The Effect of Iron Deficiency on Worker Productivity:

Evidence from a Field Experiment in India

John Yohan Rhee

Department of Policy Analysis and Management

Cornell University

Ithaca, NY 14850

Jyr5@cornell.edu

I would like to personally thank Dr. Jere D. Haas for allowing me to use his data set from India as well as for guidance and advice throughout the thesis process. I would also like to thank Dr. Michael Lovenheim for all of the guidance throughout the thesis and helping me through many difficult moments. I would also like to thank Dr. Rosemary Avery, Dr. Michael Wenger, Julie Hammons, Eric Przybyszewski, Olivia Healy, and Lianne Epstein for helping me understand the data, the feedback they gave, and the support they have given me. Lastly, I would like to thank my family and friends who have believed in me and provided me with the support at just the right times.

1. Introduction

Iron deficiency is the most common nutritional deficiency internationally. Iron is an important component in blood formation and therefore, iron deficiency can, and often does, lead to anemia, which typically is defined as hemoglobin levels less than120 g/L for adult non-pregnant females. The current study takes place in India where 20% of maternal deaths are due to anemia (Umbreit 2005). Iron deficiency anemia and iron deficiency non-anemia are both known to have cognitive and productivity effects (Horton 2003). Iron deficiency non-anemia is represented through normal hemoglobin with decreased levels of ferritin and can be accompanied by elevated levels of transferrin receptor (Haas 2006).

Iron plays an important role in neurotransmitter function throughout life, affecting cognitive function. Therefore, a deficit in iron during gestation can lead to poorer cognitive functioning later in life. Productivity is affected because iron deficiency anemia is assessed with the amount of hemoglobin in the blood, which is directly related to the amount of oxygen that is delivered to body tissues. Iron deficiency without anemia, although not directly correlated with hemoglobin levels, still has potential productivity effects, especially in endurance, due to its effects on the ability of body tissues and organs to carry out metabolic functions (Haas 2006). Although there is sizable literature on how micronutrient deficiencies affect productivity, much of the previous work in this area suffers from poor worker output information, and there has been little focus on the intermediates through which iron affects productivity. The purpose of this study is to fill some of these gaps in the understanding of how iron impacts worker productivity in developing countries.

There is an extensive literature on worker output, but most of the literature does not isolate worker output from how it is affected by other production function inputs. Hemoglobin or

iron stores do not lead directly to increased work output but rather, increases in body iron induce physiological and cognitive changes that can alter worker productivity. I hypothesize that the vehicles through which iron affect productivity are energy expenditure, cognitive function, physiological health, morbidity from infectious diseases, and labor intensity tradeoff decisions during break. I also predict that there will be behavioral changes from the women's response to iron, that I measure through changes in number of minutes spent during wage work hours as well as through changes in the number of days worked. Specifically, the research question I address in this study is, how does nutritional iron deficiency affect worker output through the production function intermediates? In other words, how does iron affect total worker output and how much of this change can be attributed to the effect of iron on the intermediate inputs to the production function for tea picking?

To identify how iron impacts worker productivity, I use data from a randomized controlled trial of tea pickers in India.[1] In this experiment, a group of 217 women from the Panighatta Tea Estate in the Darjeeling District of West Bengal, India participated in a 10-month double fortified salt intervention. The women that were recruited were 18 to 55 years old, were not pregnant, had hemoglobin levels between 80-130 g/L, and were experienced full time tea leaf pickers. The women were divided into two groups. The experiment was a randomized, double-blind salt fortification trial where the treatment group was given 7 mg/day of double fortified salt (DFS) (iodine and iron) and the control group was given iodine fortified salt. Physiological, socioeconomic, cognitive, and anthropometric data were collected at baseline and endline. At both baseline and endline, energy expenditure information was collected through heart rate monitors that women wore throughout the day. Data on cognitive function, morbidity from

[1] Dr. Jere D. Haas, Cornell University, graciously provided the data for this analysis.

infectious diseases, physiological health, and labor intensity, as measured both by worker logs and by accelerometers that women wore throughout the day.

I leverage the randomized nature of the intervention and estimate difference-in-difference models of the change in worker output and production function intermediates between control and treatment groups from baseline to endline. I first show that the randomized trial was indeed random: the treatment and control groups are balanced on all observable measures in the baseline survey. Furthermore, the treatment was successful in increasing iron levels, with increase of 1.5.

Despite the fact that treated women experienced a large increase in blood iron levels, total output of tea leaves picked did not increase in the treated versus control groups However, many of the production function intermediaries did change in response to the iron intervention. In particular, being treated with iron fortified salts. In order to relate these intermediate changes to output, I estimate Blinder-Oaxaca decompositions that decompose the total change in worker output into the part that is due to changes in each production function input and the part that is due to the changing returns to those inputs or to other, unobserved, factors. However, because there was such a small change in output and due to the large standard errors and the very small change in output, the coefficients are imprecise, and therefore, the results are only suggestive.

2. Literature Review

2.1 Background on Iron Deficiency

Iron is transported in the plasma bound to transferrin as diferric transferrin, a glycoprotein. Two-thirds of iron in the body is contained in hemoglobin in the blood and is absorbed through the intestine, where it is transported to the bone marrow and incorporated into hemoglobin, making up 75% of body iron. Iron deficiency can lead to iron deficiency anemia because of the incorporation of iron in red blood cells (Umbreit 2005). Bleeding, usually due to

parasites in the intestines in developing countries, is the most common cause of iron deficiency (Umbreit 2005).

The lower limits of what is considered iron deficiency anemia are somewhat different across different studies and vary with ethnicity and sex. Beutler and Waalen (2006) state that iron deficiency anemia is defined as a hemoglobin concentration below 13.7 grams per deciliter (g/dL) in a white man aged 20-60. The corresponding value for women of all ages is 12.2 g/dL. Severe iron deficiency anemia is defined as hemoglobin below 8 g/dL and moderate iron deficiency anemia between 8.0 and 12.0 g/dL. Iron deficiency non-anemia typically is defined as iron deficiency with normal hemoglobin, greater than12.0 g/dL and less than 18 g/dL (Haas and Brownlie 2001).

2.2 Iron Deficiency Anemia and Worker Productivity

An early experiment conducted by Edgerton et al. (1979) shows the effects of iron deficiency anemia on worker output on a set of tea leaf pickers at a plantation in Sri Lanka. The average weight in kilograms of tea picked per subject per day was measured as well as the tea per day per gram of hemoglobin per deciliter of blood. The experimental group was treated with 200 mg of ferrous sulphate, a tablet containing iron normally given to iron deficient patients, versus the control group, which was given 300 mg of calcium lactate, a tablet supporting absorption of calcium and magnesium but not iron. Physical activity was recorded through a wrist watch, belt, and ankle sensing devices. After one month's treatment, significantly more tea was picked with higher hemoglobin concentration through iron supplementation compared to the control. Subjects with initial hemoglobin concentration under 9.0 g/dl increased output from 15.6 to 17.5 kg/day, but when the initial hemoglobin was greater than 11.0 g/dl, mean output decreased with iron supplementation (Edgerton et al. 1979).

The study, however, had limited technology in measuring physical activity, and cognitive function was not measured. Furthermore, the iron treatment was given only for one month, after which both control and treatment groups were given iron. After the second month, there was no control group with which to compare the effects of iron, and therefore, the effects could not be measured beyond one month. Still, the experiment laid groundwork to show that iron deficiency anemia may affect productivity in some way.

Basta et al. (1979) analyzed iron deficiency anemia and labor productivity of adult males in Indonesia through a double-blind, randomized, placebo controlled trial. Seventy rubber tappers and eighty-three weeders received 100 mg of ferrous sulfate with glucose daily for a period of sixty days while the control group was given a placebo. Financial incentives took the form of 100% of daily wages for the group that finished the largest tract of land in one hour. During the intervention, both iron treated and placebo groups increased productivity.

It is important to note that there was a high drop-out rate in this experiment, and therefore, results may be biased from selective attrition. The subjects also reported feeling more hungry after taking the tablets and reported buying more of other types of food with the extra income supplement from more output, which could have affected iron levels. Basta et al. (1979) makes a crude estimate of a benefit cost ratio to be 260:1 for iron supplementation for tappers, one of the first papers to analyze economic benefits of an iron intervention.

Another experiment, involving less strenuous physical work, was performed on female cotton millworkers in China, where nonpregnant female workers were randomly assigned to ferrous sulphate capsules or placebo treatment for 12 weeks (Li et al. 1994). In the iron treated group, the mean hemoglobin level increased from 113 to 127 g/L ($p<0.001$), mean serum ferritin increased from 9.7 micrograms per liter (ug/L) ($p<0.001$), and mean free erythrocyte

protoporphyrin decreased from 1.01 to 0.49 micromols per liter (umol/L) ($p<0.001$). Output was measured in productivity efficiency, calculated as a ratio of output (Yuans/day) to energy expenditure (MJ/day), resulting in Yuans/MJ. Energy expenditure was estimated using minute-by-minute heart rate monitors based on the relationship between heart rate and energy expenditure. The results showed that productivity efficiency increased and that the women were able to do the same amount of pre-supplementation work at a lower energy cost. Textile work, however, is continuous and requires only moderate exertion and intensity. Therefore it is difficult to say how iron may affect productivity at higher levels of exertion or in more physically demanding jobs (Li et al. 1994).

In both the Edgerton et al. (1979) and Li et al. (1994) studies, the workers not only increased work output but also increased voluntary activity at home. The women from the Chinese mill increased time spent on leisure activities, housework, and other activities presumably because they exerted less energy at work. This finding suggests that iron deficiency, resulting in fatigue or tiredness, also can affect how people allocate their time for other activities outside of work.

Iron deficiency results in a large array of deficits in bodily functions, including tiredness, decrease in cognitive function, and decreased ability to work with increased susceptibility to illness. In the context of iron deficiency, Thomas et al (2003) reported results from an iron treatment intervention in Central Java, Indonesia on 0-70 year olds where iron supplements were given weekly to half the subjects and the other half were given placebos. After six months of intervention, there was an elevation in blood iron levels and there was evidence that those older adults who received treatment and who were iron deficient were better off in terms of physical health, psycho-social health, and economic successes. Males were working more, earning more,

and losing less work time due to illness; they were more energetic, slept less, and were able to participate in more physically taxing activities. Results were similar for females but the effects were less pronounced. These results indicate that iron supplementation leads to physiological changes that can make workers more productive and thus leads to increased work output. The goal of this analysis is to understand more fully how iron impacts these different inputs to worker productivity and how these input changes drive any worker output effects.

3. Intervention

3.1 Study Design

A group of 498 women from the Panighatta Tea Estate in the Darjeeling District of West Bengal, India were asked to participate in the study. The women that were recruited were 18 to 55 years old, were not pregnant, had hemoglobin levels between 80-130 g/L (80-130g/l is still considered anemic but not dangerously anemic to disqualify from study), and were experienced full time tea leaf pickers. From the 498 participants, 217 were dropped because they were non-residents, non-tea leaf pickers, too old or too young, had severe anemia, had poor health, were pregnant or were not interested, leaving 281 women for baseline blood data collection. From this group, 41 were not qualified due to severe anemia, poor health, injury, pregnancy, insufficient baseline data, or refusal, leaving 248 participants qualified for a follow up. These 248 women were stratified by hemoglobin level (anemic or non-anemic) and randomized into four groups. About 15% of those women (31 participants) did not return for follow up assessment after the intervention, leaving 217 women that completed the 10-month intervention.

The women were divided into two groups. The experiment was a randomized, controlled, double-blind salt fortification trial where the treatment group was given double fortified salt (iodine and iron) (DFS) which contained iodine (50mg/1000g) and ferrous fumerate (10mg/10g)

and the control group were given iodine fortified salt. The intervention was done over a 10-month period with the treated group receiving about 7mg/day of iron from ferrous fumerate added to the DFS. The salts were divided into four colors: black, blue, green, and red. Black and blue were the control salts and green and red were the double fortified salts. The salts were color coded with multiple colors to ensure that the randomization would be double blind for both the administrators of iron and the participants. Iron status indicators, hemoglobin, red blood cells, white blood cells, serum ferritin, soluble transferrin receptor, body iron, C-reactive protein, α_1-acid glycoprotein, vitamin B12, serum folate, and urinary iodine were measured a baseline and after 10 months. There were no monetary incentives to participate in the study except for healthcare, but the mini clinic with one physician was offered to all women regardless of participation on the tea estate.

Socioeconomic data were collected at baseline for marital status (currently married, widowed, separated, divorced, never married, deserted), literacy level (able to read and write or not), spouse's literacy level (spouse able to read and write or not), highest completed level of education (standard 1-2, standard 3-5, standard 6-8, secondary school, adult education, never been to school, don't know), spouse's highest completed level of education, whether the spouse works or not (or not applicable), spouse's type of work (does not work, currently unemployed, disabled, casual work, manual labor, family business, not applicable), monthly household income (less than 1000 rupees, more than 1000 rupees, unsure), number of people in the household, number of children in the household, type of family (joint or nuclear), type of diet (non-vegetarian or vegetarian), type of housing (low cost housing/traditional housing, domestic/estate quarters, other), household drinking water source (piped into yard, communal pipes, public well, well in yard, pond/lake, vendor, other), household toilet facility (flush to sewage or septic, pour

flush latrine, pit latrine, open pit, no facilities, other), whether the household has electricity or not, whether the household has appliances or not, and the type of fuel used for cooking (gas, electricity, wood, gas & wood, all three). Anthropometric data were collected at baseline, midpoint and endline as well. The data collected include age, height, weight (kg) accounting for clothing weight, mid-upper arm circumference (mm), and body mass index (weight/height^2).

The data was collected on a week-by-week basis for worker output, with a different group of women entering the experiment in different waves, starting in the end of July until end of September for baseline and starting the end of May and ending in the beginning of August for endline. There were about 25 women in each wave, some containing more and some less. There were two picking times during the day, morning picking and afternoon picking. The women would weigh their tea bags either two or three times a day, depending on how much tea they were able to pick that day. Usually, tea was weighed twice in the morning and once at the end of the day in the afternoon before leaving work. There was usually a break in between morning and afternoon picking that was two hours long, but the women could choose to start working earlier if they wished. Many times, the women took the full two hours to eat, take a nap, and talk amongst one another. More time constrained women sometimes used this time to go to the bank or do additional housework before coming back and finishing off the day. The distance to the tea weighing station was variable to where the women were picking each day. Sometimes it could be as little as a two minute walk, while other times it could be as far as a thirty minute walk.

Due to the large number of women participating in the experiment, cognitive function was only taken on a smaller subset of women, about 130 women for each cognitive measure, randomly chosen from the population.

3.2 Specific Measures

3.2.1 Iron

Most previous research has measured iron status through hemoglobin levels. This is usually done as a matter of convenience or practicality because hemoglobin is relatively easy to measure. However, hemoglobin does not measure iron deficiency in the absence of anemia, anemia can be due to other factors other than iron deficiency, and hemoglobin would not measure non-iron deficiency anemia. Therefore, I will use total body iron as the measure of the level of iron in the body. Cook, Flowers, and Skikne (2003) describes the calculation for deriving total body iron based on the logarithm of the concentrations in micrograms per liter of the ratio of serum transferrin receptor to serum ferritin[2]. This measure has been shown to increase accuracy and evaluation of iron status in the body because of its sensitivity (Cook, Flowers, & Skikne 2003). Furthermore, it has been tested as a useful clinical measure for those that may be very iron deficient. Values of sTfR from the commercial lab in India were adjusted to a calibration subsample (n=35) using the Ramco ELISA method[3] (Ramco, Stafford TX). The ferritin values were adjusted to account for acute phase inflammation effects using AGP and CRP with the Thurnham method[4]. These corrected values were then used to calculate body iron. I hypothesize a strong increase in total body iron for treated individuals at endline.

3.2.2 Energy Efficiency

[2] (-(log10 (sTfR1*1000/sFt1)-2.8229)/0.1207); sTfR is serum transferrin receptor, and sFt is serum transferrin, both measures of levels of iron in the blood. SFt holds free iron in the blood and sTfR are on the surfaces of cells to receive the sFt.

[3] ELISAs, enzyme-linked immunosorbent assay, are used to detect the presence of a certain substance in a sample. Ramco is the kit that made the ELISA toolkit to measure sTfR in the blood.

[4] The Thurnham method is an adjustment to total body iron because ferritin values increase with inflammation and therefore could underestimate iron deficiency without the correction. Thurnham DI, McCabe LD, Haldar S, Wieringa FT, Northrop-Clewes CA, McCabe GP. Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. Am J Clin Nutr. 2010;92:546–55

Energy expenditure can be predicted through a regression model based on gender, heart rate, and physical activity. Assah et al (2010) measured heart rate and physical activity of 33 adults in Cameroon and also used the doubly labeled water technique, another measure of energy expenditure that is highly accurate for long term studies but expensive to use in the field. Doubly labeled water ($^{3}H_2{}^{18}O$) is based on the principle that the process of expending energy results in a predicted loss of CO_2 and this loss can be measured by the dilution in body fluids of labeled hydrogen as deuterium (^{3}H) and oxygen (^{18}O). There were no statistically significant differences between the physical activity energy expenditure predicted models using heart rate, physical activity (through uni-axial accelerometers), and gender versus the results from the doubly labeled water (Assah et al 2010).

Polar A3 monitors, consisting of a Polar T31C Transmitter and a Polar A3 receiver, monitored heart rate for each of the subjects during tea picking. The Polar T31C Coded Transmitter Set, fastened to an elastic band, is worn around the participant's chest and does not interfere with physical activity. The transmitter, (100g), was moistened at the location in which it contacts the skin to ensure conductivity. The receiver looks like a watch and was kept in a pouch that was hung around the participant's neck.

Energy expenditure was calculated using heart rate data. There is a close relationship between energy expenditure and heart rate, where the relationship is segmented into linear functions, the abrupt change occurring at a "flex point" between levels of rest and activity, usually defined as the average of the highest resting heart rate value and the lowest exercising heart rate value (Spurr et. al., 1988). According to Leonard (2003), estimates of energy expenditure from the flex-heart rate method are highly correlated with one another ($r=0.88$, $p<0.001$). The flex point separates light activity from more strenuous dynamic muscle work and

the separation between the two levels is usually between 80 and 100 beats per minute (Hiilloskorpi 2003) Spurr et. al. (1988) used an average flex point of about 96 (SD=6) beats per minute in adult females. To be on the conservative side, I used 90 beats as the cutoff flex point for the females tea pickers in our group because the women were very physically fit and therefore most likely to have a lower flex point heart rate. Below the flex point, energy expenditure can be estimated from weight and age, and is independent of changes in heart rate. Above the flex point, the relationship between energy expenditure and heart rate is linear, but this relationship depends on sex and body weight and height (BMI) (Hiilloskorpi et. al. 2003). Above 85% of an individual's maximal heart rate, the relationship becomes, once again, nonlinear. But since tea picking does not require extreme exertion (Hiilloskorpi et. al. 2003), subjects are mostly likely not reaching that level of exertion.

Energy expenditure in this study was based on a model developed in a field-based validation study of 39 tea pluckers from Panighatta using portable indirect calorimetry, heart rate, and physical activity data to predict energy expenditure (Przybyszewski 2011). The model separates out rest versus non-rest activity and uses different regression equations for both types of activities. Below the flex point, considered rest, energy expenditure is estimated through age and weight. At non-rest, energy expenditure is estimated through heart rate and BMI and the interaction of the two.

The equation for predicting REE (resting energy expenditure for heart rates below 90 bpm) is:

$$\text{logREE} = 0.501564 + 0.011546*\text{weight} - 0.007846*\text{age}$$

The equation for predicting AEE (active energy expenditure for heart rates above 90 bpm) is:

$$\text{logAEE} = -20.948266 + 11.179237*\text{logHR} - 3.61537*\text{BMI} + 1.91938 * \text{logHR} * \text{BMI}$$

(BMI is categorical: 1, <18.5; 2, 18.5≤ BMI ≤24.5; 3, >24.5, where BMI is a set of dummy variables) (Przybyszewski 2011).

The CosMed K4b^2, portable indirect calorimeter, provided breath-by-breath output of heart rate (bpm), volume of oxygen consumption (VO_2) and CO_2 production (VCO_2) in L/min, from which energy expenditure (kJ/min) is calculated. Heart rate (bpm) was also measured through a Polar A3 heart monitor, and time of measurements recorded (hh:mm:ss). The output was summarized into minute-by-minute data. Accelerometry and physical activity data were measured through activity counts using the Actigraph GT3M triaxial accelerometer and expressed as counts per minute. Energy expenditure was calculated in kJ/minute (Przybyszewski 2011). The flex point could not be individually measured for every single person in this study because of technical constraints, therefore, the equation was used to generate a population regression model estimating the flex point. For this study population, Przybyszewski (2011) shows that considering population heterogeneity in height, weight, and age, a multivariate prediction equation which also includes heart rate produce more accurate predictions of energy expenditure than an equation based on heart rate alone. I also looked at accelerometry data during work as an alternative measure to energy expenditure.

I hypothesize that iron treated individuals would have lower energy expenditure per kg of tea leaf picked, resulting in higher work output or same work output with lower energy expenditure compared to the control group. Energy expenditure from heart rate was measured in kilocalories. Accelerometry data was measured in metabolic equivalent tasks (METs).

3.2.3 Cognitive Function

In this experiment, cognitive data was collected at baseline and endline through simple reaction time tests, contrast threshold tests, and temporal threshold tests. The test was

administered to 138 of the participants because of resource constraints, all of whom were randomly chosen from the population. These tests measured how iron affects perception. The contrast threshold, temporal threshold, and simple reaction time are used for the productivity data because they were the most relevant to worker productivity of quantity of tea leaves picked.

The contrast threshold task and the temporal threshold task are collectively known in psychology research as the method of constant stimuli. According to Ehrenstein and Ehrenstein (1999), the method of constant stimuli has been described in psychophysical tasks and methodologies as one of the most useful in sensory research. In this method, the experimenter chooses a number of stimulus values likely to encompass the threshold value. After the stimuli are presented at random, the observer reports whether or not she detected the stimulus and whether the intensity was stronger or weaker than the standard. When the stimulus intensity is presented multiple times, responses are calculated for each stimulus level and plotted on a psychometric function graph of percentage of perceived stimuli versus stimuli intensity. The graph is usually sigmoidal showing that lower stimuli intensities are detected occasionally, higher values detected more often, and intermediate values detected at some trials (Figure 1A). Variability results from fluctuations in sensitivity due to the biological sensory system. The method of constant stimuli is said to provide the most reliable threshold estimates but its weakness is that it is time-consuming and requires a patient and attentive observer because multiple trials are required (Ehrenstein & Ehrenstein 1999).

The contrast threshold task was a test given to measure the ability of participants to distinguish between varying degrees of contrast, measured on a scale from 0-80, with 0 being no clarity in the image, and 80 being an extremely clear image (Figure 1B) with a fixed exposure time. Results were recorded of mean clarity of the image where the participant most consistently

correctly detected the contrast image. This is especially important in this specific case where the tea leaf picker's productivity can be greatly increased from being able to quickly and accurately distinguish between tea leaves.

The temporal threshold task was a test given to measure the ability of participants to detect an image with a constant contrast of 30% with varying time, measured in milliseconds (Figure 1C). Results were recorded of the mean time elapsed before participants detected the image. This measure is especially important for worker productivity because how quickly a participant can detect tea leaves can affect total tea that she can pick.

In contrast to the method of constant stimuli, the simple reaction test, falling under psychophysical methods called chronometric methods, is used to describe situations in which stimuli are easily perceived or when stimuli are above a threshold (Ehrenstein & Ehrenstein 1999). Chronometric methods are used to study behavior involving cognitive processes and perceptual functions to determine the amount of information that is processed in working memory or to study the complexity of stimulus-response relations. Simple reaction time is used to study sensory performance. Simple reaction time decreases, in general, as stimulus intensity and separation increase up to a certain stimulus strength, after which there is little change in reaction time. The simple reaction test was used to measure motor reactivity. Participants were directed to look at a computer screen. At random intervals the image "*" appeared on the screen for a brief moment. Participants were directed to push a button as soon as they saw the image. I will use the median reaction time in milliseconds that a participant will be able to see and react to an image on the computer screen. This measure is another measure to complement the contrast threshold and temporal threshold tasks by seeing how quick participants can react to what they

see. This factor is important in our experiment because how quickly participants can react by picking the tea leaves can affect how much tea is actually picked.

Iron, in terms of cognitive function, affects perceptions of contrast and motor control. Therefore, they likely are strong measures of cognitive function. Furthermore, these cognitive measures should directly affect the amount of tea leaves a worker can pick, and therefore are relevant measures for explaining changes in worker output in this study. However, as the literature review shows, we do not yet understand all of the mechanisms and potential effects iron may have on the brain. While we may be missing other variables that may affect worker output through changes in cognitive function, these measures are far more extensive than have been used in previous work in this area and will provide new evidence on how iron supplementation affects worker cognition. I hypothesize that cognitive functioning will increase with quicker reaction times and better perception of contrast.

3.2.4 Physiological Health

In this study, physiological health was measured by differences in baseline and endline measures of mid-upper arm circumference (MUAC). Research done by Powell-Tuck and Hennessey (2003) show that MUAC closely correlates with body mass index (BMI), but MUAC is easier to measure and better predicts outcomes, such as mortality rate and length of hospital stays. BMI and weight loss are common measures used to measure nutrition of patients. Powell-Tuck and Hennessey's (2003) experiment was undertaken to measure how improved nutritional status may affect BMI or MUAC. MUAC better predicted outcomes compared to BMI, even if there were missing data points. Cawley and Burkhauser (2006) further summarize that there is extensive medical literature that explain that using BMI is not an accurate measure of fat because it does not distinguish fat from fat-free mass such as muscle and bone.

MUAC was measured in mm at baseline and endline. BMI (kg/m^2) was calculated through the commonly used equation [(net weight) / (height^2)] * 10,000. Increased iron can lead to decreased general malnutrition and make a person more active, directly affecting muscle mass, body weight, and vigor. Therefore, MUAC should be affected. Weight gain reflected in increase in MUAC is considered a good measure of increasing health, especially in the cases of developing countries where there is undernourishment. However, if the women are still very constrained in terms of their incomes and the additional income that they potentially make goes to purhcases other than increased food quality and quantity, MUAC may not be affected by increased iron levels. However, both BMI and MUAC measures will be used to see which is a stronger predictor of productivity because there is an argument that MUAC may be a weak measure due to its use of only one body part. I hypothesize a small increase in physiological health.

Folate and vitamin B12 were measured as controls to detect causes of non-iron deficiency anemia as baseline tests. Since this was an iron supplementation trial, folate and vitamin B12, unaffected by iron, should not change, even though both are also measures of anemia. Folate and Vitamin B12 were also measured to determine the prevalence of non-iron-deficiency anemia. Folate was measured in nanograms per milliliter (ng/mL), and Vitamin B12 was measured in picograms per milliliter (pg/mL). An individual is deficient in serum folate if folate levels fall below 3.0 ng/mL. For vitamin B12, less than 250 pg/L indicates moderately low levels, less than 170 pg/L indicates low levels, and less than 100 pg/L indicates very low levels (Gibson 1990).

3.2.5 Morbidity from Infectious Disease

Morbidity from infectious disease will be measured through white blood cell counts. Bain (2006) describes that increases in white blood cells counts are usually due to increases in the

number of neutrophils. Neutrophils are the body's natural defense against pathogens. The number of neutrophils in the blood therefore increases as a result of infections from infectious diseases. White blood cell counts were measured at baseline and endline in 1000/mL.

Alpha-1-acid glycoprotein (AGP) were measured in milligrams per deciliter (mg/dL) at both baseline and endline. AGP is a single chain peptide of 183 amino acids in humans and its serum concentration responds to systemic tissue injury, inflammation, and infection (Fournier, Medjoubi-N, and Porquet, 2000). AGP was used as another measure of infection from infectious disease burden.

C-reactive protein (CRP), measured in milligrams per liter (mg/L), is a plasma protein that increases in response to inflammation. CRP binds to surfaces of pathogens and increases within hours of tissue injury or infection, contributing to innate immune response and host defense (Black, Kushner, and Samols, 2004). Therefore, CRP was used as another measure of infectious diseases burden.

Both AGP and CRP are used as markers of inflammation from infectious diseases, and therefore are accurate measures of infectious disease burden. Furthermore, they are good measures because they are both relatively short-term inflammatory responses, and since our experiment is relatively short term, the measures can accurately reflect whether or not a person may be affected by infectious disease. AGP and CRP counts can also be high from reasons other than infectious diseases and inflammation can occur from things like cuts or fevers. Since AGP and CRP are short-term inflammatory markers, and this experiment was conducted over a ten month period, there could have been many infections during that time that did not affect baeline or endline AGP and CRP values. However, work productivity at base and endline could be influenced by concurrent infections, which could be reflected in elevated AGP, CRP, and WBC.

Therefore, changes in AGP and CRP levels should be used to complement white blood cell counts, which are a more direct way of measuring response to infectious diseases.

Cutoff values for AGP and CRP values are used to construct dummy variables, where 0 is normal and 1 is abnormal in order to account for certain extremely large variations in inflammation. Values of AGP greater than 133mg/dL are considered abnormal and values of CRP greater than 3 mg/L are considered abnormal according to values used to interpret the effect of inflammation on serum ferritin concentrations as indicators of total-body iron stores (Beard et al. 2006).

3.2.6 Labor Intensity

The change in labor intensity will be assessed through the accelerometry data, to measure the level of daily activity at baseline and endline during the day. The accelerometers measure minute-by-minute activity patterns from which one could determine whether a person was resting versus working either on the job or at home. The cutoff was set at 157.1 counts per minute to distinguish between rest and non-rest activity through an analysis of a subgroup of 39 women. By analyzing when the participant exerted less than 157.1 counts and when the participant exerted greater than 157.1 counts, I can determine the number of hours spent at rest versus in light and moderate-to-vigorous activities. The accelerometer counts were converted to Metabolic Equivalent Tasks (METs) according to a method by Crouter et. al. (2010). One MET is equal to the amount of energy expenditure at resting metabolic rate. The data was divided into the number of minutes that a participant spent in sedentary behavior (1.00 MET), light physical activity (1.01-2.99 METS), and moderate or vigorous physical activity (greater than 3.00 METS). The minutes were calculated for each day, and then they were averaged across the whole week. Each activity level was separated into before work, morning tea picking, lunch break, afternoon

tea picking, and after work. Time spent sleeping was shown through continuous zeros in the Autograph throughout the night when the ActiGraph was removed. The amount of time the women spent in each type of activity – before work, during break, and after work – was examined. By measuring the number of minutes the women spent in each type of activity, I can measure the changes in each type of activity level from baseline to endline. If the women were, for example, spending less time in moderate or vigorous physical activity at endline, then they were exerting less energy during that time, and therefore, most likely enjoying more sedentary time.

Activity times were classified into before morning work, morning picking, lunch break, afternoon picking, and after afternoon work. Of particular interest are non-work activities, before morning picking (called before work), lunch break, and after afternoon picking (called after work), where the women have more discretion on how to use their time for the labor intensity tradeoff.

I hypothesize that there will be no change in the number of hours of labor or leisure during work hours because these hours are standardized and there is not much freedom in how much the women can work. However, I predict that during non-work hours, before morning picking, during lunch, and after afternoon picking, women will adjust their time worked accordingly. Those women that are especially constrained with time will work more because they now have more energy, and women who are less constrained will work less because they will have worked more efficiently during work hours. I will be testing for with this production intermediate is whether income or substitution effects dominate in terms of the total energy output used for work which is a novel analysis in the literature on iron and worker productivity.

3.2.7 Work Output

Work output can be measured in a variety of ways. The women generally worked six days a week picking tea leaves and the tea was measured in kilograms at least twice a day. Work output was computed in three different ways to determine which may be the best way to reflect output. One measure is a simple division of total kilograms of tea picked per week over the total number of days that tea was picked. Therefore, if the participant were absent one day, that day would not be included. This measure represents the intensive margin, how much a participant produces conditional on working. A second measure is to examine the extensive margin, which is the number of days per week a participant worked, irrespective of how much was produced. Finally, total output was computed as the total kilograms of tea picked per week divided by six days. This measure includes both the extensive and intensive margins. I will analyze the effect of the intervention on total output, the intensive margin and the extensive margin separately in order to measure how output responds to iron treatment. A final measure that I use is just morning picking. This method will just be used as a check on the other measures because there was a lot of variability in the data collected for output from a strike that occurred during the data collection and other unpredictable events throughout the week, and the morning data was a lot more consistent.

3.3 Descriptive Statistics

There were 215 participants in the final data set, two of the participants had missing data. There were 109 participants in the control group and 106 participants in the treatment group (DFS). There were 109 Adivasi and 109 Nepali in the control group and 106 Adivasi and 109 Nepali in the treatment group. Eleven women made less than 1000 rupees for their monthly income and 203 women made more than 1000 rupees for their monthly income. The average

household size, the number of total people in the household, was 5.266. The average number of children per household was 2.07.

Seventy women were able to read and write and 144 were not able to read and write. Spouse literacy was 0 if she was not able to read and write and 1 if he was able to read and write.

Education was equal to 1 if the highest completed education was standard grades 1-8, 2 if completed secondary school and above, and 3 if they had never been to school or they did not know. The same coding was used for spouse education levels. Spouse type of work was coded to 0 for no work, 1 for casual work, manual, labor, and family business, and 2 for currently unemployed, disabled or not applicable. Household drinking water was coded to 0 if obtained from a public source (communal pipes, public well, pond/lake, or other) and 1 if obtained from a private source (piped into yard or well in yard). Fuel was coded to be 1 for gas, 2 for electricity, 3 for chula (a method of cooking), 4 for gas and chula, and 5 for all three.

4. Methods

4.1 Specific Measures

4.1.1 Energy Expenditure

Iron-sulfur and heme-containing enzymes participate in single-electron transfer reactions involved in energy metabolism (Beard 2001). Both iron deficiency anemia and iron deficiency nonanemia, when a person has depleted their iron stores but decreased hemoglobin levels are not yet detected, lead to decreases in work productivity and the ability to use energy efficiently. Haas and Brownlie (2001), provide a thorough outline of the biomedical consequences of iron deficiency. Iron is very important in oxidative energy production and is a key ingredient in hemoglobin, a protein that makes up red blood cells and transports oxygen to the body's cells. When hemoglobin levels decline, the amount of oxygen the body can use also declines. When

iron stores are completely used up, the amount of oxygen available to muscles also decreases, which can reduce endurance and make the heart work harder. There is evidence from animal and human studies that a causal relationship exists between iron deficiency and reduced maximum aerobic capacity (Haas & Brownlie 2001). Furthermore, iron is required for other body processes like ATP[5] production, DNA synthesis, mitochondrial function, and protection of cells from oxidative damage. In iron deficient anemic individuals, there is a 50% decrease in muscle myoglobin content, cytochrome oxidase activity, and electron transport capacity in skeletal muscle as well as 50% decrease in oxygen transport capacity from anemia (Beard 2001).

In addition, there is evidence that productivity is not just affected by iron deficiency anemia but also by iron deficiency without anemia. Beard (2001) explains that oxygen is bound to myoglobin in the muscles. Myoglobin is a single-chain hemoprotein that increases the rate of oxygen diffusion from the red blood cells to the cytoplasm and mitochondria. If body tissues are iron deficient, skeletal muscle capacity is drastically reduced from decreased diffusion of oxygen to mitochondria. This directly affects muscle and energy expenditure for iron deficient individuals whether anemic or not.

Haas (2006) explains that although the effects are less severe than iron deficiency anemia, iron deficiency non-anemia may be significant in endurance capacity and energetic efficiency through reduced oxygen transport, reduced cellular oxidative capacity, energetic efficiency, and reduced work productivity.

In summary, the literature suggests that increased intake of iron increases the amount of work that can be accomplished. This increase could raise worker production output in the current study through either increasing the amount of tea leaves picked or decreasing the amount of

[5] ATP stands for adenosine triphosphate and is a coenzyme that helps transport chemical energy in cells for metabolism

energy expended per unit of tea leaves picked. This increased work efficiency would allow for more time and effort devoted to discretionary or non-work activities.

4.1.2 Cognitive Function

Iron is the metal with the second highest concentration in the brain (McCann & Amnes 2007). Iron is rich in the basal ganglia, the substantia nigra, globus pallidus, and nucleus accumbens, which play a role in the transmission of the neurotransmitter dopamine in adult life but are less affected by iron deficiency compared to the cortex or the striatum. The basal ganglia as a whole play a role in gating information to different parts of the brain, including motor control and procedural learning (Penney & Young 1986). Iron is primarily located in the microglia and oligodendrocytes. Studies show that the effect of iron deficiency on brain iron content may depend on timing of nutritional deficiency (Beard 2001). Iron is high in content in the cerebrospinal fluid and it plays a key role in myelination of the spinal cord and white matter of cerebellar folds (Beard 2001). However, it has been difficult to tease out whether iron affects cognitive function due to iron's interaction in the brain or due to anemic affects of physical tiredness and physiological effects. There is clear evidence that iron is important with children less than 2 years of age, where nutrients are very important in brain formation and cognitive function (McCann & Amnes 2007). Iron deficiency is common during late infancy and toddler periods, which is a very important time for hippocampal and cortical regional development, myelinogenesis, dendritogenesis, and synaptogenesis in the brain. In the iron deficient rat brain, there was a loss of cytochrome-C oxidase in the highly metabolic hippocampal areas CA1 and CA3 in the frontal lobes of the cortex. Magnetic resonance spectroscopy of the brain showed that iron deficiency resulted in metabolic changes in the hippocampus at rest that lasted through

puberty. This may result in deficiencies in recognition memory and processing (Lozoff et al. 2006).

There have only been a few studies on iron deficiency and cognitive development of children older than 2 years, but results show that iron therapy over a two-month period improved IQ and that iron may reverse the IQ deficit in anemic 5-6 year old Indian boys by half a standard deviation (Horton & Ross 2003). A study that followed a group of Israeli children at 9 months and assessed their development at 2, 3, and 5 years showed that hemoglobin levels at 9 months were correlated with IQ at 5 years and that for each 10 g/L increase in hemoglobin, there was a 1.75 point increase in IQ score (Lozoff et al. 2006). Another study in France showed that mean cell hemoglobin concentration at two years was correlated positively with overall developmental, motor and social quotients at 2 years (Lozoff et al. 2006). An important study in Florida showed that for each unit decrease in hemoglobin at entry into the Women, Infant, and Children hemoglobin screening program, there was a 1.28 increased risk of being placed in special education (Lozoff et al. 2006).

There is only one study that looks at the effects of iron on children older than ten. This study, conducted in Costa Rica, found that those who were chronically iron deficient in infancy tested lower in math and writing achievement and motor function compared to those children that had good iron status as infants. Tests showed that the iron deficient children did worse on executive function tasks, specifically for inhibition and planning (Lozoff et al. 2006). I am not aware of any studies that look at the impact on adult cognitive function.

In summary, iron has a direct physiological effect on cognitive function. Improved iron status affects cognitive function, which affects worker output through its effects on memory, motor coordination, and various neurotransmitter functions that improve motor memory control.

The current study looks into the cognitive effects that iron may have on adults, an area where there is less literature available.

4.1.3 Physiological Health

Anemia can be caused by factors other than iron deficiency, including folate deficiency or vitamin B12 deficiency, and is a result of iron deficiency only about half the time. It is, therefore, important to make a distinction between iron deficiency anemia and other types of anemia common in developing countries. However, iron deficiency may not always exhibit symptoms.

Research by Kanani and Poojara (2000) showed in a randomized control trial that iron-folic acid tablets given to anemic girls increased hemoglobin levels by 17.3 g/L and weight gain by 0.83kg, a statistically significant increase. Iron and folic acid supplementation also resulted in increases in growth. A person who was previously malnourished or undernourished but gains weight as well as muscle mass during recovery likely will be stronger and therefore produce more work output because of reduced fatigue due to the body's ability to metabolize additional energy.

4.1.4 Morbidity from Infecitous Diseases

Iannotti et al. (2006) compile evidence that iron supplementation is correlated with increased morbidity under certain infectious diseases such as malaria, tuberculosis, and HIV/AIDS. Iron supplementation at higher doses can increase the incidence, duration, or severity of the infection. This may be due to the fact that iron is an important nutrient for pathogens that attack the body's tissues. Iron is important for both host survival and for the metabolism of invading pathogens. In hypoferremia, a method of immune defense, iron is bound to proteins to reduce the amount of iron available for parasites. A compilation of 28 randomized control trials

show that iron supplemented children showed no increase in the incidence rate ratio for all infectious illnesses including respiratory tract infection, diarrhea, malaria, and other infections (Iannotti et al 2006). Iannotti et al. (2006) summarize a large amount of literature that investigated whether iron may be associated with increased risk of illness and morbidity. Only one study from Indonesia shows a positive effect of iron on reduced frequency of fever, respiratory infection, and diarrhea for children ages 2-5 yeras.

Three large-scale infectious diseases prevalent in developing countries that iron may affect are malaria, HIV, and tuberculosis. Malarial infections result in red blood cell destruction and their infection with parasites. Red blood cells carry iron bound to hemoglobin, and therefore, there may be a physiological negative relationship between iron status and malaria. Infectious diseases, including HIV, on the whole are negatively correlated with increases in nutritional status. Although there are no randomized control trials relating HIV/AIDS and iron status, increased nutritional status, including iron status, would lead to stronger body defenses against HIV/AIDS and other infectious diseases. Growth of *Mycobacterium tuberculosis*, the pathogen that causes tuberculosis is enhanced by iron. The bacteria grow in macrophages, where iron is loaded. Therefore, although there are no randomized control trials as of yet examining tuberculosis and iron status, this is an area that should be researched.

Iron also is important for immune response function because it is required by the host for initiating and sustaining immune responses against infectious diseases. Because iron is essential for cell differentiation and growth, and for peroxide-generating enzymes and nitrous oxide-generating enzymes that play a role in immune cell function, this mineral plays a very important role in immunity. Nonspecific and cell-mediated immunity is decreased in iron deficient individuals. Furthermore, bactericidal activity of macrophages is decreased, neutrophil ability for

intracellular killing of pathogens is decreased, and the number of T-lymphocytes is decreased. This finding may be due to DNA synthesis being limited by iron deficiency or control of cell differentiation (Beard 2001).

Therefore, it is unclear what effects iron has on morbidity from infectious diseases and how any such interaction will affect worker output. While more research is needed to tease out more directly the degree to which micronutrient deficiencies play a role in increasing or decreasing prevalence of infectious diseases, this study will provide some evidence about these linkages.

4.1.5 Labor Intensity

Consumers maximize their welfare and utility over consumption of goods and services (including leisure) subject to a budget constraint that includes time spent working. This framework has been widely used for analyzing household labor supply, intra-household time allocation, investment in human capital and labor supply response. When wage rates increase, people may increase leisure, but this prediction depends on whether the income or the substitution effect dominates. If people have a target income as shown by Chou[6], they may adjust their behavior and reduce their willingness to work and thereby increase leisure time. They may also have a target number of minutes worked as shown by Farber. [7]

In developed countries, there is usually a clear distinction between labor and leisure. However, in developing countries or rural areas, labor and leisure are not strictly separate and the overlap between the two can be very complex. Just a minimum amount of leisure for recovery

[6] Chou, Yuan K. 2000. "Testing Alternative Models of Labor Supply: Evidence from Taxi Drivers in Singapore." Research Paper no. 768, Dept. Econ., Univ. Melbourne.

[7] Farber, Henry. 2003. "Is Tomorrow Another Day? The Labor Supply of New York Cab Drivers." Working Paper no. 9706 (May). Cambridge, Mass.: NBER.

for the next day of work may be needed, especially in farming settings. Leisure is usually measured as the number of hours in the day minus farm work time and home work, including travel time (Ruben & Ruiter 2002).

Therefore, the labor intensity (or in this case, the intensity of activity during non-tea plucking time) tradeoff can play an important role in this analysis. The women in this study were paid a piece-rate pay structure by how the number of kilograms of tea they were able to pick. Women may be spending more leisure time if they picked the tea leaves more quickly in their allotted time, so the income effect dominates. However, the women may also be picking more because they are able to pick more for less energy, which is an implicit wage increase, and they may labor more, so the substitution effect dominates. This analysis is the first to identify how iron affects the labor-leisure tradeoff, which is important because iron may impact labor supply in a manner that increases utility without fully impacting output. In paying attention to how worker behavior changes due to the iron intervention, this analysis will be able to detect changes in worker well-being that are not necessarily reflected in changes in worker output.

When a person is using less energy per output of work and has a target income, that person may make a labor intensity tradeoff decision and enjoy more leisure. Decreased energy expenditure per amount of work or increased energy from iron supplementation is like a subsidy because picking tea is less expensive after intervention than before intervention. Due to the income effect, women will consume more normal goods with the additional income, including "leisure", and so they will work less. The substitution effect predicts that, women will work more because they can pick more tea per hour with less energy expenditure, increasing the amount of earnings per hour and making leisure more expensive. What is actually measured here is labor

intensity, because the women are so constrained that even during lunch break, they are most likely doing some type of work, just not picking tea leaves.

4.1.6 Worker Productivity

$$P = \alpha + \beta E + \delta C + \mu M + \tau P + \rho L + \varepsilon$$

Figure 2 summarizes the factors that may affect the pathway from iron supplementation to work output. The hypothesis is that the production function intermediates would increase work output through either increase in the amount of tea leaves picked or from increasing the rates of returns of the production function intermediates E (energy expenditure), C (cognitive function), M (morbidity from infectious diseases), P (physiological health), and L (labor intensity). From the productivity model, I would be able to see, for example, how the coefficient β explains the effects or iron on work output through cognitive function.

4.2 Models

A difference-in-difference approach will be used to measure the change in work output at baseline and endline. Due to randomization, observables and unobservables that are not affected by iron should be the same pre and post treatment. As checks, I ran a regression of a studycompleted variable on each of the intermediate variables. There were no differences between those that stayed in the study throughout and those that dropped out except for body iron but it was only statistically significant at the 10% level (Figure 1A).

An Oaxaca-Blinder decomposition (Oaxaca 1973) will be used to decompose the total change in worker output into the part that is due to changes in each production function input and the part that is due to the changing returns to those inputs. Changes in each of the intermediate variables will be estimated from differences in baseline and endline values. Then, the production function coefficients will be used to decompose the effects that each of the intermediates has on

the final production output. I will perform this decomposition separately for all output measures discussed above.

4.2.1 Blinder-Oaxaca Decomposition Equation

The equation used is summarized below. Y_2 signifies endline data and Y_1 baseline data. *Ene* is energy expenditure, *cog* is cognitive function, *hea* is physiological health, and *ll* is the additional time spent in break.

$$Y_2 = \beta_{20} + \beta_{2ene}x_{2ene} + \beta_{2cog}x_{2cog} + \beta_{2mid}x_{2mid} + \beta_{2hea}x_{2hea} + \beta_{2ll}x_{2ll} + \varepsilon_2$$

$$Y_1 = \beta_{10} + \beta_{1ene}x_{1ene} + \beta_{1cog}x_{1cog} + \beta_{1mid}x_{1mid} + \beta_{1hea}x_{1hea} + \beta_{1ll}x_{1ll} + \varepsilon_1$$

$$\begin{aligned} E(Y_2 - Y_1) = {} & (\beta_{20} - \beta_{10}) + \beta_{1ene}(\bar{x}_{2ene} - \bar{x}_{1ene}) + \bar{x}_{2ene}(\beta_{2ene} - \beta_{1ene}) + \\ & \beta_{1cog}(\bar{x}_{2cog} - \bar{x}_{1cog}) + \bar{x}_{2cog}(\beta_{2cog} - \beta_{1cog}) + \\ & \beta_{1mid}(\bar{x}_{2mid} - \bar{x}_{1mid}) + \bar{x}_{2mid}(\beta_{2mid} - \beta_{1mid}) + \\ & \beta_{1hea}(\bar{x}_{2hea} - \bar{x}_{1hea}) + \bar{x}_{2hea}(\beta_{2hea} - \beta_{1hea}) + \\ & \beta_{1ll}(\bar{x}_{2ll} - \bar{x}_{1ll}) + \bar{x}_{2ll}(\beta_{2ll} - \beta_{1ll}) + E(\varepsilon_2 - \varepsilon_1) \end{aligned}$$

Δ kg tea picked = $E(Y_2 - Y_1)$

Δ due to energy expenditure = $\beta_{1ene}(\bar{x}_{2ene} - \bar{x}_{1ene})$

Δ due to cognitive function = $\beta_{1cog}(\bar{x}_{2cog} - \bar{x}_{1cog})$

Δ due to physiological health = $\beta_{1mid}(\bar{x}_{2mid} - \bar{x}_{1mid})$

Δ due to morbidity from infectious disease = $\beta_{1hea}(\bar{x}_{2hea} - \bar{x}_{1hea})$

Δ due to additional leisure = $\beta_{1ll}(\bar{x}_{2ll} - \bar{x}_{1ll})$

The changes are due to the baseline coefficient times the difference-in-differences of the production function intermediates. I controlled for Vitamin B12, folate, ethnicity, experience, treatment, study completed, marital, literacy, spouse literacy, education, spouse education,

spouse work, spouse worktype, income, household size, number of children, family type, diet, type of housing, water source, electricity source, appliances, fuel, age, height, and net weight.

4.3 Baseline Characteristics Check

Each baseline characteristic variable was regressed on treatment status in order to determine if there were any significant differences between treatment and control groups at baseline, to confirm the effectiveness of the randomization scheme.

The results from these balance tests are shown in Table 2A-2B. There were no statistically significant differences in any of the characteristics at baseline. At the 10% level, education level, height, and average break were statistically significant (Table 3). I also looked at the percent of the population that was deficient in certain characteristics such as BMI, hemoglobin, serum ferritin, and transferrin receptor (Table 2C). As can be seen in the Table, the a fairly large percentage of the population is deficient in all of the characteristics, going as high as 75% for transferrin receptor and 52.4% deficient for hemoglobin.

5. Results

5.1 Difference in Difference Estimates

Differences between endline and baseline for each of the intermediate production functions were estimated. Each of the differences was regressed on treatment status. At the 5% level, total body iron, simple reaction time, contrast threshold, temporal threshold were all statistically significantly different between treatment and control groups. Total body iron was actually different between control and treatment groups at the 0.001 level (Table 3A) as well as temporal threshold (Table 3B).

The changes in all of the cognitive function values are statistically significant at the 10% level. All of the cognitive function values had very strong changes, stronger than in any previous

iron treatment study including measures of cognitive function. There were also statistically significant differences in accelerometry data during work ($p<0.05$) (Table 3E). None of the other inputs were statistically significantly between treatment and control groups (Tables 3C, 3D, 3F). I looked at accelerometer activity data during break (3C) and before and after work (3D), and there were no statistically significant differences in the number of minutes spent in sedentary, moderate, or vigorous physical activity during any of those time periods.

As expected, total body iron increased significantly between treatment and control groups from baseline to endline. When looking at the Kernel Density plot, total body iron has a normal distribution and clearly shifts up for the treatment group for iron status, while the other distributions stay about the same (Figure 3A). This shows that the fortification trial was successful and that there was a strong improvement in iron status due to the randomized control trial.

All the measures of cognitive function also seem to have improved by significant amounts. Since there were potentially more outliers with this measure, the simple reaction time median was used rather than the mean. The Kernel Density Plot shows a skewed distribution to the right because there was no limit as to how long it could take for the participant could take before detecting the image. The skewed distribution resulted in a statistically significant difference ($p<0.02$) at endline treatment even though it may not be obvious from the distribution plot. Figure 3B shows a very clear shift in temporal threshold for the treated group at endline and is normally distributed. There is a shift to the end because the iron treated group was predicted to detect the images quicker, shifting the distribution to the left. The Kernel Density Plot for the contrast threshold test is shown in Figure 3C. Although the plot is not normally distributed for treatment endline, the endline treatment is clearly different than the treated baseline, control

baseline, and control endline. The plot is not normally distributed because there is a large hump at 2 because any values below a certain threshold were given a value of 2. The second hump shows the shift at endline treatment. Lastly, Figure 3D shows the Kernel Density Plot for the simple reaction test. There seems to not be a strong shift in the hump of the distribution, but there seems to be a much tighter distribution around the mean at both endline treatment and control.

There were no statistically significant changes for the other intermediate production functions. Energy expenditure was distributed normally according to the Kernel Density plot. There was no statistically significant increase in energy expenditure between treatment and control from baseline to endline (Figure 3E). There were no changes in mid-upper arm circumference (Figure 3F) and is normally distributed. There were no changes in AGP and CRP as shown in the Kernel Density Plots in Figures 3G and 3H, respectively, and they are normally distributed around 0 and 1, 0 with no inflammation and 1 for inflammation.). Figure 3I shows the Kernel Density Plot for sedentary activity during break and is normally distributed. According to the distribution, there is a slight shift for treated endline, but is not statistically significantly different from treated baseline. For the labor intensity less than 3 METS (Figure 3J), the Kernel Density plot is normally distributed as well as the labor intensity for greater than 3 METS (Figure 3K). There was also no difference seen in the Kernel Density plot for total work output (Figure 3L).

One interesting difference that I did see was between ethnic groups, although there was no difference between treatment and control groups. There were statistically significant differences between sedentary and <3 METS activity for labor intensity, where the Adivashi spent 19.85 more minutes in sedentary activity and 10.85 more minutes in <3METS labor

intensity (Table 4A). However, they compensated for this by picking about 26 more kilograms of tea leaves per week in the morning (Table 4B).

5.2 Blinder-Oaxaca Decomposition

5.2.1 Total Output

Total output was regressed on simple reaction time, contrast threshold, temporal threshold, total energy expenditure, body mass index, mid-upper arm circumference, total body iron, white blood cell count, an AGP dummy, a CRP dummy, vitamin B12, folate levels, and changes in time spent in sedentary METS, METS<3, and METS>3 at baseline controlling for all baseline characteristics (Table 5A). When all production functions were included in the regression model, all variables became not significant at the 5% level. This shows that, although iron strongly affected, for example, the cognitive function variables, the cognitive function variables did not, in turn, affect worker output at a statistically significant level.

The Blinder-Oaxaca decomposition predicts how much of a change in productivity, measured in kilograms of tea leaves picked, is explained by changes in the intermediate production functions and how much is explained by the changes in the returns to those functions (the coefficients). The rest of the change in productivity is due to unexplained factors outside of these intermediate production functions.

The results show that there is a very small increase in tea leaves picked due to energy expenditure, but it is not statistically significant. Cognitive function increases by a large amount, showing that the increases due to cognitive function resulting in 45.124 more kilograms of tea leaves picked. However, the large standard deviations made this number also not significant. Physiological health decreased total output by about 6 kilograms. Changes in morbidity resulted in about 6 fewer kilograms of tea picked. Changes in labor intensity resulted in about 3 more

kilograms of tea leaves picked. These changes left a residual of 38.6867 kilograms of tea leaves between the sum of the total changes from the production function intermediates and the actual total change in output that is unexplained.

5.2.2 Morning Output

The morning data was less variable than the afternoon data because most of the women had more discretion in the afternoon. If they picked a lot in the morning, they may pick less in the afternoon. There were also many half days, in which the women left work for a variety of reasons ranging from sickness to bank visits to weddings. Therefore, the same regression was run for only the morning output and morning energy expenditure, to explore the possibly of a stronger relationship between the output variable and the intermediate production functions. Results are summarized in Table 5B.

Results showed that there was actually a negative effect on the amount of tea leaves picked in the morning. There was a very small decrease in morning output due to energy expenditure, a small decrease of 2 kilograms of tea leaves picked due to cognitive function, an increase in about 3 kilograms due to changes in physiological health, a decrease of 4 kilograms dues to changes in morbidity from infectious diseases and a very small increase in kilograms of tea leaves picked due to changes in labor intensity. These changes left a residual of 13.00 kilograms of tea leaves between the sum of the total changes from the production function intermediates and the actual total change in output that is unexplained.

5.3 Time Spent Plucking

I looked at the amount of time the women spent picking tea leaves in minutes to see if there was any change in the time allocated to picking tea leaves. Table 6 summarizes the results. There was no statistically significant decrease in number of minutes of tea picked in the morning

of afternoon. There was no overall change in the total number of minutes of tea leaves picked. There were also no statistically significant differences in the number of days of the week spent plucking tea leaves (Table 6), with about 5.3 days picked at baseline for both control and treatment groups and about 4.6 days picked at endline for both control and treatment groups.

5.4 Robustness Checks

Differences in characteristics were examined between women who dropped out from the study from baseline to endline with the women who stayed in the study, to determine if there was not selective attrition.

Differences between treatment and control groups at baseline and endline were examined for folate and B12. Because this was an iron supplementation trial, there should have been minimal to no change in folate and B12 levels. This was used as another robustness check to ensure that there were no other changes that differentially affected subjects. Results showed that there were no significant differences between treated and control groups for both folate and B12 for baseline and endline treated and control groups (Table 2B).

I also measured changes in energy expenditure for those with hemoglobin values below anemia levels of 11 g/L to see if there were larger changes for just those anemic in treatment versus control groups. However, there were no statistically significant changes in both total and morning energy expenditures, and both values were positive.

In addition, I tried logging both total and morning energy expenditure as well as total and morning total output of tea leaves picked to try to account for outliers. However, the difference-in-difference estimates between treatment and control groups from baseline to endline were still not statistically significant at the 10% level (Table 3E).

6. Conclusion and Discussion

I predicted there to be a decrease in energy expenditure, an increase in cognitive function, a decrease in morbidity from infectious diseases, a decrease in labor intensity during non plucking, and an increased physiological health status. There was a statistically significant increase in total body iron. This shows that the fortification trial was successful and that there was a strong improvement in iron status due to the randomized control trial. In fact, the change in total body iron was very large compared to the baseline values, increasing by about 1.9 times for the treated group from baseline to endline. The difference in difference was about a 57% increase over the average of the baseline treated and control groups.

All the measures of cognitive function also improved by significant amounts. In fact, a decrease in 42.774 milliseconds for temporal threshold is about a 68% decrease from the average of the baseline values. That is a very large number considering that the baseline values were 64.3 and 60.8 for treatment and control groups, respectively. The large changes in cognitive function are most likely due to the fact that for this study, specific tasks that were predicted to be affected by iron in the hippocampus were looked at rather than general cognitive measures.

For the Blinder-Oaxaca Decomposition, due to the large standard errors and the very small change in output, the coefficients are imprecise, and therefore, the results are only suggestive. However, in the decomposition, the results are very interesting because the changes of each of the intermediate factors are fairly large except for energy expenditure for the total output data. Therefore, I should expect to see large changes in kilograms of tea leaves picked from cognitive function, morbidity from infectious diseases, physiological health, and labor intensity. However, because total output changed so little, there must be another unobserved factor that is strongly negatively affecting all of the production function intermediates.

As for just morning data, the numbers seemed a lot more reasonable. There was about a 4.5 kilogram residual that was left unexplained. The production function intermediates were also a somewhat smaller and therefore seemed to be in a more reasonable range in explaining changes in total output of tea leaves picked.

The results as a whole are interesting because the intervention had a very strong effect on total body iron and cognitive function. The measures of cognitive function indicate that there should be a change in output of tea leaves picked. Therefore, this implies that there is some very large unobserved factor that is driving up productivity, resulting in a wash effect. One possibility is that there is a social factor that prevents the women from picking more tea leaves. The women may have some goal in mind when picking the tea leaves that even if there were increased cognitive function or decreased energy expenditure, they would still pick the same amount of tea leaves. There is a lot more scope for behavioral changes from a change in iron that are not necessarily explained by the intermediate production functions.

Previous literature shows that energy expenditure should decrease with iron treatment, but this experimental trial shows that there was no statistically significant change in energy expenditure, despite a clear change in total body iron status. There could have been a change for just those that were anemic, therefore, I used the traditional hemoglobin cutoff of 11 g/dL to see if there was any change in energy expenditure between treatment and control groups from baseline to endline, and there was no statistically significant change. There are three explanations for such a result. The results may have been affected by missing data, as explained below, or changes in iron status actually do not affect energy expenditure to the degree that was previously shown in the literature. The third explanation is that these women are in very good shape. They work hard from early in the morning to late in the afternoon, sometimes carrying up to 80 kg of

tea leaves and carrying it over long distances. At such high work capacity, it may be possible that iron status does not highly affect this specific population because energy expenditure was measured through heart rate, which fluctuates less the more fit a person is.

The very interesting fact is that, although there wasn't a change in energy expenditure or output of tea picked. There seems to be some substitution going on in the number of minutes that the women spent picking tea, especially in the morning. Therefore, the iron treatment may be showing up, rather, in a substitution away from amount of time spent working rather than energy expenditure or output of tea leaves picked.

All three measures of cognitive function, the simple reaction test, the contrast threshold test, and temporal threshold test, improved at the 5% level, as predicted. There were no statistically significant changes in the morbidity measures of AGP, CRP, and WBC count, but that was expected from the literature, where iron both increases overall health, and therefore ability to fight off diseases, but also can increase bacterial burden. There were no changes in labor intensity during all of the combined non-plucking periods.

The change in labor intensity must be examined with caution. It is difficult to tell the difference between what the women may see as labor versus leisure. Furthermore, if some of the women are extremely constrained, they will not be able to adjust their labor intensity tradeoff despite improved iron status. Many times, women had a lot of in home work to do after formal tea picking labor such as picking up firewood for cooking or walking great distances for shopping and carrying water. In these cases, iron may not affect the tradeoff decision, but that in itself can be an interesting result in terms of policy implications. An interesting result from labor intensity was the fact that there were highly statistically significant differences between labor intensity between ethnic groups. This should be explored in more detail and may have policy

implications in terms of how different ethnic groups may have different behavioral responses to iron treatment and therefore may be considered differently for different sets of policies.

6.1 Limitations of the Study

The data logs for the exact time at which the women started work, ended work, and started and ended lunch break were, at times, estimations. This is because most of the women were asked at the end of the work day to recall when they finished picking and when they took breaks. This could have potentially biased the range over which heart rates were taken into account for energy expenditure calculations. Furthermore, all heart rate monitors started recording data at 7:30AM, but some women started picking before that time. I estimated the heart rates if a woman did start picking tea before 7:30AM. For example, if a woman started working at 7:00AM, I used the average heart rates for the period from 7:30AM to 8:00 AM to estimate energy expenditure from 7:00 AM to 7:30 AM. However, since heart rates were measured minute-by-minute and then collapsed to one average for morning and afternoon pickings, a few extra or fewer minutes should not have had a large impact on the total estimated energy expenditure values.

I had to drop a few subjects for the changes in labor intensity due to non-retrievable corrupted accelerometry data at endline. The data appear to be corrupted randomly, and therefore should not affect any type of selection bias for the experiment, but it did affect the sample size of the comparisons between baseline and endline samples. There was also missing heart rate data for certain subjects due to either not turning on the monitors at the right time or the data were not transmitted properly. However, there is no evidence that this occurred in a systematic manner, so it should not have biased the results. Furthermore, both heart rate monitors and accelerometers

may show differences in controlled laboratory settings, but in the field there is a lot more noise in terms of activities that are performed and data that is collected.

This is the first time that the specific model that Przybyszewski (2011) created was implemented in order to predict energy expenditure from heart rate. Although the model has high predictability and correlation, it has not yet been rigorously tested. The ideal approach would be to develop a heart rate by energy expenditure equation for each woman. However, this is not possible with the resources available for the large number of subjects in this experiment. This model tests for the ability of the group equation to predict energy expenditure in a subset of women for whom a heart rate by energy expenditure curve was developed (Przybyszewski 2011). Still, the literature shows that there is a strong correlation between heart rate and energy expenditure, so this should not pose a problem except whether or not such a model would be applicable to this the specific subset of women tea leaf pickers.

There is also the question as to whether the women were using the double fortified salts for food and that food was consumed and iron absorbed, or in other words, a measure of compliance. As mentioned above, the precaution of deworming the women was taken at the beginning of the experiment and at 6 months in order to minimize blood loss, resulting in loss of iron. Due to the large change in iron status of the treatment group, however, there is strong evidence that the women did consume and absorb the iron. The slight increase in iron status for the control probably reflects the effects of deworming the entire sample of subjects.

There was some amount of attrition from baseline to endline, and the information obtained varied slightly from test to test. Overall, there were 248 subjects at baseline and 217 subjects at endline. There is always a chance that attrition did not happen randomly, for example, if more women who became sick did not complete the experiment, biasing the effects on

productivity in a positive direction. However, most of the women that did not complete the experiment dropped due to random events such as moving to a new location, not wanting to participate any longer, illness, pregnancy, etc.

The experiment was conducted on a very specific population in West Bengal, India. There may have been specific characteristics about this population of women and environment that may not be generalized to other populations, especially for the behavioral outcomes measured in this study. For example, the women in this population were extremely active, having to carry, at times, many 35 kilograms of tea across long distances every day. This may affect their level of physical fitness, which in turn could have affected their mechanical and metabolic efficiency while performing rigorous work while plucking tea.

Future research should look into why energy expenditure or labor intensity might not change in such a population, and what are the reasons behind which iron may not affect these changes. Research should also look into what the unobserved factors are that are keeping productivity low despite large changes in production intermediate functions. Future research should also attempt to measure the costs and benefits of a double fortified iron intervention for the region of West Bengal and India as a whole.

Figures and Tables

Table 1A. Tests for attrition between participants that completed the study versus those that did not complete the study at baseline for the intermediate variables with the means and standard deviations (in parenthesis). There are no differences between the two groups except for body iron which is statistically significant at the 10% level, suggesting that there are no differences between those that did not complete the test and those that did.

Variable	Study Completed N=218	Study Not Completed N=30	Difference
Body Iron	3.091 (5.111)	4.0446 (4.268)	-1.859* (0.986)
Contrast Threshold	21.756 (21.401)	18.288 (20.151)	19.984 (22.083)
Temporal Threshold	62.346 (16.493)	82.00 (16.493)	-19.654 (16.555)
Total Energy Expenditure	1.627 (0.585)	1.714 (0.166)	-0.087 (0.415)
Morning Energy Expenditure	1.645 (0.642)	1.809 (0.218)	-0.164 (0.456)
Output Total	207.091 (92.482)	207.286 (34.4569)	-0.194 (35.032)
MUAC	23.956 (2.807)	24.57 (0.5028)	-0.614 (0.536)
CRP	1.471 (2.527)	1.467 (0.4766)	0.004 (0.508)
Average Sed Break	41.772 (22.387)	49.3143 (8.4116)	-7.543 (8.579)
Average Break <3METS	31.601 (14.509)	35.8 (5.3953)	-4.199 (5.503)
Average Break >3METS	21.172 (13.187)	15.9429 (4.9514)	5.229 (5.050)
Statistical significance: ***0.01, **0.05, *0.10			

Table 1B. Tests for attrition between participants that completed the study versus those that did not complete the study at baseline for the control variables with the means and standard deviations (in parenthesis). There are no differences between the two groups except for body iron which is statistically significant at the 10% level, suggesting that there are no differences between those that did not complete the test and those that did.

Variable	Study Completed N=218	Study Not Completed N=30	Difference
Income	0.9447 (0.229)	1.00 (0.00)	-0.055 (0.043)
Literacy	0.3194 (0.4673)	0.4828 (0.5085)	-0.163* (0.093)
Electricity	0.9861 (0.1173)	1.00 (0.00)	-0.014 (0.022)
Spouse Literacy	0.7118 (0.4543)	0.7391 (0.4490)	-0.027 (0.101)
Education	0.5115 (0.6244)	0.7241 (0.5914)	-0.213 (0.123)
Spouse Education	0.8529 (0.7188)	0.9130 (0.6683)	-0.060 (0.158)
Spouse Work Type	0.8024 (0.3994)	0.8966 (0.3444)	-0.136 (0.108)
Family	0.4047 (0.4920)	0.5517 (0.5061)	-0.147 (0.098)
Diet	0.8894 (0.3144)	0.8966 (0.3099)	-0.007 (0.062)
Housing	0.7558 (0.4306)	0.8276 (0.3844)	-0.072 (0.084)
Water	0.1244 (0.3308)	0.2069 (0.4123)	-0.082 (0.067)
Toilet	0.6313 (0.4836)	0.7241 (0.4549)	-0.093 (0.095)
Appliances	0.9355 (0.2462)	0.9310 (0.2579	0.004 (0.049)
Statistical significance: ***0.01, **0.05, *0.10			

Table 2A. Demographic breakdowns between treatment (iron and iodine) and control (iodine) groups at baseline to test the success with randomization with means and standard deviations (in parenthesis). None of the variables are statistically significantly different from one another at baseline.

	Treatment N=106	Control N=109	Difference
Household Size	5.339 (1.689)	5.162 (1.781)	0.177 (0.236)
Income	0.934 (0.250)	0.955 (0.208)	-0.021 (0.031)
Number of Children	2.119 (0.856)	2.034 (1.005)	0.085 (0.143)
Literacy	0.311 (0.465)	0.345 (0.478)	-0.034 (0.064)
Electricity	0.990 (0.098)	0.982 (0.134)	0.008 (0.016)
Spouse Literacy	0.762 (0.428)	0.678 (0.470)	0.084 (0.069)
Education	0.481 (0.651)	0.550 (0.600)	-0.068 (0.085)
Spouse Education	0.905 (0.670)	0.816 (0.755)	0.089 (0.109)
Spouse Work Type	0.607 (0.491)	0.605 (0.492)	0.002 (0.075)
Experience	0.491 (0.502)	0.431 (0.498)	0.059 (0.068)
Family	0.394 (0.491)	0.414 (0.495)	-0.020 (0.067)
Diet	0.887 (0.318)	0.892 (0.312)	-0.005 (0.043)
Housing	0.745 (0.438)	0.766 (0.425)	-0.020 (0.059)
Water	0.104 (0.306)	0.144 (0.353)	-0.040 (0.045)
Toilet	0.642 (0.482)	0.631 (0.485)	0.011 (0.066)
Appliances	0.953 (0.213)	0.919 (0.274)	0.034 (0.033)
Statistical significance: ***0.01, **0.05, *0.10			

Table 2B. Baseline tests as robustness checks to test differences between treatment (iron and iodine) and control (iodine) groups for demographic variables with their means and standard deviations (in parenthesis) None of the variables are statistically significantly different from one another at baseline, indicating randomization was successful except for height and net weight, which are statistically significant only at the 10% level.

Variable	Treatment N=106	Control N=109	Difference
Age	38.984 (8.101)	39.179 (8.014)	-0.109 (1.060)
Height	150.791 (5.211)	149.868 (5.487)	1.342* (0.716)
Net Weight	45.439 (8.127)	43.802* (7.170)	1.819 (1.052)
Mid Upper Arm Circumference	24.28 (2.905)	23.776 (2.582)	0.525 (0.378)
White Blood Cell Count	6.415 (1.656)	6.541 (1.578)	-0.070 (0.221)
Hemoglobin	11.610 (1.269)	11.715 (1.224)	-0.090 (0.171)
CRP	1.333 (2.407)	1.610 (2.795)	-0.277 (0.331)
AGP	80.173 (34.150)	77.234 (32.220)	2.939 (4.537)
Folate	3.708 (1.968)	3.694 (1.757)	0.085 (0.208)
Vitamin B12	280.217 (148.940)	257.275 (127.300)	20.911 (18.588)
Body Iron	3.213 (5.049)	3.419 (5.152)	-0.554 (0.621)
SRT Median	502.280 (318.625)	461.231 (31.2413)	41.049 (44.520)
Contrast Threshold	21.252 (20.963)	22.3855 (2.6952)	-1.134 (3.856)
Temporal Threshold	64.277 (20.547)	60.8116 (1.9851)	3.465 (2.850)
Output Total	201.434 (86.394)	212.6055 (8.7154)	-11.172 (12.412)
Average Sedentary Break	41.901 (18.541)	42.232 (2.364)	-0.331 (3.316)
Average Break < 3 METS	31.580 (14.103)	31.9528 (1.5154)	-0.372 (2.126)
Average Break > 3 METS	20.975 (12.666)	20.9640 (1.3928)	0.011 (1.954)
Statistical significance: ***0.01, **0.05, *0.10			

2C. Means and standard deviations (in parenthesis) of baseline characteristics with percentage of the population that was deficient at baseline (n-248). There were no differences between DFS and control groups at baseline.

Characteristic	**Mean (Standard Deviation)**	**% Deficient**
Age	39.081 (8.042)	
BMI (kg/m^2)	18.715 (3.052)	37.9
Hemoglobin (g/L)	116.62 (12.45)	52.4
Serum ferritin (mg/L)	32.621 (33.784)	44.8
Transferrin receptor (µg/mL)	6.387 (4.533)	75.0
Deficiency = BMI < 18.5 kg/m^2; Hb < 12.0 g/L; sFt < 20 mg/L; sTfR < 8.0µg/mL		

Table 3A. Summary of means and standard deviations (in parenthesis) for baseline and endline broken down by treatment (iron and iodine) and control (iodine) groups for body iron and output variables.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Body Iron	2.19468 (4.5987)	2.7491 (4.553)	4.2886 (3.6743)	3.4665 (3.5965)
Difference	-0.55442		0.8221	
Diff in Diff	1.4097*** (0.3769)			
Mean Total Output	201.434 (86.3942)	212.6055 (95.2484)	129.4717 (62.2964)	141.4862 (70.3351)
Difference	11.1715		-12.0145	
Diff in Diff	-0.8430 (14.1511)			
Mean Morning Output	106.5421 (33.4002)	111.4865 (41.69388)	62.16981 (43.3663)	70.3578 (47.4327)
Difference	-4.9444		-8.188	
Diff in Diff	-3.7854 (7.8587)			
Log Total Output	5.2212 (0.4115)	5.2617 (0.4478)	4.7484 (0.4926)	4.8456 (0.4893)
Difference	-0.0405		-0.0972	
Diff in Diff	-0.0592 (0.0813)			
Log Morning Total Output	4.6158 (0.3372)	4.6472 (0.3714)	4.1999 (0.4942)	4.2832 (0.5375)
Difference	-0.0314		-0.0833	
Diff in Diff	-0.0878 (0.0853)			
Statistical significance: ***0.01, **0.05, *0.10				

Table 3B. Summary of means and standard deviations for baseline and endline broken down by treatment (iron and iodine) and control (iodine) groups for cognitive function and infectious disease variables.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean SRT Median	502.2803 (318.6253)	461.231 (174.2385)	353.6236 (90.89868)	434.0741 (179.9507)
Difference	41.0493		-80.4505	
Diff in Diff	-121.1213** (49.4243)			
Mean Contrast Threshold	21.25185 (20.96333)	22.3855 (23.05996)	14.44031 (13.7758)	25.35865 (23.7921)
Difference	-1.13365		-10.91834	
Diff in Diff	-9.7847* (4.8064)			
Mean Temporal Threshold	64.2769 (20.5468)	60.8116 (11.4214)	17.9231 (15.1578)	57.2319 (11.6469)
Difference	3.4653		-39.3088	
Diff in Diff	-42.7741*** (3.2766)			
Mean CRP Dummy	0.06542 (0.2484)	0.1261 (0.3335)	0.1214 (0.3283)	0.12613 (0.3335)
Difference	-0.06068		-0.00473	
Diff in Diff	0.05607 (0.4320)			
Mean AGP	0.10280 (0.3051)	0.0631 (0.2442)	0.06542 (0.2484)	0.0631 (0.24417)
Difference	0.0397		0.00232	
Diff in Diff	-0.03738 (0.5213)			
Mean WBC	6.4243 (1.682)	6.4946 (1.5760)	6.1217 (1.7072)	6.1560 (1.6041)
Difference	-0.0703		-0.0343	
Diff in Diff	0.0981 (0.1667)			
Mean MUAC	24.2365 (2.9627)	23.7117 (2.6172)	25.5543 (3.27065)	25.39074 (3.0904)
Difference	0.5248		0.16356	
Diff in Diff	-0.3535 (0.4006)			
Statistical significance: ***0.01, **0.05, *0.10				

Table 3C. Summary of means and standard deviations for baseline and endline broken down by treatment (DFS) and control (iodine) groups for labor intensity variable during lunch break.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Sedentary Break	41.9006 (18.5408)	42.23165 (25.6178)	39.5333 (19.3715)	34.13623 (14.3208)
Difference	-0.33105		5.39707	
Diff in Diff	6.1413 (4.4997)			
Mean <3 METS Break	31.5804 (14.1030)	31.9528 (14.49417)	33.39722 (14.3736)	30.50145 (11.42833)
Difference	-0.3724		2.8958	
Diff in Diff	1.9337 (2.8366)			
Mean >3 METS Break	20.9746 (12.66562)	20.96404 (13.6119)	21.975 (13.2201)	21.3768 (13.5835)
Difference	0.01056		0.5982	
Diff in Diff	-0.588 (3.340)			
Statistical significance: ***0.001, **0.01, *0.05				

Table 3D. Summary of means and standard deviations for baseline and endline broken down by treatment (DFS) and control (iodine) groups for labor intensity variable before and after work.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Sedentary Before Work	38.859 (25.680)	37.339 (18.493)	41.858 (54.929)	36.471 (20.662)
Difference	1.52		5.387	
Diff in Diff	4.389 (8.690)			
Mean <3 METS Before Work	62.938 (21.977)	64.733 (20.312)	68.494 (43.578)	61.19 (27.639)
Difference	-1.795		7.304	
Diff in Diff	8.359 (6.440)			
Mean >3 METS Before Work	49.338 (19.287)	51.027 (19.247)	50.488 (25.794)	50.286 (24.585)
Difference	-1.689		0.202	
Diff in Diff	1.490 (4.730)			
Mean Sedentary After Work	119.714 (36.809)	110.097 (36.841)	127.031 (108.347)	105.365 (35.873)
Difference	9.617		21.666	
Diff in Diff	9.904 (16.312)			
Mean <3 METS After Work	103.883 (25.950)	102.121 (30.387)	102.786 (39.924)	102.846 (31.266)
Difference	1.762		-0.06	
Diff in Diff	-4.647 (6.954)			
Mean >3 METS After Work	68.723 (23.330)	70.774 (22.763)	67.469 (24.040)	75.870 (24.936)
Difference	-2.051		-8.401	
Diff in Diff	-5.791 (5.142)			
Statistical significance: ***0.001, **0.01, *0.05				

Table 3E. Summary of means and standard deviations for baseline and endline broken down by treatment (DFS) and control (iodine) groups for labor intensity variable during work.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Sedentary During Work	91.967 (35.216)	101.802 (74.148)	116.655 (117.714)	91.346 (42.730)
Difference	-9.835		25.309	
Diff in Diff	49.092* (19.010)			
Mean <3 METS During Work	223.485 (42.598)	239.302 (46.160)	208.058 (70.092)	212.404 (52.425)
Difference	-15.817		-4.346	
Diff in Diff	11.471 (12.989)			
Mean >3 METS During Work	54.331 (15.917)	58.50 (26.263)	55.836 (32.374)	50.929 (19.593)
Difference	-4.169		4.907	
Diff in Diff	9.383 (5.937)			
Mean Total During Work	381.194 (27.798)	408.264 (104.780)	379.373 (189.600)	354.679 (56.825)
Difference	-27.07		24.694	
Diff in Diff	64.609* (30.664)			
Statistical significance: ***0.001, **0.01, *0.05				

Table 3F. Summary of means and standard deviations for baseline and endline broken down by treatment (DFS) and control (iodine only) groups for energy expenditure variables. Energy expenditure is measured in kilocalories.

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Energy Expend Total	1.61094 (0.536924)	1.64395 (0.62549)	1.6755 (0.61024)	1.6276 (0.646814)
Difference	-0.03301		0.0479	
Diff in Diff	0.0689 (0.1038)			
Mean Energy Expend Morning	1.64231 (0.60091)	1.64997 (0.67891)	1.6613 (0.59526)	1.6357 (0.68522)
Difference	-0.00766		0.0256	
Diff in Diff	-0.0347 (0.1105)			
Mean EE Total Anemia	1.523 (0.470)	1.380 (0.558)	1.581 (0.483)	1.383 (0.505)
Difference	0.143		0.198	
Diff in Diff	0.101 (0.205)			
Mean EE Morning Anemia	1.513 (0.526)	1.328 (0.569)	1.555 (0.502)	1.338 (0.548)
Difference	0.185		0.217	
Diff in Diff	0.033 (0.172)			
Log EE Total	0.4243 (0.3264)	0.4182 (0.4153)	0.4495 (0.3756)	0.4091 (0.4042)
Difference	0.0061		0.0404	
Diff in Diff	0.0450 (0.0613)			
Log EE Morning Total	0.4349 (0.3488)	0.4076 (0.4520)	0.4398 (0.3832)	0.4040 (0.4306)
Difference	0.0273		0.0358	
Diff in Diff	0.0203 (0.0640)			
Statistical significance: ***0.01, **0.05, *0.10				

Table 4A. Differences in labor intensity between treatment and control groups and ethnic groups, where 0 indicated Nepali and 1 indicated Adivashi including a treatment*ethnicity interaction variable.

Variable	Mean (Standard Error)
Average Sedentary Difference	
Treatment	13.142** (6.642)
Ethnicity	19.853*** (6.494)
Treatment*Ethnicity	-8.777 (8.783)
Constant	-21.074 (5.161)
Average <3METS Difference	
Treatment	6.690 (4.268)
Ethnicity	10.852*** (4.173)
Treatment*Ethnicity	-6.658 (5.644)
Constant	-6.790 (3.317)
Average >3METS Difference	
Treatment	-4.244 (2.937)
Ethnicity	2.978 (2.817)
Treatment*Ethnicity	4.800 (3.852)
Constant	-11.067 (2.275)
Statistical significance: ***0.01, **0.05, *0.10	

Table 4B. Differences in output tea leaves picked between treatment and control groups and ethnic groups, where 0 indicated Nepali and 1 indicated Adivashi including a treatment*ethnicity interaction variable.

Variable	Mean (Standard Error)
Output Morning Difference	
Treatment	6.439 (11.841)
Ethnicity	26.408** (11.130)
Treatment*Ethnicity	-16.136 (15.724)
Constant	-56.500 (8.595)
Output Afternoon Difference	
Treatment	2.462 (15.145)
Ethnicity	-0.514 (14.236)
Treatment*Ethnicity	0.598 (20.112)
Constant	-29.932 (10.994)
Total Output Difference	
Treatment	9.219 (21.549)
Ethnicity	26.212 (20.255)
Treatment*Ethnicity	-15.856 (28.615)
Constant	-86.750 (15.641)
Statistical significance: ***0.01, **0.05, *0.10	

Table 5A. Baseline regression for Blinder-Oaxaca Decomposition with total output controlling for Vitamin B12, folate, ethnicity, experience, treatment, study completed, marital, literacy, spouse literacy, education, spouse education, spouse work, spouse worktype, income, household size, number of children, family type, diet, type of housing, water source, electricity source, appliances, fuel, age, height, and net weight. There was no statistically significant change in output due to changes in the intermediate production functions.

Variable	**Coefficient (Standard Error)**	**Difference in Difference**	**Change in Total Output**
Output		0.08101	
Energy Expenditure			
Energy Expenditure	3.0146	0.0689	0.2077
Total Change			0.2077
Cognitive Function			
SRT Median	0.04846	-121.1213	-5.8623
Contrast Threshold	-0.60761	-9.7847	5.9453
Temporal Threshold	-1.054	-42.7741	45.041
Total Change			45.124
Physiological Health			
Mid Upper Arm Circumference	18.2212	-0.3535	-3.3973
Total Change			-3.3973
Morbidity from Infectious Disease			
AGP	35.3634	-0.03738	-1.3219
CRP	-72.1783	0.05607	-7.10506
WBC	14.3341	0.0981	1.406
Total Change			-6.0797
Labor Intensity			
Break <3METS	0.7694	1.9337	1.488
Break >3METS	-2.6794	-2.2976	6.156
Sed Break	-0.77036	6.1413	-4.731
Total Change			2.913
Total Change From Inputs			38.7677
Residual			-38.6867

Table 5B. Baseline regression for Blinder-Oaxaca Decomposition for morning output controlling for Vitamin B12, folate, ethnicity, experience, treatment, study completed, marital, literacy, spouse literacy, education, spouse education, spouse work, spouse worktype, income, household size, number of children, family type, diet, type of housing, water source, electricity source, appliances, fuel, age, height, and net weight. There was no statistically significant change in output due to changes in the intermediate production functions.

Variable	**Coefficient (Standard Error)**	**Difference in Difference**	**Change in Total Output**
Output		0.2326	
Energy Expenditure			
Energy Expenditure	10.7742	-0.0347	-0.374
Total Change			-0.374
Cognitive Function			
SRT Median	0.03898	-121.1213	-9.0524
Contrast Threshold	-0.05526	-9.7847	0.5407
Temporal Threshold	-0.1415	-42.7741	6.053
Total Change			-2.4587
Physiological Health			
Mid Upper Arm Circumference	9.6105	-0.3535	-6.441
Total Change			-6.441
Morbidity from Infectious Disease			
AGP	35.3634	-0.03738	-1.3219
CRP	-72.1783	0.05607	-4.0470
WBC	9.4179	0.0981	0.9239
Total Change			-4.445
Labor Intensity			
Break <3METS	-0.3028	1.9337	-0.5856
Break >3METS	-0.8215	-2.2976	1.887
Sed Break	-0.0570	6.1413	-0.3502
Total Change			0.9512
Total Change From Inputs			-12.7675
Residual			13.0001

Table 6. Summary of means and standard deviations for baseline and endline broken down by number of minutes worked and the number of days worked

Variable	Baseline		Endline	
	Treatment	Control	Treatment	Control
Mean Minutes Work Morning	225.018 (18.146)	228.237 (20.500)	218.653 (15.581)	219.873 (15.885)
Difference	-3.219		-1.22	
Diff in Diff	0.624 (4.253)			
Mean Minutes Work Afternoon	147.798 (36.076)	156.619 (30.263)	155.302 (23.586)	159.935 (14.979)
Difference	-8.821		-4.633	
Diff in Diff	3.061 (6.492)			
Mean Minutes Work Total	371.74 (41.100)	383.614 (38.012)	373.798 (30.933)	380.255 (25.303)
Difference	-11.874		-6.457	
Diff in Diff	3.145 (8.754)			
No. Days worked during week	5.372 (0.684)	5.361 (0.696)	4.557 (1.398)	4.689 (1.315)
Difference	0.011		-0.132	
Diff in Diff	-0.119 (0.214)			
Statistical significance: ****0.001, ***0.01, **0.02, *0.05				

Figure 1A. Relationship between percentage of times that a stimulus is perceived and the corresponding stimulus intensity. The threshold is defined as the intensity at which the stimulus is detected 50% of the time (Ehrenstein & Ehrenstein 1999)

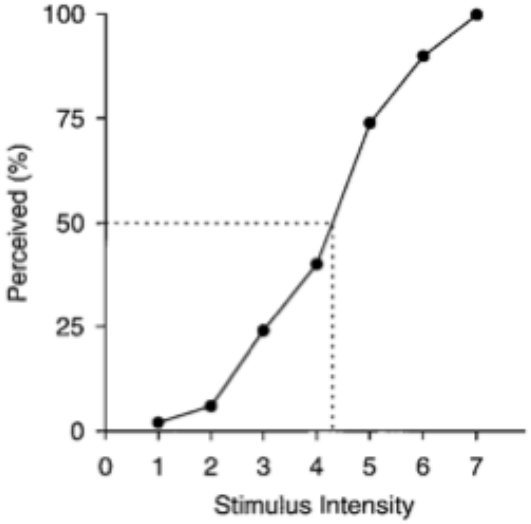


Figure 1B. Example stimulus of the contrast threshold task, the Gabor Patch at 15% contrast

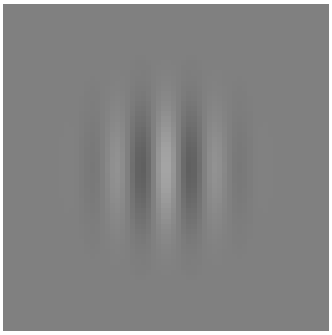


Figure 1C. Example stimulus of the temporal threshold task, the Gabor patch at 30% contrast.

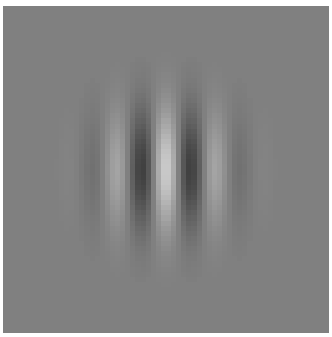
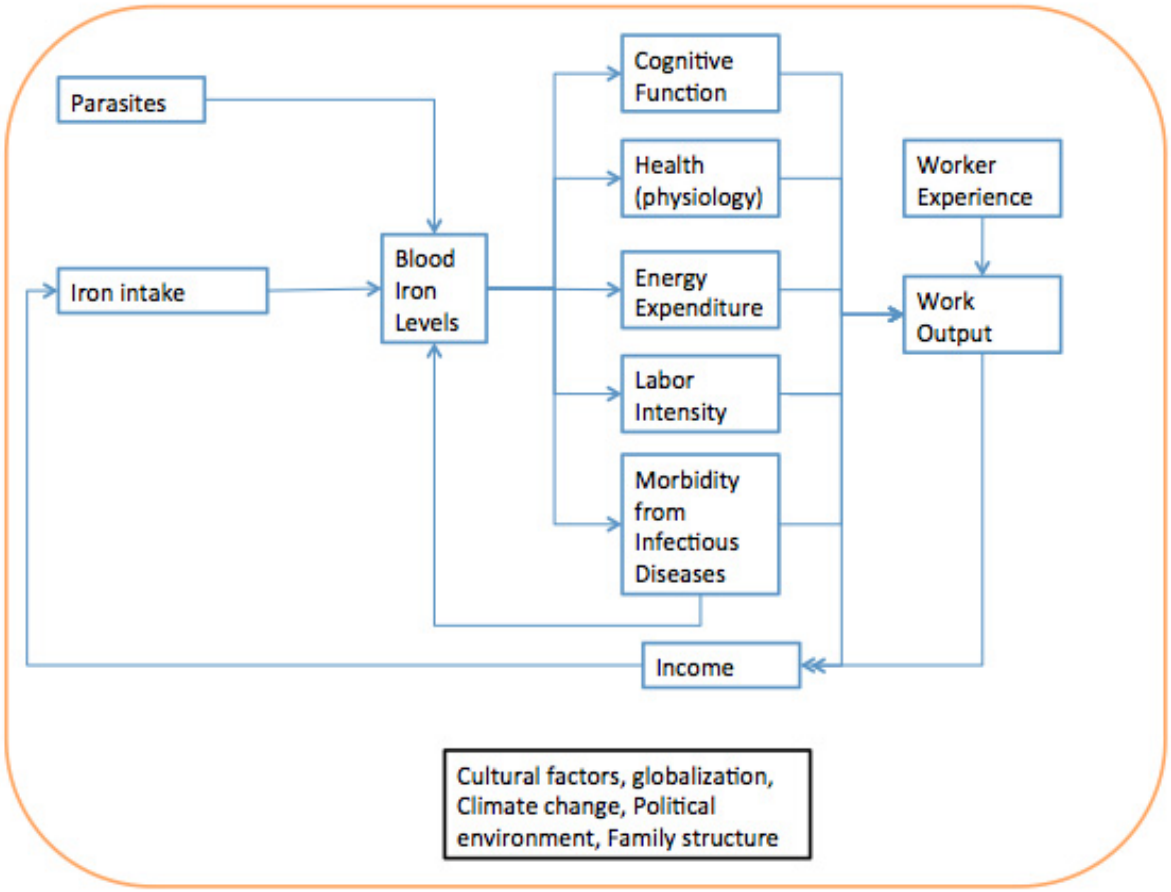


Figure 2. Pathway from iron supplementation to work output

Iron supplementation should directly affect blood iron levels. I (Blood iron levels), which in turn, affects the production function intermediates, E (energy expenditure), C (cognitive function), H (physiological health), D (morbidity from infectious disease burden), and L (labor intensity). The production function intermediates, in turn, affects worker output, measured in T (tea leaves picked). Iron (I) affects the intermediates and the rates of return of those intermediates.



T = F (E[I], C[I], H[I], D[I], L[I])

Figure 3A. Total Body Iron K-Density Plot

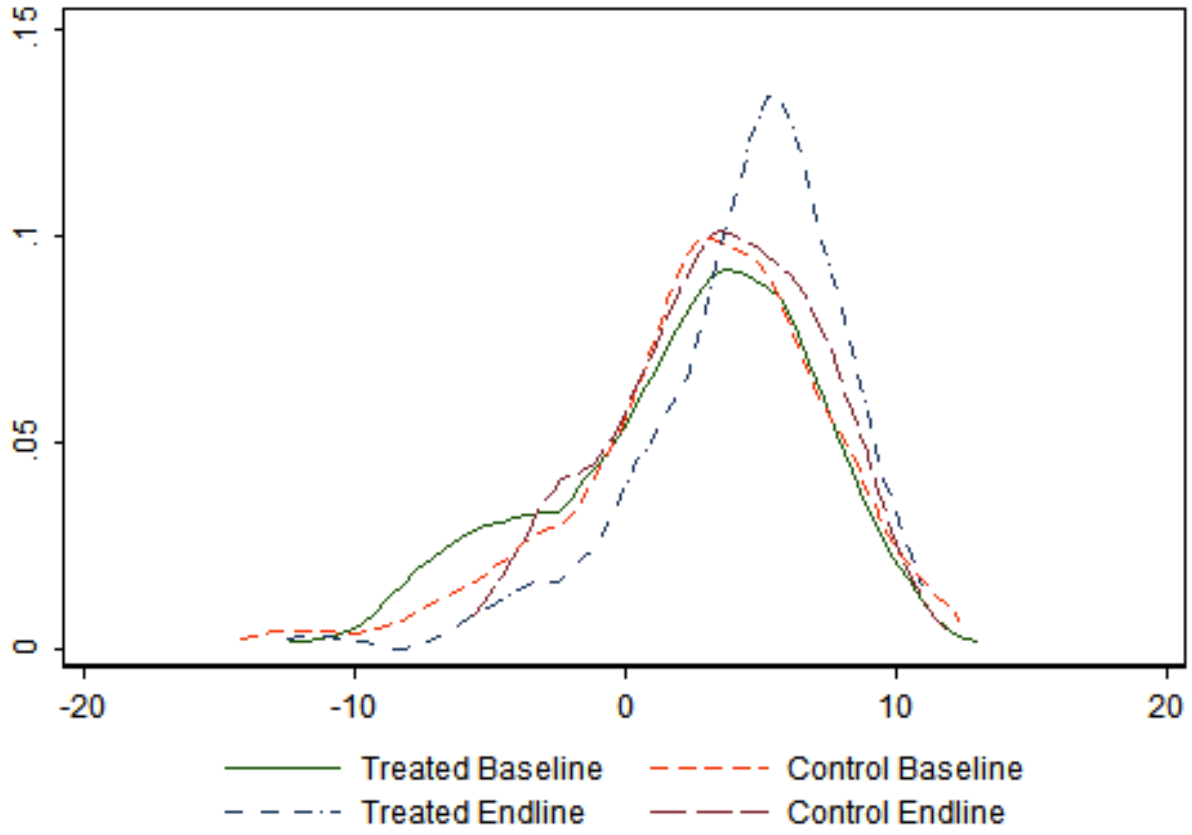


Figure 3B. Temporal Threshold K-Density Plot

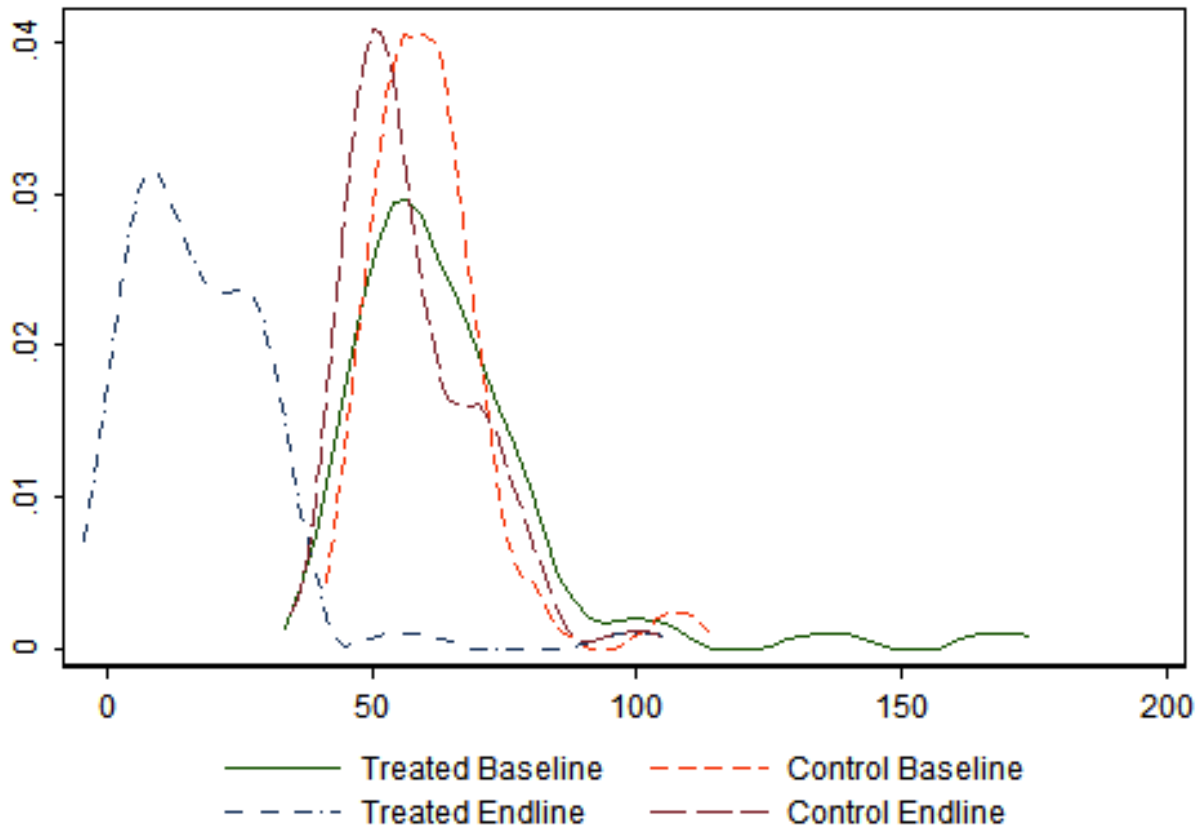


Figure 3C. Contrast Threshold K-Density Plot

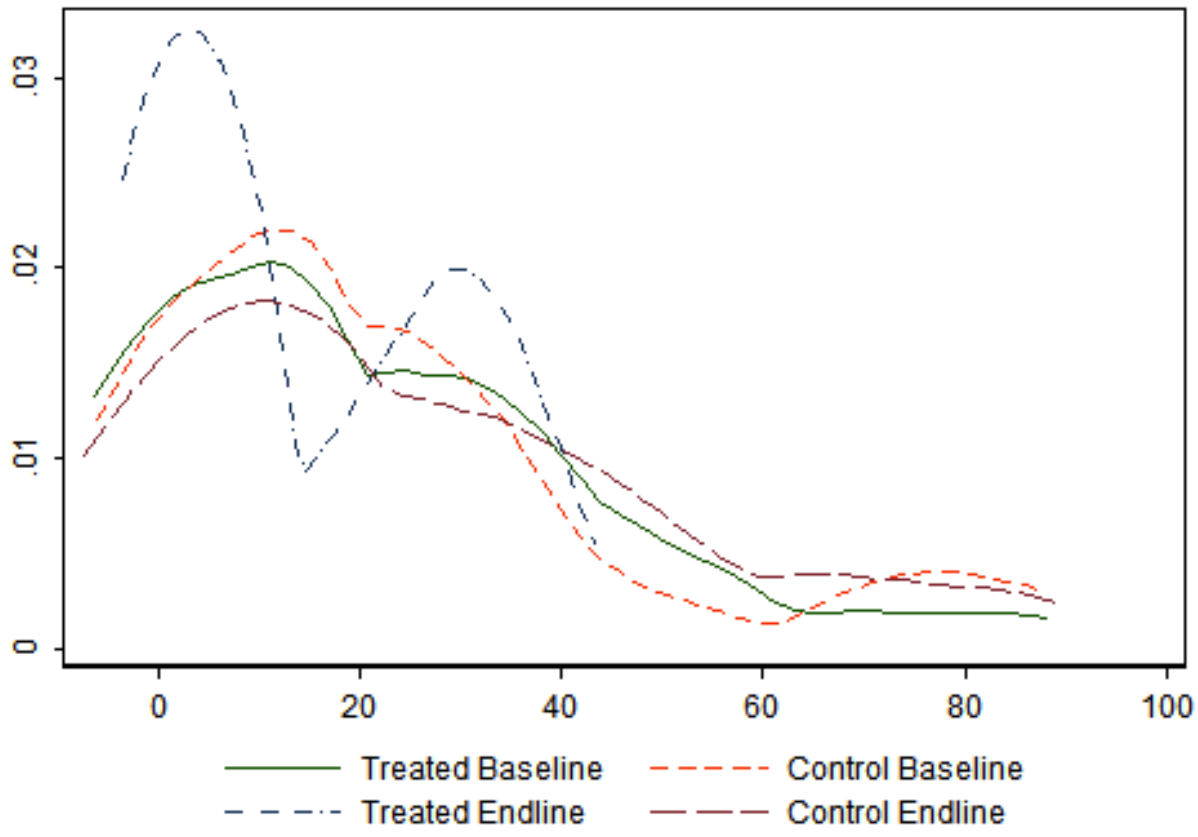


Figure 3D. Simple Reaction Time Median K-Density Plot

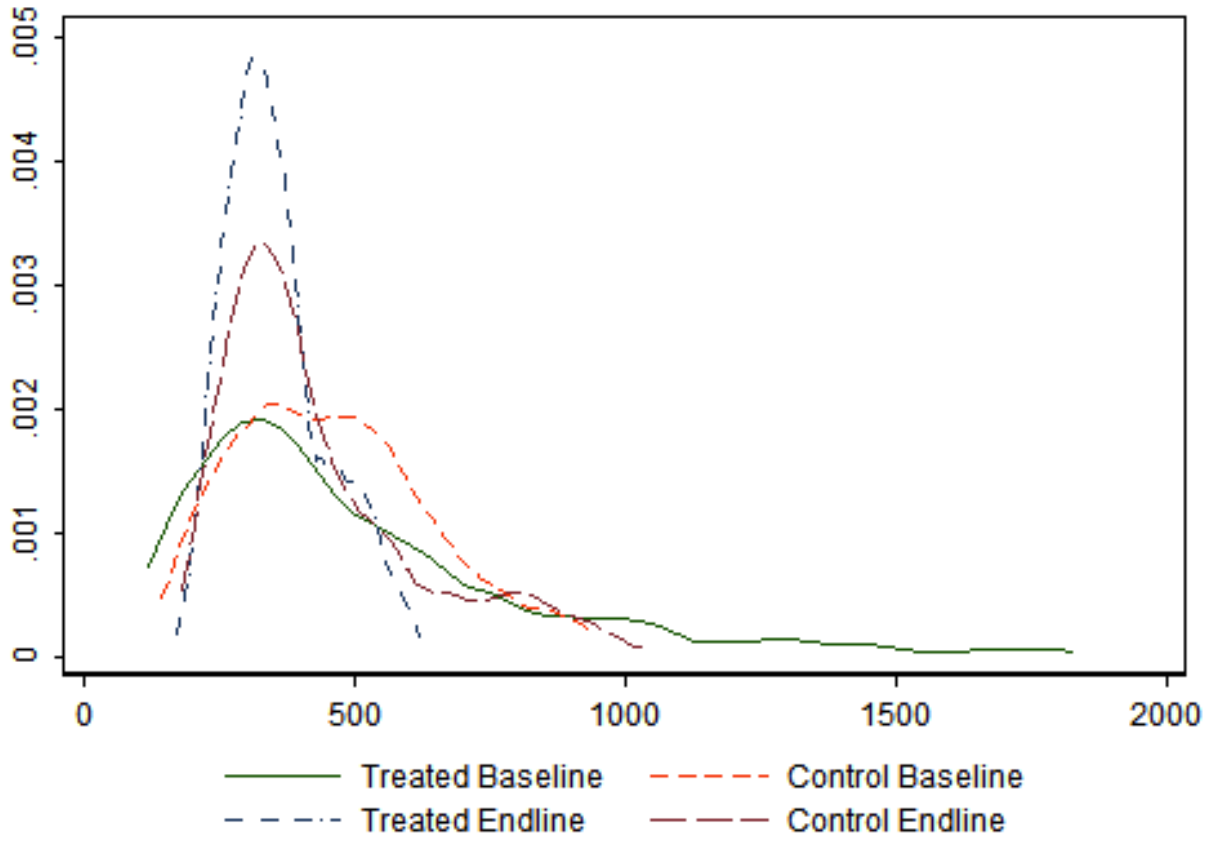


Figure 3E. Energy Expenditure Total K-Density Plot

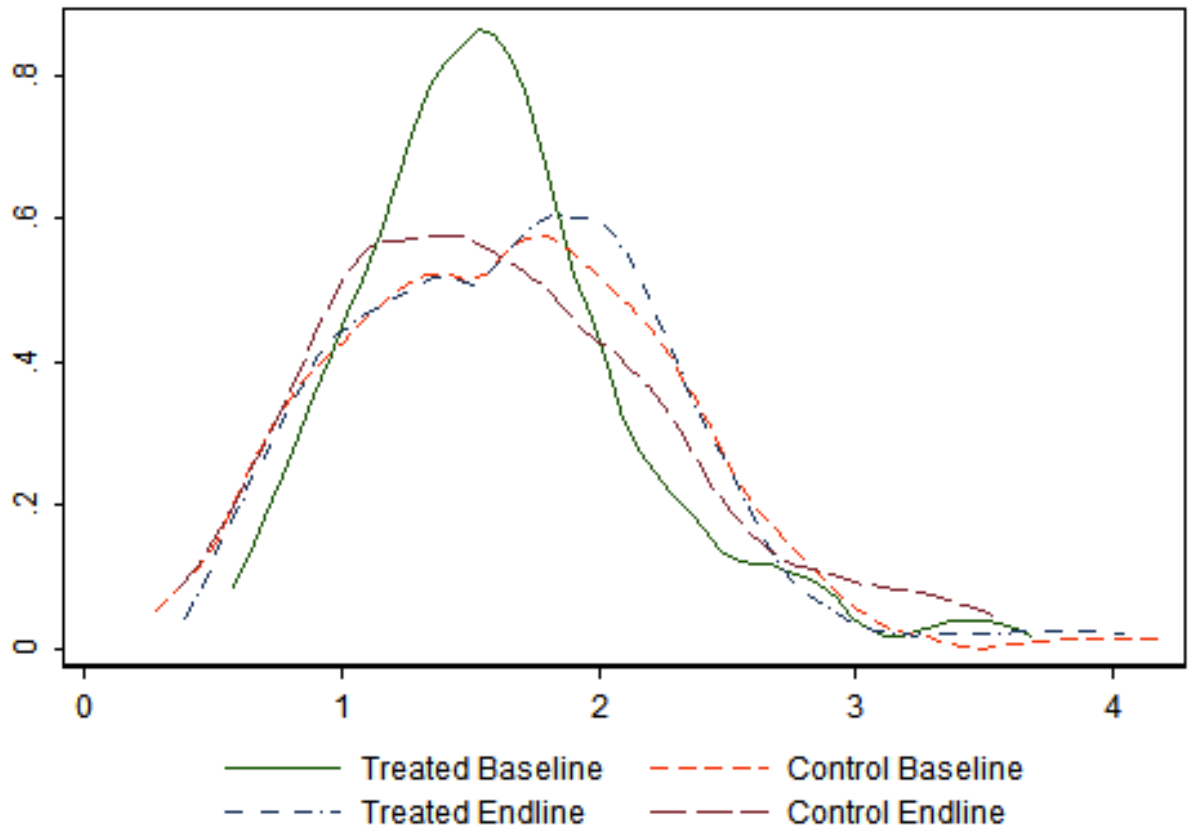


Figure 3F. MUAC K-Density Plot

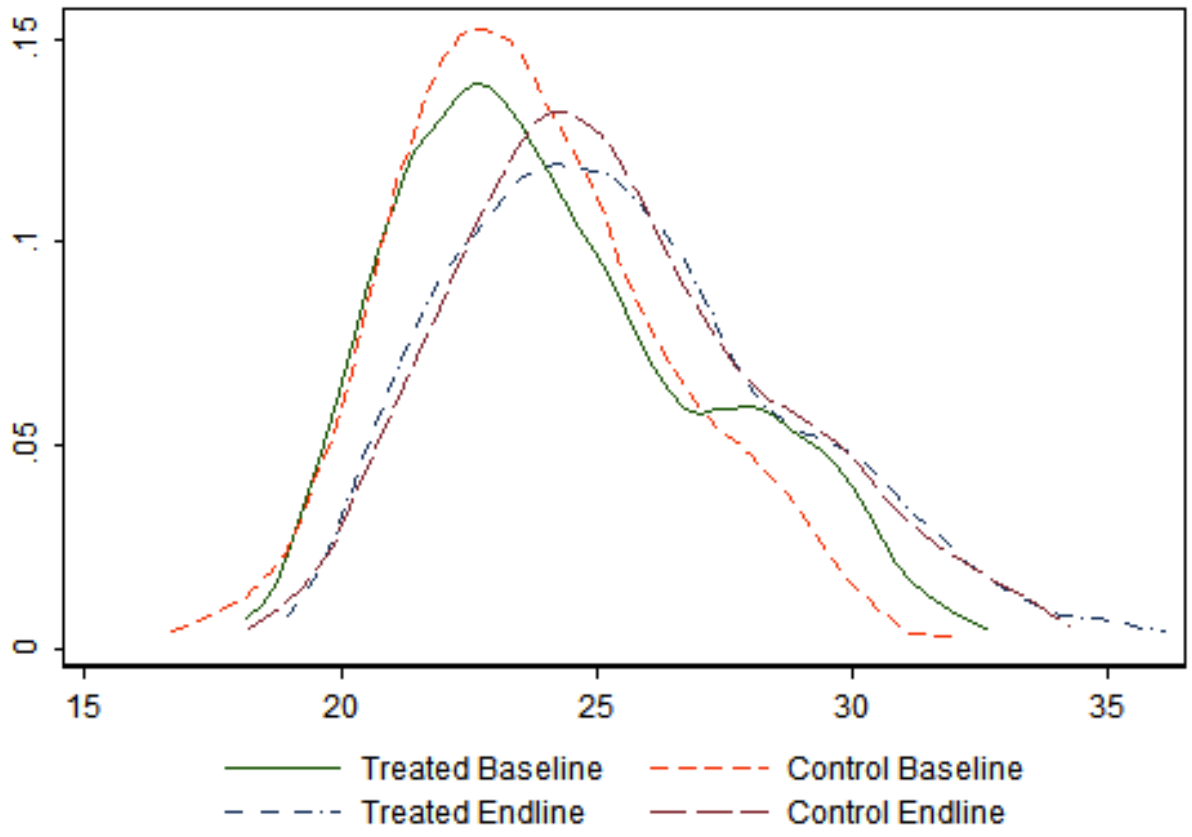


Figure 3G. AGP K-Density Plot

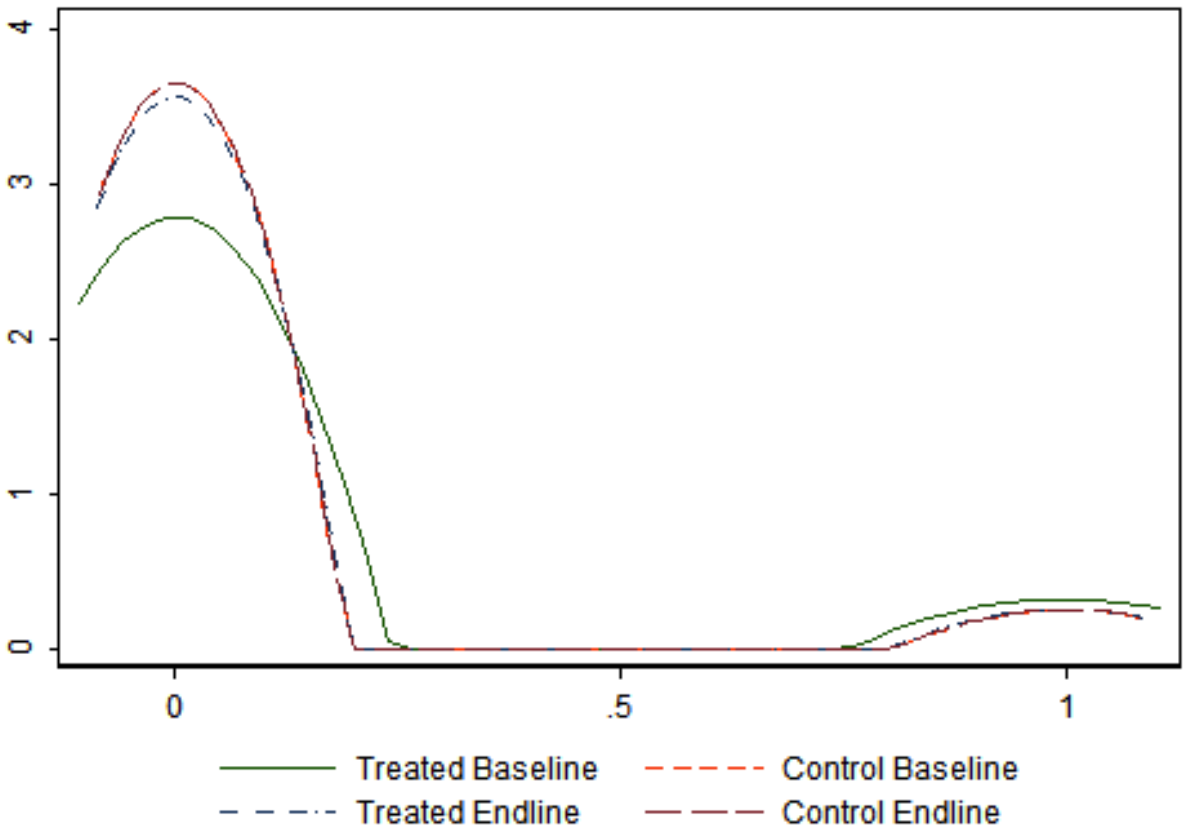


Figure 3H. CRP K-Density Plot

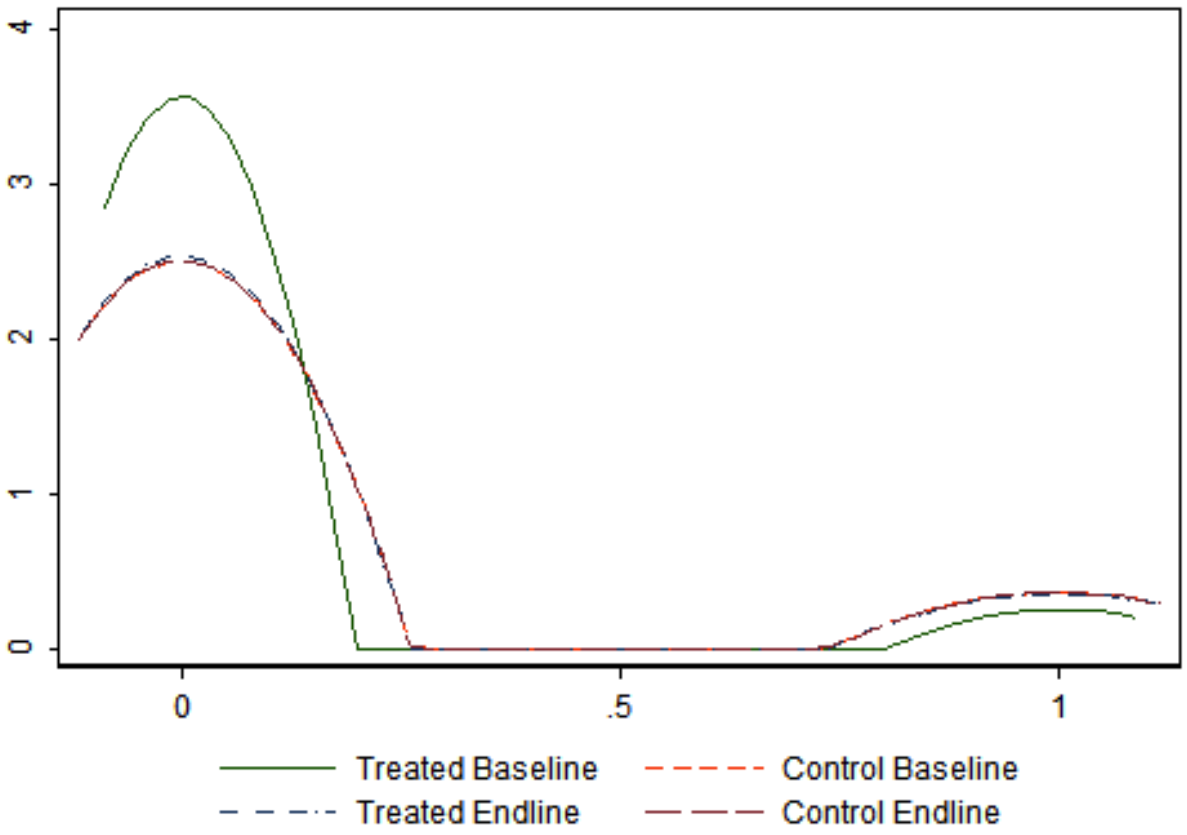


Figure 3I. Labor Intensity Sedentary Activity K-Density Plot

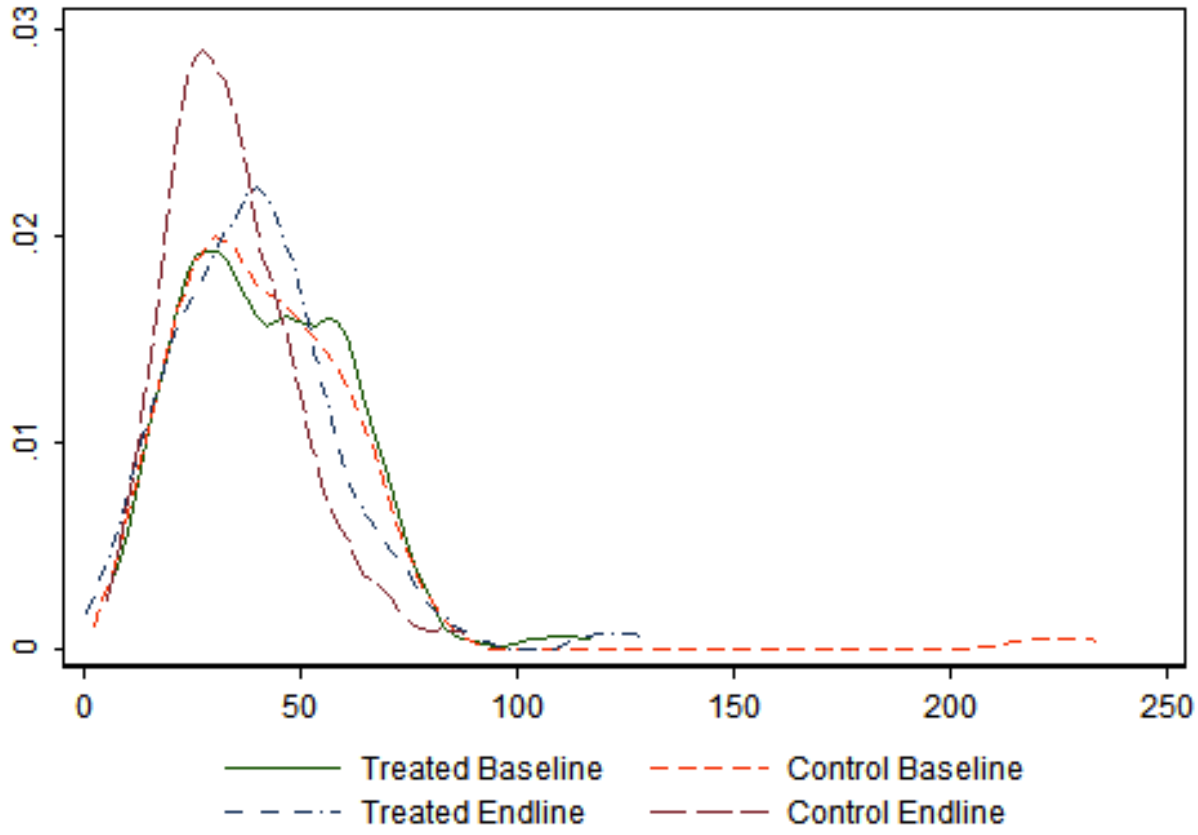


Figure 3J. Labor Intensity Less Than 3 METS K-Density Plot

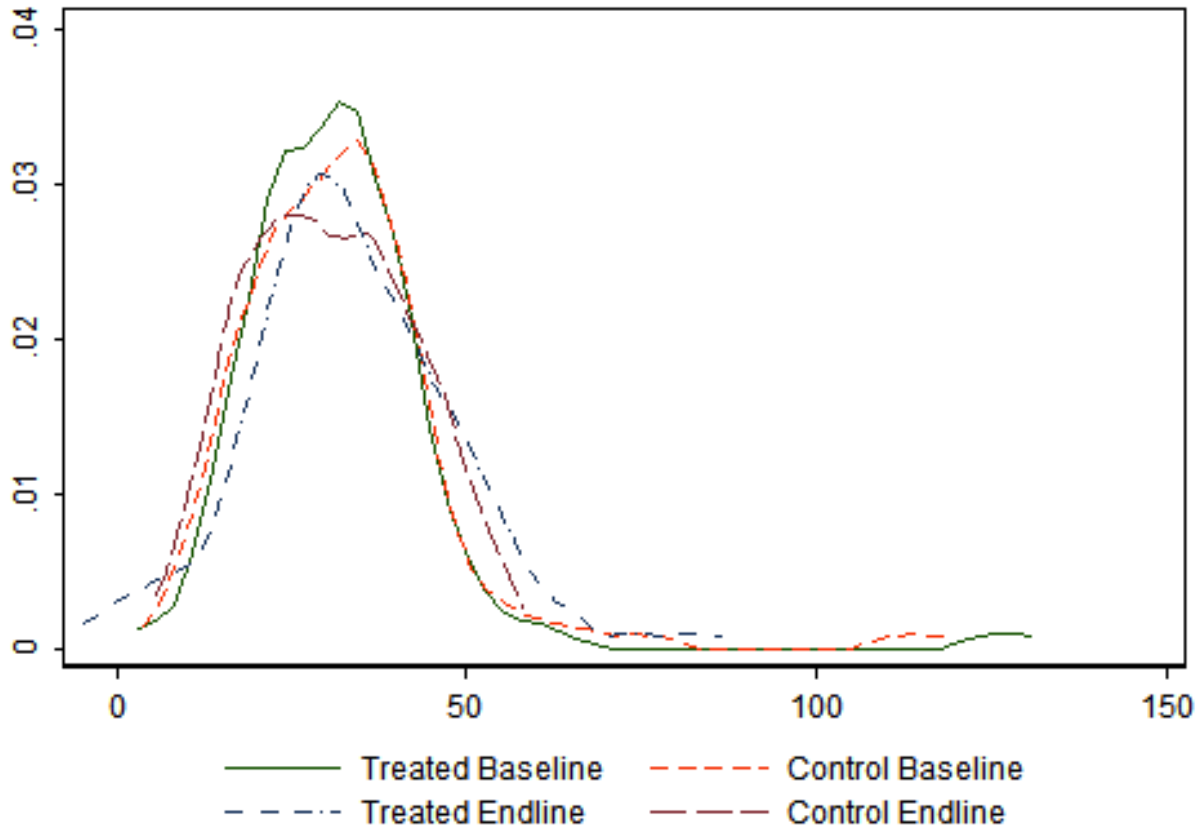


Figure 3K. Leisure Intensity Greater Than 3 METS K-Density Plot

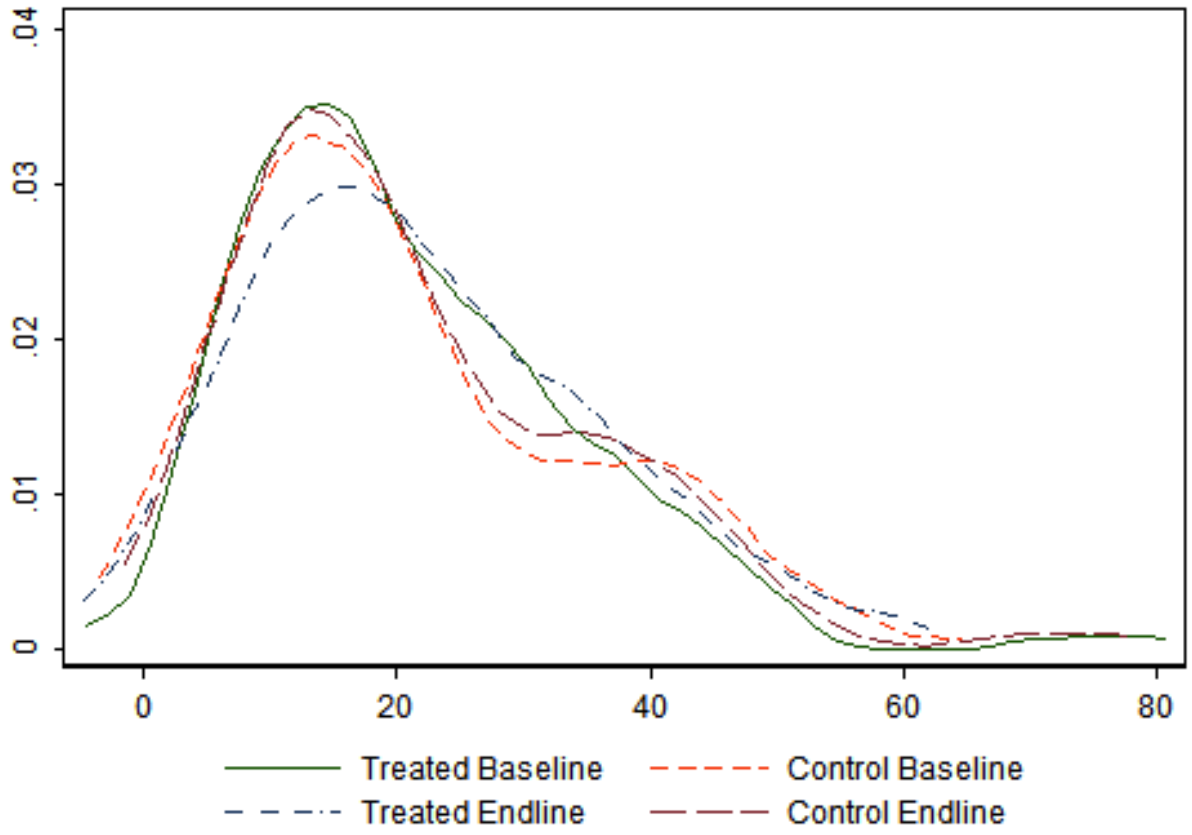
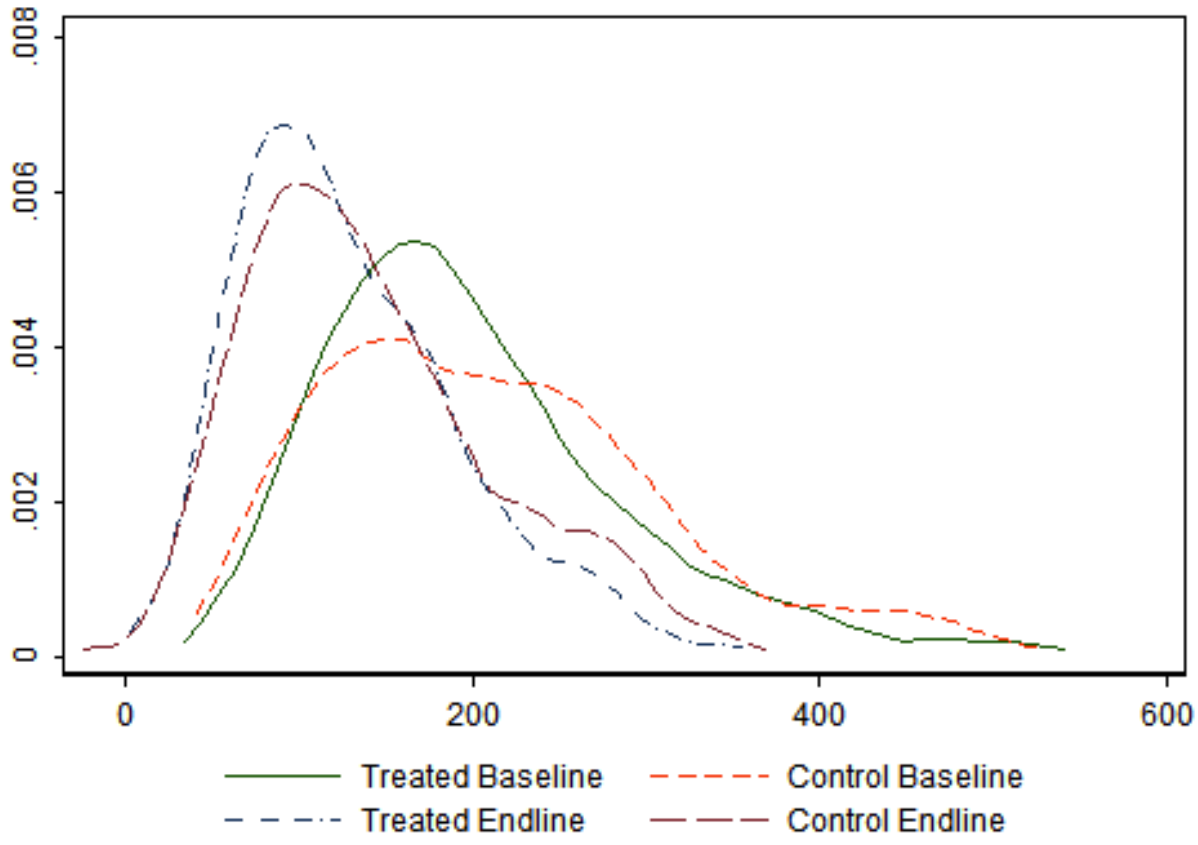


Figure 3L. Kilograms of Tea Leaves Picked K-Density Plot



References

Assah, Felix K., et al. "Accuracy and Validity of a Combined Heart Rate and Motion Sensor for the Measurement of Free-Living Physical Activity Energy Expenditure in Adults in Cameroon." International journal of epidemiology 40.1 (2011): 112-20.

Bain, Barbara J. "Blood Cells : A Practical Guide." Blackwell Publishing Ltd. 2006.

Basta, Samir S., et al. "Iron Deficiency Anemia and the Productivity of Adult Males in Indonesia." The American Journal of Clinical Nutrition 32.4 (1979): 916-25.

Beard, John L., et al. "Interpretation of Serum Ferritin Concentrations as Indicators of Total-Body Iron Stores in Survey Populations: The Role of Biomarkers for the Acute Phase Response." American Journal of Clinical Nutrition, 84.6 (2006): 1498-1505.

Beutler, Ernest, and Jill Waalen. "The Definition of Anemia: What is the Lower Limit of Normal of the Blood Hemoglobin Concentration?" Blood 107.5 (2006): 1747-50.

Black, Steven, Irving Kushner, and David Samols. "C-Reactive Protein." Journal of Biological Chemistry 279.47 (2004): 48487-90.

Cawley, John H., Richard V. Burkhauser, and National Bureau of Economic Research. Beyond BMI : The Value of More Accurate Measures of Fatness and Obesity in Social Science Research. Cambridge, MA: National Bureau of Economic Research, 2006.

Cole, Matthew A., and Eric Neumayer. "The Impact of Poor Health on Total Factor Productivity." The Journal of Development Studies 42.6 (2006): 918-38.

Committee on Mineral Requirements for Cognitive and Physical Performance of Military Personnel, and Committee on Military Nutrition Research. Mineral Requirements for Military Personnel : Levels Needed for Cognitive and Physical Performance during Garrison Training. Washington, DC, USA: National Academies Press, 2006.

Cook, James D., Flowers, Carol H., and Barry S. Skikne. "The Quantitative Assessment of Body Iron." Blood 101.9 (2003): 3359-64.

Crichton, Robert R. Iron Metabolism : From Molecular Mechanisms to Clinical Consequences. Chichester, UK: John Wiley & Sons, 2009.

Crouter, S.E., Kuffel E., Haas J.D., Frongillo E.A., and D.R. Bassett Jr. A refined 2-regression model for the ActiGraph accelerometer. *Med Sci Spots Exerc.* 2010; 42: 1029-1037.

Edgerton V.R., et al. "Iron-Deficiency Anaemia and its Effect on Worker Productivity and Activity Patterns." British medical journal 2.6204 (1979): 1546-9.

Ehrenstein, Walter H. and Addie Ehrenstein. Psychophysical methods. In Windhorst, U. and Johansson, H. (eds), Modern Techniques in Neuroscience Research. Springer, Berlin, 1999.

Fan, Jin, et al. "Testing the Efficiency and Independence of Attentional Networks." Journal of cognitive neuroscience 14.3 (2002): 340-7.

Fournier, Thierry, Najet Medjoubi-N, and Dominique Porquet. "Alpha-1-Acid Glycoprotein." Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1482.1-2 (2000): 157-71.

Gibson, R.S. Principles of Nutritional Assessment. Oxford University Press, Oxford, UK, 1990.

Haas, Jere D. “The effects of iron deficiency on physical performance.” *Mineral Requirements f Military Personnel: Levels Needed for Cognitive and Physical Performance during Garrison Training*, edited by the Food and Nutrition Board, Institute of Medicine (IOM), The National Academies Press, Washington D.C, 451-61, 2006.

Haas, Jere D., and Thomas Brownlie. "Iron Deficiency and Reduced Work Capacity: A Critical Review of the Research to Determine a Causal Relationship." The Journal of nutrition 131.2 (2001): 676S-90S.

Hiilloskorpi, H.K., Pasanen, M.E., Fogelholm, M.G., Laukkanen, R.M., Manttari, A.T. (2003). Use of heart rate to predict energy expenditure from low to high activity levels. *Int J Sports Med, 24*, 332-335.

Hoddinott, John, et al. "Effect of a Nutrition Intervention during Early Childhood on Economic Productivity in Guatemalan Adults." The Lancet 371.9610 (2008): 411-6.

Horton, Sue, and Jay Ross. "The Economics of Iron Deficiency." Food Policy 28.1 (2003): 51-75.

Iannotti, Lora L., Tielsch, James M., Black, Maureen M., and Robert E. Black. "Iron supplementation in early childhood: health benefits and risks." The American Journal of Clinical Nutrition. 84 (2006): 1261-76.

Kanani Shubhada J., and Rajan H. Poojara. "Supplementation with Iron and Folic Acid Enhances Growth in Adolescent Indian Girls." The Journal of Nutrition 130.2S (2000).

Li, Ruowei, et al. "Functional Consequences of Iron Supplementation in Iron-Deficient Female Cotton Mill Workers in Beijing, China." The American Journal of Clinical Nutrition 59.4 (1994): 908-13.

Lozoff, Betsy, et al. "Long-Lasting Neural and Behavioral Effects of Iron Deficiency in Infancy." Nutrition reviews 64 (2006): S34-43.

McCann, Joyce C., and Bruce N. Ames. "An Overview of Evidence for a Causal Relation between Iron Deficiency during Development and Deficits in Cognitive Or Behavioral Function." The American Journal of Clinical Nutrition 85.4 (2007): 931-45.

Meenakshi, J. V., et al. "How Cost-Effective is Biofortification in Combating Micronutrient Malnutrition? an Ex Ante Assessment." World Development 38.1 (2010): 64-75.

Mielczarek, Eugenie V., and Sharon Bertsch McGrayne. Iron, Nature's Universal Element : Why People Need Iron & Animals make Magnets. New Brunswick, N.J.: Rutgers University Press, 2000.

O'Donnell, Owen, et al. "Analyzing health equity using household survey data: A guide to techniques and their implementation." Washington, DC: The World Bank, 2008

Oaxaca, Ronald. "Male-female Wage Differentials in Urban Labor Markets." International Economics Review. 9:693-709.

Penney John B. Jr, and Anne B. Young. "Striatal Inhomogeneities and Basal Ganglia Function." Movement disorders : official journal of the Movement Disorder Society 1.1 (1986): 3-15.

Powell-Tuck Jeremy, and Enid M. Hennessy. "A Comparison of Mid Upper Arm Circumference, Body Mass Index and Weight Loss as Indices of Undernutrition in Acutely Hospitalized Patients." Clinical nutrition (Edinburgh, Scotland) 22.3 (2003): 307-12.

Przybyszewski, Eric. "Predicting Energy Expenditure from Physical Activity, Heart Rate, and Anthropometry in Female Indian Tea Pluckers." Working Paper.

Rajagopalan, S, and Melvin A. Vinodkumar. "Effects of Salt Fortified with Iron and Iodine on the Haemoglobin Levels and Productivity of Tea Pickers." Food and Nutrition Bulletin 21.3 (2000): 323-9.

Ruben, Ruerd, and Annemarie de Ruiter. "Labour, Leisure and Household Income in Rural Settlements: An Empirical Assessment in the Atlantic Zone of Costa Rica." LEISURE STUDIES 21 (2002): 201-20.

Scholz, Barbara D., et al. "Anemia is Associated with Reduced Productivity of Women Workers Even in Less Physically Strenuous Tasks." British Journal of Nutrition 77 (1997): 47-57.

Skikne, Barry S., Flowers, Carol H., and James D. Cook. "Serum Transferrin Receptor: A Quantitative Measure of Tissue Iron Deficiency." Blood 75.9 (1990): 1870-6.

Spurr G.B., Prentice A.M., Murgatroyed P.R. Energy expenditure from minute by minute heart rate recording: comparison with indirect calorimetry. American Journal Clinical Nutrition 48.3 (1998), 552-59.

Stein, Alexander J., et al. "Analyzing the Health Benefits of Biofortified Staple Crops by Means of the DALY Approach: A Handbook Focusing on Iron, Zinc, and Vitamin A." HarvestPlus (2005).

Strauss, John, and Duncan Thomas. "Health, Nutrition, and Economic Development." Journal of Economic Literature 36.2 (1998): 766-817.

Thomas, Duncan, et al. “Iron Deficiency and the Well-Being of Older Adults: Early Results from a Randomized Nutrition Intervention.” Population Association of America Annual Meetings. National Institutes of Aging, 2003.

Umbreit, Jay. "Iron Deficiency: A Concise Review." American Journal of Hematology 78.3 (2005): 225-31.

Zivin, Joshua S. Graff, and Matthew J. Neidell. "The Impact of Pollution on Worker Productivity." NBER Working Paper No. 17004 (April 2011).