UPPSALA
UNIVERSITET

Uppsala Center for Labor Studies
Department of Economics

Working Paper 2013:13

Revisiting the Educational Effects of Fetal Iodine Deficiency

Niklas Bengtsson, Stefan Peterson and Fredrik Sävje

Uppsala Center for Labor Studies
Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2013:13
October 2013

Revisiting the Educational Effects of Fetal Iodine Deficiency

Niklas Bengtsson, Stefan Peterson and Fredrik Sävje

Papers in the Working Paper Series are published on internet in PDF formats.
Download from http://ucls.nek.uu.se/

Revisiting the Educational Effects of Fetal Iodine Deficiency*

Niklas Bengtsson† Stefan Peterson‡ Fredrik Sävje§

October 25, 2013

Abstract

Recent research has reported positive effects on schooling due to in utero protection from iodine deficiency resulting from iodized oil capsule distribution in Tanzania. We revisit the Tanzanian experience by investigating how these effects differ over time and across surveys; across different treatment specifications; and across additional educational outcome measures. Contrary to previous studies, we find that the estimated effects tend to be small and not robust across specifications or samples. Using all available data and a medically motivated iodine depletion function, we find no evidence of a positive long-run effect of iodine deficiency protection on educational attainment.

Keywords: *Iodine deficiency, Education, Prenatal exposure, Multiple outcomes, Replication, Field, Robles and Torero*

JEL: I12, I21, J16, O15

*We would like to extend our gratitude to Douglas Almond, Sebastian Axbard, Rita Ginja, Linna Martén, Eva Mörk and seminar and conference participants at Uppsala University and EEA Gothenburg 2013, for comments and suggestions. Bengtsson and Sävje acknowledge financial support from the Swedish International Development Cooperation Agency through grant Uforsk SWE-2012-109. Please direct correspondence to Fredrik Sävje.

†Department of Economics, Uppsala University and UCLS. E-mail: niklas.bengtsson@nek.uu.se.

‡Department of Public Health Sciences, Karolinska Institutet and Department of Women's and Children's Health, Uppsala University and School of Public Health, Makerere University. E-mail: stefan.peterson@ki.se.

§Department of Economics, Uppsala University and UCLS. E-mail: fredrik.savje@nek.uu.se.

1 Introduction

Iodine deficiency (ID) is known to have significant effects on brain development and cognitive ability. Iodine exists naturally in the soil, but its natural prevalence varies across regions, and iodine deficiency disorders (IDD) remain endemic in many developing countries. Preventing ID is relatively inexpensive and is expected to have substantial economic and social benefits (World Health Organization 1992). Since the 1970s, iodine supplementation programs have been implemented on a large scale throughout Africa and Asia, and many of these programs have been followed by countrywide increases in economic growth and educational attainment.

In an important contribution, Field, Robles and Torero (2009) use the natural variation in fetal ID protection arising from an iodized oil capsule (IOC) distribution program in Tanzania to estimate the impact of ID protection on educational attainment. The Tanzanian IOC distribution program was launched in 1986 and is noteworthy for its sheer size: the IOC programs targeted districts that contain a quarter of the Tanzanian population and a total of 6 million IOCs were distributed (Peterson et al. 1999). To isolate the causal effect of ID protection, Field et al. (2009) use the roll-out of the programs and the timing of births as the source of identification. The main identifying assumption is that fetal brain development is particularly sensitive to iodine deficiency during the first trimester. This assumption is supported by previous medical research, which has established that iodine availability in the early prenatal period is of great importance for future development (Pharoah and Connolly 1987; Cao et al. 1994). As mothers are (arguably) unable to perfectly time their pregnancies and the exact timing of the programs differ in an unsystematic way, exposure to iodine supplementation during the first trimester is as-if random, given fixed district and cohort effects.

Field et al. (2009) focus on educational attainment among children aged 10-13 and find that the iodine supplementation programs increased educational attainment by 0.35 years on average. They find no effects on health, suggesting the mechanism operates chiefly through cognitive ability. Using the IOC programs' coverage rates in Tanzania to calculate the hypothetical case of achieving full IOC protection, they conclude that internationally, "the expected increase in grade attainment for a child protected from fetal ID is a minimum of 0.73 years" (p. 165). Moreover, they find that the impact of ID protection is larger and more robust for girls, which would provide an explanation for gender differentials in higher education since "lower female attainment is almost entirely accounted for by the extremely low rate at which girls pass the national secondary school qualifying exam" (p. 142). Using the same identifying strategy, Adhvaryu and Nyshadham (2011) find that ID protection also affects parental investments in child health, suggesting that behavioral responses can be of importance in addition to the direct biological link.

In this paper, we exploit the availability of new data to reexamine the link between ID protection and educational attainment in Tanzania, using longer-run outcomes and complementary educational outcomes. A first motivation for our study is that the

medical literature is not fully informative about the effect of mild iodine deficiency in utero on educational gains (Zimmermann 2009a). While there is a documented link between iodine deficiency and cognitive ability among children (although it is not fully explored), educational attainment does not necessarily reflect cognitive ability, particularly in Africa (Chaudhury et al. 2006). Other indicators of cognitive ability, such as literacy, might be more informative regarding the mechanisms involved. Second, long-term data are necessary to assess the proposition that ID protection increased secondary school enrollment among females, thereby helping to close the gap in male-female secondary education. Finally, the robustness of past work with respect to sample, outcomes and model specification needs to be assessed to fully characterize its significance, in particular given the strong assumptions on the timing of birth and the timing of the IOC program.

To serve this aim, we revisit an extended number of cohorts exposed to the IOC program in utero. In particular, we include data from individuals born between 1986 (when the first IOC began) and 1994 (when salt iodization was nearly universal) collected from five household surveys conducted between 1999 and 2010. In total, we increase the sample size by more than 600% relative to the past literature. The additional data not only increase precision and the scope for long-term analysis but also contain better data for the identification of treatment effects.

We improve the treatment assignment primarily by employing an updated depletion model that is more relevant to the Tanzanian context, and by using more detailed birth date information. This improvement is important because the identification strategy requires exact information on the biological properties of iodine depletion, the roll-out of programs and timing of birth. As this information cannot be directly observed, additional assumptions and approximations must be made, which are likely to introduce measurement error in the treatment assignment variable. Due to this imprecision and because protection is likely to spill over across trimesters and even to infancy (Zoeller and Rovet 2004), the estimates should be interpreted as lower bounds of the treatment effects. However, due to the improved treatment assignment used in this study, attenuation bias is expected to be lower than in Field et al. (2009), which would result in larger estimates if the true effect is positive (i.e., tighter lower bounds).

On the full sample, using specifications that from an a priori perspective are sufficient to identify causal effects, we find no evidence that fetal ID protection affects educational attainment or literacy. We find indicative effects when using a particular depletion function, on some particular subsamples, and on some particular outcomes. However, using our preferred iodine depletion model on the pooled set of data from all surveys, adjusting inference for multiple outcome indicators, and combining all available sources of information to maximize precision (but disregarding potential attenuation bias due to measurement error), we find no long-run effects and can also reject economically significant effects. There are no significant differences across genders. Our conclusion is that the identification method used in this study, and previously in the literature, on the existing data cannot establish whether the IOC programs in Tanzania had an

impact on educational attainment.

The discrepancy between our conclusions and those of the existing literature lies primarily in the use of new data and new outcomes. The Tanzanian Household Budget Survey from 2000 (THBS 2000) is the only data source in which we find statistically significant effects on educational attainment on the full sample, consistent with the claims in Field et al. (2009), who focus on that data source. However, when employing our preferred specification on the THBS 2000 data, the estimated effects are lower and exhibit no significant gender differences. This difference can be attributed to slight differences in the employed specifications, of no obvious a priori significance.

In addition to our main analysis, the alternative data sources allow us to compare how the proxies used to identify birthplaces (based on the current place of residence) relate to the true place of birth. While we find that migration might have an attenuating effect, we also find indicative evidence that endogenous migration result in an overestimation of the effect for women and an underestimation for men—thereby providing a possible explanation for the findings reported in previous studies. Migration alone is, however, unlikely to explain the lack of significant effects or the differences in estimates among surveys. As the diluting effect of migration is expected to increase over time, we would expect (significant) positive effects in the early surveys that over time tend towards zero. Instead, the estimates indicate positive effects in the surveys from 2000 and 2008 and negative or zero effects in 1999, 2004 and 2010.

The paper is structured as follows. In the next section, we outline the background of the intervention and iodine deficiency. Section 3 discusses the data and the empirical framework. The results are presented in Section 4. Section 5 investigates whether the lack of evidence can be explained by attenuation bias due to migration or imprecise birth information, and Section 6 concludes.

2 Background

2.1 Iodine and iodine deficiency disorders

The proposed link between in utero exposure to iodine deficiency and educational attainment is expected to primarily operate through the biological link between iodine deficiency and cognitive ability. Iodine is a chemical element and a micronutrient important for the synthesis of thyroid hormones. Thyroid hormones play a vital role in the regulation of metabolism and, more important for this study, are essential for growth and development in humans. The conditions resulting from low levels of thyroid hormones are generally labeled *iodine deficiency disorders* (IDD). We will use the abbreviation ID for *iodine deficiency*. While these deficiencies might be reinforced or counteracted by social mechanisms during childhood, in this paper, we will primarily focus on the reduced-form relationship between ID exposure and educational attainment.

The most common cause of IDD in developing countries is low dietary iodine intake, which over time leads to the depletion of iodine stores and thereby the decreased pro-

duction of thyroid hormones. The availability of dietary iodine varies with geography. Particularly inland regions with old, non-clay soil surfaces tend to have low soil iodine availability and low availability of iodine-rich sea-food, which result in low dietary iodine intake (Fuge and Johnson 1986). However, there are several interacting factors affecting the intake of iodine or its synthesis into thyroid hormones. Among these is the availability of other micronutrients vital for thyroid synthesis, predominately iron (Zimmermann et al. 2000) and selenium (Vanderpas et al. 1990). Further, goitrogens, which are substances that interfere with the uptake and metabolism of iodine, have a detrimental effect. Two common sources of goitrogens are cassava consumption and tobacco use.[1] The prevalence of IDD thus varies greatly both between and within countries depending on the exact geography and dietary conditions. All these factors are relevant in the current setting. An early survey estimates that 40% of the Tanzanian population lived in areas with low iodine availability and 25% suffered from iodine deficiency (Haar et al. 1988).

The most common and noticeable of the IDDs is goiter, which is the enlargement of the thyroid gland. The enlargement is a compensatory response to low iodine intake that increases hormone production efficiency. Although the effects of goiter is mild relative to other IDDs, it can lead to irreversible thyroid dysfunction and is important in tracking the overall prevalence of ID due to its noticeability. Some of the most severe IDDs are the result of ID during the pregnancy and childhood, due to the sensitivity of brain development to deviations from optimal thyroid hormone levels during that period. Severe instances of ID during pregnancy are associated with stillbirth, spontaneous abortion and severe mental retardation—cretinism (Zimmermann 2009b).

Although the more severe negative health effects of ID have long been known, more recent studies have highlighted that also mild ID during pregnancy is associated with hindered development (Lavado-Autric et al. 2003; Pop et al. 1999; Haddow et al. 1999). While there is an ongoing discussion concerning the exact period of greatest sensitivity to ID and how persistent the effects of mild ID are, there is substantial evidence that the early prenatal period is particularly sensitive. Notably, during the first trimester, the fetus cannot itself synthesize thyroid hormones and is completely reliant on the mother's hormone production and is thus arguably more sensitive to ID. Further, experimental studies have demonstrated that cretinism can be prevented in an iodine deficient population only if supplementation is given before conception (Pharoah et al. 1971), and that iodine supplements fail to prevent hindered cognitive development to a measurable level if given in the third trimester (Cao et al. 1994).

The previous literature has primarily focused on the effects of ID in early childhood. The effect on outcomes in later childhood and adulthood, including the effect on higher cognitive abilities, has been less studied. There is evidence that different sets of abilities are sensitive to ID during separate periods, and only a subset of these is confined to

[1] See, for example, Gaitan (1990); Bourdoux et al. (1978); Bertelsen and Hegedüs (1994). Particularly, the exact means of preparing cassava for consumption substantially affects the amount of goitrogens (Peterson 2000).

the prenatal period (Zoeller and Rovet 2004 provide an overview). These other abilities might also affect educational attainment later in life. The estimates resulting from the use of the first-trimester identification strategy (outlined below) are thus likely to be a lower bound of the effects of ID on educational outcomes.

The persistence of severe cognitive disability from early ID is well established; for example, cretinism is a permanent disorder and thus the effects of supplementation preventing cretinism continue long after birth (Pharoah and Connolly 1987). The evidence regarding milder ID is more mixed. While there are indications that sufficient iodine can have a compensating effect on spells of ID, at least during prenatal development (Pop et al. 2003), those treated for existing disorders were found to have only partial recovery (see Dugbartey 1998 for a review).

The discrepancy between girls and boys observed in Field et al. (2009) is a somewhat unexpected result. There are two laboratory studies on rodents that provide some indications of gender differentials. In particular, Chan et al. (2005) report different responses between male and female guinea pigs in the regulation of thyroid hormone receptors. However, Chan et al. note that this may be the result of different compensating mechanisms rather than a manifestation of increased sensitivity. Further, Friedhoff et al. (2000) observe behavioral differences between male and female rats when their mothers' thyroid glands were completely removed, thereby inducing severe ID during the prenatal period. There are, however, virtually no studies on human subjects have been able to establish a gender difference with respect to the effects of prenatal ID protection. Other than Field et al. (2009), the one exception to our knowledge is a recent observational study of one-year olds in Spain that finds a correlation between a diet low in iodine (proxied by self-reported fish consumption and mineral supplement intake) among pregnant mothers and infant neurodevelopment (Murcia et al. 2011).

2.2 The IOC programs

Following the increased awareness of the benefits of IDD prevention, several large-scale IOC supplementation programs were introduced in the late 1980s in Tanzania. The intended purpose of the IOC programs was to target the populations most affected with IDD until universal salt iodization (USI) began in the early 1990s. The IOC programs were thus directed to districts with high goiter prevalence, and a total of 27 districts were selected for inclusion.

The intended structure of the programs was to distribute iodized oil capsules every second year. At each distribution round, each male and female aged from 2 to 45 years was to be given IOC containing 400 mg of iodine, and children aged from 12 to 23 months were to be given a dose of 200 mg of iodine (Peterson et al. 1999). Due to administrative problems, delays in both initial and repeated distribution rounds were common, with only 10 districts receiving their initial round in 1988 or before.[2] The average coverage rate was 64%, and full coverage was never reached in any district.

[2]Refer to Appendix A for details on the timing of the programs.

Despite the distributional problems, the programs reached a substantial number of individuals. A conservative estimate is that the programs provided 12 million person-years of protection from ID (Peterson et al. 1999).

In the early 1990s, the USI program had begun. By that time the focus of the IOC programs shifted from districts with high IDD prevalence to districts not yet reached by the USI program, namely districts where less than 75% of households had access to iodized salt. Thus during this period the absence of an IOC program does not necessary indicate that the population is unprotected from ID, even in districts with previously very high levels of IDD.

2.3 Schooling

The educational system in Tanzania is divided into three levels—primary education, lower secondary ("ordinary level") and higher secondary ("advanced level"). Primary education consists of the first seven grades of formal education, of which each grade is expected to take one year to complete. Secondary education encompasses six grades in total, divided into four grades of lower secondary and the following two in upper secondary. Children are expected to begin the first grade of primary education at age seven, thus completing primary school at age 13 (Ministry of Education and Culture 1995). During primary education two mandatory exams are given—the first in the fourth grade and the second at the end of primary education.

Considerable advances have been made to achieve universal primary education in Tanzania in recent years.[3] However, during the IOC program, Tanzania struggled with late school entry and high drop-out rates. The gross enrollment rate in 1998 was 76.4%, but due to late entry and repeated grades, the net enrollment rate was only 56.7%. The transition between grades was low: in 1998 the probability that a student prematurely left education, on a yearly basis averaged over all primary grades, was 6.6% and the probability of repeating the previous grade was 3.1%. One large contributing factor is the fourth grade exam, after which 8.8% of the students drop out and 11.6% of the students repeat the fourth grade. The exam at the end of primary education, the *Primary School Leaving Examination*, which confers eligibility for secondary school, had a pass rate of 27.1% in 2002 (Ministry of Education and Culture 2002). For cohorts relevant to this study we therefore expect a considerable margin of improvement.

3 Empirical analysis

3.1 Data

We investigate the effects of protection from ID using data from a collection of five household surveys conducted in Tanzania between 1999 and 2010. Three are the *Demographic*

[3] In the early 2000s, the Tanzanian government launched a large-scale program, the *Primary Education Development Program* (PEDP), with the purpose of increasing enrollment and educational quality. Subsequently, the net enrollment rate rose to 80.7% in 2002. This program targeted cohorts younger than those studied here.

and Health Surveys (DHS) for the years 1999, 2004-2005 and 2009-2010 (hereafter labeled DHS 1999, DHS 2004 and DHS 2010). The fourth is the first wave of the *National Panel Survey* conducted in 2008-2009 (labeled NPS 2008), and the fifth is the *Tanzanian Household Budget Survey* of 2000-2001 (labeled THBS 2000).

The surveys differ somewhat in their construction and focus, but all of them are representative of the whole of Tanzania and contain the relevant data needed for identification.[4] All were conducted by the National Bureau of Statistics in Tanzania with the involvement of international organizations.

The main outcome of interest is educational attainment, measured by the number of completed grades. As non-standard grades are reported differently across surveys, we restrict our attention to established grades in formal education (namely the seven grades in primary school, the four grades in lower secondary and the two in upper secondary). We code the educational attainment variable such that it takes the value of the highest completed standard grade (where 0 is no grade, 1 is the first grade in primary school and 13 is the last grade in upper secondary school). In addition, we will consider alternative definitions of educational attainment, such as binary indicators of milestone educational achievements (e.g., completed primary school), as well as a literacy indicator. The definitions are discussed in detail when the results are presented in Section 4.

The analysis is restricted to the sample of respondents who are permanent residents of the surveyed households (unless otherwise noted) and are related to the household head. This differs from previous studies, which further restrict the sample to children of the household head. The changed sample restriction is made in an effort to increase sample size and improve generality of the results (as all children in the included sample was targeted by the IOC programs). When we perform our analysis on the subsample of children of the household head, as presented in Appendix D, there are only slight differences compared to our main specification and only when analyzing the genders separately. The overall conclusion are thus unaltered by this restriction. This specification choice is further discussed in Appendix B.

The cohorts potentially benefiting from protection from ID in utero were born between 1986, when the first program began, and 1994, when universal salt iodization had begun in nearly all districts.[5] We include all districts targeted by at least one IOC program, net of two targeted districts where the first program began in 1994 or later (and as a result there is essentially is no treatment variation). The two districts are the Mbinga

[4]The main differences are that the DHS surveys also conduct a more in-depth interview with a selected number of respondents in each household, mainly women. While the additional information available for these respondents is not vital to the analysis, it aids us by providing more accurate measurements of the birth date (and in the DHS 1999 and DHS 2004 also revealed literacy). The NPS 2008 is the first wave in a long-running panel data set under construction, with highly detailed data. The sample size is subsequently lower than in the cross-sectional surveys. The THBS 2000 is, instead, a large survey with less detailed information.

[5]Salt iodization began prior to 1994 in some districts. For later cohorts in the sample, the absence of an IOC program does therefore not necessarily imply vulnerability to ID. As this would tend to contaminate the control group, albeit only for a small subset of the sample, in Appendix E we investigate the effect for the cohorts born before any salt iodization (cohorts born 1986 to 1990). The effects then tend to be negative but do not change the overall conclusions.

district in the Ruvuma region and the Bukoba Rural district in the Kagera region.[6] On the one hand, their inclusion could increase the precision in the control variables and the time fixed effects. On the other hand, as the objective of the later IOC programs was to complement the salt iodization programs, we have reason to believe that these districts were targeted on different grounds than districts with early programs.[7] As a precaution, we therefore exclude them from the main analysis. However, the results are not sensitive to their inclusion, as presented in Appendix F.

In our primary treatment specification, but not in the four alternative specifications, we use information from a number of goiter surveys that were conducted in Tanzania during the 1980s. The purpose of these surveys was to estimate how widespread and severe IDD was in Tanzania. Due to the complexity and diversity of the disorders following from iodine deficiency, a common procedure was to use goiter prevalence as a proxy for other disorders. In particular, we use the district average *total goiter rate*—the ratio between individuals with an enlarged thyroid gland and the total population—proxied by goiter prevalence amongst school children when calculating the probability of treatment.[8]

3.2 Empirical strategy

We follow Field et al. (2009) and adopt a fixed effects approach where treatment is considered to be as-if randomly assigned conditional on district and birth date effects. The exact specification is described by:

$$Y_{idt} = \beta_0 + \beta_1 T_{dt} + \boldsymbol{\beta}_2 \mathbf{X}_{idt} + \mu_d + \lambda_t + \varepsilon_{idt},, \tag{1}$$

where Y_{idt} denotes the studied outcome for respondent i born in district d with birth date t. The treatment variable, T_{dt}, is the calculated probability of in utero protection from ID for an individual born in district d at date t (see the following section for details of this calculation), thus β_1 is coefficient of interest. μ_d and λ_t are district and birth date fixed effects and $\mathbf{X}_{idt}$ is a vector of control variables measured at either the individual or household level. We allow the error term ε_{idt} to be correlated within each cohort in a district (by clustering the standard errors at the district-birth year level).

The main difference from previous studies is that we use a more parsimonious specification with respect to the control variables included in $\mathbf{X}_{idt}$. In particular, on the individual level we only include indicators of the respondent's age, sex, relation to the household head and date of interview. On the household level we include indicators

[6] In Mbinga, the first program was in 1995, thus no one in the relevant cohorts could have benefited from the program in the first trimester. In Bukoba Rural, the first program was in 1994, thus only a very small fraction of the respondents here is considered treated (1.5%) and those with a non-zero probability are assigned a small probability (on average 11.2%). See Appendix A for further details.

[7] Some indication of this is given when comparing goiter rates: Bukoba Rural had a total goiter rate, before the IOC programs began, of 28.0%, which is the lowest rate among the targeted districts. Mbinga had a rate of 35.9%, which is the fourth lowest. The average rate among all districts with IOC programs is 60.2%.

[8] One of the treated districts, Kasulu (Kigoma), was not included in the goiter surveys. We impute the missing value with the average of the treated districts in the Kigoma region.

of the educational levels of the household head and spouse, and whether the household resides in an urban area.[9] These controls are important predictors of educational attainment and are included to increase precision. This is especially true for the age variable, which in the current setting is not perfectly correlated with birth year due to the overlapping surveys. We can therefore, unlike previous studies, differentiate between age and cohort effects. Although the surveys are reasonably similar, we interact the time of interview indicators with survey indicators to capture eventual differences in survey design.

Given the identification strategy, the control variables should not contribute to identification, although they might increase the precision of the estimates. Notably, their inclusion comes at the cost of increasing the risk of biased estimates. Due to the nature of the treatment, which occurred before birth, the majority of the potential control variables run the risk of being themselves affected by treatment. Some of the control variables used in previous studies are subject to this risk.[10] For example, fertility decisions could be endogenous if, in economic jargon, the quality of children increases the shadow price of the quantity of children. This would suggest that an exogenous increase in a child's health could induce parents to revise their "stopping rule" of fertility (Becker and Tomes 1976; Rosenzweig and Wolpin 1980) and would thereby imply that healthy children, or children with higher cognitive ability, are more likely to have a high birth order. In addition, as a response to a better-endowed child, parents may act to compensate for household inequality by investing more in the endowed child's siblings.[11] Parents of more healthy children might thus locate nearer to school centers and health clinics to compensate the siblings. Notably, some of these potentially problematic variables, particularly birth order, are correlated with treatment, consistent with the concerns that these outcomes might reflect parental responses to child endowments at birth.

The specification we ultimately employ only includes variables that are likely to be important determinants of schooling but deemed unlikely to be affected by treatment, thus minimizing this problem. Using an even more parsimonious specification that, in addition to the identifying variables, only includes controls more or less guaranteed to be unaffected by treatment (namely indicators of age, sex and date of interview), as presented in Table 14 in Appendix G, only slightly alter the results. Neither are the overall conclusions sensitive to the inclusion of a set of controls similar to those used in

[9]The educational level of the household head and spouse are defined in six categories, corresponding to no education, some primary, completed primary, some secondary, completed secondary, and higher education. For households without a spouse (2,326 respondents or 23.3% of the sample), the educational level of the spouse will naturally be unreported. In order not to exclude these observations from the analysis, we form a separate category for them. Refer to Appendix B for a discussion of the importance of the inclusion of these respondents. For an additional 81 respondents (0.8% of the sample), the educational level is unreported for either the head or the spouse. Similarly, we include a separate category for them. A third group are respondents who themselves are either a household head or spouse (in total 374 respondents, or 3.8% of the sample, almost exclusively in the later surveys). Including the educational level as a control variable would, for these respondents, be to include the outcome variable as an independent variable. Thus we form a separate category for these respondents.

[10]In particular, compared to Field et al. (2009), we exclude controls for fertility decisions (birth order and number of children), access to health and schooling facilities (distance to secondary education and health clinic) and the different poverty measures (food security, home ownership, and house quality).

[11]See Almond and Currie (2011) for a review of behavioral responses to fetal exposure to shocks.

previous studies, as presented in Table 15 in Appendix G.

In the presence of migration and endogenous family composition, this reasoning could also apply to the fixed effects. This motivates the use of fixed effects at the district level rather than the village or household level. Lower level fixed effects could, however, increase precision by controlling for factors varying within the districts. To investigate this aspect we, in Appendix H, also present the results when using ward fixed effects (each district contains up to eight wards). With this specification, the estimated effects generally increase (in a few instances double), but from a low initial level. The qualitative results thus remain intact independent of the level of the fixed effects. In light of the high risk of endogeneity, we conduct the remaining analysis with district fixed effects.[12]

3.3 Treatment assignment

The available data do not allow us, or the previous studies, to assign treatment perfectly and any assignment faces a number of uncertainties. These uncertainties are mainly: (1) The precise details concerning the biological properties of iodine uptake, depletion and need during pregnancy have not been settled in the medical literature. There is uncertainty regarding the stage of pregnancy at which protection is most needed and hence when iodine supplements are the most effective. There is also substantial individual heterogeneity, especially concerning the depletion rate. This problem is potentially aggravated in districts with overlapping programs because, to our knowledge, that particular setting has not been studied previously. (2) No program achieved full coverage, thus only a subset of the individuals in the targeted districts received IOCs. (3) The precise dates when the IOC programs begun are not known; the available data only provide starting years. (4) The length of each program is unknown. They generally ran over a period of several months (according to some reports, up to two years), but the details on each program is not available. Furthermore, the relative distribution intensity during the program could vary but is not observed. (5) The data, depending on the survey, only provide incomplete information of the respondents' birth dates. In some instances both birth year and month are reported, while in others only age in years is provided. As treatment is assigned relative to when the respondent was in utero, this could introduce substantial imprecision. (6) Whether the respondent's mother received the IOC supplement is governed by whether she resided in a targeted district at the time of distribution. This information is not available in any data set. In general, the data only provide the district in which the respondent currently resides. Inter-district migration would therefore be problematic.

We begin by discussing the first point, namely the biological properties of the iodine metabolism. An oft-used model for the depletion of micronutrients or the excretion of

[12]Notably, the results are not robust between the 1986-1994 and 1986-1990 samples discussed above, with positive and slightly significant result in the former and negative and slightly significant results in the later. The result is mainly driven by the NPS 2008 sample, which is a clear outlier (as will be apparent). The pooled estimates that exclude that survey are approximately half the size.

toxicants is the exponential function. However, several studies have found that iodine supplements do not deplete exponentially. Instead the rate of depletion diminishes over time after administration (see, e.g., Untoro et al. 2006). A common alternative is the hyperbolic function, which allows for this type of diminishing depletion rate. The main advantage of the hyperbolic function when studying depletion patterns, is that it, while accounting for the depletion pattern fairly well, does not increase the number of free parameters that must be estimated. However, the hyperbolic function should not be regarded as a representation of the structural depletion process. As discussed in, for example, Furne et al. (1995), the hyperbolic function is instead best understood as an approximation of an underlying multi-compartment model, where iodine is stored in several places in the body and each store depletes exponentially at a different rate.

Given its prevalence in the medical literature, the hyperbolic depletion function would seem to be a natural choice to base the calculation of treatment assignment on (as in the previous studies), but its use entails certain disadvantages. First, the hyperbolic model has exclusively been used to study the depletion pattern of a single supplementation invention. The extent to which the hyperbolic model is able to account for overlapping interventions (as in the Tanzanian setting) is unclear. As the depletion rate in the hyperbolic model is given as a function of the time elapsed since administration, overlapping inventions would provide conflicting depletion rates (the depletion rate will depend on which of the interventions we use as a basis for the calculations). A possible solution would be to let the iodine from each intervention deplete separately with the rate calculated from its respective administration date. This would, however, result in unreasonably low depletion rates for certain levels of stored iodine (and thus overstating the length of protection).[13]

Second, it is not obvious how to incorporate depletion heterogeneity in the hyperbolic model. While altering the depletion parameter is straightforward, it is unclear how a particular setting would translate to a particular parameter value. An important explanation for the observed diminishing depletion rate is that there is continuous dietary iodine intake after supplementation. As mentioned above, the natural availability of iodine varies regionally, which affects dietary iodine intake. If the baseline iodine intake level was high (relative to other districts with endemic goiter) we would expect the supplement to offer protection for a longer period—thereby partly explaining the heterogeneity. In the current setting, we have access to a proxy for baseline iodine intake

[13]The route taken by Field et al. (2009), who use the hyperbolic model, is to calculate the probability of protection given by each intervention separately, as if there were no overlap, and then sum the probabilities in a second step. This yields a probability of protection higher than 100% in some instances (which is rounded down to 100%). We also note that Field et al. (2009) use the hyperbolic model for the stocks of iodine in the body, although the hyperbolic model has been exclusively used to model urine iodine concentration in the existing literature, which is a flow measure. As a possible consequence, their model may overstate the length of protection. Notably, it assigns full protection of the fetus for 24 months and partial protection for an additional 16 months. Most studies have found that administering 400 mg of oral iodine, in the form of fortified poppyseed oil (e.g. Lipiodol, which was used in the Tanzanian programs), offers protection from ID for at most 24 months (Wolff 2001), and some studies have estimated the period to be less than a year (Ingenbleek et al. 1997). Considering the increased requirements during pregnancy, this would indicate that the length, as anticipated, is overstated.

in the form of pre-intervention goiter rates.[14] However, the hyperbolic model does not allow for the explicit inclusion of continuous iodine intake and must rely on parameter adjustments. A model that directly accounts for intake would instead endogenously assign higher depletion rates in districts where iodine intake are low.

The specifics of the Tanzanian setting thus warrant improvements. We will base our depletion model on the discussion in Furne et al. (1995) and use a *multi-compartment* specification. This specification adjusts for depletion heterogeneity by explicitly modeling intake. Furthermore, with this model, there is no need to consider overlapping programs separately, thereby directly accounting for interaction effects. In short, we presume there to be two compartments where iodine can be stored in the body, one with a low depletion rate (representing the thyroid gland) and the second with a high rate (representing all other storage mechanisms). Given baseline iodine intake and eventual IOC supplementation, we can directly model the stores and flows of iodine (albeit in an approximate sense) and thereby calculate the probability of in utero protection. The precise details of this depletion model are explained in Appendix C.

While this model, like any depletion model, encompasses some strong assumptions and while we cannot confirm that it characterizes the true depletion process, we argue that it should be preferred in light of its merits. Notably, the multi-compartment model predicts urine iodine concentration levels that follow the characteristic hyperbolic functional form when shocked with an intake higher than baseline. Further, the model predicts goiter protection ranging from one to two years depending on baseline intake when shocked with a 400 mg IOC supplement, in line with the existing studies. However, to guard against the possibility of misspecification, we will also investigate the effects using four alternative models of iodine depletion.

The first alternative specification is that used in Field et al. (2009), namely a hyperbolic depletion model with an initial half-life of three months.[15] When using this specification, we also include the "correction factor" used in the specification of Field et al. (2009). This variable captures the portion of the assigned treatment probability that relies on the assumptions of partial protection and is intended to counteract eventual misspecification in depletion. Additionally, we investigate three cruder models that

[14] In addition to dietary iodine intake, goiter rates would arguably capture other factors affecting iodine availability that varies regionally, for example exposure to goitrogens. An analysis using this proxy would therefore benefit from the possibility of accounting for a wide range of possible sources of heterogeneity in depletion.

[15] There are some slight differences in the specifications. First, Field et al. (2009) assume that the IOC dose was 380 mg, while the reports by Peterson et al. (1999) and others report that it was 400 mg. However, due to the very high initial depletion rate in this model this makes only a very slight difference. Second, Field et al. (2009) implicitly assume that all respondents are born on the first day of the month. We instead use the expected birth date, i.e. the middle of the month. This again has but a slight impact on the calculated probability. Third, Field et al. (2009) assign the maximum monthly protection probability in the first trimester, implicitly assuming that any month can fully compensate for low protection in the others. We find no support for *full* compensation in the medical literature and therefore average over the first trimester. Fourth, we assign treatment using the full set of birth date information available, as documented below, which differs from the procedure in Field et al. (2009). Fifth, the "correction factor" discussed below cannot be specified in the same way as in Field et al. (2009), as their definition relies heavily on their assumptions regarding birth dates and the following assignment procedure. In Appendix B, we investigate the effects using the exact specification used by Field et al. (2009).

do not make any explicit assumptions regarding the depletion rates. In these models we will assume that the stores of iodine offer protection during pregnancy for 12, 18 and 24 months, respectively, after IOC administration.

The remaining calculations of the treatment probability are common across all five models. For any of the specifications, let $S(t^*, M_1, \cdots, M_K)$ denote the protection probability of a fetus in month t^* when its mother was given IOC in months $M_1, \cdots, M_K$, where K denotes the number of relevant IOC programs. As there was incomplete coverage in all programs, the true K will generally be less than the total number of programs in the district. However, we cannot observe which programs are relevant for a specific respondent (corresponding to the second uncertainty listed above). We follow the past studies and assign treatment as if there were full coverage—that is, we set K equal to the total number of programs in the district.

Since we are interested in the probability of protection during the complete prenatal period, not a specific month, the measure above must be aggregated to an overall probability of in utero protection of a child born on a certain date. Although clear evidence exists that the development of the fetus is highly sensitive to deviations from optimal iodine levels, as previously discussed, the relative importance during the pregnancy is not entirely settled. There are indications that the first trimester is especially important. However, the late prenatal stage are not completely insensitive to deviations (especially for the development of higher cognitive ability, see for example Zoeller and Rovet (2004)). Nevertheless, we will follow the previous studies and assume that *only* the first trimester is of importance. A child born in month t would, on average, been conceived in the middle of month $t-9$. The probability of fetal protection for that child, denoted $P(t, M_1, \cdots, M_K)$, is thus derived by averaging over the first three months of pregnancy:

$$P(t, M_1, \cdots, M_K) = \frac{0.5S(t-9, \cdot) + S(t-8, \cdot) + S(t-7, \cdot) + 0.5S(t-6, \cdot)}{3}.$$

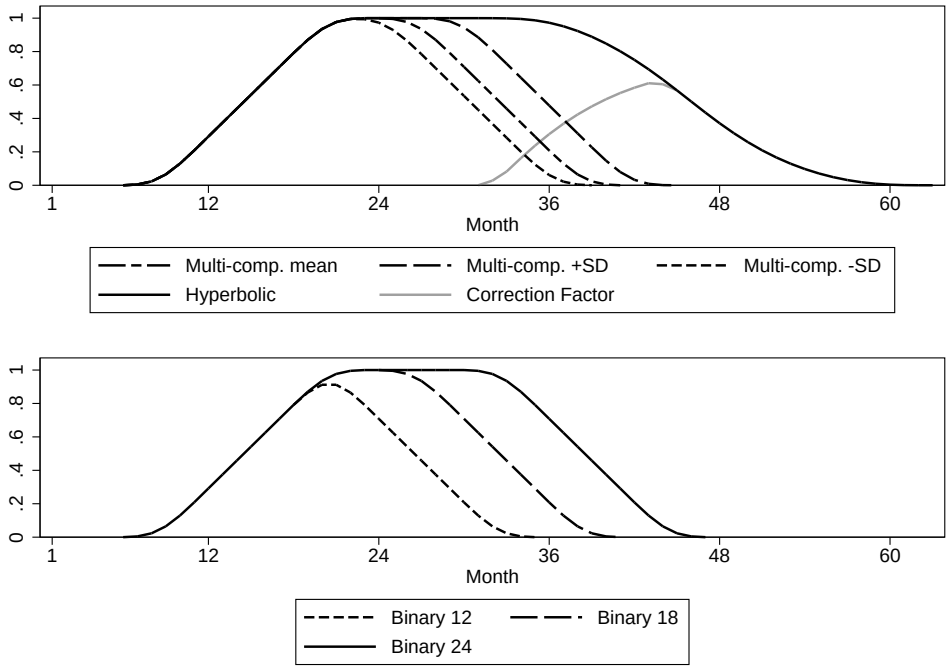
This function requires that we know the exact month when the mothers were given the IOC supplement. However, as listed in the third and fourth points above, we do not know the exact starting date of the programs or their length. We assume that they began uniformly during the specified year with a length of three months. This implies that a resident in a targeted district could have received the IOC at any time between January the program year and February the following year. As the interaction effect of overlapping programs could be of importance, we will consider all possible combinations of distribution dates. For a district with IOC programs in the years $y_1, \cdots, y_K$, the resulting probability of protection for an individual born in month t is given by:

$$T(t, y_1, \cdots, y_K) = \frac{1}{36^K} \sum_{m_1=0}^{11} \sum_{m_1'=0}^{2} \cdots \sum_{m_K=0}^{11} \sum_{m_K'=0}^{2} P(t, M_1, \cdots, M_K),$$

where each M_n is given by $(12y_n + m_n + m_n')$.

The calculated monthly probability resulting from each depletion model when there is a single IOC program ($K = 1$) is presented in Figure 1. Notably, the multi-compartment model varies with baseline iodine intake, while the hyperbolic model predicts protection for considerably longer time than the other models. The inclusion of the "correction factor" (also presented in the figure) in the hyperbolic specification, however, mitigates this problem, resulting in an assignment similar to the cruder model that assigns protection for 24 months.

Figure 1: Treatment probability by birth month



Notes: The two panels present the calculated probability of being protected from ID in utero by birth month relative to January the year of the IOC program (labeled with 1). In the first panel the two specifications that make explicit assumptions on the depletion pattern are presented. The multi-compartment model is presented with three different baseline iodine intake levels, corresponding to the district mean and with one standard deviation from the mean. For the hyperbolic model the base probability is presented, as well as the "correction factor" discussed in the text. The second panel presents the models that only implicitly make assumptions on the depletion pattern, representing 12, 18 and 24 months of protection.

The function $T(t, y_1, \cdots, y_K)$ assigns probability of protection based on the respondent's birth month. This information is only available for a fraction of the respondents, leading to the fifth uncertainty. The data present three levels of birth date information. At the most detailed level, both birth year and month are reported. At the next level, birth year is known, while at the lowest level only the age of the respondent is reported (which is also reported at the two higher levels). The year and month when surveyed are available for all respondents. We will exploit this full set of information to assign treatment as accurately as possible. Notice that when the birth month is reported, we

can use $T(t, y_1, \cdots, y_K)$ directly, while in the other two cases we can only derive intervals of possible birth months.

In the first case with limited information, where age and birth year are known, we can calculate the "predicted" age based on the year of interview and birth year. A respondent born in year B and interviewed in year Y and month M will have a predicted age of $\hat{A} = Y - B$. If $\hat{A}$ is one year greater than the reported age, A, we know that the respondent has not yet had his or her birthday during the current year.[16] This implies, given the month of interview, that the respondent was born in month M or later. Subsequently, if $\hat{A} = A$ then the birthday must have been in M or earlier. We assign the average probability of these possible birth months to the respondent (the month of interview is always possible but less likely relative to the other months and is thus included with a weight of one half).

In the second case, where we only know the age of the respondent, even the birth year is uncertain. Depending on whether the respondent has had his or her birthday in the months preceding the interview, which is unknown, the birth year will differ. Assume that a respondent with an age of A was born at a date preceding the interview date, then the birth year would be $Y - A$ with a birth month less than or equal to M. If the respondent was born at a later date, then the birth year is $Y - A - 1$ and the birth month greater than or equal to M. For the respondents where the most detailed birth information is age, we thus average over all 13 possible months (as above, month M in $Y - A$ and M in $Y - A - 1$ are less likely than the other months and are therefore each included with a weight of one half).[17] Whenever we average over several possible birth months, we adjust for birth seasonality differences at the regional level.

This leaves the final uncertainty—where the mother resided during the IOC programs. In general, we cannot observe the migration history of the mother or the respondent. We here follow the previous studies and assign treatment using the respondent's current place of residence. Inter-district migration could therefore pose a serious problem. A recent study of migration patterns in the Kagera region of Tanzania, however, found that while inter-district migration exists, most migration occurs within the district (Beegle et al. 2011). Although we cannot mitigate this problem, some of the data sources provide information that allows for a more in-depth investigation of its consequences, as presented in Section 5.

4 Results

The main results are presented in Tables 1, 2 and 3. All three tables follow the same structure. Each cell represents a separate regression. The reported statistics are the

[16] In the few cases with impossible values, for example if the reported and predicted ages differ by more than one year, we set the birth year as missing.

[17] This procedure differs from that in Field et al. (2009), in that they derive the birth year, for the THBS 2000 survey, as 2000 minus the current age and then calculating treatment probability by averaging over all months that year. Their procedure thereby neglects the information provided by the interview date. Most notably, it ignores that 25.5% of the sample was interviewed in 2001.

point estimates, clustered standard errors and the number of observations. Columns indicate different subsamples (females, males and both), and rows indicate the separate specifications. The estimated coefficients should be interpreted as the impact of ID protection in utero on educational attainment.

Table 1: Educational attainment by treatment specification

	All	Females	Males
Multi-Comp.	0.052	0.063	0.037
	(0.062)	(0.086)	(0.083)
Hyperbolic	0.090	0.169*	0.003
	(0.066)	(0.090)	(0.091)
Binary 12	-0.009	0.066	-0.088
	(0.076)	(0.106)	(0.103)
Binary 18	0.016	0.037	-0.010
	(0.064)	(0.087)	(0.084)
Binary 24	0.045	0.078	0.008
	(0.060)	(0.084)	(0.079)
Observations	9972	5016	4956

Notes: Each cell presents the result from a separate regression where the columns indicate different gender samples (female, male and both) and rows indicate different treatment specifications. The specific details for each specification are outlined in Section 3.3. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on the district-cohort level within parenthesis. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 1 presents the effects of ID protection on educational attainment (grades completed) across different depletion models. Overall, the estimated effects are slightly positive but close to zero. The preferred multi-compartment specification suggests that in utero protection from ID leads to an increase in completed grades by 0.05 grades, which is not statistically different from zero. The effects are slightly higher for females than for males, but the difference is not statistically or economically significant. The binary specifications differ somewhat from the multi-compartment specification but yield qualitatively similar estimates. By contrast, the hyperbolic specification produces large differences between males and females. The effect for males is almost exactly zero, while the effect for females is approximately three times larger than in the multi-compartment specification. Under the hyperbolic depletion model, the effect for females is also significant at the 10% level. Notably, the hyperbolic specification from Field et al. (2009) stands out as the only specification that produces any economically significant treatment effect, albeit still statistically insignificant at conventional levels.

Table 2 separates the estimation by data source. For these estimates, we use the multi-compartment depletion model. Positive effects are found using the samples from the THBS 2000 and NPS 2008 surveys, where they seem to be stronger for females. However, for the other three surveys, the effects tend to be negative but never significantly different from zero at conventional levels. Noteworthy is the NPS 2008 survey that, in light of the other surveys, stands out with considerably higher estimated effects, especially for females. The slight positive effect observed in the pooled sample with all

Table 2: Educational attainment by data source

	All	Females	Males
Pooled			
Full	0.052	0.063	0.037
	(0.062) [9972]	(0.086) [5016]	(0.083) [4956]
Excl. NPS 2008	0.023	-0.006	0.032
	(0.060) [9468]	(0.085) [4768]	(0.078) [4700]
By Survey			
DHS 1999	-0.055	-0.008	-0.084
	(0.092) [871]	(0.114) [452]	(0.131) [419]
THBS 2000	0.141**	0.162*	0.121
	(0.064) [5490]	(0.085) [2751]	(0.082) [2739]
DHS 2004	-0.133	-0.366*	0.124
	(0.125) [1895]	(0.201) [951]	(0.151) [944]
NPS 2008	0.685	1.583**	0.216
	(0.470) [504]	(0.620) [248]	(0.691) [256]
DHS 2010	-0.075	0.253	-0.357
	(0.207) [1212]	(0.334) [614]	(0.280) [598]

Notes: Each cell presents the result from a separate regression where the columns indicate different gender samples (female, male and both) and rows indicate different surveys. The two first rows present the results from a pooled analysis, either including or excluding the NPS 2008 survey. Each subsequent row is a separate survey. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

surveys is, in fact, to a large part driven by NPS 2008—particularly for females. While this survey only amounts to 5% of the total sample size, its exclusion, as presented in the second row of Table 2, changes the effect for females to almost exactly zero.

The lack of robustness between the different surveys is troubling. While this could be interpreted as an indication that the identification strategy is not completely sound, there are other possible explanations. As the surveys were conducted in different years, the estimates could in principle capture some true underlying heterogeneity in the treatment effect over time. This seems however highly unlikely in the current context. If ID protection increased cognitive ability we would not expect the sign of the effects to alternate between years in the way displayed, but instead change monotonically as the cohorts became older. Another possible explanation is potential differences in sampling methods across surveys, which would results in that the samples were drawn from different (but partially overlapping) populations. However, since all surveys were intended to be representative of the Tanzanian population at large and were conducted by the same agency, this would also seem unlikely to account for the observed pattern.

A third, and more likely, explanation is that the reduction in sample size that follows when estimating the effect for each survey separately is expected to decrease precision.[18]

[18]There are three reasons why precision declines with fewer individual observations, even when the standard errors are cluster-adjusted. The first is that the number of observations in some district-cohort cells is sometimes reduced to zero. The second is that treatment is assigned as a function of both birth information and interview date, which varies within cluster cells. The third is that the precision of the

We can therefore not disregard the possibility that the large variation, and particularly the surprisingly large effect in the NPS 2008 sample, is due to random sampling error. This would, however, not explain the significance levels observed in some of the disaggregated samples. While not at a remarkable level, they are slightly higher than expected by chance. This would indicate that we have underestimated the standard errors.[19] In any case, considering the large number of estimates, and the multiple comparison problem that follows, we are reluctant to read too much into the estimates for single surveys.

The variability between surveys, independent of reason, motivates us to focus on an aggregate sample as the interpretation as the average treatment effect over the included years is maintained. Although the NPS 2008 is an outlier, there are no a priori reasons for its exclusion. All surveys reported in Table 2 meet existing international reliability standards and have been widely used by scholars. Our preferred estimate is therefore based on the pooled sample using all surveys, which suggests that the overall the effect appears to be slightly positive but close to zero and statistically insignificant.

The final component of the main analysis concerns multiple outcomes. The data offer several variables that are closely related to educational attainment and, more generally, cognitive ability. The statistical advantage of educational attainment, relative to other proxies for cognitive skill, is that it varies at all age categories. However, equating an additional grade linearly makes interpretation somewhat delicate. Starting school or graduating from primary school are arguably more important educational advances than, for example, the transition from second to third grade—the current measure would equate them all. Furthermore, important educational milestones are arguably less affected by measurement error than educational attainment. Specifically, we analyze (1) whether the respondent ever attended primary school, (2) whether the respondent has completed primary school, and (3) whether the respondent ever attended secondary school. In addition, literacy is reported in a few surveys and is analyzed when available.[20] Since the hypothesized mechanism, by which educational attainment is affected, is that protection from ID increases cognitive ability (and not by an indirect effect through, e.g., improved health and attendance), the ability to read appears to be a more direct measure of that mechanism.

When using these alternative outcomes, the effects are again close to zero and insignificant, except for the effect on having some primary education, which is positive for females. The effect on completed primary school is negative for both females and males, albeit close to zero and statistically insignificant.

While it is occasionally customary to interpret significant effects on some outcomes

individual controls increases with the pooled sample, thus reducing within-cluster variation.

[19] We have, for example, not adjusted the standard errors for serial correlation (as discussed by Bertrand et al. (2004)), which to some degree would be relevant in the current setting.

[20] Approximately one third of the respondents in the DHS surveys were asked to read a short text out loud. The interviewers then judged the respondents' reading proficiency. In the THBS 2000 survey, revealed literacy was not measured; however, the respondents were asked whether they are able to read and write. The constructed literacy indicator takes the value one if the respondent can read in at least one language.

Table 3: The effect on additional outcomes

	All	Females	Males
Individual			
Highest Grade	0.052	0.063	0.037
	(0.062) [9972]	(0.086) [5016]	(0.083) [4956]
Some Primary	0.021*	0.037**	0.007
	(0.012) [10031]	(0.018) [5046]	(0.016) [4985]
Comp. Primary	-0.011	-0.010	-0.014
	(0.009) [9963]	(0.012) [5013]	(0.012) [4950]
Some Secondary	-0.001	0.002	-0.002
	(0.007) [9963]	(0.008) [5013]	(0.010) [4950]
Literacy	0.015	0.008	0.024
	(0.019) [6659]	(0.026) [3654]	(0.028) [3005]
Average effects			
AES, excl. lit.	0.011	0.024	-0.001
	(0.019) [9963]	(0.024) [5013]	(0.026) [4950]
AES, incl. lit.	0.018	0.018	0.025
	(0.022) [6575]	(0.030) [3609]	(0.033) [2966]

Notes: Each cell presents the result from a separate regression where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in this section) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

as evidence of a partial treatment effect on some margins, this practice is not always desirable. On the one hand, such interpretations suffer from multiple hypothesis bias which tend to overstate significance levels.[21] On the other hand, if the outcomes are not perfectly correlated, many insignificant results pointing in the same direction could be indicative of an underlying overall non-zero effect (in this case a possible underlying variable would be cognitive ability). Investigating each outcome separately would thus fail to detect the overall effect. Following the procedure proposed by Kling et al. (2007), we will estimate the *average effect size* (AES) and can thereby test for the joint effect of a family of outcomes.[22] This procedure has predominantly been used in designed randomized experiments (see, e.g., Duflo et al. 2007 and Clingingsmith et al. 2009). However, the procedure has recently spread to purely observational studies with several related outcomes present (see, e.g., Alesina et al. 2013). Its main advantage is a stronger power than the conventional joint significance test because it assumes that treatment affects the outcomes unidirectionally, which is reasonable in the current setting (i.e.,

[21] For instance, the probability that we observe at least one estimate significant at the 5% level among 10 (uncorrelated) estimates, corresponding to the five outcomes for female and males, when the null of no treatment effect is true is approximately 40%.

[22] In particular, the procedure estimates the average standardized effect, $\beta_{AES} = \frac{1}{N}\sum(\beta_n/\sigma_n)$, where β_n is the estimated effect of ID protection on outcome n, σ_n is the standard deviation of that outcome in the untreated population (here defined as having a treatment probability of less than 10%) and N is the total number of outcomes. In practice, the sample standard deviation of the outcomes are used and β_n is estimated using seemingly unrelated regression to account for possible error correlation across estimations. Refer to Kling and Liebman (2004) for a more detailed discussion.

that protection from ID in utero affects the outcomes in the same direction).

We construct two different families of outcomes, with and without literacy. The results are also presented in Table 3. The average effect size is close to zero. The effect of ID protection on the averaged educational outcome is between 0.01 and 0.02 standard deviations. The effect tends to be higher for females than males, but the difference is not statistically significant. Including literacy, and thereby reducing the sample, the average effect size estimate increases slightly, which is driven by an increase in the estimate for males. We conclude that the average educational effects are small and statistically insignificant. Furthermore, the precision of the estimated average effects is sufficiently large to exclude the possibility of economically significant effects (disregarding potential attenuation bias).

5 Investigating attenuation bias

5.1 Birth month

A potential issue with the current analysis, inherent to the available data, is that we do not observe exactly when the respondents are born. For the majority of the sample (69.4%), we only know the birth year or the respondent's age. This poses two problems. First, treatment must be assigned by averaging over several months, thus decreasing assignment precision. This is especially problematic because the probability of protection in certain years varies considerably (as seen in Figure 1). Second, the censoring prohibits us from using time fixed effects at a monthly level, which may be desirable from the perspective of both identification and precision. With the exception of the THBS 2000 survey (in which the most detailed level of birth information is age), a subset of the respondents has their birth month reported. This subsample is therefore not affected by lack of birth date precision and could be informative of the magnitude of these issues in the full sample.

Table 4: Investigating bias due to imprecise birth information

	All	Females	Males
Age only	-0.042	0.098	-0.114
	(0.135) [3050]	(0.169) [1661]	(0.209) [1389]
Birth month	-0.028	0.039	-0.064
	(0.118) [3050]	(0.153) [1661]	(0.193) [1389]

Notes: Each cell presents the result from a separate regression on the subsample of respondents that have reported birth month and year information, where the columns indicate different gender samples (female, male and both) and rows indicate different levels of used birth date information. The first row only utilizes the reported age when assigning treatment, and thereby disregards more detailed level of birth date information. The second row instead utilizes this information and assigns treatment by birth month. Birth month-year fixed effects are also included. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 4 presents the results when we exclude all respondents with missing birth month information. In a first analysis, we assign treatment as if birth year and month were unknown, thus we only consider the reported age in years. This therefore imitates the procedure used for the majority of the respondents in the excluded sample and is affected by both issues discussed above. The results from this specification are presented in the first row. In the next analysis, as presented in the second row, we instead utilize the full set of birth date information but keep the exact set of observations as in the first analysis. Treatment is now assigned by birth month and we include birth year-month fixed effects in the regressions. This specification is thereby rendered unaffected by the two problems arising from the cruder birth measures. Comparing the two specifications, we see that the one with more accurate birth information drives the estimates towards zero. If there is a non-zero true effect, this is unexpected as we would anticipate the effects to increase in absolute terms with higher treatment precision. While the results are consistent with a true effect of zero and improved identification provided by the monthly fixed effects, the remaining imprecision do not allow that conclusion.

Even if the evidence is only indicative, this analysis provides us with reason to believe that the absence of an estimated significant effect in the full sample is not driven by the incomplete birth information.

5.2 Migration

Another major issue is migration. The relevant information for treatment assignment is where the mother of the respondent lived during the IOC programs. This information is not available; instead, we assign treatment based on where the respondent currently resides. If the respondent migrated out of his or her district of birth, or if the respondent's mother migrated between IOC distribution and birth, then the treatment assignment would be incorrect and attenuate the estimated effect. While we expect imprecise treatment assignment to be the most important consequence of migration, treatment in itself could affect the propensity to migrate. For example, if, in a given district-cohort, the treated individuals are more likely to migrate then we would have a higher tendency to sample untreated respondents (both exist due to incomplete coverage). This would bias the estimates towards zero to an even greater extent than for the same level of "random" migration.

Furthermore, when considering the sampling design of the surveys, where established and stable households tend to be included, selective migration would be even more problematic, as movement within the district could also introduce bias. As an illustration, if treated individuals (or individuals who respond particularly well to treatment) enroll in boarding school to a higher degree, then they would be excluded from the surveys and bias the estimates downwards even without inter-district migration. In fact, without any assumptions on migration patterns, not even the direction of the bias is known. If treated individuals who do not respond well to treatment have a lower tendency to be sampled (for whatever reason), we would overstate the true effect. While this would

necessitate the existence of a non-zero treatment effect for some individuals, the true average effect could nonetheless be zero. In this section, we exploit more detailed information reported in some of the surveys to investigate how migration affects the results.

First, the DHS 1999, DHS 2004 and NPS 2008 surveys report the number of years the respondent has lived in the current place of residence.[23] For the respondents that never moved, we know that treatment assignment is unaffected by attenuation due to inter-district migration. However, selective migration remains unresolved (and is potentially exacerbated, as we now also exclude all intra-district migration).

Table 5: Investigating migration, excluding movers

	All	Females	Males
All	0.006	-0.032	0.066
	(0.103) [3270]	(0.142) [1651]	(0.151) [1619]
Excl. Movers	0.066	0.088	0.110
	(0.116) [2490]	(0.151) [1269]	(0.169) [1221]

Notes: Each cell presents the result from a separate regression on the full sample from the DHS 1999, DHS 2004 and NPS 2008 surveys and from the restricted sample of never-movers. The columns indicate different gender samples (female, male and both) and rows indicate the different samples based on migration history. The first row includes all respondent while the second row restricts the sample to non-movers. Besides the estimated coefficient for the treatment variable, each cell reports standard errors clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

As shown in Table 5, restricting the sample to non-movers increases the point estimates somewhat. While this indicates that attenuation due to imprecise treatment assignment could provide a partial explanation for the lack of evidence of an positive effect, it cannot explain it alone—the effects are only slightly greater than in the full sample and still insignificant. As noted, this analysis cannot account for selective migration. To investigate this issue, we must restrict our attention to the NPS 2008 survey, which contains additional information that can be used to characterize the nature and impact of migration in greater detail.

In the NPS 2008 survey, respondents that reported having moved since birth were asked in which district they were born. Disregarding recall bias, this survey provides nearly prefect district information for treatment assignment (the only remaining source of dilution is mothers migrating before birth). Unlike the previous analysis, where we restrict the sample to respondents that never moved, in this subsample we can assess the consequences of migration without imposing any further sample restrictions. While this may not entirely resolve the issue of selective migration, it would at least not be exacerbated. The first row in Table 6 presents the estimates when treatment is assigned using the current district of residence, as in the main analysis. The next row utilizes the provided birth district information. We see that the estimates increase, in line with a true positive effect, but only for males. This would lead us to believe that increase

[23] This is unreported for children younger than 12 years (NPS) or 15 years (DHS). In these cases we use the time the mother has lived in the current place of residence. If the information still is missing, we use the time the household has existed at the current place of residence.

for females found in Table 5, at least to some extent, could be explained by the added sample restrictions rather than increased precision. While the estimated effect is large and significant, the NPS 2008 survey is an outlier, as is apparent from Table 2, and we would hesitate to draw any definite conclusion based on it alone. Nevertheless, this analysis still indicates that attenuation bias caused by migration may exist, at least for males.

Table 6: Investigating migration, using reported birth district

	All	Females	Males
Curr. District	0.685	1.583**	0.216
	(0.470) [504]	(0.620) [248]	(0.691) [256]
True District	0.938**	1.529***	0.527
	(0.425) [527]	(0.553) [264]	(0.726) [263]

Notes: Each cell presents the result from a separate regression on the sample from the NPS 2008 survey. The columns indicate different gender samples (female, male and both) and rows indicate the different method of assigning treatment. In the first row treatment is assigned by where respondents currently resides, while in the second row by their reported birth district. Besides the estimated coefficient for the treatment variable, each cell reports standard errors clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

In a final analysis, we attempt to investigate how selective migration affects the estimates. In the NPS 2008 survey, women were asked to report any children of hers living outside the household, including their educational attainment. If some types of individuals are less likely to be included in the surveys, adding children outside the household to the analysis would mitigate the selective migration issue. However, it would not be a complete solution, as there could still be selectivity at the household level or in the women's reports.

Table 7: Investigating migration, including movers

	All	Females	Males
Residents Only	0.685	1.583**	0.216
	(0.470) [504]	(0.620) [248]	(0.691) [256]
Incl. Movers	0.579	0.964**	0.238
	(0.370) [679]	(0.462) [347]	(0.591) [332]

Notes: Each cell presents the result from a separate regression on the sample from the NPS 2008 survey. The columns indicate different gender samples (female, male and both) and rows indicate specification changes. In the first row we use the sample as in the main analysis, while in the second row we include children living outside the household (as reported by mothers). Besides the estimated coefficient for the treatment variable, each cell reports standard errors clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

The first row in Table 7 focuses on those currently residing in the household (these are the same estimates reported in the baseline results). The second row also includes reported children residing elsewhere. The estimates for males increase only slightly, indicating that selective migration does not pose a large problem for this group. For

females, however, the estimate decreases substantially, to nearly 60% of the estimate in the first row. Thus, selective migration seems to overestimate the IOC protection effect for females. A speculative interpretation consistent with these results is that women who respond well to treatment leave the household later than otherwise (e.g., postponing marriage due to studies), resulting in that this type of women are oversampled, while this effect is not present for men. It is, however, important to note that the NPS 2008 is one of the later surveys in which mobility is highest. Thus it is unclear how large these problems are in earlier surveys.

In summary, we find some indications that are consistent with the existence of a positive treatment effect and attenuation bias. However, any eventual bias does not seem sufficiently large to alone explain the lack of significant effects in the main analysis. While the remaining uncertainties could nonetheless be influential (which ultimately is unknown given the current data), the factors analyzed in this section do not change the overall picture painted by the main analysis.

6 Concluding remarks

In this paper, we revisit the Tanzanian experience with iodized oil capsule distribution and ask whether increased protection from iodine deficiency in utero in goiter-prevalent districts increases educational attainment and literacy in the long-run. We find no evidence of increased educational attainment among these cohorts.

Numerous medical studies have established the existence of a causal effect of prenatal iodine deficiency on cognitive ability among children. The unresolved questions are to what extent the natural shortage of iodine in the soil in Africa has severely impeded educational attainment at the national level and to what extent the policy instruments undertaken to address this shortage contributed to the subsequent gains in educational attainment. If such a link exists, we would be able to cite evidence of how the natural health environment affects human development in general (consistent with, e.g., Sachs 2001 and Diamond 1997). It is thus the "field relevance" of iodine deficiency that we consider is the open question, not whether the marginal overall benefit of protection against iodine deficiency is strictly positive, all else being equal.

There are at least three explanations for why an effect might not exist for longer-run outcomes, while maintaining the existence of a short-run effect as established in the medical literature. The first is that the documented effects of iodine supplements on cognitive ability do not persist until the school-going age. The second explanation is that for the specific outcomes examined here, and at the margins of possible improvement, sufficient iodine after pregnancy has a compensatory effect. ID was nearly eradicated in Tanzania during the period of early childhood of these cohorts, due to the salt iodization programs (Assey et al. 2009). As this is a potential compensating factor for ID during pregnancy, it could overshadow the potential effect that would have been observed if ID had remained endemic. The last explanation is that the educational system in Tanzania does not capture cognitive ability; grades may reflect other fundamentals. However, the

lack of an observed effect using literacy and other proxies for schooling indicate that the last explanation is unlikely.

In addition, we highlight numerous elements of uncertainty concerning treatment assignment in existing work, which could potentially mask a true positive effect of the programs. We attempt to address these uncertainties by using additional data, better information on the timing and place of birth, alternative outcomes, and a more context-specific iodine depletion function. When combining these measures to maximize precision, we still do not find support for an effect of ID protection on cognitive ability among the samples exposed to treatment. In particular, the differential impact across males and females, advanced in the past literature, arises in but a limited number of specifications, which is difficult to rationalize ex ante. While a possible explanation for the lack of significant treatment effects is the imprecision in the treatment assignment variable, it is, however, unlikely to be entirely explained by imprecision due to post-treatment migration or imprecise birth information.

We conclude that the available data together with the identification strategy employed in this study, and the previously studies, are unable to establish the effect of the IOC programs. While there are indicative evidence pointing towards positive effects, the lack of significance and robustness does not motivate any such conclusions. However, the epigram "the absence of evidence is not the evidence of absence" should in this context be remembered. Studies from the medical literature would make us suspect potentially large effects, which might very-well have been the case. The analyses using the timing of the IOC programs as identification do, however, not provide additional support for that proposition—and we must conclude that we ultimately do not know.

References

Adhvaryu, A. and A. Nyshadham (2011) "Endowments and investments within the household: Evidence from iodine supplementation in Tanzania,"Technical report, Discussion Paper, Economic Growth Center.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013) "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, Vol. 128, No. 2, pp. 469–530.

Almond, Douglas and Janet Currie (2011) "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, Vol. 25, No. 3, pp. 153–72.

Assey, Vincent, Stefan Peterson, Sabas Kimboka, Daniel Ngemera, Celestin Mgoba, Deusdedit Ruhiye, Godwin Ndossi, Ted Greiner, and Thorkild Tylleskar (2009) "Tanzania national survey on iodine deficiency: impact after twelve years of salt iodation," *BMC Public Health*, Vol. 9, No. 1, p. 319.

Becker, Gary S. and Nigel Tomes (1976) "Child Endowments and the Quantity and

Quality of Children," *Journal of Political Economy*, Vol. 84, No. 4, pp. pp. S143–S162.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon (2011) "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," *The Review of Economics and Statistics*, Vol. 93, No. 3.

Bertelsen, Jette B. and Laszlo Hegedüs (1994) "Cigarette Smoking and the Thyroid," *Thyroid*, Vol. 4, No. 3, pp. pp. 327–331.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, Vol. 119, No. 1, pp. 249–275.

Bourdoux, P., F. Delange, M. Gerard, M. Mafuta, A. Hanson, and A. M. Ermans (1978) "Evidence that Cassava Ingestion Increases Thiocyanate Formation: A Possible Etiologic Factor in Endemic Goiter," *Journal of Clinical Endocrinology & Metabolism*, Vol. 46, No. 4, pp. pp. 613–621.

Cao, Xue-Yi, Xin-Min Jiang, Zhi-Hong Dou, Murdon Abdul Rakeman, Ming-Li Zhang, Karen O'Donnell, Tai Ma, Kareem Amette, Nancy DeLong, and G. Robert DeLong (1994) "Timing of Vulnerability of the Brain to Iodine Deficiency in Endemic Cretinism," *New England Journal of Medicine*, Vol. 331, No. 26, pp. 1739–1744.

Chan, Shiao Y., Marcus H. Andrews, Rania Lingas, Chris J. McCabe, Jayne A. Franklyn, Mark D. Kilby, and Stephen G. Matthews (2005) "Maternal nutrient deprivation induces sex-specific changes in thyroid hormone receptor and deiodinase expression in the fetal guinea pig brain," *The Journal of Physiology*, Vol. 566, No. 2, pp. 467–480.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers (2006) "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *The Journal of Economic Perspectives*, Vol. 20, No. 1, pp. pp. 91–116.

Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer (2009) "Estimating the Impact of The Hajj: Religion and Tolerance in Islam's Global Gathering," *The Quarterly Journal of Economics*, Vol. 124, No. 3, pp. 1133–1170.

Diamond, Jared (1997) *Guns, germs, and steel*: Norton New York.

Duflo, Esther, Rachel Glennerster, and Michael Kremer (2007) "Using randomization in development economics research: A toolkit," *Handbook of development economics*, Vol. 4, pp. 3895–3962.

Dugbartey, Anthony T. (1998) "Neurocognitive aspects of hypothyroidism," *Archives of Internal Medicine*, Vol. 158, No. 13, pp. 1413–1418.

Field, Erica, Omar Robles, and Maximo Torero (2009) "Iodine Deficiency and Schooling Attainment in Tanzania," *American Economic Journal: Applied Economics*, Vol. 1, No. 4, pp. 140–69.

Friedhoff, Arnold J, Jeannette C Miller, Mary Armour, Jack W Schweitzer, and Sandhya Mohan (2000) "Role of maternal biochemistry in fetal brain development: effect of maternal thyroidectomy on behaviour and biogenic amine metabolism in rat progeny," *The International Journal of Neuropsychopharmacology*, Vol. 3, No. 2, pp. 89–97.

Fuge, Ronald and Christopher C. Johnson (1986) "The geochemistry of iodine – a review," *Environmental Geochemistry and Health*, Vol. 8, No. 2, pp. 31–54.

Furne, C A, G A Pfann, C E West, F van der Haar, D van der Heide, and J G Hautvast (1995) "New model for describing urinary iodine excretion: its use for comparing different oral preparations of iodized oil.," *The American Journal of Clinical Nutrition*, Vol. 61, No. 6, pp. 1257–1262.

Gaitan, Eduardo (1990) "Goitrogens in Food and Water," *Annual Review of Nutrition*, Vol. 10, pp. pp. 21–37.

van der Haar, Frits, P Kavishe, and Mehari Gehbri Medhin (1988) "The public health importance of IDD in Tanzania," *The Central African Journal of Medicine*, Vol. 34, No. 3.

Haddow, James E., Glenn E. Palomaki, Walter C. Allan, Josephine R. Williams, George J. Knight, June Gagnon, Cheryl E. O'Heir, Marvin L. Mitchell, Rosalie J. Hermos, Susan E. Waisbren, James D. Faix, and Robert Z. Klein (1999) "Maternal Thyroid Deficiency during Pregnancy and Subsequent Neuropsychological Development of the Child," *New England Journal of Medicine*, Vol. 341, No. 8, pp. 549–555.

Hassanien, Mohammed H, Laila A Hussein, Erica N Robinson, and L Preston Mercer (2003) "Human iodine requirements determined by the saturation kinetics model," *The Journal of Nutritional Biochemistry*, Vol. 14, No. 5, pp. 280–287.

Ingenbleek, Y, L Jung, G Frard, F Bordet, A M Goncalves, and L Dechoux (1997) "Iodised rapeseed oil for eradication of severe endemic goitre," *The Lancet*, Vol. 350, No. 9090, pp. 1542–1545.

Kling, Jeffrey R. and Jeffrey B. Liebman (2004) "Experimental Analysis of Neighborhood Effects on Youth,"Technical report.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007) "Experimental Analysis of Neighborhood Effects," *Econometrica*, Vol. 75, No. 1, pp. 83–119.

Lavado-Autric, Rosala, Eva Aus, Jos Victor Garca-Velasco, Mara del Carmen Arufe, Francisco Escobar del Rey, Pere Berbel, and Gabriella Morreale de Escobar (2003)

"Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny," *The Journal of Clinical Investigation*, Vol. 111, No. 7, pp. 1073–1082.

Ministry of Education and Culture, The (1995) *Education and Training Policy*, The United Republic of Tanzania, Dar es Salaam.

——— (2002) *Basic Education Statistics in Tanzania 1998-2002*, The United Republic of Tanzania, Dar es Salaam.

Murcia, Mario, Marisa Rebagliato, Carmen Iiguez, Maria-Jose Lopez-Espinosa, Marisa Estarlich, Beln Plaza, Carmen Barona-Vilar, Mercedes Espada, Jess Vioque, and Ferran Ballester (2011) "Effect of Iodine Supplementation During Pregnancy on Infant Neurodevelopment at 1 Year of Age," *American Journal of Epidemiology*, Vol. 173, No. 7, pp. 804–812.

Peterson, Stefan (2000) *Controlling Iodine Deficiency Disorders: Studies for Program Management in Sub-Saharan Africa*: Uppsala Dissertations from the Faculty of Medecine 943.

Peterson, Stefan, Vincent Assey, Birger Carl Forsberg, Ted Greiner, Festo P. Kavishe, Benedicta Mduma, Hans Rosling, Alfred B. Sanga, and Mehari Gebre-Medhin (1999) "Coverage and cost of iodized oil capsule distribution in Tanzania," *Health Policy and Planning*, Vol. 14, No. 4, pp. 390–399.

Pharoah, P.O.D., I.H. Buttfield, and B.S. Hetzel (1971) "Neurological Damage to the Fetus Resulting from Severe Iodine Deficiency During Pregnancy," *The Lancet*, Vol. 297, No. 7694, pp. 308–310.

Pharoah, POD and KJ Connolly (1987) "A Controlled Trial of lodinated Oil for the Prevention of Endemic Cretinism: A Long-Term Follow-Up," *International journal of epidemiology*, Vol. 16, No. 1, pp. 68–73.

Pop, Victor J., Evelien P. Brouwers, Huib L. Vader, Thomas Vulsma, Anneloes L. Van Baar, and Jan J. De Vijlder (2003) "Maternal hypothyroxinaemia during early pregnancy and subsequent child development: a 3-year follow-up study," *Clinical Endocrinology*, Vol. 59, No. 3, pp. 282–288.

Pop, Victor J., Johannes L. Kuijpens, Anneloes L. van Baar, Gerda Verkerk, Maarten M. van Son, Jan J. de Vijlder, Thomas Vulsma, Wilmar M. Wiersinga, Hemmo A. Drexhage, and Huib L. Vader (1999) "Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy," *Clinical Endocrinology*, Vol. 50, No. 2, pp. 149–155.

Rosenzweig, Mark R. and Kenneth I. Wolpin (1980) "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, Vol. 48, No. 1, pp. pp. 227–240.

Sachs, Jeffrey D. (2001) "Tropical Underdevelopment," Working Paper 8119, National Bureau of Economic Research.

Untoro, Juliawati, Werner Schultink, Clive E West, Rainer Gross, and Joseph GAJ Hautvast (2006) "Efficacy of oral iodized peanut oil is greater than that of iodized poppy seed oil among Indonesian schoolchildren," *The American Journal of Clinical Nutrition*, Vol. 84, No. 5, pp. 1208–1214.

Vanderpas, J B, B Contempr, N L Duale, W Goossens, N Bebe, R Thorpe, K Ntambue, J Dumont, C H Thilly, and A T Diplock (1990) "Iodine and selenium deficiency associated with cretinism in northern Zaire.," *The American Journal of Clinical Nutrition*, Vol. 52, No. 6, pp. 1087–93.

Wolff, J (2001) "Physiology and pharmacology of iodized oil in goiter prophylaxis," *Medicine*, Vol. 80, No. 1, pp. 20–36.

World Health Organization (1992) "National Strategies for Overcoming Micronutrient Malnutrition.," Unpublished.

World Health Organization, ICCIDD, and UNICEF (2001) "Assessment of the Iodine Deficiency Disorders and monitoring their elimination,"Technical report, WHO publication, WHO/NHD/01.1.

Zimmermann, M, P Adou, T Torresani, C Zeder, and R Hurrell (2000) "Iron supplementation in goitrous, iron-deficient children improves their response to oral iodized oil," *European Journal of Endocrinology*, Vol. 142, No. 3, pp. 217–223.

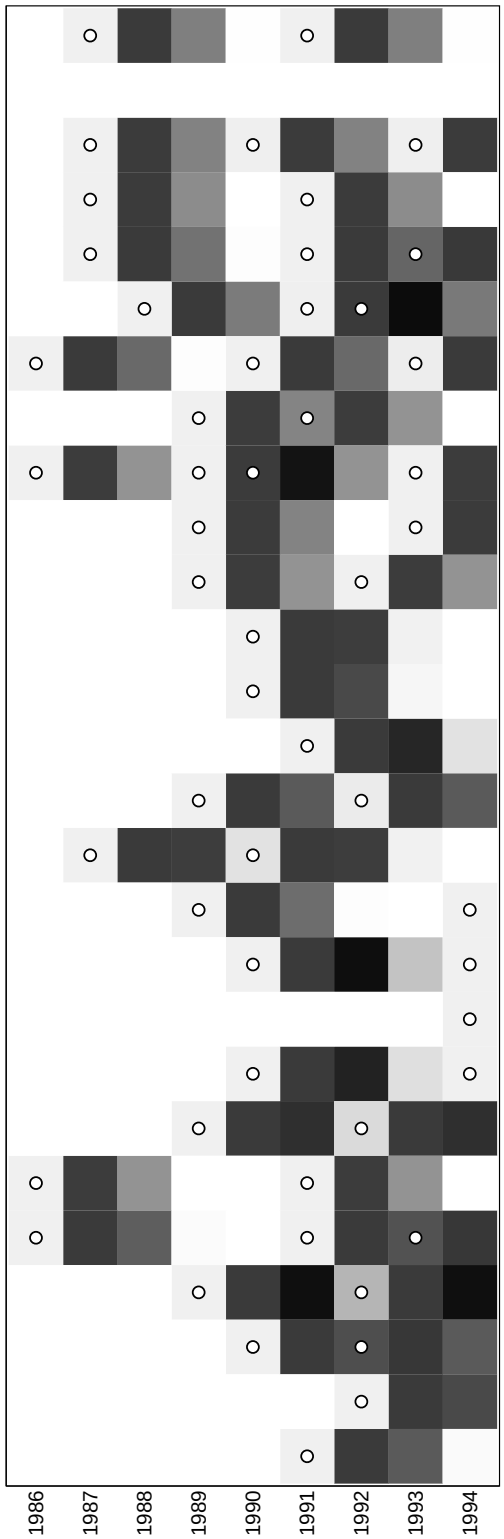
Zimmermann, Michael B (2009a) "Iodine deficiency," *Endocrine reviews*, Vol. 30, No. 4, pp. 376–408.

——— (2009b) "Iodine deficiency in pregnancy and the effects of maternal iodine supplementation on the offspring: a review," *The American Journal of Clinical Nutrition*, Vol. 89, No. 2, pp. 668–672.

Zoeller, R. T. and J. Rovet (2004) "Timing of Thyroid Hormone Action in the Developing Brain: Clinical Observations and Experimental Findings," *Journal of Neuroendocrinology*, Vol. 16, No. 10, pp. 809–818.

A Treatment assignment by year and district

Figure 2: Average treatment probability by year and district



Notes: Each cell presents the assigned treatment probability, using the multi-compartment model, by birth month for each year and district (averaged over the months of that year). The vertical axis denotes each year from 1986 to 1994, and the horizontal axis denotes each district. The background color in each cell represent the treatment probability, where white is 0% probability, black is 100% and all shades of gray the proportional values in-between the two extremes. A white circle designate that an IOC program commenced in the indicated district and year. District (9) and (26) are the two districts with late programs that are excluded from the main analysis. The numbers denotes the following districts: (1) Arumeru (2) Monduli (3) Mpwapwa (4) Ludewa (5) Makete (6) Mufindi (7) Njombe (8) Biharamulo (9) Bukoba Rural (10) Karagwe (11) Ngara (12) Kasulu (13) Kibondo (14) Kigoma Rural (15) Rombo (16) Chunya (17) Ileje (18) Kyela (19) Mbeya Rural (20) Mbozi (21) Rungwe (22) Ulanga (23) Mpanda (24) Nkansi (25) Sumbawanga Rural (26) Mbinga (27) Songea Rural.

B Comparison with previous studies

Where Field et al. (2009) estimate large and highly significant effects, we observe effects that tend to be low and insignificant. For the regressions run using the same data source (THBS 2000), we find marginally statistically significant effects but they are generally much lower (particularly for females). Our identification and data sampling strategy follows theirs closely, but there are some notable differences. In this section, we detail exactly what data and model decisions explain these differences.

In sum, the diverging results are primarily explained by Field et al.'s (2009, hereafter FRT) decision to exclude the 1986 cohort and a non-negligible number of observations due to missing values in background controls. Most notable of these variables is the exclusion of households without a spouse present. If we rerun the analysis with the exact specification used by FRT but include the cohort born in 1986 and households without a spouse, as presented in Column 8 in Table 8, the effects are more than halved and not significant at conventional levels.

In greater detail, Table 8 presents the estimated effects from ten specifications. Beginning with our preferred specification, we move, step-by-step, towards the specification used by FRT. We restrict our attention to the cohorts born from 1986 to 1990 from the THBS 2000 survey, which roughly correspond to the sample used by FRT (net of the additional sample restrictions discussed below). In the first specification, in the first column, we use the same specification as in the main analysis—the only difference is the added cohort restriction (refer to Appendix E for a discussion of this restriction). The last column corresponds exactly to the specification used by FRT and thus exactly replicates their results. All columns in-between gradually introduce changes in the specifications to highlight the important differences. All regressions, including those in the tenth column, are run on data sets we constructed using the original data files to the greatest extent possible. We also reconstructed the exact treatment specification used in FRT. This procedure ensures that we do not make any unintended changes when changing specifications. Our reconstructed data set has been compared, observation by observation, to the data set constructed by FRT. All variables are equal between the sets, except for the treatment variable. The difference is, however, extremely small (on average a difference of less than 2×10^{-8}, and the maximum difference is 1.2×10^{-7}) and is most likely due to differences in the floating point data structure in the softwares used for the analyses.

The first specification change, presented in column two in Table 8, addresses some minor differences between our data strategy and that of FRT. First, the specification only considers reported age, thus sets the birth year at 2000 minus the current age, as in FRT. The information provided by the date of the interview is thus disregarded, and treatment is not assigned as precisely. Second, the definition of the education variables is changed using the THBS 2000 definition of attainment (where non-standard grades are coded as full grades). Additionally, where in the main analysis we use indicators for different levels of the education level of the household head and spouse, in this

Table 8: Stepwise specification change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All	0.195*	0.118	0.173*	0.107	0.120	0.147	0.174	0.151	0.236*	0.347**
	(0.104)	(0.097)	(0.096)	(0.115)	(0.116)	(0.113)	(0.119)	(0.115)	(0.130)	(0.148)
	[3040]	[3017]	[3017]	[2436]	[2149]	[2149]	[2056]	[2056]	[1705]	[1395]
Females	0.183	0.062	0.095	0.162	0.187	0.219	0.252	0.245*	0.329**	0.594***
	(0.160)	(0.154)	(0.151)	(0.166)	(0.156)	(0.151)	(0.155)	(0.147)	(0.158)	(0.170)
	[1522]	[1512]	[1512]	[1194]	[1055]	[1055]	[1009]	[1009]	[839]	[678]
Males	0.190	0.162	0.226*	0.068	0.079	0.107	0.131	0.095	0.169	0.190
	(0.137)	(0.135)	(0.131)	(0.148)	(0.159)	(0.156)	(0.164)	(0.158)	(0.179)	(0.160)
	[1518]	[1505]	[1505]	[1242]	[1094]	[1094]	[1047]	[1047]	[866]	[717]

Notes: Each cell presents the result from a separate regression where the columns indicate different specifications and rows indicate different gender samples (female, male and both). The first column uses the specification from the main analysis, while the last column uses the exact specification used in FRT. All intermediate columns stepwise change the specification to highlight important differences (as detailed in the text). Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

specification we use the number of completed grades linearly, as in FRT. Finally, we add family controls—birth order and number of children—which (as discussed in Section 3.2) might be endogenous. As indicated in the second column, the estimated effects are reduced due to these changes, suggesting that they are not important in explaining the observed difference.

The next change (column three) changes the treatment specification to the exact specification used in FRT. An increase in the estimates is observed, which is expected since their treatment specification very closely resembles the hyperbolic depletion model, as presented in Table 1. However, the results do not differ substantially from the estimates obtained with the multi-compartment model, leading us to conclude that the treatment specification alone cannot explain the differences.

In column four, we restrict the sample to children of the household head and spouse. FRT argue that matching children to their mothers is essential, motivated by mitigating problems with migration and the possibility that young mothers received a lower IOC dose. We are, however, not aware of a procedure stipulating that young mothers were to receive a lower dose and, as discussed by FRT, that part of the protocol was likely abandoned before any program commenced. Although the sample resulting from this restriction arguably is less mobile than other children in the household, it implies that the analysis is restricted to a particular subset of individuals. The manipulation creates a notable difference between boys and girls, which is difficult to rationalize. The resulting estimates almost double for females and shrinks to less than a third for males, although the estimates are not statically different from zero. This change in the estimated effect seems however to be confined to the THBS 2000 survey, as it is not present when the same sample restriction is imposed on the pooled sample (as presented in Appendix D).

The next step—column five—further reduces the sample by dropping all children that cannot be directly matched to a unique mother. As the previous step restricted the sample to children of the household head or spouse, this restriction would only be motivated if we would deem the characteristics of the mothers to be important (e.g., their age). Children are matched to a woman in the household if she is the only woman within a reasonable age span among the household head and spouses. Implicitly, this restriction produces a rather particular subsample, in that all households with a single father and nearly all polygamous households are dropped. Continuing to column six, we also add the controls that are made possible due to matching to mothers, namely an indicator of whether the mother was younger than 23 at birth and that indicator interacted with the treatment "correction factor" (as detailed in FRT). By dividing this specification change into two steps, we can investigate the importance of the implied sample restriction and the included controls separately. We see that, while both changes increase the estimates slightly, they do not differ qualitatively from the main analysis.

Columns seven and eight investigate the influence of the five household level control variables used by FRT that, for some respondents, have unreported values and are possibly endogenous.[24] They are included in two steps. First (column seven), we impose

[24]The variables are indicators of food security, home ownership and house quality, and distance to

the sample restriction implied by their inclusion, but without actually adding them to the regression. Then, in a second step, we include them in the analysis, as presented in column eight. This, again, enables us to separate the changes due to the implicit sample restriction from the changes due to the inclusion of the controls. As seen, the sample restriction tends to increase the estimates while the controls themselves tend to lower the estimated effects slightly. While these estimates are slightly higher than those in the main analysis for the relevant cohorts, they would still not support the rejection of no effect. Notably, the precision of the estimates changes only very slightly with the inclusion of the additional controls.

The current specification is very close to that of FRT; the last two columns of Table 8 investigate the remaining differences. In all analyses, controls of the educational level of the household head and spouse are included, as these are strong predictors of the child's educational attainment. When there is no spouse in the household, the educational level of the spouse is unreported. If not properly attended to, including this control variable would, implicitly, restrict the sample in a rather non-obvious way. In effect, this would exclude all households with a single mother (all single-father households are already dropped due to the matching to mothers). In column nine, we drop all respondents residing in a household without a spouse. Observe that the control variable of the educational level of the spouse is included in both specifications (but with an indicator for missing values in column eight), the only difference is thus the sample restriction. This turns out to have a large effect on the estimates, leading to an increase of more than 50%.

In the last column, we exclude those born 1986 (i.e., those assigned an age of 14) from the analysis and thereby arrive at the exact specification used by FRT. This restriction nearly doubles the estimates for girls, from an already high level, while the estimates for boys only increase slightly. FRT explain the focus on those born between 1987 and 1990 by observing that ages 10-13 are "the modal age of enrollment."[25] While this motivation appears reasonable, the lack of robustness is troubling. If a lack of variation in the dependent variable were the main concern, this manipulation would primarily affect precision, not the point estimates. Moreover, on closer examination, the variation in educational attainment for the cohort born 1986 is sufficiently large to warrant inclusion. The 1986 cohort average is 4.1 grades with a standard deviation of 2.6, compared to an average of 3.2 grades with a standard deviation of 1.8 for the 1987 cohort.

While each of the discussed steps appear to have sound motivations, many have unexpected implications that restrict the analysis to a particular subsample of children. In the end, the analysis in FRT effectively draws inference to a population consisting of only sons and daughters of household heads in non-polygamous households where both the mother and father are alive and present. While one could argue that this is an

nearest secondary education and health clinic.

[25] In Appendix I, we estimate the effect for different age groups, including those aged 10-13, using our preferred specification. The overall conclusion remains unchanged also in light of this analysis.

interesting subpopulation, we still face with issue of the 1986 cohort. With a specification that only differs from FRT in the, arguably, two least controversial aspects—the inclusion of the 1986 cohort and households without a spouse—the estimates do not allow for the rejection of zero effects. In particular, the differential impact across genders appears highly non-robust and seems to be driven to a large extent by the exclusion of the 1986 cohort.

C The multi-compartment depletion model

We here detail the multi-compartment depletion model used in our preferred specification. Let I_t denote the total amount of stored iodine in the body in month t. This amount can be stored in either of two compartments (or types of compartments).[26] The first compartment represents the pool of stored iodine in follicular cells in the thyroid. This compartment has a good ability to retain iodine over a long period of time, which is modeled with a low depletion rate. However, the thyroid can only store a limited amount of iodine, which we set to 15 mg based on its estimated maximum storage capacity (Hassanien et al. 2003). The second compartment represents all other storage mechanisms in the body, which are less suited to the purpose and thus modeled with a higher depletion rate. We assume that iodine can be transported between the compartments freely and without cost,[27] but because the thyroid is the preferred location of storage, it will be filled first. The second compartment is assumed to have unlimited capacity (which is reasonable in the relevant interval considering the high depletion rate).

In addition to depletion, iodine stores are affected by consumption and intake. The thyroid can maintain a euthyroid state when at least 50 μg of iodine can be utilized daily to synthesize thyroid hormones (Zimmermann 2009a). In the current context, a reasonable approximation is that *at most* 50 μg per day (or 1.5 mg per month) will be used if the stores last. This implies that monthly consumption is given by $C_t = \min(I_t, 1.5)$. Intake is dietary iodine from food consumption and any eventual IOC supplements. Their sum in mg on a monthly basis is denoted N_t. Approximately 30% of orally administered iodine is subject to instantaneous fecal excretion (Hassanien et al. 2003), and thus never enters the blood stream, implying that $0.7N_t$ mg are added to the current stores in t.

We set the first compartment to deplete exponentially with a half-life of 12 months, corresponding to a monthly depletion factor of 0.056.[28] The second compartment depletes at a very high rate, some studies indicate a half-life of less than one month (Wolff 2001). We will set the depletion rate to exactly one month, or a monthly depletion factor of 0.5. Thus the iodine retained in the first compartment from t to the following month is $0.944 \min(I_t - C_t, 15)$, while $0.5 \max(I_t - C_t - 15, 0)$ is retained in the second. This yields the following process of the iodine storage and consumption:

$$\begin{aligned} I_t &= 0.944 \min(I_{t-1} - C_{t-1}, 15) + 0.5 \max(I_{t-1} - C_{t-1} - 15, 0) + 0.7N_t \\ C_t &= \min(I_t, 1.5) \end{aligned}$$

[26] While we model the stored iodine with two compartments, the flows would be described by a semi-exponential, three-compartment model due to how we model consumption.

[27] Under normal circumstances, the thyroid is limited in how much iodine it can trap per day, which would motivate us to also limit monthly transport ability. However, iodine deficiency induces the pituitary gland to produce thyroid stimulating hormones that increases uptake. Thus given the upper limit of 15 mg and the availability in the second compartment, modeling the transport, on a monthly basis, as unlimited is reasonable.

[28] A higher depletion rate is often discussed for this compartment (see, e.g., Wolff 2001). However, this includes consumption, which we model separately.

Notice that if uptake is less than the required 1.5 mg, the stored iodine will diminish over time until the stores are completely depleted, in which case consumption will only consist of the current uptake. Thus in steady state (where $I_t = I_{t-1}$), consumption and uptake must be equal if consumption is less than 1.5 mg. We will use this fact to derive the baseline iodine intake level through the goiter surveys. In addition to assuming that the population is in steady state at the time of observation, we will assume that the goiter rate is proportional to the average iodine uptake. Thus if there is virtually no goiter (induced by iodine deficiency) in a given population, uptake must be at least 1.5 mg per month (i.e., an intake of at least 2.14 mg); if the entire population suffers from goiter, we assume average uptake is so low that it can be approximated by zero. Consequently, average monthly intake in district d would be $2.14(1 - g_d)$, where g_d is the goiter rate in the district.[29]

During pregnancy, iodine consumption is increased. One could therefore argue that the limit of 1.5 mg is not an adequate level in order to offer complete protection for the fetus. The necessary level of *stored* iodine has, to our knowledge, not been studied. Required *intake*, however, has been studied extensively. Based on the ratio between the recommended intake for pregnant women and the recommend intake levels for adults of 4/3 (World Health Organization et al. 2001), we will consider women with a calculated level of stored iodine higher than 2 mg to offer full fetal protection, while women with stores between 1.5 and 2 mg offer only partial protection (proportionally within that interval).[30] As detailed in Section 3.3, we denote the level of fetal protection offered in month t^* with $S(t^*, M_1, \cdots, M_K)$, where $M_1, \cdots, M_K$ describe the history of IOC administration in the district. We thereby have:

$$S(t^*, M_1, \cdots, M_K) = \begin{cases} 1 & \text{if } I_{t^*} \geq 2 \\ (I_{t^*} - 1.5)/0.5 & \text{if } 2 > I_{t^*} \geq 1.5 \\ 0 & \text{else} \end{cases} \tag{2}$$

where the stored iodine, I_{t^*}, is given by the complete iodine intake process prior to t^* implied by $M_1, \cdots, M_K$ and the relevant g_d (i.e., N_t in $t \leq t^*$).

[29] If goiter rates are very low, we only know that intake is at least 2.14 mg, naturally it could be higher. This is not relevant in the current context, as the lowest goiter rate among the treated districts is 28%. Thus no district falls outside the interval that can be predicted by the formula.

[30] The cut-off levels used by Field et al. (2009)—full protection above 6.5 mg of stored iodine and partial protection down to 4.2 mg—are based on the supposedly daily recommended iodine intake for pregnant women of 1.4 to 2.1 mg per day. However, they do not provide the source of these recommendations. Notably this level is 10 times the intake of 200 μg per day that is recommended for pregnant women by the World Health Organization et al. (2001).

D Results for children of head or spouse

Table 9: Effects on the subsample of children of the household head or spouse

	All	Females	Males
Individual			
Highest Grade	0.049	0.104	0.010
	(0.068) [7782]	(0.092) [3776]	(0.092) [4006]
Some Primary	0.019	0.037**	0.005
	(0.013) [7841]	(0.018) [3805]	(0.019) [4036]
Comp. Primary	-0.009	-0.011	-0.006
	(0.010) [7780]	(0.013) [3776]	(0.013) [4004]
Some Secondary	0.002	0.003	0.003
	(0.008) [7780]	(0.009) [3776]	(0.012) [4004]
Literacy	0.015	0.016	0.015
	(0.021) [5240]	(0.028) [2760]	(0.028) [2480]
Average effects			
AES, excl. lit.	0.014	0.029	0.006
	(0.021) [7780]	(0.026) [3776]	(0.030) [4004]
AES, incl. lit.	0.008	0.003	0.038
	(0.025) [5165]	(0.033) [2720]	(0.035) [2445]

Notes: Each cell presents the result from a separate regression on the subsample of children to the household head or spouse, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

E Results for cohorts born in the 1986-1990 period

Our baseline specification includes all cohorts born between 1986 and 1994, with the purpose of maximizing the sample size. There are, however, two caveats that speak against the inclusion of the later cohorts. First, as previously discussed, the salt iodization programs had already begun in some districts by the early 1990s (which districts that were targeted is not documented). During this time, the program gradually shifted from targeting districts that suffered from particularly high goiter rates to complementing salt iodization efforts (Peterson et al. 1999). Therefore, in the later cohorts, the respondents not affected by the IOC programs could still be protected from ID through their mothers' salt consumption. As salt iodization is often preferred to IOC distribution, due to iodine intake in small doses continuously over time and higher coverage, the worst case scenario is that our treatment variable actually captures the inverse probability of protection from iodine for these cohorts—where the "untreated" are better protected from ID. However, since this only applies to a few districts, or parts of districts, the problem should not be particularly severe. Second, in the period 1990-1994, iodine programs had been initiated in all investigated districts and were repeated on a more regular basis. Thus, for the cohorts born in 1991 and later, the absolute majority is considered treated (see Appendix A) and the effective control group (with respect to the birth fixed effects) is very small.

In this appendix, we therefore present the results using only the 1986-1990 cohorts. The effect now tends to be negative but close to zero for females (with the exception of the hyperbolic specification). For males, they tend to be negative to a degree that would be economically significant, but they remain statistically insignificant. It is not obvious how to interpret this change. On the one hand, the reduced sample is a priori preferred, as there is less risk of contamination of the control group. This would indicate that the estimates in the main analysis overstate the true effect, which might even be negative. On the other hand, the standard errors are larger due to the reduced sample size, and in all cases the confidence intervals include rather substantial positive effects.

Table 10: Educational attainment by treatment specification, 1986-1990 cohorts

	All	Females	Males
Multi-Comp.	-0.068	-0.016	-0.143
	(0.111)	(0.170)	(0.150)
Hyperbolic	0.005	0.139	-0.188
	(0.112)	(0.166)	(0.144)
Binary 12	-0.150	-0.066	-0.225
	(0.126)	(0.203)	(0.173)
Binary 18	-0.066	-0.014	-0.137
	(0.109)	(0.168)	(0.146)
Binary 24	-0.018	0.062	-0.137
	(0.105)	(0.158)	(0.137)
Observations	5204	2654	2550

Notes: Each cell presents the result from a separate regression on the sample restricted to cohorts born between 1986 and 1990, where the columns indicate different gender samples (female, male and both) and rows indicate different treatment specifications. The specific details for each specification are outlined in Section 3.3. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on the district-cohort level within parenthesis. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 11: Educational attainment by data source, 1986-1990 cohorts

	All	Females	Males
Pooled			
Full	-0.068	-0.016	-0.143
	(0.111) [5204]	(0.170) [2654]	(0.150) [2550]
Excl. NPS 2008	-0.059	-0.082	-0.059
	(0.108) [4980]	(0.180) [2526]	(0.146) [2454]
By Survey			
DHS 1999	-0.264	-0.399	0.001
	(0.176) [462]	(0.310) [230]	(0.257) [232]
THBS 2000	0.195*	0.183	0.190
	(0.104) [3040]	(0.160) [1522]	(0.137) [1518]
DHS 2004	-0.414	-0.712	-0.144
	(0.257) [910]	(0.480) [457]	(0.305) [453]
NPS 2008	0.494	1.484	-2.250
	(0.889) [224]	(1.091) [128]	(1.432) [96]
DHS 2010	-0.316	0.367	-1.236*
	(0.420) [568]	(0.816) [317]	(0.694) [251]

Notes: Each cell presents the result from a separate regression on the sample restricted to cohorts born between 1986 and 1990, where the columns indicate different gender samples (female, male and both) and rows indicate different surveys. The two first rows present the results from a pooled analysis, either including or excluding the NPS 2008 survey. Each subsequent row is a separate survey. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 12: The effect on additional outcomes, 1986-1990 cohorts

	All	Females	Males
Individual			
Highest Grade	-0.068	-0.016	-0.143
	(0.111) [5204]	(0.170) [2654]	(0.150) [2550]
Some Primary	0.003	0.035	-0.022
	(0.018) [5235]	(0.022) [2673]	(0.023) [2562]
Comp. Primary	-0.038***	-0.049**	-0.034*
	(0.014) [5203]	(0.021) [2654]	(0.021) [2549]
Some Secondary	0.004	0.014	-0.007
	(0.009) [5203]	(0.011) [2654]	(0.012) [2549]
Literacy	0.032	0.009	0.052
	(0.028) [3834]	(0.038) [2152]	(0.035) [1682]
Average effects			
AES, excl. lit.	-0.023	0.005	-0.056
	(0.031) [5203]	(0.039) [2654]	(0.043) [2549]
AES, incl. lit.	0.030	0.004	0.059
	(0.031) [3800]	(0.042) [2131]	(0.053) [1669]

Notes: Each cell presents the result from a separate regression on the sample restricted to cohorts born between 1986 and 1990, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

F Specification including district with late programs

Table 13: Effects when including the two districts with late treatment

	All	Females	Males
Individual			
Highest Grade	0.084	0.094	0.069
	(0.062) [10907]	(0.084) [5492]	(0.083) [5415]
Some Primary	0.021*	0.036**	0.007
	(0.012) [10968]	(0.017) [5523]	(0.016) [5445]
Comp. Primary	-0.005	-0.000	-0.010
	(0.009) [10896]	(0.012) [5488]	(0.012) [5408]
Some Secondary	0.002	0.002	0.003
	(0.007) [10896]	(0.008) [5488]	(0.010) [5408]
Literacy	0.018	0.003	0.032
	(0.019) [7225]	(0.026) [3971]	(0.027) [3254]
Average effects			
AES, excl. lit.	0.020	0.032	0.009
	(0.018) [10896]	(0.023) [5488]	(0.025) [5408]
AES, incl. lit.	0.027	0.020	0.038
	(0.021) [7137]	(0.029) [3924]	(0.032) [3213]

Notes: Each cell presents the result from a separate regression when the two districts with late treatment are included (Mbinga and Bukoba Rural), where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

G Specifications with less and more controls

Table 14: Effects when using the most parsimonious specification

	All	Females	Males
Individual			
Highest Grade	0.050	0.026	0.062
	(0.066) [9972]	(0.089) [5016]	(0.085) [4956]
Some Primary	0.023*	0.035*	0.012
	(0.013) [10031]	(0.018) [5046]	(0.016) [4985]
Comp. Primary	-0.012	-0.014	-0.011
	(0.009) [9963]	(0.013) [5013]	(0.012) [4950]
Some Secondary	-0.002	-0.003	0.001
	(0.007) [9963]	(0.008) [5013]	(0.010) [4950]
Literacy	0.019	0.007	0.030
	(0.021) [6659]	(0.029) [3654]	(0.029) [3005]
Average effects			
AES, excl. lit.	0.010	0.011	0.008
	(0.019) [9963]	(0.025) [5013]	(0.027) [4950]
AES, incl. lit.	0.018	0.008	0.037
	(0.024) [6575]	(0.033) [3609]	(0.034) [2966]

Notes: Each cell presents the result from a separate regression when only controls of age, sex and time of interview is used, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 15: Effects when adding additional controls

	All	Females	Males
Individual			
Highest Grade	0.064	0.073	0.053
	(0.061) [9972]	(0.083) [5016]	(0.082) [4956]
Some Primary	0.021	0.036**	0.007
	(0.013) [10031]	(0.018) [5046]	(0.016) [4985]
Comp. Primary	-0.010	-0.009	-0.011
	(0.009) [9963]	(0.012) [5013]	(0.012) [4950]
Some Secondary	0.000	0.003	-0.001
	(0.007) [9963]	(0.008) [5013]	(0.010) [4950]
Literacy	0.019	0.010	0.030
	(0.020) [6659]	(0.027) [3654]	(0.028) [3005]
Average effects			
AES, excl. lit.	0.013	0.026	0.003
	(0.018) [9963]	(0.023) [5013]	(0.026) [4950]
AES, incl. lit.	0.023	0.019	0.032
	(0.023) [6575]	(0.030) [3609]	(0.034) [2966]

Notes: Each cell presents the result from a separate regression when additional controls are added, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The added controls are number of children and number of boys in the household, overall and sex specific birth order, indicators of house ownership, grass roof and food shortages, and distances to nearest secondary school and health clinic. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

H Specification using ward fixed effects

Table 16: The effect on additional outcomes with ward fixed effects, 1986-1994 cohorts

	All	Females	Males
Individual			
Highest Grade	0.103*	0.150*	0.088
	(0.060) [9972]	(0.087) [5016]	(0.082) [4956]
Some Primary	0.020	0.048***	0.002
	(0.013) [10031]	(0.018) [5046]	(0.016) [4985]
Comp. Primary	-0.005	-0.002	-0.005
	(0.009) [9963]	(0.013) [5013]	(0.012) [4950]
Some Secondary	0.004	0.009	-0.000
	(0.006) [9963]	(0.008) [5013]	(0.009) [4950]
Literacy	0.020	0.032	0.018
	(0.019) [6659]	(0.028) [3654]	(0.027) [3005]
Average effects			
AES, excl. lit.	0.024	0.052**	0.008
	(0.017) [9963]	(0.024) [5013]	(0.024) [4950]
AES, incl. lit.	0.027	0.047	0.039
	(0.020) [6575]	(0.029) [3609]	(0.026) [2966]

Notes: Each cell presents the result from a separate regression when using ward (rather than district) fixed effects with cohorts born 1986-1994, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

Table 17: The effect on additional outcomes with ward fixed effects, 1986-1990 cohorts

	All	Females	Males
Individual			
Highest Grade	-0.005	-0.041	-0.004
	(0.101) [5204]	(0.189) [2654]	(0.149) [2550]
Some Primary	-0.004	0.036	-0.034
	(0.017) [5235]	(0.023) [2673]	(0.023) [2562]
Comp. Primary	-0.028**	-0.050**	-0.022
	(0.014) [5203]	(0.023) [2654]	(0.022) [2549]
Some Secondary	0.012	0.007	0.012
	(0.008) [5203]	(0.013) [2654]	(0.014) [2549]
Literacy	0.022	0.016	0.028
	(0.026) [3834]	(0.042) [2152]	(0.036) [1682]
Average effects			
AES, excl. lit.	-0.007	-0.005	-0.023
	(0.026) [5203]	(0.040) [2654]	(0.040) [2549]
AES, incl. lit.	0.016	-0.011	0.054
	(0.027) [3800]	(0.043) [2131]	(0.046) [1669]

Notes: Each cell presents the result from a separate regression when using ward (rather than district) fixed effects with cohorts born 1986-1990, where the columns indicate different gender samples (female, male and both) and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

I Results by age group

Table 18: The effect in four age groups

	Ages 6-9	Ages 10-13	Ages 14-17	Ages 18-21
Individual				
Highest Grade	0.048	0.109	0.154	-0.042
	(0.052) [2863]	(0.089) [3812]	(0.185) [1940]	(0.321) [885]
Some Primary	0.012	0.001	0.040	-0.003
	(0.025) [2894]	(0.018) [3836]	(0.025) [1944]	(0.041) [885]
Comp. Primary			-0.020	-0.038
			(0.035) [1940]	(0.047) [885]
Some Secondary			-0.019	0.025
			(0.027) [1940]	(0.042) [885]
Literacy	-0.012	0.044	0.026	0.074
	(0.028) [2500]	(0.030) [2498]	(0.057) [1034]	(0.068) [396]
Average effects				
AES, excl. lit.	0.048	0.035	0.014	-0.011
	(0.040) [2859]	(0.046) [3811]	(0.063) [1940]	(0.079) [885]
AES, incl. lit.	0.001	0.106	-0.098	0.033
	(0.040) [2453]	(0.052) [2472]	(0.108) [1029]	(0.113) [396]

Notes: Each cell presents the result from a separate regression where the columns indicate different age groups and rows indicate different outcome variables. The first five rows present the result for each outcome separately while the two last rows present the average effect size (as described in detail in Section 4) when either including or excluding literacy. Blank cells indicate that there is no variation in the relevant outcome for the studied age group. Besides the estimated coefficient for the treatment variable, each cell reports the standard error clustered on district-cohort level within parenthesis and number of observation in the particular regression in square brackets. Asterisks indicate significant estimates at the 10% (*), 5% (**) or 1% (***) level.

WORKING PAPERS
Uppsala Center for Labor Studies
Editor: Per-Anders Edin

2010:1 Susanne Ek and Bertil Holmlund, Family Job Search, Wage Bargaining, and Optimal Unemployment Insurance. 33 pp.

2010:2 Torberg Falch and Sofia Sandgren Massih, The effect of early cognitive ability on earnings over the life-cycle. 37 pp.

2010:3 Anders Forslund and Peter Fredriksson, Income support systems, labour supply incentives and employment – some cross-country evidence. 23 pp.

2010:4 Per Engström and Per Johansson, The medical doctors as gatekeepers in the sickness insurance?. 32 pp.

2010:5 Jonas Malmberg, Posting Post Laval, International and National Responses. 26 pp.

2010:6 Johanna Rickne, Gender, Wages and Social Security in China's Industrial Sector. 48 pp.

2010:7 Ulrika Vikman, Does Providing Childcare to Unemployed Affect Unemployment Duration? 43 pp.

2010:8 Helena Holmlund, Mikael Lindahl and Erik Plug, The Causal Effect of Parent's Schooling on Children's Schooling A Comparison of Estimation Methods. 56 pp.

2010:9 Jonathan Gemus, College Achievement and Earnings. 43 pp.

2010:10 Sara Pinoli, Rational Expectations and the Puzzling No-Effect of the Minimum Wage. 56 pp.

2010:11 Anna Persson and Ulrika Vikman, Dynamic effects of mandatory activation of welfare participants. 37 pp.

2010:12 Lena Hensvik, Competition, wages and teacher sorting: four lessons learned from a voucher reform. 41 pp.

2010:13 Magnus Gustavsson and Pär Österholm, Labor-Force Participation Rates and the Informational Value of Unemployment Rates: Evidence from Disaggregated US Data. 10 pp.

2010:14 Heléne Lundqvist, Matz Dahlberg and Eva Mörk, Stimulating Local Public Employment: Do General Grants Work? 35 pp.

2010:15 Susanne Ek and Bertil Holmlund, Deltidsarbetslöshet och arbetslöshetsförsäkring i Sverige. 75 pp.

2010:16 Jonathan Gemus, The Distributional Effects of Direct College Costs. 34 pp.

2010:17 Kent Eliasson, Pär Hansson and Markus Lindvert, Jobs and Exposure to International Trade within the Service Sector in Sweden. 34 pp.

2010:18 Olof Åslund, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist, Peers, neighborhoods and immigrant student achievement - evidence from a placement policy. 42 pp.

2010:19 Magnus Gustavsson and Pär Österholm, Does the Labor-Income Process Contain a Unit Root? Evidence from Individual-Specific Time Series. 26 pp.

2011:1 Matz Dahlberg, Karin Edmark and Heléne Lundqvist, Ethnic Diversity and Preferences for Redistribution. 43 pp.

2011:2 Susanne Ek and Bertil Holmlund, Part-Time Unemployment and Optimal Unemployment Insurance. 32 pp.

2011:3 Geert Ridder and Johan Vikström, Bounds On Treatment Effects On Transitions. 55 pp.

2011:4 Johan Vikström, Michael Rosholm and Michael Svarer, The Relative Efficiency of Active Labour Market Policies: Evidence From a Social Experiment and Non-Parametric Methods. 34 pp.

2011:5 Daniel Hallberg, Is Early Retirement Encouraged by the Employer? Labor-Demand Effects of Age-Related Collective Fees. 31 pp.

2011:6 Niklas Bengtsson and Per Engström, Control and Efficiency in the Nonprofit Sector - Evidence from a Randomized Policy Experiment. 26 pp.

2011:7 Niklas Bengtsson, The marginal propensity to earn and consume out of unearned income: Evidence using an unusually large cash grant reform. 27pp.

2011:8 Jesper Roine and Daniel Waldenström, On the Role of Capital Gains in Swedish Income Inequality. 32 pp.

2011:9 Andrea Ichino, Elly-Ann Lindström and Eliana Viviano, Hidden consequences of a first-born boy for mothers. 17 pp.

2011:10 Jonas Kolsrud, Consumption Smoothing during Unemployment. 45 pp.

2011:11 Adrian Adermon and Magnus Gustavsson, Job Polarization and Task-Biased Technological Change: Sweden, 1975–2005. 33 pp.

2011:12 Bertil Holmlund, Svensk arbetsmarknad under två kriser. 15 pp.

2011:13 Niklas Bengtsson, Regulation Failure and CO2-emissions: An Experimental Investigation of the Cape Town Taxi Market. 26 pp.

2011:14 Mikael Carlsson, Julián Messina and Oskar Nordström Skans, Wage adjustment and productivity shocks. 57 pp.

2011:15 Daniel Avdic and Marie Gartell, The study pace among college students before and after a student aid reform: some Swedish results. 43 pp.

2011:16 Daniel Hallberg, Thomas Lindh and Jovan Žamac, Study achievement for students with kids. 32 pp.

2011:17 Klaus J. Brösamle and Oskar Nordström Skans, Paths to higher office: evidence from the Swedish Civil Service. 40 pp.

2011:18 Francis Kramarz and Oskar Nordström Skans, When strong ties are strong Networks and youth labor market entry. 69 pp.

2011:19 Petter Lundborg, Martin Nilsson and Johan Vikström, Socioeconomic Heterogeneity in the Effect of Health Shocks on Earnings. Evidence from Population-Wide Data on Swedish Workers. 42 pp.

2011:20 Hans Grönqvist and Caroline Hall, Education policy and early fertility: Lessons from an expansion of upper secondary schooling. 43 pp.

2011:21 Mikael Lindahl, Mårten Palme, Sofia Sandgren Massih and Anna Sjögren, Transmission of Human Capital across Four Generations: Intergenerational Correlations and a Test of the Becker-Tomes Model. 27 pp.

2011:22 Krzysztof Karbownik and Michał Myck, Mommies' Girls Get Dresses, Daddies' Boys Get Toys. Gender Preferences in Poland and their Implications. 49 pp.

2012:1 Eva Mörk och Linus Liljeberg, Fattig, sjuk och arbetslös – en beskrivning av personer i kläm mellan stat och kommun. 23 pp.

2012:2 Karin Edmark and Kajsa Hanspers, s welfare dependency inherited? Estimating the causal welfare transmission effects using Swedish sibling data. 41 pp.

2012:3 Karin Edmark, Che-Yuan Liang, Eva Mörk and Håkan Selin, Evaluation of the Swedish earned income tax credit. 39 pp.

2012:4 Susanne Ek, Unemployment benefits or taxes: How should policy makers redistribute income over the business cycle? 30 pp.

2012:5 Simona Bejenariu and Andreea Mitrut, Save Some, Lose Some: Biological Consequences of an Unexpected Wage Cut. 67 pp.

2012:6 Pedro Carneiro and Rita Ginja, Long Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. 82 pp.

2012:7 Bertil Holmlund, Wage and Employment Determination in Volatile Times: Sweden 1913–1939. 43 pp.

2012:8 Helge Bennmarker, Oskar Nordström Skans and Ulrika Vikman, Workfare for the old and long-term unemployed. 34 pp.

2012:9 Olle Folke and Johanna Rickne, Female representation but male rule? Party competition and the political glass ceiling. 28 pp.

2012:10 Per Engström, Pathric Hägglund and Per Johansson, Early interventions and disability insurance: Experience from a field experiment. 40 pp.

2012:11 Tommy Ferrarini, Kenneth Nelson, Joakim Palme and Ola Sjöberg, Sveriges socialförsäkringar i jämförande perspektiv. En institutionell analys av sjuk-, arbetsskade- och arbetslöshetsförsäkringarna i 18 OECD-länder 1930 till 2010. 71 pp.

2012:12 José I. Silva and Javier Vázquez-Grenno, The Ins and Outs of Unemployment in a Two-Tier Labor Market. 24pp.

2012:13 Niklas Bengtsson, Bertil Holmlund and Daniel Waldenström, Lifetime versus Annual Tax Progressivity: Sweden, 1968–2009. 56 pp.

2012:14 Oana Borcan, Mikael Lindahl and Andreea Mitrut, The Impact of an Unexpected Wage Cut on Corruption: Evidence from a "Xeroxed" Exam. 36 pp.

2012:15 Anders Böhlmark and Mikael Lindahl, Independent Schools and Long-Run Educational Outcomes - Evidence from Sweden's Large Scale Voucher Reform. 46 pp.

2012:16 Stefan Eriksson, Per Johansson and Sophie Langenskiöld, What is the right profile for getting a job? A stated choice experiment of the recruitment process. 52 pp.

2012:17 Krzysztof Karbownik and Michal Myck, For some mothers more than others: how children matter for labour market outcomes when both fertility and female employment are low. 28 pp.

2012:18 Niklas Bengtsson and Jan Pettersson, The Outreach and Sustainability of Microfinance: Is There a Tradeoff? 27 pp.

2012:19 Nikolay Angelov and Arizo Karimi, Mothers' Income Recovery after Childbearing. 34 pp.

2012:20 Arizo Karimi, Erica Lindahl and Peter Skogman Thoursie, Labour Supply Responses to Paid Parental Leave. 38 pp.

2012:21 Helge Bennmarker, Erik Grönqvist and Björn Öckert, Effects of contracting out employment services: Evidence from a randomized experiment. 55 pp.

2012:22 Pedro Carneiro and Rita Ginja, Partial Insurance and Investments in Children. 32 pp.

2013:1 Olof Åslund and Mattias Engdahl, The value of earning for learning: Performance bonuses in immigrant language training. 52 pp.

2013:2 Matz Dahlberg, Karin Edmark and Heléne Lundqvist, Ethnic Diversity and Preferences for Redistribution: Reply. 23 pp.

2013:3 Nikolay Angelov, Per Johansson and Erica Lindahl, Is the persistent gender gap in income and wages due to unequal family responsibilities? 39 pp.

2013:4 Ulrika Vikman, Paid parental leave to immigrants: An obstacle to labor market entrance? 51 pp.

2013:5 Nikolay Angelov, Per Johansson and Erica Lindahl, Gender differences in sickness absence and the gender division of family responsibilities. 51 pp

2013:6 Tove Eliasson, Decomposing immigrant wage assimilation - the role of workplaces and occupations. 40 pp.

2013:7 Olof Åslund and Mattias Engdahl, Open borders, transport links and local labor markets. 41 pp.

2013:8 Magne K. Asphjell, Lena Hensvik and J Peter Nilsson, Businesses, Buddies, and Babies: Fertility and Social Interactions at Work. 55 pp.

2013:9 Helena, Svaleryd, Self-employment and the local business cycle. 28 pp.

2013:10 Caroline Hall, Does more general education reduce the risk of future unemployment? Evidence from labor market experiences during the Great Recession. 51 pp.

2013:11 The Short- and Long-term Effects of School Choice on Student Outcomes – Evidence from a School Choice Reform in Sweden. Verena Wondratschek, Karin Edmark and Markus Frölich. 98 pp.

2013:12 Magnus Gustavsson, Permanent versus Transitory Wage Differentials and the Inequality-Hours Hypothesis. 11 pp.

2013:13 Niklas Bengtsson, Stefan Peterson and Fredrik Sävje, Revisiting the Educational Effects of Fetal Iodine Deficiency. 48 pp.