wealth index ≤2nd quintile) with Hindu or Muslim head	0.360
Ba	No antenatal visits	0.323
Bb	<=4 antenatal visits	0.269
Bc	> 4 antenatal visits	0.562
C	Girls born to mothers with < 8 years of schooling in richer HHs (wealth index > 2nd quintile) with Hindu or Muslim head	0.485
D	Infants born to mothers with more than 8 years of schooling and aged ≤ 24 years in HHs with Hindu or Muslim head	0.549
Da	Unaware of pregnancy complications	0.522
Db	Aware of pregnancy complications and availed ≤ 4 antenatal visits	0.517
Dc	Aware of pregnancy complications and availed > 4 antenatal visits	0.574
E	Boys born to mothers aged >24 years with > 8 years of schooling in HHs with Hindu or Muslim head	0.600
Ea	Availed ≤4 antenatal care visits	0.486
Eb	Availed > 4 antenatal visits	0.673
F	Girls born to mothers aged >24 years with > 8 years of schooling in HHs with Hindu or Muslim head	0.393
Fa	Not received supplementary food during pregnancy	0.502
Fb	Aware of pregnancy complications	0.076
Fc	Unaware of pregnancy complications	0.376
G	HHs with non-Hindu or non-Muslim head	0.306
Ga	Received either no antenatal care or care from CHWs and no supplementary food	0.152
Gb	Received antenatal care from CHWs and received supplementary food	0.294
Gc	Received antenatal care from a combination of CHWs and doctors	0.257
Gd	Received antenatal care from only doctors	0.150

Figure 3: Main Tree built by Model Based Recursive Partitioning for Intergenerational Transmission of Mother's Height on Infant Birthweight

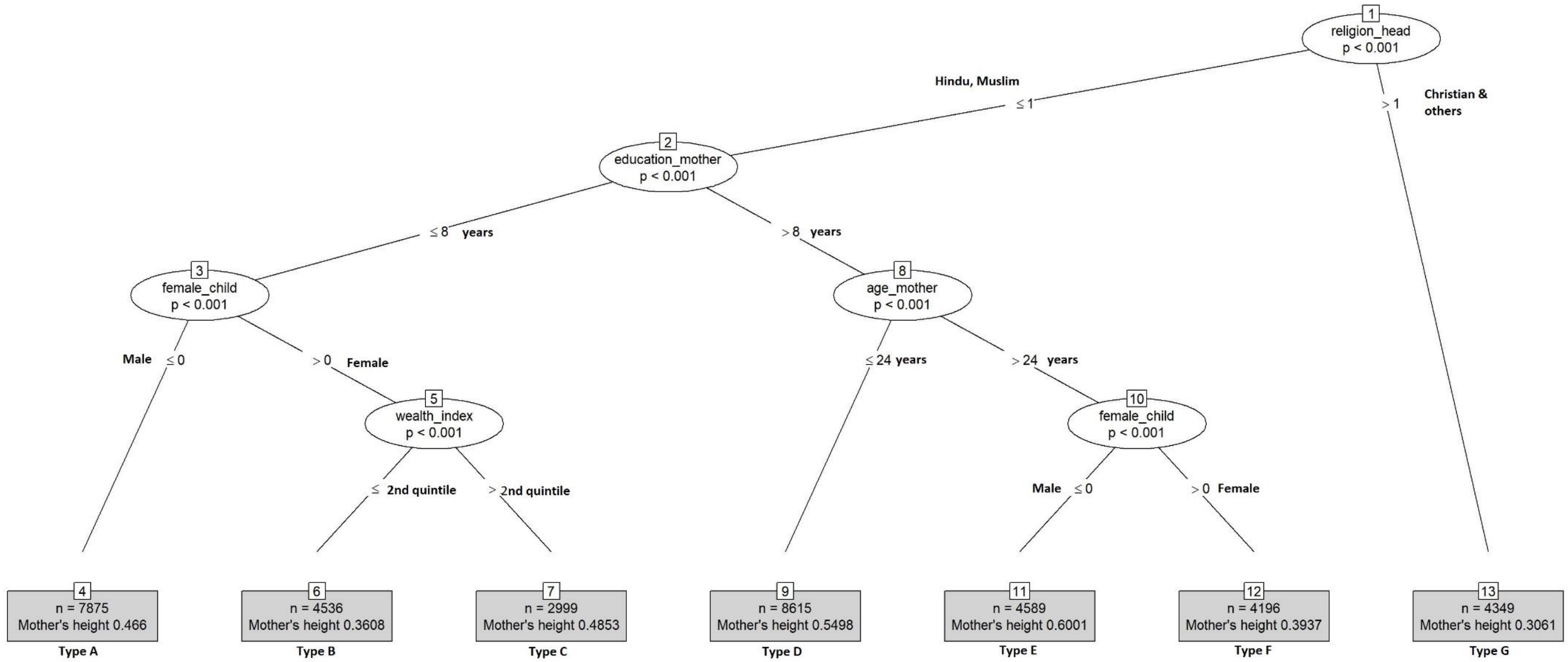

Figure 4: Type A subtree

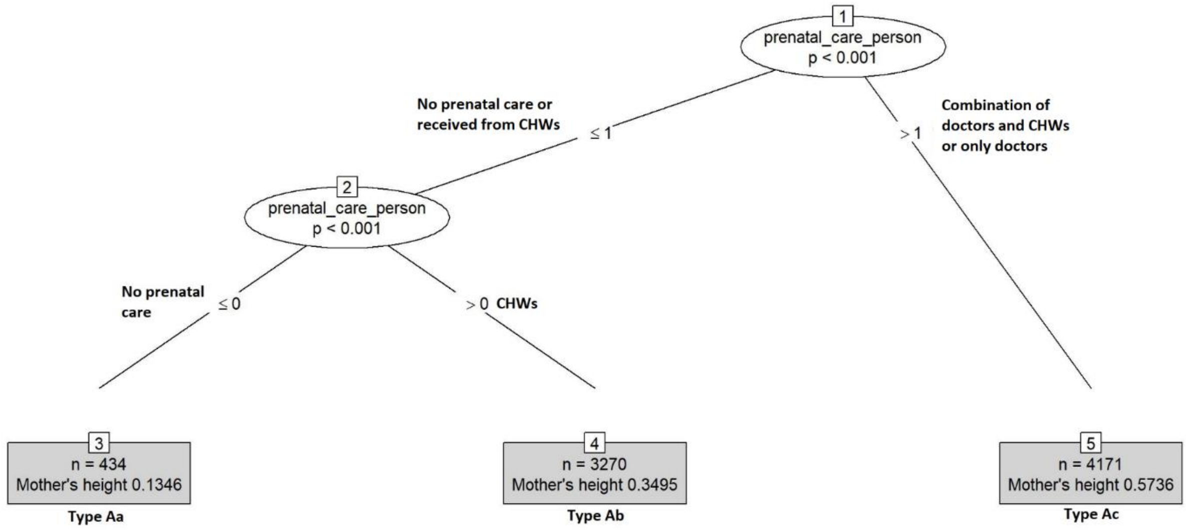

Figure 5: Type B Subtree

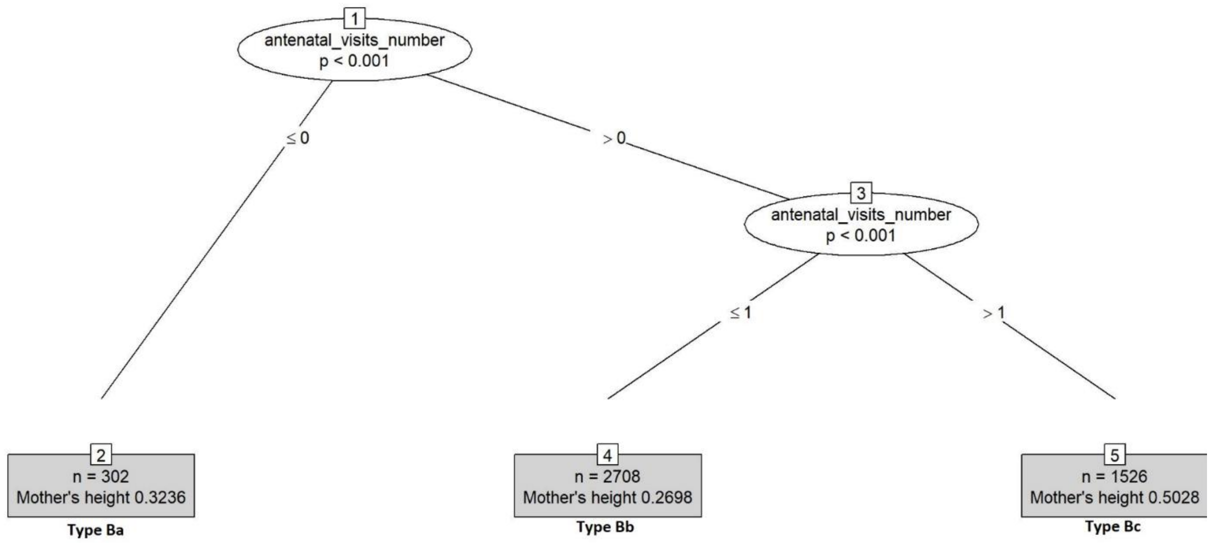

Figure 6:Type D Subtree

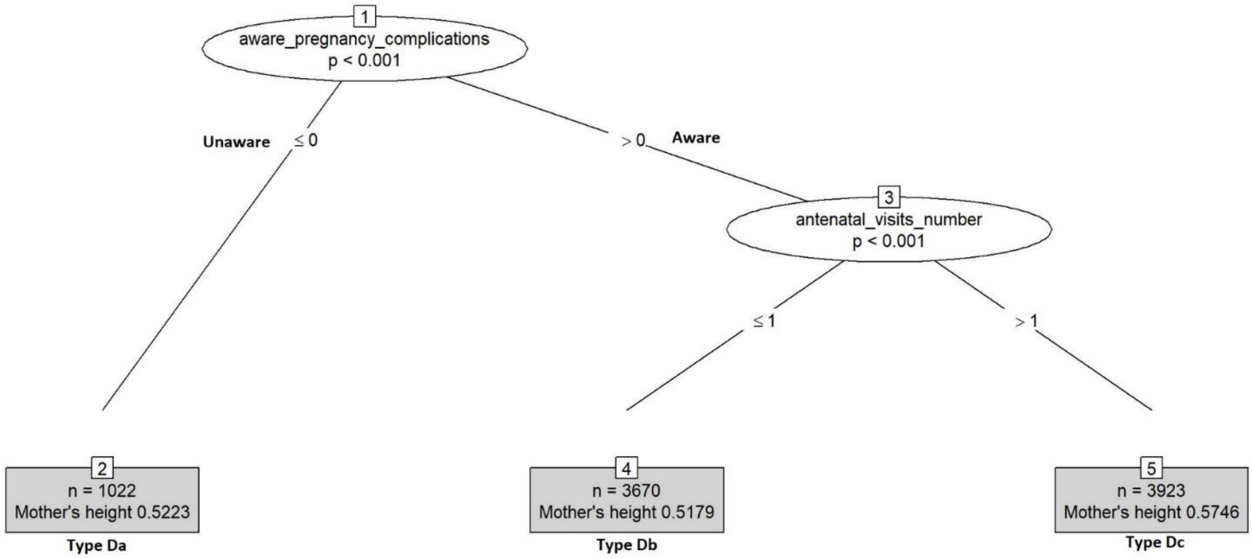

Figure 7:Type E Subtree

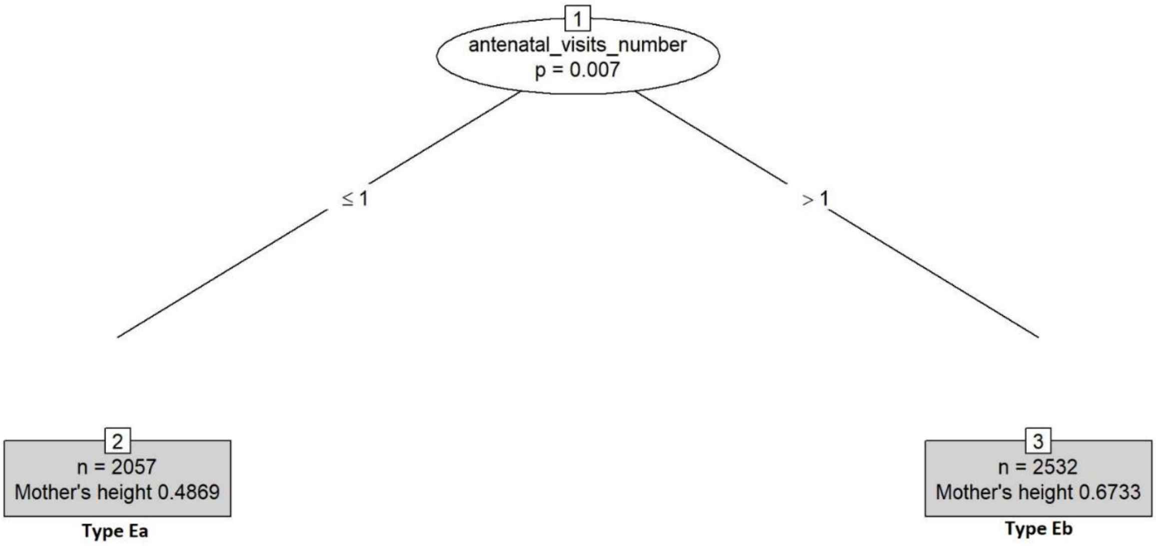

Figure 8: Type F Subtree

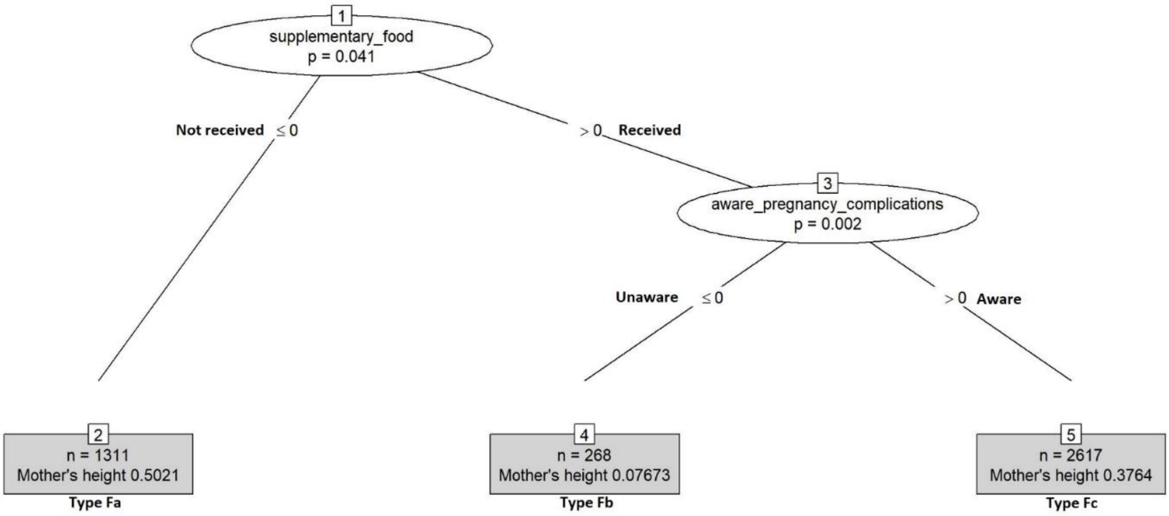

Figure 9: Type G Subtree

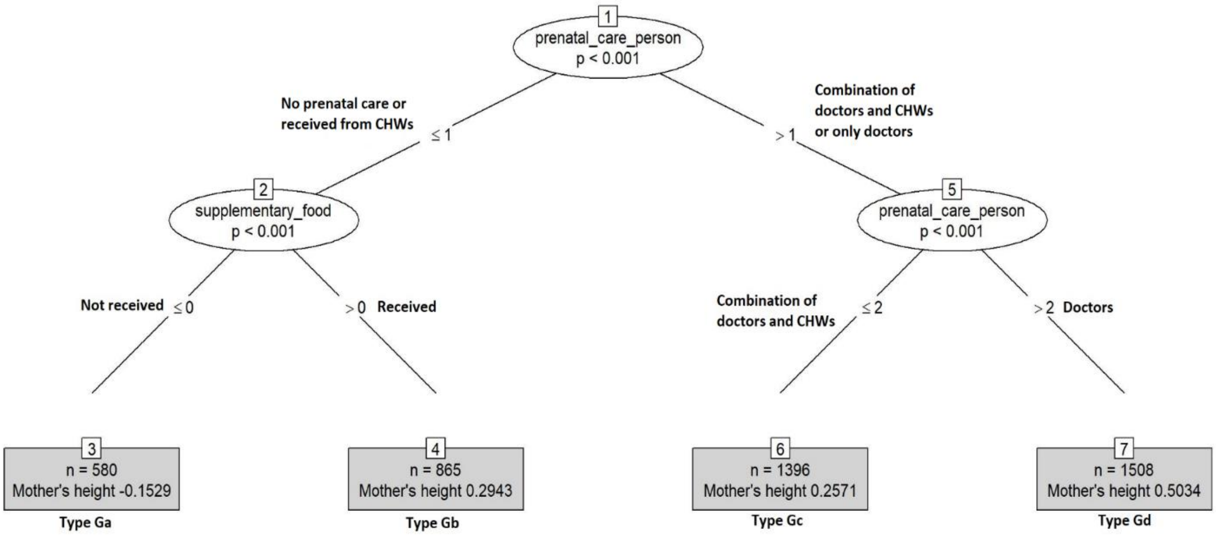

Figure 10: Main Tree built by Model Based Recursive Partitioning for Intergenerational Transmission of Mother's BMI on Infant Birthweight

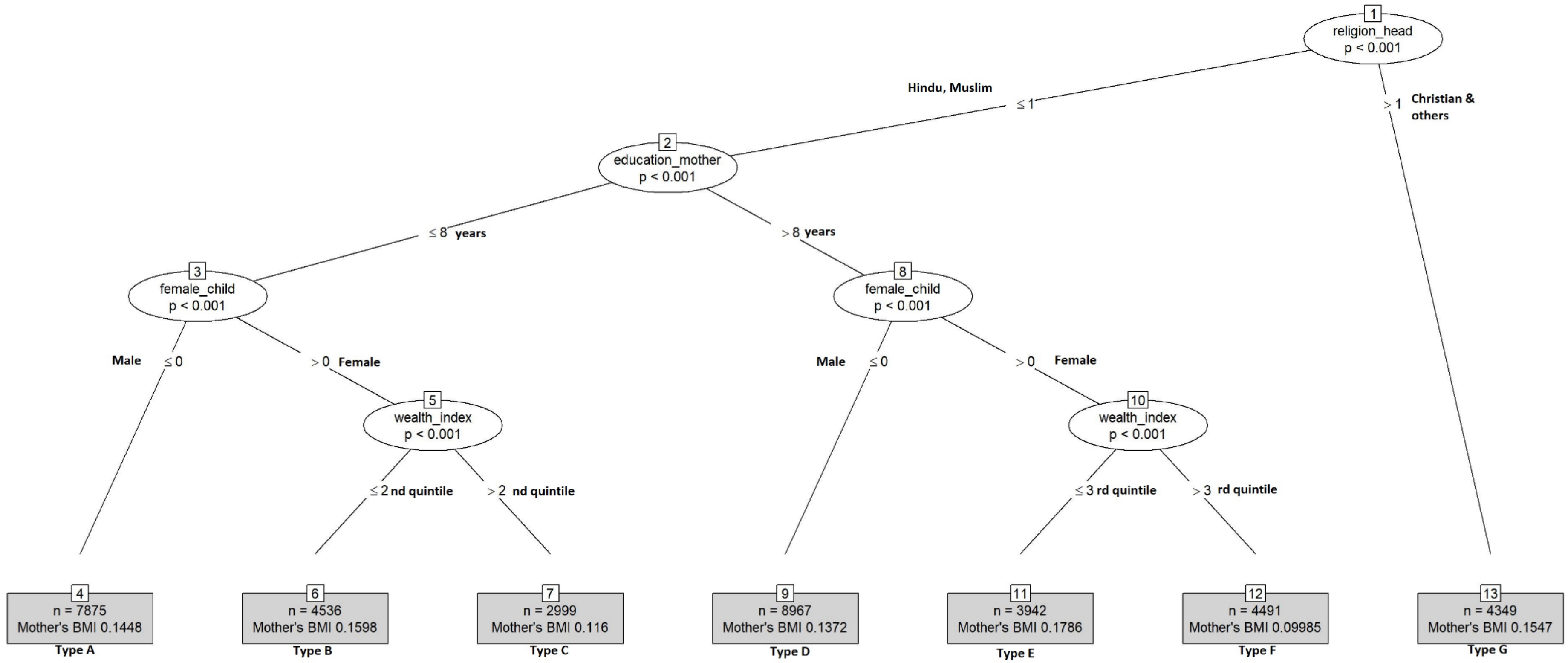

Figure 11: Type A subtree

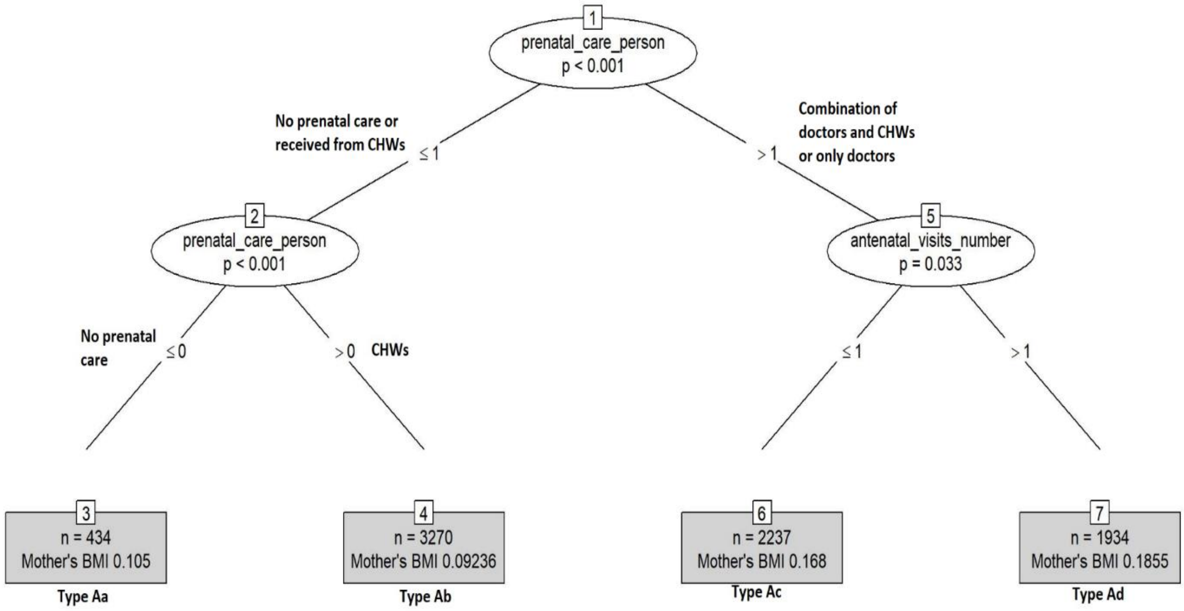

Figure 12: Type B Subtree

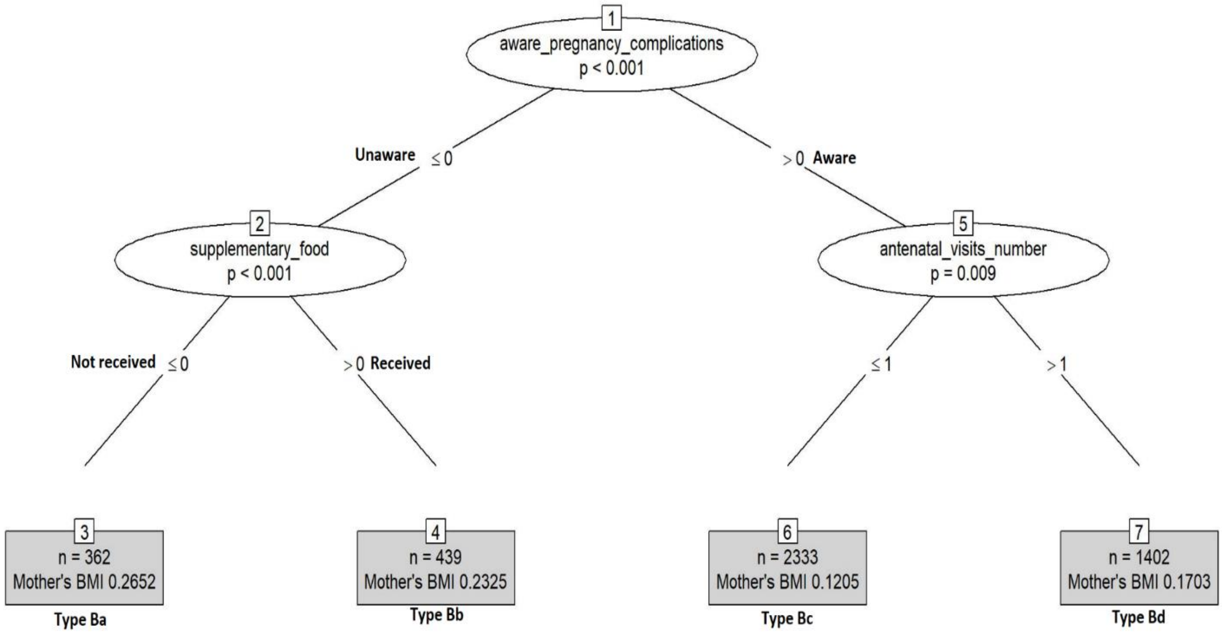

Figure 13: Type D Subtree

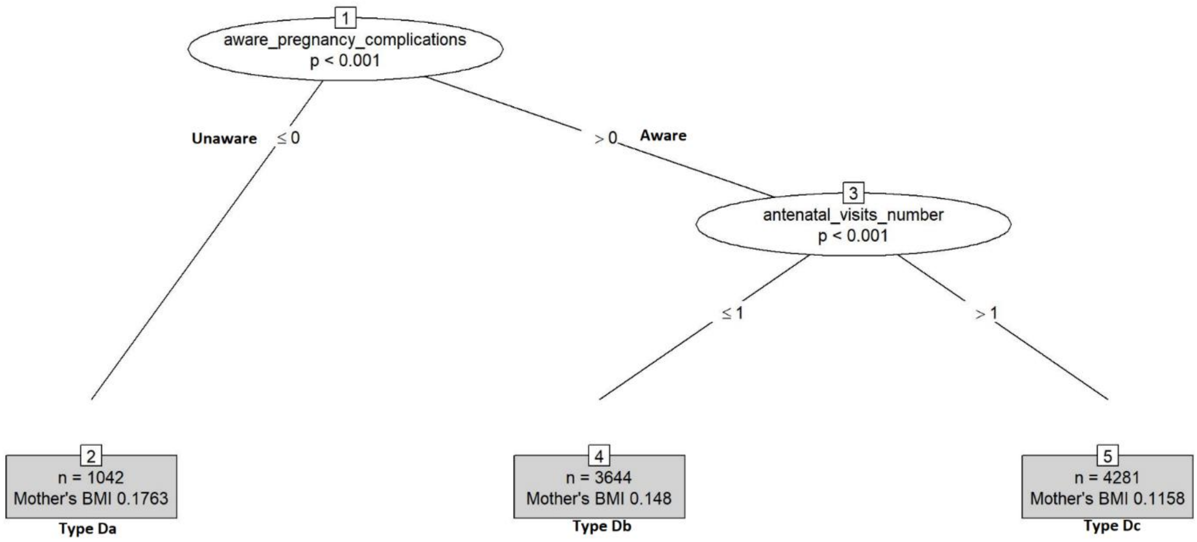

Figure 14: Type E Subtree

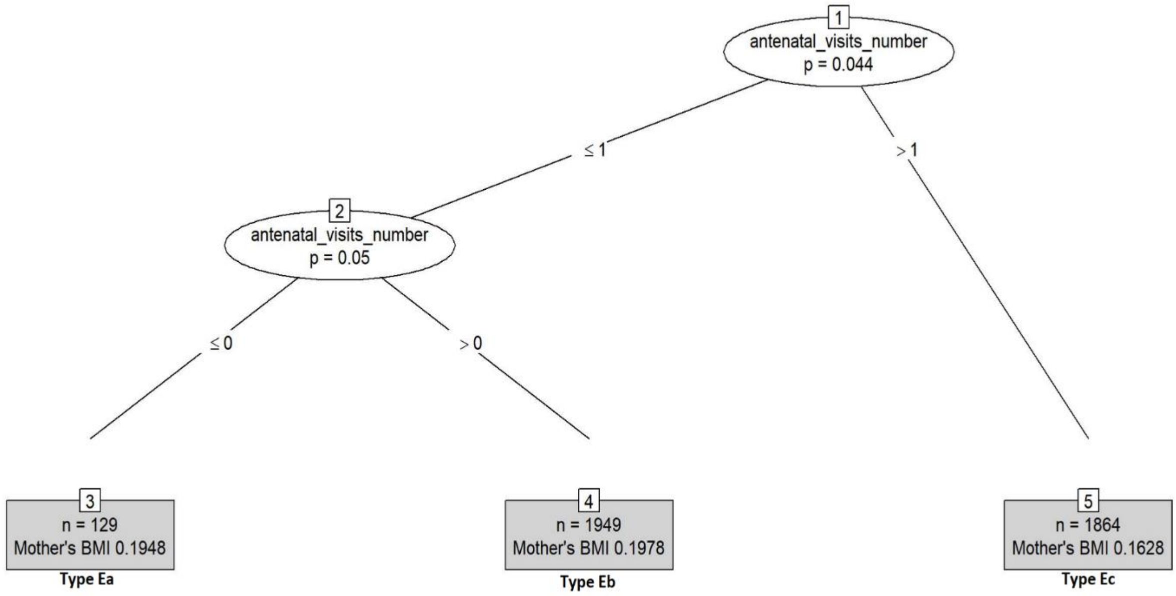

Figure 15: Type F Subtree

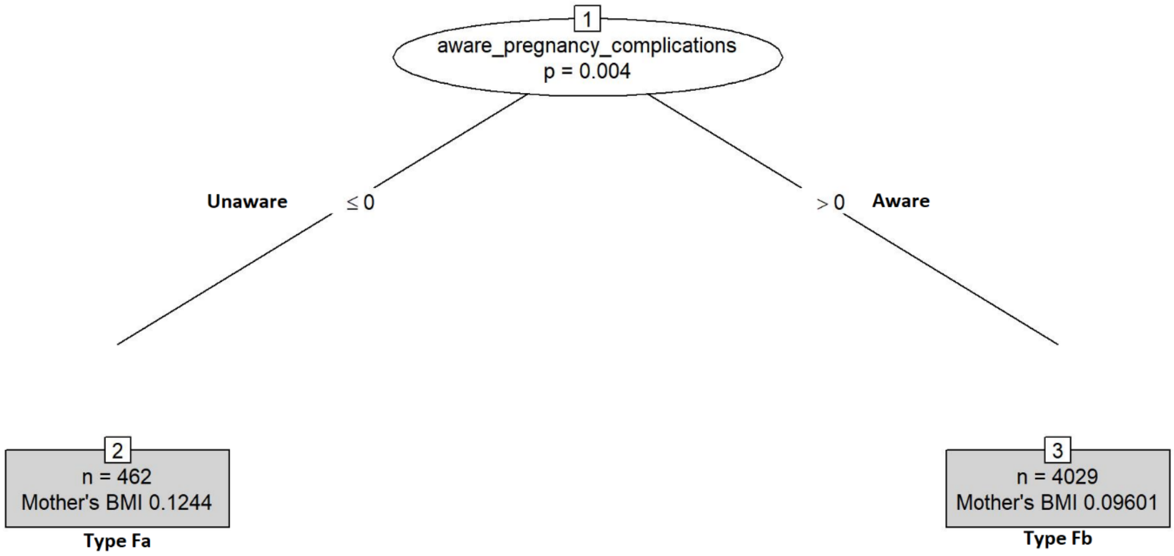

Figure 16: Type G Subtree