Association between Iron Deficiency Anemia and Wages in India: A secondary data analysis

Sanchit Makkar, St. John's Research Institute, sanchit.m@sjri.res.in

Selected Paper prepared for presentation at the 2019 Agricultural & Applied Economics Association Annual Meeting, Atlanta, GA, July 21 – July 23

Association between Iron Deficiency Anemia and Wages in India: A secondary data analysis

Sanchit Makkar[1], Sumedha Minocha[1], Kishor Bhat[1], Anjaly Teresa John[1], Sumathi Swaminathan [1], Tinku Thomas[2], Venkatesh Mannar[3], Anura V Kurpad[1,4]

1. Division of Nutrition, St. John's Research Institute, Bengaluru, India
2. Department of Bio-statistics, St. John's Medical College, Bengaluru, India
3. University of Toronto, Canada
4. Department of Physiology, St. John's Medical College, Bengaluru, India

Abstract

Lack of diet diversification coupled with inadequate environmental conditions have resulted in numerous health issues in India. Anemia is one such problem that has continuously affected the wellbeing, particularly productivity and cognitive level of the population. Policy level interventions such as iron supplementation have not proved to be sufficiently effective largely due to compliance issues. Another alternative is Double Fortified Salt (DFS), which is fortified with iodine and iron. For the effective implementation of any food and health policies, it becomes important to understand its inclusiveness. But prior to that, analyses of economic benefits further provides impetus for improvement and successful running of any programme. The present study was divided into two phases. The first phase, presented in this paper, explored the association between iron deficiency anemia (IDA) and wages of adult men and women between 15- 49 years in India. In the second phase, these results were used to evaluate monetary benefits by the roll out of an iron fortification programme, and were compared with its cost. The analysis was performed by deploying statistical, mathematical and economic modelling techniques on available nationally representative datasets. The results suggest that the presence of IDA resulted in a decline of 19%, 12.1% and 6% in wages of regular salaried employed men& women and male casual laborers respectively as compared to those without IDA. However, the presence of IDA had no significant impact on the wages of female casual labourers. As the earnings of self-employed were not reported in the survey, the association between anemia and the earnings of individuals in this category could not be determined. Thus, addressing anemia through a universal iron fortification program such as DFS could possibly change living standards of the population through improvement in earnings in the short run. The monetary benefits associated with DFS would not only depend on the expected improvement in wages but also the associated reduction in IDA through the consumption of DFS. However, the long-term implications of DFS need to be carefully evaluated as excessive consumption of iron can lead to toxicity in men and women both.

Keywords:

NSSO, Iron deficiency Anemia, Wages, Fortification

1. Introduction

Anemia continues to be an extensive public health problem in India despite years of policy efforts. The groups most severely affected are women and children with prevalence rates greater than 50% in these groups, although, in Indian men too, a moderate level of prevalence at 23% has been reported (1). It is estimated that about 50% of anemia cases reported worldwide are due to iron deficiency, referred to as iron deficiency anemia (IDA) (2,3), but there are various other factors such as poor socio-economic environment, lack of diet diversification and awareness about the local availability of other erythropoietic micronutrient rich foods, and hereditary disorders which together add up to the total burden of anemia.

Iron deficiency is of a concern because it decreases the circulation of oxygen in the blood affecting metabolic pathways, increasing the susceptibility to infections and the likelihood of experiencing weakness and fatigue, thus effecting other downstream functions (4). The consequences are much more severe in children because it can even cause a long- term negative impact on cognition (5–7). In women, anemia is associated with increased risk of low birthweight and prematurity, perinatal and neonatal mortality, inadequate iron stores for the new born, and maternal morbidity and mortality (8,9). As a consequence, in adults, it can cause a reduction in productivity and work capacity, therefore, effecting the earnings (10,11). The Global Burden of Diseases Study (GBD, 2016) ranked iron deficiency anemia (IDA) to be the leading cause of "years lived with disability" (YLD), specifically in India. In fact, for India, IDA accounted for 3.46% of total burden of diseases, as measured by disability adjusted life years (DALY)(12).

Even in the case of IDA, it is important to understand that it is not simply caused by inadequate intake of iron- rich foods in isolation from other factors. Rather, there are various nutrients which inhibit or enhance the absorption of iron as well as there are other causes of iron loss such as that due to parasites. In this context, diet diversification plays an important role as it not only increases the intake of iron but also makes it more bioavailable with an increase in the intake of other nutrients such as vitamin C, vitamin B12, vitamin A and folate that help in the absorption of iron. Equally, some nutrients such as phytate and phytic acid inhibit iron absorption. The presence of inhibitors is common in developing countries such as India, where diets are more uniform, focused on cereal-based foods and less on other plant and animal source foods (13).

Further, the burden of anemia is unlikely to decrease in regions with poor sanitation and hygiene practices (14,15). At the national level, about 62% of households reported that they did not treat water prior to drinking, and the problem is higher in rural areas as compared to urban areas (1). Of the rural households, around 54% reported that they practiced open defecation and among the household where place for handwashing was observed, 50% household in reported that they did not practice handwashing with both soap and water (1).

The multi-faceted nature of the problem means that programs should not only be focused on improving the quality of the diet but also addressing other environmental determinants such as improvement in water and sanitation. In terms of improving the quality of diets, however, production pattern and consumption habits are unlikely to change in the short run and could be addressed through supplementation and food fortification. Indeed, World Health Organization (7,16) recommends to adopt large-scale Weekly Iron and Folic Acid Supplementation (WIFS) programs for vulnerable groups, and several countries including India have embarked upon such efforts over the past decade (17). These programs have had limited success in India due to compliance issues and poor implementation. (18).

An alternative solution is enhancing the iron content of a commonly consumed food item through fortification; this solution was ranked amongst the top three international development priorities at the Copenhagen Consensus (2008) and generated much excitement in policy circles (19). With India being a culturally diverse country there are very few foods that can penetrate all sections of the diverse population. Equally important is the fact that the fortified food should not be consumed in large quantities, given the risks associated with overconsumption of iron. In this context, salt is used universally by all sections of the population The focus of the present study is on DFS(20) which in addition to iodine, is fortified with iron. The government of India is likely to roll out DFS at a large scale in several states in India not only through governmental programs but also through the commercial route. Based on National Sample Survey Organization (NSSO) household consumer expenditure survey, the per capita intake of salt is around 8 gms (21). Given the high coverage of iodized salt in India (93%), minimal operational bottlenecks can be expected in the distribution of DFS (1). Hence, DFS is considered a potentially valuable vehicle for providing iron and reducing the burden of IDA in India.

The objective of the present study was to understand the economic implications of DFS in reducing the burden of anemia though a program with universal coverage, specifically iron deficiency anemia, through a cost- benefit analysis using secondary data analyses. In doing so, the aim was to evaluate the benefits of providing DFS which extended beyond reducing the burden of the given micronutrient deficiency and potentially helped in improving the quality of life through the enhancement of the wages of the target population. The study was divided into 2 phases. The present paper presents the results from the first phase in which the association between IDA and wages was modelled by deploying mathematical, statistical and econometric techniques on nationally representative survey data. The results of the model were used in the next phase of the study in which the potential benefits of DFS were calculated by simulating the mathematical model to establish the reduction in the distribution of IDA by increased intake of iron through DFS, and the expected improvement in wages through the intake of DFS.

This paper contains a section describing the data used; a methodological framework section that discusses all the models used; the results obtained; and concluding with a discussion.

2. Data

Nationally representative survey data were used for the analysis. The primary source of data used was 68th round of unit level data of consumer expenditure survey and the employment-unemployment survey collected by National Sample Survey Organization (NSSO) in 2011-12.

The consumer expenditure survey provided information on quantity and value of commodities consumed by the household for both food and non-food items with a recall period of 30 days for most items. It provided estimates of consumption of different food items in Indian diets across the country. About 223 foods were listed of which 197 were raw foods. The consumption of these foods was linked to food composition tables to get estimates of intake of iron (total and heme and non-heme iron), calcium, phytates, polyphenols, ascorbic acid, at the household level. The data sources used were the Indian Food Composition Tables (IFCT) 2017(22) and 1996(23), and United States Department of Agriculture (USDA) Standard Reference Release 28 (24).

The employment-unemployment survey provided statistical indicators of various characteristics of labour force related to employment status, unemployment status, type of employment and other socio-economic indicators through which the labour market in India was understood. The survey also provided information on earnings, specifically of casual labors and regular/ salaried employees. Though the survey identified self- employed, the information on the earnings of this category was not provided.

Further, national level prevalence of anemia from the National Family Health Survey 2015-16 (NFHS-4) was used to fit the IDA predicted using the mathematical model. For the purpose of validation, an external primary dataset collected from a baseline survey in Jharkhand was used which had information on food consumption and measured level of haemoglobin (Hb) and ferritin for a sample of 374 women. The survey was for the purpose of evaluating the impact of distribution of DFS through the Public Distribution System (PDS).

3. Methodological Framework for the analysis

A mathematical model was used to predict the status of IDA in individuals (25). This model used individuals' intake of nutrients, specifically calcium, phytate, polyphenols, ascorbic acid, heme and non-heme iron from the NSSO consumer expenditure survey adjusting for bioavailability, sanitation and menstrual losses in women.

A statistical matching technique was performed to triangulate the NSSO employment-unemployment unit level data with the NSSO consumer expenditure unit level data to create the final dataset. The final data included individual level information on wages, nutrient intake, and the predicted individual levels of IDA from the mathematical model. The Heckman Selection Model was used to establish the association between IDA and wages from this dataset.

The description of the framework is represented diagrammatically in **Figure 1**. The analysis was performed using R version 3.4.1, Stata version 14 and Python 3.

3.1 Mathematical model to predict the prevalence of iron deficiency anemia

The model used was based on the assumption that the iron status (Hb and iron stores) of an individual regulates the fractional absorption of iron (25). The degree to which this regulation occurs was taken as the marker of iron status. The model used was developed based on a previous model (25), where body's own regulation of iron absorption was incorporated to model intervention effects of iron supplementation.

Iron in the body store is considered to be in homeostasis when intake and excretion of iron are matched. The excretion of iron is taken as a constant for every individual based on sex.

The absorption of iron is regulated dynamically such that iron balance can be achieved at any intake. The model first calculated the fractional rate of iron absorption by comparing iron absorbed with iron consumed. Next, the measured rate was compared with a normative fractional rate of iron absorption as derived using the Hallberg algorithm (26). This algorithm takes dietary factors as input to calculate absorption in a subject with adequate iron stores. The Hallberg Algorithm was then inverted to determine body iron stores. Finally, the calculations of body iron stores were used to determine IDA status in the presence of environmental factors.

3.1.1 *Calculating fractional rate of iron absorption*

It was assumed that a normal individual loses 1 mg/day in non-menstrual losses (27) and 0.3 mg/day (28) in menstrual losses (for women), and that the body upregulates or down regulates the amount of iron it absorbs to match the average amount lost. Homeostasis in the samples, whereby the amount excreted by the sample was equal to the amount of iron absorbed, was assumed. This meant that healthy men and women who were neither gaining nor losing iron were assumed to be absorbing 1 mg/day and 1.3mg/day respectively.

3.1.2 *Normative fractional absorption and Calculating Iron Stores*

Bioavailability of iron was calculated as the ratio of total body iron loss and daily dietary iron intake, using an algorithm developed by Hallberg and Hulthen (29) which predicted fractional iron absorption percentage as a function of dietary inputs and serum ferritin. The Hallberg formula considered consumption of phytates, polyphenols, Vitamin C, meat, fish and seafood, calcium, egg, soy protein, and alcohol. Details of the algorithm are given as supplementary material in the appendix (26,30,31). The algorithm was reverse engineered to derive serum ferritin from iron bio-availability and dietary intakes. The relevant formulae were inverted using the Newton-Raphson method. The Newton-Raphson method is a commonly-used mathematical algorithm for finding approximate solutions to equations valued at zero.

3.1.3 *Determining IDA status*

Calculation of a proxy value for Hb was done using a previous model(32) to arrive at:

$$Hb = a{*}OBI{*}(log(\frac{(Ab - Ab_{min})(Ab_{max} - Ab_{HB})}{(Ab_{HB} - Ab_{min})(Ab_{max} - Ab)}) + BV{*}b$$

Where, Hb is in g/dl and OBI stands for Other Body Iron in grams. Ab_{HB}, Ab, Ab_{max}, and Ab_{min}, all stand for the absorption rate at normal Hb levels, the current absorption rate, the maximum absorption rate and the minimum absorption rate respectively. BV stands for body volume, and *a* and *b* are constants.

The model was modified by making the following substitutions: Estimated serum ferritin was taken as a proxy for other body iron stores; absorption level at serum ferritin level 23 ng/ml was taken in place of Ab_{HB} and 80% and 0% were taken as the maximum and minimum absorption. Thus, Ab_{23} denotes expected percentage iron absorption for person with Serum Ferritin of 23 ng/ml, Ab_m stands for percentage iron absorption for person with our estimated value for Serum Ferritin, and log is taken as to the base 10. Constants were fit to match the prevalence rate of IDA to 50% (2,3) of the total rate of anemia as measured in the NHFS-4 in both adult men and women, and using the assumption of ratio of the body weight between men and women as the ratio of the reference body weights in kilograms of 60:50 (24), and an increase of hematocrit by 5% in men.

The model was modified as:

1) $13 = 2.7F(log(\frac{Ab_{23}(0.8-Ab_m)}{Ab_m(0.8-Ab_{23})}) + 44.6$ for men and
2) $12 = 2.7F(log(\frac{Ab_{23}(0.8-Ab_m)}{Ab_m(0.8-Ab_{23})}) + 35.0$ for women.

Where, F represents serum ferritin concentration in ng/ml, Ab_{23} represents expected percentage iron absorption for person with serum ferritin of 23 ng/ml, Ab_m represents percentage iron absorption for person with our estimated value for Serum Ferritin, and the logarithm is to the base 10.

Subjects were then assigned a value for additional loss based on their likely access to sanitation facilities. This was done by comparing the quintile income ranking of the subject with the percentage of the population with access to sanitation in that district based on NFHS 4. For example, if the subject was in the second lowest quintile income group, and less than 60% of the population had access to sanitation facilities in that district, they would be classified as unable to access sanitation facilities. Subjects who were classified as unable to access sanitation facilities, were estimated to have an additional iron loss, which amounted to dropping their estimated Hb levels by 0.98.

Further, a sensitivity analysis was undertaken where all nutrient inputs were systematically both increased and decreased by 10% of the input parameters so as to check the robustness of the model. The model was also tested on an external data collected in Jharkhand where the same algorithm was used for generating predictions and comparing with the actual (measured) IDA status based on Hb and ferritin measurements.

3.2 Statistical Matching to create a synthetic dataset on nutrient intakes and wages in India

The NSSO consumer expenditure data was triangulated with the NSSO employment-unemployment data at an individual level for the year 2011- 12. The objective was to construct a synthetic dataset by merging two survey data sets with a similar target population. The fundamental principle of this statistical matching technique was to identify a set of common variables (X) between the data sets that explained the variation of both the outcome and exposure. The employment- unemployment data was chosen as the recipient which contained information on earnings and occupation while the consumer expenditure was chosen as the donor. A set of common/matching variables which were considered were household type, religion, social group, total value of the land possessed, per capita monthly expenditure, age, sex, marital status and the education. The non-parametric technique called nearest neighbor distance hot deck was implemented to perform the statistical matching (33). The final dataset provided overall information on individual expenditure on food and non-food items along with the earnings which consisted of 126,671 men and 124,515 women in the age group of 15 to 49. For validation of statistical matching as well as Heckman Selection model, missing values were dropped as per the requirement. It should be noted that while the NSSO employment-unemployment data also contained information on food expenditure, it was not as detailed as that provided in consumer expenditure data.

The validation of the dataset was done by deploying two different methods. The first method focused on analyzing the associations between the consumer expenditure in both the data sets. The second method was used to develop a log regression model in which the value of individual expenditure from both the datasets separately was regressed on the set of matching variables from the employment- unemployment data. To understand variation, the results of the models were compared with each other. Here, the dataset was divided into 3 different sub-datasets i.e. training, cross validation and test data sets. Originally, the model was built on the training dataset and then the trained model was used to make predictions on cross validation dataset and finally was run on test datasets. The best fit model selection was based on the combination of how well variations were explained with the lowest RMSE (Root mean squared error) and MAE (Mean absolute error).

3.3 Effect of anemia on wages: Heckman selection model

While Randomized Controlled Trials are commonly used to capture the effect of anemia on wages (34), the present study modelled this effect using regression techniques on secondary data available through national level surveys. The Mincer Earnings Function (35,36) has been extensively used in determining the effect of age, experience and education on wages, where the log of wages is regressed on the given determinants.

The effect of health status (for example IDA) on wages can be evaluated by adding it as an independent variable in the regression model. However, a severe limitation arises in the Ordinary Least Square (OLS) estimation. An individual's choice of labour force participation depends on his/her reservation wage; running an OLS regression on wages would imply inclusion of only those individuals in the sample who have been offered wages greater than their reservation wages, thus leading to a non-random sample and a violation of the Gauss Markov assumption. To control for the selection bias in labour force participation, the present study used the Heckman Selection Model (37).

The Heckman Selection Model tries to correct for inherent biases in the choice to participate in the labour market by first estimating the labour force participation equation through a probit model given as follows:

$$Pr(LFP_i = 1) = \Phi(\beta_1 + \beta_2 no\_dependentchildren_i + \beta_2 pc\_monthlyexp_i \ + \underline{\gamma}\underline{Z} + \mu_i)$$

Where the dependent variable is the probability of participating in the labour force (0- not employed, 1- employed) where the employment status was based on an individual being self-employed, regular salaried employed or wage employed. Φ is the cumulative distribution function of the standard normal distribution, the number of dependent children less than five years of age and the per capita monthly expenditure are the exclusion restrictions and $\underline{Z}$ is the vector of independent variables that affect labour force participation such as land owned (in hectares), marital status (0- not married, 1- married), education attainment level (where 1- illiterate, 2- literate without formal education or below primary education, 3- primary education or above), age, regional zones (1- north India, 2- west India, 3- central India, 4- east India, 5- south India, 6- north- east India) and μ_i is the error term. The present study modelled the participation equation separately for men and women. The exclusion restriction in the participation equation helped in preventing the problem of multicollinearity in the outcome model, explained below.

The Inverse Mills Ratio, which is the ratio of the probability density function over the cumulative distribution function of the distribution was then calculated from the probit model, and inserted into the outcome model as an additional regressor to control for the bias in the mincer equation or the outcome model, which is given as follows:

$$logwages_i = \alpha_i + \beta IDA_i + \gamma \underline{X} + \delta IMR_i + \varepsilon_i$$

Where the dependent variable is the log of wages calculated as average daily earnings (in INR), the independent variable of interest is the dummy variable IDA (0 – non-IDA, 1-IDA), $\underline{X}$ is a vector of other determinants of wages such as land owned, marital status, education attainment level, age, regional zones, physical activity status (0 – sedentary, 1 – moderate or heavy), social group (0 – others, 1 – general category), location of residence (0 – urban, 1- rural). IMR is the Inverse Mills Ratio from the probit model and ε is the error term. The outcome equations were determined separately for the regular salaried employed and the casual labourers and the coefficient of the dummy variable for IDA was used to calculate the monetary benefits. Specifically, the percentage equivalent of the coefficient was used as the expected changes in wages of the IDA population if they became non- IDA, as explained in the next section. The data on earnings of self-employed was not reported, and hence, the effect of IDA on the earnings of self-employed could not be estimated.

The error terms in the selection and outcome model (μ and ε respectively) are distributed as:

$$(\varepsilon, \mu) \sim N\left(0,0,\sigma_\varepsilon^2, \sigma_\mu^2, \rho_{\varepsilon\mu}\right)$$

i.e., the error terms are normally distributed with mean 0, variances as indicated above, and these correlate with each other. The error terms are also assumed to be independent of both sets of explanatory variables.

4. Results

The final results on all the methods employed for conducting the analysis has been presented and summarized.

4.1 Mathematical model to predict the prevalence of iron deficiency anemia

The prevalence rate of IDA estimated from the mathematical model was predicted to be 10.6% and 23.6% for men and women respectively.

When the model was compared with the Jharkhand dataset, the model predicted similar IDA prevalence where the predicted IDA in the sample was 42.9% and the measured IDA (based on Hb and ferritin measurement) was 31.8%. A sensitivity analysis, whereby all nutritional inputs were systematically both increased and decreased by 10% of the input parameters showed the model was robust against all nutritional inputs, barring, of course, dietary iron.

4.2 Statistical Matching to create a synthetic dataset on nutrient intakes and wages in India

The correlation coefficient between the values of individual expenditure across both the datasets was about 0.76 and 0.70 for men and women respectively. The distribution of the imputed variable i.e. value of household expenditure from intake dataset before and after statistical matching was similar across all the regions. The results of the log regression model for men and women is presented in **Appendix Table 1** and **Appendix Table 2**. The R-square of the model was around 38% and 37% for men and women respectively and the coefficients were statistically significant.

4.3 Heckman Selection Model – Effect of Iron deficiency Anemia on Earnings

The results for the labour force participation equation and sample-selection-bias-corrected wage outcome models for men is presented in **Appendix Table 3**. The results suggest that keeping everything else constant, marital status, age (proxy for experience), education level, monthly household expenditure and the number of dependent children less than five years of age in the household had a significant effect on labour force participation for men.

Sample-selection-bias-corrected wage outcome models were estimated separately for regular salaried employed men and for male casual labour and are presented in Appendix Table 3. The results indicate that on an average, keeping everything else constant, having IDA caused a fall of 19% and 6% in the earnings of a regular salaried employed man and a male casual laborer respectively, as compared to their non-IDA counterpart. The significance of the coefficient of the Inverse Mills Ratio in the outcome equations for both regular salaried men and casual labour indicated that the null hypothesis that the error terms in the participation model and the outcome model are not correlated was rejected, thus indicating evidence of a sample selection bias in both these cases.

From the model, we can also infer that on an average, rural men earned lower than their urban counterpart for both the employment categories. Conditional upon having a job, the average earnings of both regular salaried men as well as a male casual laborers increased with educational attainment. While the dummy for religion as Hindu was positive for regular salaried employees, it was negative for casual laborers. On the other hand, age, which was used as a proxy for experience, had a positive and significant effect only on the wages of regular salaried employees. The dummy for marital status as well as social group was also positive and significant for this employment category indicating that married regular salaried men earn more than their unmarried counterpart and regular salaried men from general caste category earned more than other castes. On the other hand, a regular salaried man engaged in an occupation that entailed moderate or heavy physical work was likely to earn lower than a man engaged in a sedentary occupation. Total land owned was also positively and significantly associated with wages for regular salaried employees. The regional dummies for both the employment categories were significant indicating regional variation in wages. The outcome model for regular salaried men explained 29% of the variability in average daily earnings of regular salaried employed men around its mean, whereas the model for male casual labour explained only 19% of the variability in average daily earnings of regular salaried employed men around its mean.

The results from the labour force participation equation and Sample-selection-bias-corrected wage outcome models for women is presented in **Appendix Table 4**. The results indicate that on an average, keeping everything else constant, having IDA caused a fall of 12.1% in the earnings of a regular salaried employed woman, as compared to that of her non-IDA counterpart, whereas the results for female casual labourer indicated that having IDA had no significant impact on the earnings. The significance of the coefficients of the Inverse Mills Ratio in the outcome models for both regular salaried women and female casual labourers indicates a rejection of the null hypothesis that the error terms in the participation model and the outcome model are not correlated, thus indicating evidence of a sample selection bias in both these cases.

The results also indicate that age, which is a proxy for experience, had a positive and significant impact on the wages of both regular salaried and casual labourers. While the dummy for social group was positive and significant in the case of regular salaried, it was negative in the case of casual labourers. This could possibly be due to disproportionately high observations for non-general category in the female casual labour group which represented 88% of this population; cultural barriers often prevent women from higher- caste households to undertake menial jobs. For both the employment categories, in comparison to illiterate women, a literate women without formal education or below primary education had no significant effect on earnings whereas education above primary level had a negative and significant effect. This result is counter-intuitive to theory. Further, specifically the results of the regular salaried women suggest that not only was a married woman less likely to participate in the labour force but also, would earn wages lower than her unmarried counterpart. The dummy for physical activity status and the dummy for sector for this employment category had a negative and significant coefficient indicating that women engaged in moderate or heavy physical activity and women residing in rural areas, on an average, earned less than women engaged in sedentary activity and urban area respectively. The total land owned had a positive and significant effect on the earnings for regular salaried women. The significant coefficients for regional dummies for both the employment categories indicated regional variation in wages. The outcome model for regular salaried women explained 36% of the variability in average daily earnings of regular salaried employed women around its mean, whereas the model for female casual labour explained 5% of the variability in average daily earnings of regular salaried employed women around its mean.

5. Discussion:

The present study establishes the linkages between IDA and wages through survey datasets, a method which has not been applied in the past research on this topic in India and can help in understanding the possible effects of fortification in monetary terms. This is because a comprehensive single nationally representative survey which collects information on nutrient data, anemia status and wage data is presently not available in India. Hence, a mathematical model predicting the IDA status using national level data on nutrient intake was triangulated with the wage data to create a synthetic data to understand these linkages. The statistical matching was possible because both the datasets were collected by a common agency, NSSO, and hence, had a similar sampling technique.

Even though the benefits of a food programme can be assessed by its impact on health outcomes, an additional understanding of how changes in health outcomes could translate into economic outcomes, add more value to the programme. The broad area in which iron deficiency is considered to have important functional impact on humans, where economic consequences can be estimated is in the work capacity and wages of adults (38). As a first step, it was thus important to establish an association between IDA and wages. In the next phase of the study, this was used to evaluate the economic impact of DFS, in terms of reducing the burden of IDA and the associated changes in wages of the population.

The economic burden of anemia has been documented in various studies. A study which assessed the effect of IDA on the productivity of tea plantation workers in Sri Lanka through a randomized controlled trial, with the treatment group being given iron supplements and the control group a placebo (34). The study found that the worker productivity of those in the treatment group had increased and that the treatment group members were significantly more active 2 and 3 weeks after the start of treatment than their matched controls. The study also indicated that the most significant gains in productivity were accrued to those with the lowest pre-treatment Hb levels. In one of the systematic reviews, 29 research reports were examined and established a strong casual effect of severe iron-deficiency anemia and moderate iron deficiency anemia on aerobic capacity in animals and humans (39).

Some solutions that have been explored to reduce the burden of anemia have provided iron supplements to anemic individuals, and more recently, have provided DFS to households. There are several efficacy studies which have established the association between DFS and productivity. One such study found that in rural women of reproductive age working as tea pickers in Northern West Bengal showed significant improvements when compared with controls in hemoglobin (+2.4 g/L), ferritin (+0.13 $\log_{10} \mu$g/L) and other factors after 9 months of using DFS (40). However, very limited effectiveness studies, especially in the context of India, have tested this relationship (41,42). A randomized control trial on DFS conducted in rural Bihar found that there was an insignificant impact on anemia and other related factors, with an exception of reduction in anemia amongst the adolescent population (42). However, the uptake of DFS was very low in the study.

As done in the past, one method to calculate the monetary gains could be to use these studies in assuming an average value of the expected impact, and use this value to generate country level estimates of the monetary gains. These studies conducted a literature review on the association between DFS and Hb levels and then used various assumptions and algorithms to calculate possible gains at a country level (20,43,44). Assumptions were made on the values of wage share in GDP, share of blue collar employment in total employment, prevalence of anemia and other factors. The authors found that the benefit to cost ratio varied from 2:1 to 5:1 depending upon the population group.

There are several limitations of the present analysis. Nutrients which affect red blood cell production, such as folic acid and Vitamin B12 have not been considered in the mathematical model. Sanitation loss was taken as a constant and was estimated by proxy indicators, even though losses can vary. The loss of iron from the body was taken as a constant, though daily losses are usually expected to vary by 20% across individuals and ages. With no identifiable macroeconomic variables to attribute loss, the model is limited in terms of predicting an individual's hemoglobin-health status, though it is instructive on a macro-level. The model also assumed a constant volume of distribution of iron in the body, as there were no data available that could be used to determine the plasma volume of individuals. It is expected that people with high body plasma volume would require slightly more iron than those with lower body plasma volume owing to a higher volume of distribution of iron. Owing to these causes of variability, the mathematical model was limited to assessing whether subjects were IDA or not rather than the precise value of Hb in the subject. Also, the NSSO employment- unemployment surveys are not perfect data, they are nevertheless the best available data that provide information on earnings at a national level. However, the survey only provides the earnings for certain types of employment. The wage outcome models were estimated only for regular salaried employees and casual labourers due to the paucity of data on earnings of self-employed individuals.

On the other hand, Heckman selection model incorporating the Mincer wage function has been extensively used to study the determinants of wages in past research. In the context of India, a study used the Heckman Selection Model to understand the days of labour supplied by female labour force (45). Another study incorporated the Mincer wage Function into the Heckman Selection Model to estimate the effects of health and education on wages and productivity in Australia (46). One study also estimated the impact of height, body mass index, per capita calorie intake and per capita protein intake on wages of urban workers in Brazil by incorporating a Heckman Selection Model in Instrumental Variables Estimation Technique to account for the sample selection bias and the potential endogeneity of health and wages (47). The biasness that occurs with choice of employment or unemployment has been addressed through the Heckman Sample Selection model in these studies.

The present study indicates that at a national level, there was a significant and negative impact of IDA on the earnings of employed men and women, with men being worse-off than women. Being anemic, on an average, caused a 6% decline in the earnings of male casual labourers, and a 19% decline in the earnings of regular salaried men in India. On the other hand, regular salaried anemic women earned 12.1% less as compared to their non- anaemic counterparts. However, there was no observed impact of anemia on the wages of female casual labour which indicates that there are other factors than food intake and physiology as well as socio- economic factors that determine the wages of female casual labourers. Although the prevalence of anemia is higher amongst women than men, the economic burden is higher from men than women because not only is the share of men greater in workforce, the scale of the impact of IDA on wages is greater on men than women.

Given the negative and significant impact of IDA on wages, it can be conjectured that an intervention programme focused on improving iron intakes like DFS would help in reducing the burden of IDA and improve the wages of the population. The total monetary benefits for the country would not only depend on the scale of improvement in wages but also the reduction in IDA. Therefore, the next second phase of the study established these changes through the mathematical model by simulating new iron intakes and new prevalence rates of IDA if DFS was hypothetically introduced in the country through a national program. The benefits were compared with the cost of DFS to calculate benefit-cost ratio.

Equally, cautionary steps need to be undertaken to avoid overconsumption of iron intake through over fortification especially for the population group which already consume sufficient amounts of iron. Currently, there is much enthusiasm to fortify even rice and wheat with iron. Overconsumption can trigger various chronic diseases. Men are at higher risk of consuming toxic levels since the losses in this group are low. For women also, if iron fortification is combined with targeted iron folic acid supplementation, there is a potential risk of overdose of iron, especially in non-pregnant and non-lactating women taking supplements. The intake of iron supplements particularly in non-anemic pregnant women with higher compliance was found to be associated with the increased risk of low birthweight and decreased mean birth weight by 72 grams (48). More research is required on this front but till then, a holistic approach needs to be undertaken before introducing any national policy which directly affects the physiology. On the other hand, a carefully planned strategic approach on fortification can contribute to the economy through improved productivity and wages of not only the present generation but also the future generations through improved physical and mental development of child at early stages. It can also help in saving resources allocated towards combating anemia through various programs and can be allocated to other sectors which are important for the development of the economy.

References:

1. International Institute for Population Sciences. National Family Health Survey-4. Mumbai: International Institute for Population Sciences; 2016.

2. World Health Organization, Centres for Disease Control and Prevention. Report of a Joint World Health Organization/ Centers for Disease Control and Prevention Technical Consultation on the Assessment of Iron Status at the Population Level [Internet]. Geneva; 2004. Available from: http://apps.who.int/iris/bitstream/10665/75368/1/9789241596107_eng.pdf?ua=1

3. Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995–2011: a systematic analysis of population-representative data. Lancet Glob Heal. 2013 Jul;1(1):e16–25.

4. Haas JD, Brownlie T. Iron Deficiency and Reduced Work Capacity: A Critical Review of the Research to Determine a Causal Relationship. J Nutr. 2001 Feb;131(2):676S–690S.

5. Seshadri S, Gopaldas T. Impact of iron supplementation on cognitive functions in preschool and school-aged children: the Indian experience. Am J Clin Nutr. 1989 Sep;50(3):675–86.

6. Lozoff B. Iron Deficiency and Child Development. Food Nutr Bull. 2007 Dec;28(4_suppl4):S560–71.

7. Chong A, Cohen I, Field E, Nakasone E, Torero M. Iron Deficiency and Schooling Attainment in Peru. Am Econ J Appl Econ. 2016 Oct;8(4):222–55.

8. Rush D. Nutrition and maternal mortality in the developing world. Am J Clin Nutr. 2000 Jul;72(1):212S–240S.

9. Allen LH. Anemia and iron deficiency: effects on pregnancy outcome. Am J Clin Nutr. 2000 May;71(5):1280S–1284S.

10. Thomas D, Elizabeth Frankenberg U, Jed Friedman U, World Bank Jean-Pierre Habicht T, Jones N, Sikoki B, et al. Causal effect of health on labor market outcomes: Experimental evidence. 2006.

11. Lozoff B, Beard J, Connor J, Barbara F, Georgieff M, Schallert T. Long-lasting neural and behavioral effects of iron deficiency in infancy. Nutr Rev. 2006 May;64(5 Pt 2):S34-43; discussion S72-91.

12. Indian Council of Medical Research, Public Health Foundation of India, Institute for Health Metrics and Evaluation. India : Health of the Nation's States: The India State-Level Disease Burden Initiative. New Delhi; 2017.

13. Zijp IM, Korver O, Tijburg LBM. Effect of Tea and Other Dietary Factors on Iron Absorption. Crit Rev Food Sci Nutr. 2000 Sep;40(5):371–98.

14. Coffey D, Geruso M, Spears D. Sanitation, Disease Externalities and Anaemia: Evidence From Nepal. Econ J. 2018;128:1395–432.

15. Baranwal A, Baranwal A, Roy N. Association of Household Environment and Prevalence of Anemia Among Children Under-5 in India. Front Public Heal [Internet]. 2014;2(October):1–7. Available from: http://journal.frontiersin.org/article/10.3389/fpubh.2014.00196/abstract

16. Stoltzfus RJ, Dreyfuss ML. Guidelines for the Use of Iron Supplements to Prevent and Treat Iron Deficiency Anemia [Internet]. Available from: http://www.who.int/nutrition/publications/micronutrients/guidelines_for_Iron_supplementation.pdf

17. Ministry of Health and Family Welfare. Anemia Mukt Bharat, Operational guidelines for programme managers,Intensified national iron plus initiative (I-NIPI). 2018.

18. Mora JO. Iron Supplementation: Overcoming Technical and Practical Barriers. 2002.

19. Horton S, Mannar V, Wesley A. MICRONUTRIENT FORTIFICATION (IRON AND SALT IODIZATION). 2008;

20. Horton S, Wesley A, Venkatesh Mannar MG. Double-fortified salt reduces anemia, benefit:cost ratio is modestly favorable. Food Policy. 2011;36:581–7.

21. National Sample Survey Organisation. Household Consumer Expenditure Survey 2011-12 Schedule 1 Type 1, National Sample Survey Organization. New Delhi: National Sample Survey Organization, Ministry of Statistics & Programme Implementation, Government of India;

22. Longvah T, Ananthan R, Bhaskarachary K, Venkaiah K. Indian Food Composition Tables [Internet]. Hyderabad; 2017. Available from: http://ninindia.org/ifct_2017.htm

23. Gopalan C, Rama Sastri B, Balasubramanian S. Nutritive value of Indian foods. Updated by Narasinga Rao BS, Deosthale YG, Pant KC: National Institute of Nutrition. Indian Council of Medical Research. Hyderabad; 1996.

24. US Department of Agriculture, Agriculture Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Legacy. Version Release 28. Washington DC;

25. Thankachan, P Kalasuramath S, Hill A, Thomas T, Bhat K, Kurpad A. A mathematical model for the hemoglobin response to iron intake, based on iron absorption measurements from habitually consumed Indian meals. Eur J Clin Nutr. 2012;66(4):481.

26. Hallberg L, Hulthen L. Prediction of dietary iron absorption : an algorithm for calculating absorption and bioavailability of dietary iron. Am J Clin Nutr. 2000;

27. Hunt JR, Zito CA, Johnson LK. Body iron excretion by healthy men and women. Am J Clin Nutr. 2009 Jun;89(6):1792–8.

28. Ghosh S, Sinha S, Thomas T, Singh H, Kurpad A. Revisiting dietary iron requirement for women: Implications for iron fortification and supplementation. J Nutr. (Accepted for publication).

29. Hallberg L, Hulthén L. Prediction of dietary iron absorption: an algorithm for calculating absorption and bioavailability of dietary iron. Am J Clin Nutr. 2000 May;71(5):1147–60.

30. Hoppe M, Hulthén L, Hallberg L. The importance of bioavailability of dietary iron in relation to the expected effect from iron fortification. Eur J Clin Nutr. 2008 Jun;62(6):761–9.

31. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical Recipes in FORTRAN: The Art of Scientific Computing. 2nd ed. Ca. England: Cambridge University Press; 1992.

32. Thankachan P, Walczyk T, Muthayya S, Kurpad A V, Hurrell RF. Iron absorption in young Indian women: the interaction of iron status with the influence of tea and ascorbic acid. Am J Clin Nutr. 2008 Apr;87(4):881–6.

33. D'Orazio M. Statistical Matching and Imputation of Survey Data with StatMatch. 2017. 1-36 p.

34. Edgerton VR, Gardner GW, Ohira Y, Gunawardena KA, Senewiratne B. Iron-deficiency anaemia and its effect on worker productivity and activity patterns. Br Med J. 1979 Dec;2(6204):1546–9.

35. Mincer JA. Investment in Human Capital and Personal Income Distribution. J Polit Econ. 1958;66(4):281–302.

36. Mincer JA. Schooling, Experience, and Earnings. National Bureau of Economic Research. New York: Columbia University Press; 1974.

37. Heckman JJ. Sample Selection Bias as a Specification Error. Econometrica. 1979;47(1):153–61.

38. Stein AJ, Meenakshi JV, Qaim M, Nestel P, Sachdev HPS, Bhutta ZA. Analyzing the Health Benefits of Biofortified Staple Crops by Means of the Disability-Adjusted Life Years Approach: a Handbook Focusing on Iron, Zinc and Vitamin A. Washington DC; Cali: HarvestPlus; 2005.

39. Haas JD, Brownlie T. Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. J Nutr. 2001;131(2S–2):676S–688S.

40. Haas JD, Rahn M, Venkatramanan S, Marquis GS, Wenger MJ, Murray-Kolb LE, et al. Double-Fortified Salt Is Efficacious in Improving Indicators of Iron Deficiency in Female Indian Tea Pickers. J Nutr. 2014;144(6):957–64.

41. Thomas D, Frankenberg E, Friedman J, Habicht J-P, Ingwersen N, McKelvey C, et al. Causal effect of health on labor market outcomes: Experimental evidence. Los Angeles; 2006.

42. Banerjee A, Barnhardt S, Duflo E. Can Iron-Fortified Salt Control Anemia? Evidence From

Two Experiments in Rural Bihar. NBER Working Paper. 2016.

43. Horton S, Ross J. The economics of iron deficiency. Food Policy. 2003;28:51–75.

44. Horton S, Ross J. Corrigendum. Food Policy. 32:141–3.

45. Dasgupta P, Goldar B. Female Labour Supply in Rural India: An Econometric Analysis. 2005.

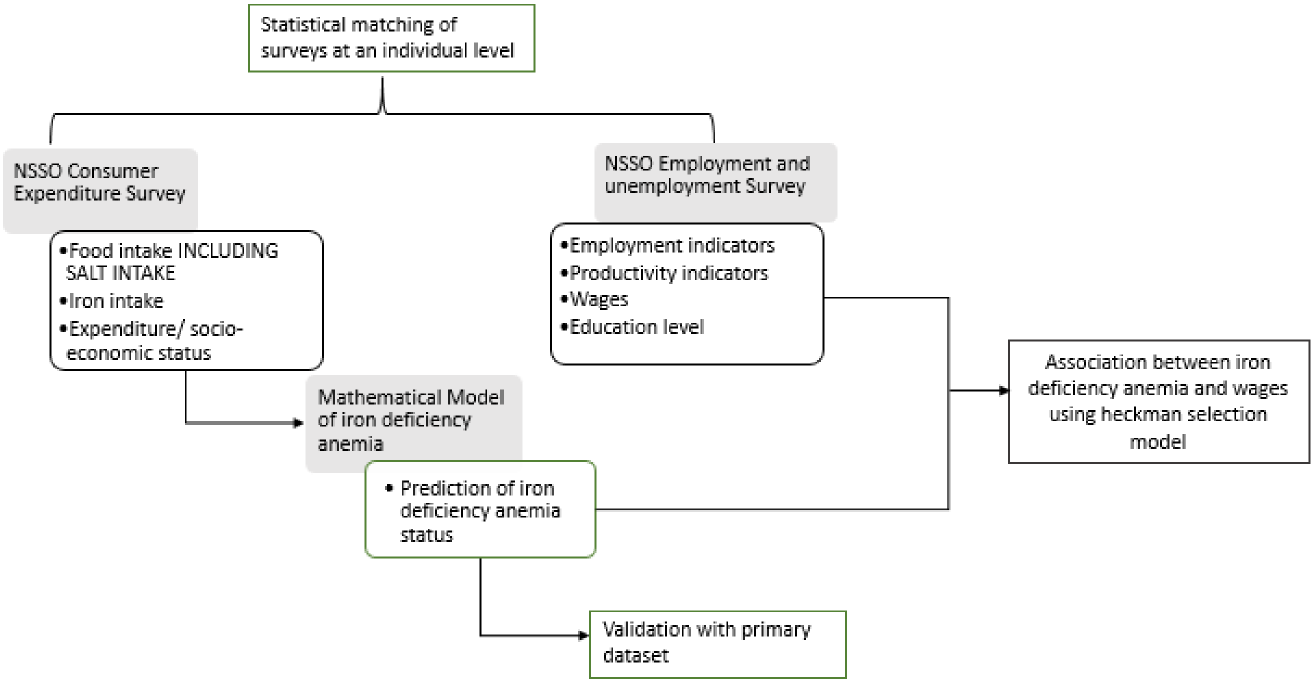
46. Forbes M, Barker A, Turner S. The Effects of Education and Health on Wages and Productivity, Productivity Commission Staff Working Paper. 2010.

47. Thomas D, Strauss J. Health and wages: evidence on men and women in urban Brazil. J Econom. 1997;77:159–85.

48. Shastri L, Mishra PE, Dwarkanath P, Thomas T, Duggan C, Bosch R, et al. Association of oral iron supplementation with birth outcomes in non-anaemic South Indian pregnant women. Eur J Clin Nutr [Internet]. 2015;69:609–13. Available from: http://dx.doi.org/10.1038/ejcn.2014.248

Figures:

Figure 1: Analysis framework



Appendix

Supplementary Information:

1. Description of Hallberg algorithm:

In order to account for the relationship between Iron status and regulation of absorption, estimates of iron absorption had to be adjusted for dietary factors that enhance and inhibit iron absorption. Our estimates of absorption were based on an algorithm by Hallberg and Hulthen, which accounted for the effect of phytate, ascorbic acid, tannin, meat, alcohol, calcium and eggs for non-heme iron absorption and calcium for heme iron absorption. Due to reporting issues, alcohol intake an adjustment for alcohol intake was not considered. The equations used in estimating bioavailable iron were (29,30):

Absorption ratio from Phytate ($F_{phytate}$):

$$10^{-0.30 log_{10}(1+phytate)}$$

where *phytate* is phytate-phosphorus in milligrams.

Absorption ratio from Ascorbic acid and phytate interaction (F_{AA}):

$$1 + 0.01AA + log_{10}(phytate + 1) \times 0.01 \times 10^{0.8875 log_{10}(AA+1)}$$

where AA is the ascorbic acid (more commonly known as Vitamin C) content of the meal and *phytate* is the phytate-phosphorus content of the meal, both in milligrams.

Absorption ratio from tannic acid (polyphenols), ascorbic acid and meat interaction (F_{TA}):

$$(1 + 0.01M) \times 10^{0.4515-[0.715-0.1825 log_{10}(1+AA)] log_{10}(1+TA)}$$

where TA is tannic acid in mg, AA is ascorbic acid in mg, M is the meat, poultry, fish or seafood in gm in the meal.

Absorption ratio from Calcium (F_{Ca}):

$$0.4081 + \frac{0.5919}{1 + 10^{-(2.022 - log_{10}(Ca+1)) \times 2.919}}$$

Where Ca stands for calcium content of the meal in mg.

Absorption ratio from Phytate and meat interaction (F_M):

$$1 + 0.00628 \times M \times [1 + 0.006 phytate]$$

where M (meat, poultry, fish and seafood) are expressed as grams of uncooked food and phytate is in milligrams.

Absorption ratio from eggs (F_{egg}):

$$1 - 0.27 \times numberofeggs$$

The equation is valid if consumption of eggs ≤ 3 per meal.

Iron absorption for a day's meal:

The nonheme iron absorption (%) was derived as ($F_{nonheme}$)

$$22.1 \times F_{phytate} \times F_{AA} \times F_{TA} \times F_{Ca} \times F_{M} \times F_{egg}$$

where 22.1 is basal absorption factor for iron adjusted to a 40% reference dose absorption.

The percentage absorption of heme iron was adjusted to iron status corresponding to a reference dose absorption of 40%, and then corrected for calcium intake in the same meal using the following formula (F_{heme}):

$$F_{ca} \times 10^{1.9897-0.3092\times log_{10} SF}$$

where SF is the serum ferritin level of 23ng/ml for a reference absorption of 40%. Finally, Iron absorption from a day's diet was derived by:

$$(F_{nonheme}(\%) \times NonhemeIronintake + F_{heme}(\%) \times HemeIronintake) \times \left(\frac{23^{0.9409}}{SF}\right)$$

2. Estimation of Serum Ferritin:

The Hallberg algorithm estimates the Bioavailability of iron from Serum Ferritin and dietary intakes. If the bioavailability of iron is represented as:

$$f(SF) = \frac{Amount\ of\ iron\ absorbed\ daily}{Amount\ of\ iron\ consumed\ daily}$$

Based on the assumption that iron absorbed was 1 mg/day for men, and 1.3 mg/day for women, the Newton-Raphson method was applied to estimate serum ferritin concentration. The Newton-Raphson method is a technique in numerical analysis which is used to estimate the inputs of a real-valued mathematical function when the output is 0 (37). The method begins as an arbitrary input and incrementally improves the approximation of the input. Using the following function:

$$f(SF) - \frac{Amount\ of\ iron\ absorbed\ daily}{Amount\ of\ iron\ consumed\ daily} = 0$$

The technique is an iterative one, where subsequent values are calculated as improvements over the previous ones as follows:

$$SF_{n+1} = SF_n - \frac{f(SF) - \frac{Amount\ of\ iron\ absorbed\ daily}{Amount\ of\ iron\ consumed\ daily}}{f'(SF)}$$

Where SF_{n+1} is an improved estimate of Serum Ferritin over SF_n.
The initial condition should be chosen as a reasonable proximity to the final solution. For the purposes of the analysis, we began with a starting Serum Ferritin of 1 ng/ml, and continued iterating the function 300 times. We found that improvements after 300 iterations to be less than 1%.

Supplementary Tables:

Table 1: *Results for validation of statistical matching for men*

	Value of consumer expenditure from consumer expenditure dataset regressed on set of variables in employment-unemployment dataset		**Value of consumer expenditure from employment-unemployment dataset regressed on set of variables in employment-unemployment dataset**	
	Train Model	**Test Model**	**Train Model**	**Test Model**
Variables	**Log expenditure**	**Log expenditure**	**Log expenditure**	**Log expenditure**
Intercept	7.905*** (0.017)	7.724*** (0.037)	8.021*** (0.019)	7.744*** (0.041)
Household size	-0.049*** (0.001)	-0.044*** (0.001)	-0.071*** (0.001)	-0.068*** (0.002)
Dummy for household type (Individual belongs to either non-agriculture(self-employed) in rural area or regular wage/salary earning in urban area=1)	0.032*** (0.005)	0.023** (0.01)	0.034*** (0.005)	0.057*** (0.011)
Dummy for household type (Individual belongs to regular wage/salary earning in rural area and casual labour in urban area=1)	-0.074*** (0.006)	-0.073*** (0.012)	-0.081*** (0.006)	-0.037*** (0.014)
Dummy for household type (Individual belongs to casual labour in agriculture in rural area=1)	-0.31*** (0.006)	-0.339*** (0.012)	-0.375*** (0.006)	-0.325*** (0.012)
Dummy for household type (Individual belongs to casual labour in non-agriculture in rural area=1)	-0.257*** (0.006)	-0.259*** (0.013)	-0.275*** (0.007)	-0.299*** (0.014)
Dummy for household type (Individual belongs to others=1)	0.181*** (0.011)	0.119*** (0.025)	0.19*** (0.012)	0.032 (0.028)
Dummy for religion (Individual belongs to Hinduism=1)	-0.169*** (0.012)	-0.133*** (0.026)	-0.18*** (0.013)	-0.113*** (0.03)
Dummy for religion (Individual belongs to Islam=1)	-0.266***	-0.223***	-0.27***	-0.217***

	(0.013)	(0.028)	(0.014)	(0.032)
Dummy for religion (Individual belongs to others=1)	0.157*** (0.015)	0.208*** (0.033)	0.152*** (0.016)	0.169*** (0.037)
Dummy for social group (Individual belongs to schedule caste=1)	0.119*** (0.007)	0.103*** (0.014)	0.115*** (0.008)	0.106*** (0.016)
Dummy for social group (Individual belongs to other backward class=1)	0.216*** (0.006)	0.201*** (0.013)	0.231*** (0.007)	0.225*** (0.015)
Dummy for social group (Individual belongs to others=1)	0.387*** (0.007)	0.397*** (0.014)	0.384*** (0.007)	0.314*** (0.016)
Total hectares of land possessed	-0.036*** (0.001)	-0.035*** (0.002)	-0.035*** (0.001)	-0.028*** (0.002)
Age	0.002*** (0)	0.004*** (0.001)	0.003*** (0)	0.004*** (0.001)
Dummy for marital status (Individual belongs to one of the categories i.e. never married, widowed and divorced/separated=1)	0.071*** (0.005)	0.091*** (0.01)	0.082*** (0.005)	0.107*** (0.011)
Dummy for education status (Individual is literate up to higher secondary=1)	-0.414*** (0.005)	-0.335*** (0.011)	-0.47*** (0.006)	-0.355*** (0.013)
Dummy for education status (Individual is literate without formal schooling=1)	-0.644*** (0.027)	-0.614*** (0.053)	-0.614*** (0.032)	-0.588*** (0.074)
Dummy for education status (Individual is illiterate=1)	-0.598*** (0.007)	-0.486*** (0.014)	-0.653*** (0.007)	-0.505*** (0.016)
Number of Observations Adjusted R-squared RMSE(Root mean squared error)[1] MAE(Mean absolute error)[1]	73421 0.3781 2111.49 1703.382	14461 0.3778 1937.01 1648.391	73627 0.4019 2224.342 1712.985	14355 0.3496 1881.72 1623.244

Note- Figures in the parenthesis are standard errors;
***significant at 1%; ** significant at 5%; * significant at 10%

[1] values of RMSE and MAE on cross validation dataset have an observation of 34,062 and 33,962 respectively for both the log regressions.

Table 2: *Results for validation of statistical matching for women*

Variables	Value of consumer expenditure from consumer expenditure dataset regressed on set of variables in employment-unemployment dataset		Value of consumer expenditure from employment-unemployment dataset regressed on set of variables in employment-unemployment dataset	
	Train Model	Test Model	Train Model	Test Model
	Log expenditure	Log expenditure	Log expenditure	Log expenditure
Intercept	7.788*** (0.016)	7.816*** (0.035)	7.881*** (0.018)	7.782*** (0.037)
Household size	-0.042*** (0.001)	-0.041*** (0.002)	-0.064*** (0.001)	-0.062*** (0.002)
Dummy for household type (Individual belongs to either non-agriculture(self-employed) in rural area or regular wage/salary earning in urban area=1)	0.043*** (0.005)	0 (0.01)	0.048*** (0.005)	0.031*** (0.011)
Dummy for household type (Individual belongs to regular wage/salary earning in rural area and casual labour in urban area=1)	-0.064*** (0.006)	-0.066*** (0.013)	-0.062*** (0.007)	-0.042*** (0.013)
Dummy for household type (Individual belongs to casual labour in agriculture in rural area=1)	-0.285*** (0.006)	-0.307*** (0.012)	-0.334*** (0.006)	-0.366*** (0.012)
Dummy for household type (Individual belongs to casual labour in non-agriculture in rural area=1)	-0.253*** (0.006)	-0.248*** (0.013)	-0.264*** (0.007)	-0.283*** (0.014)
Dummy for household type (Individual belongs to others=1)	0.065*** (0.009)	-0.018 (0.019)	0.075*** (0.01)	-0.026 (0.021)
Dummy for religion (Individual belongs to Hinduism=1)	-0.152*** (0.011)	-0.103*** (0.025)	-0.139*** (0.013)	-0.157*** (0.026)
Dummy for religion (Individual belongs to Islam=1)	-0.26***	-0.188***	-0.228***	-0.218***

	(0.012)	(0.027)	(0.014)	(0.028)
Dummy for religion (Individual belongs to others=1)	0.122*** (0.015)	0.156*** (0.032)	0.155*** (0.017)	0.16*** (0.034)
Dummy for social group (Individual belongs to schedule caste=1)	0.118*** (0.007)	0.132*** (0.015)	0.129*** (0.008)	0.109*** (0.016)
Dummy for social group (Individual belongs to other backward class=1)	0.211*** (0.006)	0.225*** (0.014)	0.243*** (0.007)	0.209*** (0.014)
Dummy for social group (Individual belongs to others=1)	0.383*** (0.007)	0.372*** (0.014)	0.384*** (0.007)	0.333*** (0.015)
Total hectares of land possessed	-0.028*** (0.001)	-0.031*** (0.002)	-0.024*** (0.001)	-0.026*** (0.002)
Age	0.007*** (0)	0.005*** (0)	0.007*** (0)	0.006*** (0)
Dummy for marital status (Individual belongs to one of the categories i.e. never married, widowed and divorced/separated=1)	0.047*** (0.004)	0.036*** (0.009)	0.051*** (0.005)	0.03*** (0.01)
Dummy for education status (Individual is literate up to higher secondary=1)	-0.445*** (0.006)	-0.467*** (0.014)	-0.523*** (0.007)	-0.358*** (0.015)
Dummy for education status (Individual is literate without formal schooling=1)	-0.668*** (0.027)	-0.728*** (0.064)	-0.81*** (0.032)	-0.569*** (0.063)
Dummy for education status (Individual is illiterate=1)	-0.675*** (0.007)	-0.69*** (0.015)	-0.764*** (0.008)	-0.574*** (0.016)
Number of Observations	72345	14247	72507	14101
Adjusted R-squared	0.3715	0.3711	0.3828	0.3621
RMSE(Root mean squared error)[1]	2074.04	1910.915	2303.627	1883.08
MAE(Mean absolute error)[1]	1683.402	1631.653	1714.491	1598.665

Note- Figures in the parenthesis are standard errors;
***significant at 1%; ** significant at 5%; * significant at 10%

[1] values of RMSE and MAE on cross validation dataset have an observation of 33,555 and 33,539 respectively for both the log regressions.

Table 3: *Heckman Selection Model – Men*

Heckman Results of Effect of Anemia on Earnings of Men			
	Labour Force Participation Model	**Outcome Model for Regular Salaried Employed**	**Outcome Model for Casual Labour**
	Dummy for employment status (employed=1)	**Log of wages**	**Log of wages**
Total hectares of land owned	0.00 (0.00)	0.00*** (0.00)	0.00 (0.00)
Dummy for marital status (Individual is married=1)	0.83*** (0.03)	0.36*** (0.05)	-0.03 (0.03)
Dummy for education status (Individual is literate without formal education or below primary education=1)	0.38*** (0.07)	0.23*** (0.06)	0.02 (0.02)
Dummy for education status (Individual is educated above primary education=1)	-0.27*** (0.04)	0.50*** (0.05)	0.10*** (0.01)
Number of dependent children less than five years	0.16*** (0.02)		
Per capita monthly expenditure	-0.00*** (0.00)		
Age	0.07*** (0.00)	0.039*** (0.002)	0.00 (0.00)
Dummy for IDA (Anaemic=1)		-0.17*** (0.03)	-0.058*** (0.01)
Dummy for physical activity status (Individual engages in moderate or heavy work=1)		-0.31*** (0.02)	-0.00 (0.04)
Dummy for social group (Individual belongs to general category=1)		0.20*** (0.02)	0.05*** (0.01)

Dummy for sector (rural sector=1)		-0.27*** (0.02)	-0.12*** (0.01)
Inverse Mills Ratio		0.63*** (0.11)	-0.25*** (0.05)
Dummy for west zone	0.06** (0.03)	-0.001 (0.03)	-0.10*** (0.02)
Dummy for central zone	0.03 (0.04)	-0.22*** (0.04)	-0.34*** (0.02)
Dummy for east zone	-0.00 (0.03)	-0.14*** (0.04)	-0.16*** (0.01)
Dummy for south zone	0.04 (0.03)	0.07** (0.03)	0.25*** (0.02)
Dummy for north east zone	0.05 (0.03)	0.12*** (0.04)	0.03 (0.02)
Constant	-1.33*** (0.06)	3.55*** (0.12)	5.12*** (0.08)
No. of Observations	**111512**	**19604**	**18295**
Adjusted R-square		**0.29**	**0.19**
pseudo R-square	**0.39**		

Note- Figures in the parenthesis are standard errors;
***significant at 1%; ** significant at 5%; * significant at 10%

Table 4: *Heckman Selection Model – Women*

Heckman Results of Effect of Anemia on Earnings of Women			
	Labour Force Participation Model	**Outcome Model for Regular Salaried Employed**	**Outcome Model for Casual Labour**
	Dummy for employment status (employed=1)	**Log of wages**	**Log of wages**
Total hectares of land owned	0.00*** (0.00)	0.00*** (0.00)	0.00 (0.00)
Dummy for marital status (Individual is married=1)	-0.09*** (0.02)	-0.17*** (0.06)	-0.03 (0.03)
Dummy for education status (Individual is literate without formal education or below primary education=1)	-0.15*** (0.03)	-0.10(0.11)	-0.03 (0.04)
Dummy for education status (Individual is educated above primary education=1)	-0.37*** (0.02)	-0.32** (0.15)	-0.18** (0.08)
Number of dependent children less than five years	-0.04*** (0.01)		
Per capita monthly expenditure	-0.00*** (0.00)		
Age	0.02*** (0.00)	0.08*** (0.01)	0.02*** (0.00)
Dummy for IDA (Anaemic=1)		-0.11** (0.05)	-0.004 (0.02)
Dummy for physical activity status (Individual engages in moderate or heavy work=1)		-0.50*** (0.05)	0.10 (0.07)
Dummy for social group (Individual belongs to general category=1)		0.19*** (0.05)	-0.05* (0.03)
		-0.35***	-0.02

Dummy for sector (rural sector=1)		(0.05)	(0.03)
inverse mills ratio		3.02*** (0.49)	0.93*** (0.31)
Dummy for west zone	0.41*** (0.02)	0.86*** (0.17)	0.21** (0.10)
Dummy for central zone	0.22*** (0.03)	0.27** (0.13)	-0.005 (0.07)
Dummy for east zone	-0.24*** (0.03)	-0.82*** (0.12)	-0.28*** (0.07)
Dummy for south zone	0.5*** (0.02)	1.18*** (0.19)	0.34*** (0.11)
Dummy for north east zone	-0.05* (0.03)	0.23** (0.10)	0.17*** (0.05)
constant	-1.17*** (0.04)	-1.02 (0.86)	2.85*** (0.52)
No. of Observations	**111245**	**5228**	**5398**
adj. R-square		**0.36**	**0.05**
pseudo R-square	**0.08**		

Note- Figures in the parenthesis are standard errors;
***significant at 1%; ** significant at 5%; * significant at 10%