Effects of Early Childhood Intervention on Child Development and Early Skill Formation. Evidence from a Randomized Controlled Trial

Malte Sandner*

Leibniz Universität Hannover
Wirtschaftswissenschaftliche Fakultät
August 2013

Abstract

This paper presents results from a randomized evaluation of a home visiting program implemented in Germany for disadvantaged first-time mothers and their families. The intervention increases infants' cognitive development by 0.18 SD, about 12 months after birth. However, the effect fades out after 24 months. Gender analyses reveal that the intervention was more beneficial for girls. Furthermore, sensitivity analyses show that the estimated effects seem biased downward by additional treatment for the control families. Analyzing the infant skill formation process reveals self-productivity of skills; however, the magnitude differed for boys and girls.

JEL-Classification: J13, J12, I21, H52
Keywords: Early Childhood Intervention, Randomized Experiment

*Financial support by the German Federal Ministry for Family, Seniors, Women and Youth (BMFSFJ) and the Saxony Social Ministry is gratefully acknowledged. I would like to thank all the participants of the following conferences and seminars for helpful comments: The Annual Conference of the Scottish Economic Society 2012, the Spring Meeting of Young Economists (SMYE) 2012, the Lower Saxony Workshop in Applied Economics 2012, the Conference of the European Society of Population Economics (ESPE) 2012, the Seminar of the Center for Economics and Neuroscience (CENs) 2012, the Seminar of the Lower Saxony Institute for Economic Research 2012, the Annual Conference of the European Economic Association (EEA) 2012, the Annual Meeting of the *Verein für Socialpolitik* (VfS) 2012, the 24th Annual Conference of the European Association of Labour Economists (EALE) 2012, the International IAB Conference "Field Experiments in Policy Evaluation" 2013, the Research Committee for Economic of Education of the VfS 2013, the Annual Conference of Royal Economic Society (RES) 2013 and the Annual Conference of the Society of Labor Economics (SOLE) 2013.

1 Introduction

In recent years, interdisciplinary research has emphasized the negative impact of adverse early childhood conditions for lifelong human capital accumulation. This research is based upon the following aspects: First, poor maternal health, dysfunctional families, adverse childhood environments and low parenting skills have detrimental effects for child development (see Almond and Currie, 2011, for a literature overview). Second, due to the dynamic nature of the skill formation process, the earlier these adverse childhood conditions occur, the bigger the cumulative lifelong harm (Cunha and Heckman, 2007). Third, to prevent these negative conditions, parents who play an essential role for child well being must be targeted (Heckman, 2011). Therefore, policy interventions which concentrate on children from disadvantaged families, which start early enough in life, particularly antenatal, and which alter parenting behavior are supposed to have a lasting effect on children's life outcomes and can produce high benefit-cost ratios.

Home visiting is a type of early intervention, which can fulfill these requirements. In home visiting program trained midwives, nurses, or social pedagogues visit disadvantaged families at their own homes, starting at the beginning of pregnancy. These home visitors typically interact with the parents to encourage them and train them in how to raise their children. Evidence from meta-analyses, including all varieties of home visiting (e.g., programs which start after birth), documents that home visiting has modest effects on parenting behavior and attitudes, and on cognitive and socio-emotional development of the child at preschool and school age (Sweet and Appelbaum, 2004). High-quality home visiting programs – concentrating on disadvantaged families that start during pregnancy – appear to be even more effective on these outcomes (Olds et al., 1999; Gomby, 2005). The few existing studies on long-term effects indicate that these effects are stable over time (Eckenrode et al., 2010).

The greater effects of interventions that start during pregnancy, led to the *antenatal investment hypothesis* suggesting that these interventions affect cognitive development early in life (Doyle et al., 2009). Nevertheless, research on antenatal

home visits found only small direct effects on cognitive development in a child's early life (Doyle et al., 2013). Furthermore, an explicit test that the dynamic nature of the skill formation process already begins in the first two years of life is lacking. However, previous studies that investigated the effects and mechanisms of prenatal home visiting programs on early child development have methodological limitations. They measured children's cognitive development by screening methods based on maternal ratings. These methods possess little sensitivity to developmental advances and delays in the first two years of life (e.g. Florida Institute of Education, 2004; McAffee and Leong, 1997). Furthermore, these studies often used only one point of observation at infancy age and had small samples sizes.

This paper uses the first randomized controlled experiment on antenatal home visiting conducted in Germany, the *Pro Kind* Project, to examine its effect on child development in the first two years of life. Video-taped child development assessments give information about mental and psychomotor child development at the ages of 6, 12, and 24 months. Personal interviews provide information about birth outcomes and investments into the child. The *Pro Kind* sample consists of 755 disadvantaged first-time mothers in three federal states, who are randomly assigned to either the treatment group with home visits during pregnancy and the following two years or the control group. The obtained data set is unique as it includes objective child assessment data for a large sample of disadvantaged children, which gives detailed information about child development in the first two years. Therefore, the data gives the opportunity not only to evaluate the intervention, but also to shed light on the skill formation process in the first two years of life.

The *Pro Kind* data has been examined by a team of developmental psychologists before. This analysis found that children in home visited families tend to have better birth outcomes, and achieve higher cognitive development test scores (Jungmann et al., 2009, 2010). However, this past research primarily consists of comparisons of means and has paid little attention to potential threats to the validity of the experiment, the longitudinal structure of the data, or the dynamic process of skill formation. Furthermore, treatment effect heterogeneity by gender, the efficiency of home visiting and investments into the child has received no attention. Additionally,

there were deviations from the ideal experimental design in the actual implementation of Project *Pro Kind*. First, randomization was done at a state level and not at a community level, although it was stratified for community level. Nevertheless, due to the high heterogeneity between communities in the same federal state, biases could occur. Secondly, as in most longitudinal studies with disadvantaged participants, attrition is a common problem. One third of the infants whose mothers were randomized missed at least one developmental test. These limitations have not been adequately addressed in previous work.

I find that the *Pro Kind* Project was effective in improving children's mental development. At the end of 12 months, children from home visited families performed significantly better than those in control families by 0.18 standard deviations (SD) on the Mental Developmental Index. This treatment effect is equal to 2.5 percentage points at the median of a normal distribution. The effects are smaller at 6 months, and they almost vanish at 24 months. The *Pro Kind* Project fails to significantly improve the psychomotor skills, birth outcomes and the language skills of the children. However, most of the coefficients for these outcomes are positive. The program has differential impacts on girls and boys. For girls, I find significant effects on mental development with an effect size of around 0.30 SD at 6 and 12 months and of 0.20 SD at 24 months. Additionally, girls from home visited families produce more words and sentences than their counterparts from control families, with an effect size of 0.25 SD. In contrast, boys do not benefit from treatment in any of these outcomes. As an explanation for the gender specific outcomes, I find that the treatment enhances parental investments of differential magnitude for boys and girls. Investigating the skill formation process in the first few years of life reveals that self-productivity of skills already occurs in the first two years of life, but in different degree for boys and girls.

There is no indication of selective attrition between control and treatment group in respect to the baseline characteristics. However, in the control group, the test scores of the children who quit participating in the research are lower than those in the treatment group. This may be caused by the fact that mothers in both groups receive feedback about the test results. Imputing missing test scores leads to much

higher treatment effects which range between 0.2 and 0.3 SD for the whole sample.

The rest of the paper is organized as follows: Section 2 provides a description of the *Pro Kind* Project. Section 3 describes the experimental design and the data. Section 4 presents the results of the impact of the home visiting program on birth outcomes, mental and psychomotor development and language development. Section 5 conducts robustness checks and presents evidence that the main effect of the intervention might be biased downward. Section 6 analyzes the dynamics of the skill development. Section 7 presents the conclusions.

2 Background and Description of the Pro Kind Project

Pro Kind is a home visiting program for disadvantaged first-time mothers and their families. The intervention started between the 12th and 28th week of pregnancy and ended at the second birthday of the child. The program was conducted in three German federal states, two in West and one in East Germany. Families were inducted into the program between November 2006 and December 2009. Midwives, nurses, or social pedagogues conducted the home visits alone or in a team. The frequency of the home visits varied by age of the child between weekly, bi-weekly and monthly visits, with the highest frequency directly before and after birth. Overall, 52 home visits are scheduled between pregnancy and the child's second birthday. A regular home visit has a duration of 90 minutes. Teaching materials and guidebooks structure the theme and the aim of each home visit. Nevertheless, the home visitors have the flexibility to improvise the content to the needs of the mothers and their families. All home visitors regularly receive feedback, encouragement, reflection, and support from their nurse supervisors. These supervisors have an academic qualification, and they do not consult more than ten home visitors each. *Pro Kind* is an adaptation of the Nurse Family Partnership (NFP) program, which provided instructions for home visit frequency, employee selection, teaching material and guidebooks (see Jungmann et al., 2009; Olds, 2006, for more information about the *Pro Kind* program and NFP).

Improving birth outcomes and child development are the major goals of *Pro Kind*. For birth outcomes, the health of the mother during pregnancy is vital, while

parental skills and parental investments into the child are crucial for child development. To generate a healthier environment, issues like smoking and having a balanced diet are covered in the home visits. To enhance parental skills and investments, the home visitors train the parents to perceive children's signals accurately and to answer them sensitively. In order to be successful in sensitive topics like smoking or parental behavior, *Pro Kind* refers to different psychological theories like the ecological theory, attachment theory, and self-efficacy theory (Bronfenbrenner, 1992; Bowlby, 1969; Bandura, 1982, 1997).

Because of the *Pro Kind* affiliation criteria only first-time mothers during their 12th and 28th weeks of gestation were registered. Additionally all participating mothers had to be financially or socially disadvantaged. Financial disadvantage is defined as recipient of social welfare benefits, unemployment benefits, an income that is as low as social welfare benefits or over-indebtedness. The considered social risk factors included the following: low education, teenage pregnancy, isolation, experienced violence or health problems. Project partners, like gynecologists, job centers, pregnancy information centers and youth welfare offices referred three quarters of the participants to *Pro Kind*. About one quarter of the participants volunteered to register themselves to the program.

To monitor the program fidelity, the home visitors documented each home visit (e.g., duration, covered topics, maternal interest). This documentation reports that on an average, a family got 32.7 home visits with a minimum of 0, a maximum of 94 and a standard deviation of 19 home visits. During pregnancy, the families received nine home visits on an average. Because participation in *Pro Kind* is voluntary, 166 (42.2%) mothers decided to leave the program before the child's second birthday (main reasons: no further interest [n=68], not reachable [n=37] and moving away from a *Pro Kind* community [n=28]). Considering only families who received the full program dosage increases the average number of home visits to 45.3 (SD= 10.7) with a minimum of 11 visits. The average duration of a home visit was 82 minutes.

3 Experimental Design and Data

3.1 Randomization Process and Sample Baseline Characteristics

The causal effects of the *Pro Kind* intervention are evaluated using a randomized controlled trial. In the beginning of the randomization process, all women, who were referred or registered at *Pro Kind* themselves, answered a short screening questionnaire to check if the affiliation criteria were fulfilled. Most of the time, this screening questionnaire was conducted over telephone. If the affiliation criteria were met, the supervisor visited the mother at her home. At this visit, first of all, participants or, if they were underage, their parents signed an informed consent for participating in the study. Participants then, answered a baseline questionnaire to obtain socio-demographic and psychological characteristics, and risk factors. Up to this moment, the mothers only received information about the research study and as little information as possible about the home visits in order to minimize the "John Henry" effect for those mothers in the control group.[1] After answering the baseline questionnaire, women received the results of the randomization which sorted them into a home visiting or a control group.

After randomization, mothers in both, the control and home visiting groups had access to the regular welfare state services. Both groups received an address list with support services and monetary incentives for participating in the research. Additionally, the research provided feedback to mothers in both the groups about the development of their children. However, only the home visiting group was eligible for the *Pro Kind* home visits. Overall, 394 mothers were allocated to the treatment group, and 361 to the control group. Table 7 gives an overview of the randomization outcomes in each state and community.

The randomization process should ensure that two groups are equal on their baseline characteristics. To prove if this condition is fulfilled, I use the following basic model:

$$h_{ic} = \beta_0 + \beta_1 HV_{ic} + \alpha_c + \epsilon_i \tag{1}$$

[1]The "John Henry" effect explains the unexpected outcome of an experiment caused by the control group's knowledge of its role within the experiment. This knowledge causes the group to perform differently and often better than usual, eliminating the effect of the experimental manipulation (Salkind, 2010).

where h_{ic} is a risk factor or characteristic at baseline for mother i in community c and HV_{ic} is an indicator variable for whether the mother received the home visiting program. Hence, the estimate of the coefficient β_1 indicates the differences between treatment and control group mothers. Additionally, I include a community fixed effect estimator α_c in Equation 1 because the randomization results in Table 7 reveal that the number of participants in the treatment and control groups are not equally distributed in all communities.

If the randomization process worked well, no coefficients of β_1 would be significantly correlated with characteristic h_{ic} in any model specification. I present the comparison of mother and family characteristics at baseline in Table 1. Column (1) contains β_0, which gives the average of characteristic h_{ic} in the control group. Columns (2) and (3) present the estimated differences between the treatment and control groups for demographic characteristics and selected psychological and physical baseline characteristics. The model in Column (2) does not include any controls, while the model in Column (3) controls for the community fixed effects.

If a missing value occurred in one of the baseline variables, I include sample means or values from a multivariate imputation procedure for the missing value. However, for most variables, complete data is available. The percentage of missing values is higher than three percent only in the income variables (see Appendices B and C). The results hardly change if the missing values are used instead of the sample means or imputed values. In almost all variables, the missing values are equally distributed between control and treatment groups.

The difference in the average characteristics between the control and treatment groups are small and mostly statistically not significant. Migration status, defined as women who do not have German citizenship or who are not born in Germany, is the only demographic characteristic that is significantly different with a higher proportion of immigrants in the control group. None of the differences in psychological or physical risk characteristics are statistically significant and including community fixed effects does not change the results. Furthermore, I conduct a test of joint significance of all the baseline characteristics. The F-statistic is 1.19; thus the possibility that the characteristics in the treatment and control groups are the same could

Table 1: Sample Balance Across Treatments

	Control Mean	Treatment Difference No Controls	Treatment Difference Community Fixed Effects
	(1)	(2)	(3)
Demographic Characteristics			
Age in Years	21.53	-0.263 (0.316)	-0.274 (0.313)
Week in Pregnancy	20.30	-0.540 (0.420)	-0.528 (0.423)
Underage	0.177	0.033 (0.029)	0.035 (0.028)
Migration	0.177	-0.053** (0.026)	-0.049* (0.025)
Monthly HH-Income (€)	916.6	20.66 (41.78)	17.54 (40.60)
Debt over €3000	0.168	0.021 (0.027)	0.020 (0.028)
Education Risk	0.748	0.054 (0.038)	0.055 (0.038)
Income Risk	0.809	0.011 (0.028)	0.012 (0.028)
Employment Risk	0.856	-0.036 (0.027)	-0.040 (0.027)
No Partner	0.283	0.009 (0.033)	0.004 (0.033)
Living with Parents	0.267	0.014 (0.033)	0.011 (0.033)
Persons in HH	2.451	0.102 (0.120)	0.089 (0.120)
Selected Psychological and Physical Characteristics			
Unwanted Pregnancy	0.166	0.014 (0.028)	0.012 (0.028)
Daily Smoking	0.340	-0.003 (0.034)	-0.003 (0.034)
Isolation	0.080	-0.019 (0.019)	-0.020 (0.019)
Foster Care Experience	0.194	0.039 (0.030)	0.041 (0.030)
Neglect Experience	0.385	-0.009 (0.035)	-0.012 (0.036)
Lost Experience	0.539	-0.045 (0.036)	-0.048 (0.036)
Violence Experience	0.551	0.002 (0.036)	-0.001 (0.037)
Depression	0.133	-0.031 (0.023)	-0.031 (0.024)
Anxiety	0.177	-0.007 (0.028)	-0.008 (0.028)
Stress	0.288	0.027 (0.033)	0.028 (0.034)
Aggression	0.186	-0.041 (0.027)	-0.039 (0.027)
Medically Indicated Risk Preg.	0.113	0.000 (0.023)	-0.005 (0.023)
Body-Mass-Index	23.22	0.150 (0.394)	0.160 (0.394)
Sum Risk Factors	5.864	-0.131 (0.178)	0.035 (0.028)
Observations	361	755	755

Notes: Robust standard errors are shown in parentheses. The first column indicates the dependent variable. Column (1) indicates the mean of the characteristic in the control group. The variables in Columns (2) and (3) have the value one if the mother is in the treatment group. They contain estimates of the average difference in characteristics between the control and treatment participants, without controls and with community fixed effects, respectively. See Appendices B and C for variable definitions.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

not be rejected. Hence, overall, the randomization appears to have been successful in creating comparable treatment and control groups.

Analyzing the demographic and psychological characteristics of the participants reveals that women in both groups are highly disadvantaged. For example, over one third of the mothers has experienced neglect in their lifetimes and over half of the women have lost an important person during childhood. Both is related to attachment problems with their own children, and increases the probability of child maltreatment (Berlin et al., 2011; Wu et al., 2004). Furthermore, the average household income is €928.6. Considering that the average household size is 2.49 persons, this average income is below the poverty line in Germany. These are just

two examples of many characteristics, which underline the disadvantaged status of the *Pro Kind* participants. Therefore, one can conclude that *Pro Kind* was successful in acquiring high-burdened women and families, who are the target population of the intervention.

3.2 Research Design

For mothers in both the groups, the research started with a telephone interview and a personal interview during pregnancy. Telephone interviews continued at six months interval until the child's third birthday. They contain questions about birth outcomes, labor market participation, and other socio-economic outcomes of the mother and the family. Personal interviews, including child development tests, are conducted at 6, 12, and 24 months after birth. At each personal interview, cognitive abilities (IQ) were measured using the Mental Developmental Index (MDI) of the Bayley Scales of Infant Development (BSID) (Bayley, 1969). The fine and gross motor abilities, called the motor quotient (MQ), were also assessed at each personal interview using the Psychomotor Developmental Index (PDI) of the Bayley Scales. Additionally, at 24 months, a language test for two-year-old children (*Sprachentwicklungstest für zweijährige Kinder*, SETK-2) was conducted. The BSID and the SETK-2 tests are videotaped and assessed after the interview by a developmental psychologist, who was blind to the treatment group of the child. An important advantage of the BSID and the SETK-2 is that they provide observed data as opposed to parent-reported measures of child development.

The MDI and PDI test scores are normed on one hundred with an SD of 15 by an average population. A test score below 85 points indicates developmental delay. A test score below 70 points indicates serious developmental delay and need for medical assistance. If a child in the home visiting or control group scored below these cut-offs, the research sent a letter with special information and advice to the mother, in addition to the regular feedback. For my regression analysis, I standardized the test scores and birth outcomes with a mean of 0 and an SD of 1. Standardization allows comparison of effects on birth outcomes and test scores, and facilitates comparison to other home visiting interventions. MDI and PDI tests consist of different tasks.

Sometimes, the children refuse to do certain tasks of the full test battery. If the refusal or interruption rate in one test exceeds 20 percent, the overall test result is not considered reliable. Birth outcomes data are collected at two times, during the telephone interviews and the personal interviews. The data is used only when the mothers give identical information in both interviews. Additionally, part of the birth outcomes are checked against medical records, which revealed a high reliability of the self-reported measures.

Table 2: Sample Composition Developmental Tests

	Control	**Home Visiting**	**Total**
Allocated to Treatment	361	394	755
Completed 3-Month Telephone Interview	286 (79.2%)	317 (80.5%)	603
Boys	*130*	*150*	*280*
Girls	*153*	*167*	*321*
Completed 6-Month Development-Test	237 (65.7%)	265 (67.3%)	502
Boys	*110*	*125*	*235*
Girls	*127*	*140*	*267*
Completed 12-Month Development-Test	205 (56.8%)	225 (57.1%)	430
Boys	*94*	*105*	*199*
Girls	*111*	*120*	*231*
Completed 24-Month Development-Test	167 (45.7%)	180 (46.3%)	347
Boys	*76*	*83*	*159*
Girls	*91*	*97*	*188*

Table 2 reports the rates of developmental tests for treatment and control group and as per child's gender. Although the baseline comparisons presented in Table 1 show that the treatment and control groups were similar at the baseline, it is possible that the two groups become incomparable if more participants drop out from one group than from the other. However, the rates of completed developmental tests indicate that there are no statistically significant differences between the control and treatment groups and on considering children's gender. In both groups, about one third of the 6-month tests are not available. The attrition rate for the 12-month test is about 45 percent of the baseline participants, and for the 24-month test, 55 percent are missing.

Data is missing for the developmental tests for the treatment and control groups because the mothers moved away from a *Pro Kind* municipality or were difficult to engage at the assessment point within a time span of two months. Since these

tests are sensitive to age of the child, the results are not reliable if the test is not conducted within this time span. This time sensitivity of the tests also leads to cases in which children miss only one test but take part in the next test. Altogether, 71 percent of the randomized families and their children participated in at least one development test. In the treatment group, almost all families who participated in the developmental tests received at least some treatment.

3.3 Attrition

Table 2 shows that the attrition rates for the tests and interviews were similar in treatment and control groups. However, the baseline characteristics of the participants who left ("attritors") and those who remained ("non-attritors") still could have differed between the two groups. I investigate this possibility for the three months telephone interview in Column 1 of Table 3. The results for the 6-, 12- and 24-month tests are shown in Column 2, 3 and 4, respectively. Again, I use Equation 1, only including the mothers and children who participated in the interviews or tests.

All the differences between treatment and control groups are statistically insignificant, with the exception of the proportion of mothers with risk of aggression and loss experience at 24 months. The difference in mothers with immigration background becomes insignificant at the 24-month interview, which shows that even this imbalance in the randomization process remains almost stable. I conclude, therefore, that the control and home visited families are comparable throughout the follow-up tests.

Nevertheless, it might be that more or less disadvantaged mothers in the treatment and control groups refuse to participate in the interviews and tests. Table 4 compares maternal baseline characteristics of attritors with those of non-attritors. The results reveal that younger mothers and those with demographic risk factors like low education or income have a higher risk of refusing participation in the research. Psychological characteristics are less likely to be correlated with attrition. The differences in the demographic risk factors mainly observed between the baseline and the three month interview. However, the differences between attritors and non-attritors remained constant later. The only characteristic that continuously de-

creases in the attritors group is age. At the 24-month assessment, the participating mothers are more than two years older than their counterparts. If the treatment has higher effects for younger mothers, this might cause a fade-out of the effects. Nevertheless, this is a problem of program implementation, but it does not violate the internal validity of the treatment effects. Additionally, it is important to note that the remaining sample is still disadvantaged. For example, after 24 months, the cumulative sum of risk factors is 5.45 in the non-attritor group, in contrast to 6.08 in the attritor group.

MDI and PDI tests consist of different tasks. Sometimes, the infants and toddlers refuse to do certain tasks of the full test battery. If the refusal or interruption exceeds 20 percent in one test, the test results are not considered reliable. The Bayley Scale guidelines recommend not to use these results (Bayley, 1993). Therefore, 38 (Treatment group [TG]: 18, Control Group [CG]: 20) MDI test results at 6 months, 37 (TG: 22, CG: 15) at 12 months, and 48 (TG: 23, CG: 25) at 24 months cannot be used. The numbers of unreliable PDI tests are 18 (TG: 7, CG:11), 56 (TG: 22, CG: 34) and 85 (TG: 43, CG: 42). Appendices E and F demonstrate that also without these unreliable MDI and PDI tests the two groups only slightly differ with respect to the baseline characteristics.

Table 3: Selective Attrition between TG and CG - Development Tests

	Difference TG/CG			
	3 Months	6 Months	12 Months	24 Months
	(1)	(2)	(3)	(4)
Demographic Characteristics				
Age in Years	-0.078 (0.356)	-0.024 (0.400)	0.009 (0.443)	0.015 (0.505)
Week in Pregnancy	-0.613 (0.462)	-0.291 (0.513)	-0.209 (0.555)	0.034 (0.634)
Underage	0.014 (0.031)	0.005 (0.033)	0.021 (0.034)	0.039 (0.037)
Migration	-0.059** (0.028)	-0.062** (0.031)	-0.065* (0.053)	-0.020 (0.038)
Monthly HH-income (€)	19.45 (47.01)	-2.99 (52.87)	-13.15 (54.06)	27.98 (54.80)
Debt Over €3000	0.023 (0.032)	0.020 (0.035)	0.024 (0.038)	0.039 (0.043)
Education Risk	0.020 (0.032)	0.028 (0.039)	0.045 (0.042)	0.041 (0.049)
Income Risk	0.013 (0.032)	0.038 (0.036)	0.030 (0.040)	0.017 (0.043)
Employment Risk	-0.038 (0.031)	-0.024 (0.034)	-0.048 (0.038)	-0.028 (0.043)
No Partner	0.016 (0.036)	0.005 (0.040)	0.047 (0.043)	0.037 (0.048)
Living with Parents	-0.004 (0.036)	-0.022 (0.039)	-0.002 (0.042)	-0.001 (0.047)
Persons in HH	0.071 (0.135)	0.058 (0.148)	0.065 (0.155)	0.064 (0.160)
Selected Psychological and Physical Characteristics				
Unwanted Pregnancy	0.019 (0.030)	0.010 (0.033)	0.041 (0.035)	0.043 (0.039)
Daily Smoking	-0.011 (0.038)	-0.027 (0.041)	0.026 (0.045)	0.000 (0.050)
Isolation	-0.016 (0.020)	-0.017 (0.022)	0.000 (0.025)	0.021 (0.028)
Foster Care Exper.	0.044 (0.032)	0.026 (0.035)	0.045 (0.036)	0.054 (0.039)
Neglect Experience	0.006 (0.039)	0.008 (0.043)	-0.003 (0.047)	0.007 (0.053)
Lost Experience	-0.061 (0.040)	-0.044 (0.044)	-0.051 (0.048)	-0.098* (0.053)
Violence Ever	-0.019 (0.022)	-0.015 (0.023)	-0.028 (0.025)	-0.030 (0.027)
Depression	-0.010 (0.025)	-0.012 (0.026)	0.019 (0.029)	0.026 (0.033)
Anxiety	0.014 (0.030)	0.025 (0.033)	0.027 (0.036)	-0.006 (0.038)
Stress	0.039 (0.037)	0.036 (0.041)	0.048 (0.045)	0.032 (0.050)
Aggression	-0.057 (0.030)	-0.057* (0.033)	-0.068* (0.035)	-0.070* (0.040)
Medic. Indic. Risk Preg.	0.005 (0.024)	-0.008 (0.027)	-0.010 (0.029)	-0.015 (0.031)
Body-Mass-Index	-0.298 (0.449)	-0.065 (0.506)	0.356 (0.540)	0.531 (0.591)
Sum Risk Factors	-0.124 (0.193)	-0.115 (0.213)	-0.022 (0.230)	-0.081 (0.252)
Observations	603	502	430	346

Robust standard errors are shown in parentheses. Estimates include community fixed effects. See Appendices B and C for variable definitions.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Selective Attrition to Baseline - Development Tests

	Difference between Attritors/ Non-Attritors			
	3 months	6 months	12 months	24 months
	(1)	(2)	(3)	(4)
Demographic Characteristics				
Age in Years	-.986*** (0.349)	-.911*** (0.324)	-1.77*** (0.308)	-2.136*** (0.315)
Week in Pregnancy	-1.495*** (.538)	-1.720*** (0.449)	-0.856** (0.427)	-0.803* (0.429)
Underage	0.044 (0.037)	0.049 (0.031)	0.085*** (0.030)	0.104*** (0.029)
Migration	-0.026 (0.031)	-0.019 (0.027)	0.000 (0.026)	-0.015 (0.025)
Monthly HH-income (€)	-192.1*** (42.60)	-158.2*** (40.91)	-156.8*** (43.72)	-94.06** (43.97)
Debt over €3000	-0.051 (0.032)	0.001 (0.030)	-0.019 (0.028)	-0.023 (0.028)
Education Risk	0.097*** (0.034)	0.081** (0.034)	0.136*** (0.033)	0.146*** (0.037)
Income Risk	0.107*** (0.030)	0.092*** (0.028)	0.113*** (0.028)	0.057** (0.029)
Employment Risk	0.082*** (0.028)	0.055** (0.027)	0.099*** (0.027)	0.072*** (0.027)
No Partner	-0.010 (0.042)	0.021 (0.036)	-0.016 (0.034)	0.000 (0.033)
Living with Parents	-0.018 (0.041)	-0.016 (0.035)	-0.014 (0.034)	0.033 (0.033)
Persons in HH	-0.043 (0.145)	-0.016 (0.127)	0.120 (0.125)	0.264** (0.124)
Selected Psychological and Physical Characteristics				
Unwanted Pregnancy	0.018 (0.036)	0.018 (0.029)	0.015 (0.029)	0.002 (0.028)
Daily Smoking	0.043 (0.043)	0.068 (0.037)	0.062* (0.035)	0.066* (0.035)
Isolation	0.026 (0.026)	0.015 (0.022)	-0.006 (0.019)	-0.007 (0.019)
Foster Care Exper.	0.076* (0.040)	0.088*** (0.033)	0.106*** (0.031)	0.112*** (0.030)
Neglect Experience	0.050 (0.045)	0.049 (0.038)	0.040 (0.036)	0.035 (0.036)
Lost Experience	-0.037 (0.046)	0.002 (0.039)	0.028 (0.038)	0.040 (0.037)
Violence Ever	0.038 (0.028)	0.050** (0.024)	0.044** (0.022)	0.047** (0.020)
Depression	0.049 (0.033)	0.051* (0.026)	0.025 (0.024)	0.013 (0.023)
Anxiety	0.028 (0.036)	0.031 (0.029)	0.020 (0.027)	0.034 (0.027)
Stress	0.013 (0.042)	0.003 (0.036)	0.000 (0.034)	0.022 (0.034)
Aggression	0.034 (0.035)	0.026 (0.030)	0.040 (0.028)	0.009 (0.028)
Medic. Indic. Risk Preg.	0.036 (0.031)	0.002 (0.026)	-0.001 (0.025)	0.013 (0.024)
Body-Mass-Index	-0.458 (0.477)	-1.065** (0.420)	-1.109*** (0.397)	-1.423*** (0.399)
Sum Risk Factors	0.530** (0.236)	0.666*** (0.192)	0.724*** (0.181)	0.624*** (0.177)
Observations	755	755	755	755

Robust standard errors are shown in parentheses. Estimates include community fixed effects. See Appendices B and C for variable definitions.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4 Estimating Program Effects

4.1 Specification Model for Estimating Treatment Effects

I estimate the *Pro Kind* program effects on child development by OLS-regression analysis using Equation 2:

$$Y_{ic} = \beta_0 + \beta_1 HV_{ic} + \beta_2 h_{ic} + \alpha_c + \epsilon_{ic}, \tag{2}$$

where Y_{ic} is the outcome of child i in community c. HV_{ic} is a dummy variable indicating whether the child's family was home visited. h_{ic} is a vector of demographic and psychological family characteristics at baseline. I also include a dummy variable, α_c for each community to absorb the community effects. All outcomes of interest are objective measures (birth weight, birth height, and birth head circumference) or test scores by experimental assessments (MDI and PDI test scores at 6, 12, and 24 months, as well as the results of the SETK-2 at 24 months). The coefficient of interest is β_1, which indicates the size of the causal effect of the *Pro Kind* intervention. The first model in each analysis includes no controls. The second model is estimated with community fixed effects and controls for most available baseline characteristics. The results are also robust for including more or fewer control variables.

In those cases where there are missing values in the covariates, I include sample means or imputed values. However, the results are also robust if cases with missing values are excluded. For the analyses, I standardized the test scores and birth outcomes with a mean of 0 and an SD of 1. The standardization allows comparison of effects on birth outcomes and test scores, and facilitates the comparison to other early childhood interventions. I decided against clustering the standard errors at the community level, due to concerns that clustering would produce biased standard errors as the observation sizes of the clusters is greatly unbalanced.[2]

I run separate regressions for boys and girls because gender is a child's characteristic, which is unlikely to be correlated to any family characteristic. Therefore, different intervention effects between boys and girls can be fully attributed to gen-

[2] If I cluster standard errors at the community level, they are smaller at 6 months and slightly larger at 12 and 24 months.

der. Furthermore, reevaluations of preschool programs suggest that these programs benefit girls but not boys (Anderson, 2008). Such gender reevaluations are absent so far for home visiting programs.

4.2 Descriptive Statistics

In order to allow for a better interpretation of the intervention outcomes, Table 5 gives a combined overview of the birth outcomes and the results of the reliable tests for treatment and control group members. The *Pro Kind* birth outcomes are slightly lower compared to nationally representative samples of newborns (Bergmann et al., 2007). The developmental test scores reveal that the *Pro Kind* average is below the population norm of 100 points in all the tests. As expected, the *Pro Kind* eligibility criteria seem to be negatively related with test score results. At 12 months, all the test scores are closer to the norm of 100 points than at 6 months. However, at 24 months, the mean of MDI declines again. Girls score better than boys in almost all tests. However, the difference is statically significant only in MDI at 6 months at a 5-percent level (T=2.1). Using the Levene test, the variance of the test scores is not significantly different between the genders at any point. Appendices G and H present density graphs of birth outcomes and child development test scores by gender.

Table 5: Descriptive Statistics Child Outcomes

	Total		Boys		Girls	
	Mean	N	Mean	N	Mean	N
Birth Outcomes Pro Kind						
Weight (grams)	3283 (540.7)	603	3370 (526.2)	280	3210 (544.3)	321
Height (cm)	50.49 (3.17)	602	50.83 (3.15)	280	50.20 (3.18)	320
Head Circumference (cm)	34.28 (1.85)	588	34.51 (1.71)	272	34.10 (1.94)	314
6-Month Test Scores Pro Kind						
MDI	92.82 (7.91)	464	91.96 (8.45)	219	93.59 (7.32)	245
PDI	82.41 (12.35)	481	82.04 (12.88)	223	82.74 (11.90)	258
12-Month Test Scores Pro Kind						
MDI	94.22 (12.64)	393	93.90 (12.58)	187	94.50 (12.71)	206
PDI	92.67 (16.01)	374	92.75 (16.13)	169	92.61 (15.93)	205
24-Month Test Scores Pro Kind						
MDI	88.66 (14.56)	299	87.20 (14.46)	133	89.83 (14.58)	166
PDI	95.63 (13.94)	262	93.84 (14.34)	113	96.99 (13.52)	149

Standard deviation in parentheses.

4.3 Impact of *Pro Kind* on Birth Outcomes

Table 6: Impact of *Pro Kind* on Birth Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Birth Weight		Birth Height		Birth Head Circumference	
Home Visiting	0.129	0.125	0.077	0.085	0.071	0.075
	(0.081)	(0.080)	(0.082)	(0.080)	(0.083)	(0.084)
Community Fixed Effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	603	600	602	599	588	585
R^2	0.00	0.13	0.00	0.08	0.00	0.08

Notes: Robust standard errors are in parentheses. Controls include demographic, psychological, and physical baseline characteristics.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I do not find any significant effect of *Pro Kind* on birth outcomes for the whole sample (Table 6). Nevertheless, the home visiting coefficient has a positive sign for all outcomes and is close to significance at a 10-percent level for birth weight. The size of the coefficients varies only slightly with the model specifications showing that control variables are independent of the home visiting variable. Analyzing the effects separately for boys and girls reveals that boys in the home visiting group have a significantly higher birth weight than boys in the control group. However, this effect becomes insignificant when controls and mainly maternal smoking are included in the analysis. Appendix I presents density graphs of birth outcomes in the treatment and control groups.

4.4 Impact of *Pro Kind* on Child Development

My analysis of home visiting effects on cognitive abilities (MDI) or fine and gross motor abilities (PDI) begins with the whole sample (Table 7). At 6 months, all MDI coefficients are positive and get significant when controls are included. The coefficients have similar sizes for MDI and PDI. At 12 months, the MDI coefficient increases and is significant without controls. The effect for PDI is smaller than that at 6 months. At 24 months, the effect sizes decline for both MDI and PDI. While the effect for MDI is still positive, the effect for PDI becomes negative with an effect size close to zero. At all assessment points, the coefficients change only slightly when controls are included, thus confirming the validity of the randomization. Appendix J

Table 7: Impact of *Pro Kind* on Child Development

	6 Months		**12 Months**		**24 Months**	
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
Home visiting	0.141	0.173*	0.180*	0.241**	0.032	0.080
	(0.093)	(0.094)	(0.101)	(0.100)	(0.116)	(0.117)
Community Fixed Effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	464	464	393	393	299	299
R^2	0.00	0.10	0.01	0.08	0.00	0.13
	B. Psychomotor Developmental Index (PDI)					
Home visiting	0.100	0.135	0.084	0.074	-0.022	-0.014
	(0.091)	(0.092)	(0.104)	(0.106)	(0.123)	(0.129)
Community Fixed Effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	481	480	374	374	262	262
R^2	0.00	0.10	0.00	0.07	0.00	0.13

Notes: Robust standard errors in parentheses. Controls include demographic, psychological and physical baseline characteristics.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

shows the density graphs for MDI and PDI at 6, 12, and 24 months in the treatment and control groups.

Splitting the sample by gender reveals that the home visiting coefficients for the boys are close to zero or even negative for MDI and PDI at any assessment point (Table 8). In contrast, girls in the treatment group benefit strongly for MDI with an effect size of 0.3 SD at 6 and 12 months and with 0.2 SD at 24 months. The effect on PDI for girls is significant if control variables are included at 6 months, but disappears after 12 months, and becomes negative after 24 months. The differences between the models with and without controls are small for girls, but larger for boys. Appendices K and L present the density graphs for MDI and PDI at 6, 12 and 24 months in the treatment and control groups separated by gender.

4.5 Impact of *Pro Kind* on Language

The SETK-2 results (Table 9) reveal no effects of the home visiting on the language development of the whole sample. The coefficients are both, positive and negative, but they are always below 0.10 SD. However, in the production of words and sentences category, girls in the home visiting group score 0.25 SD higher than girls in the control group. This effect is significant at a 10-percent level without including

Table 8: Impact of *Pro Kind* on Child Development (Boys and Girls)

	6 Months		12 Months		24 Months	
	Basic	All Controls	Basic	All Controls	Basic	All Controls
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
	Boys					
Home Visiting	-0.027	-0.017	0.049	0.120	-0.202	-0.105
	(0.145)	(0.149)	(0.147)	(0.155)	(0.172)	(0.209)
Observations	219	219	187	187	133	133
R^2	0.00	0.15	0.00	0.12	0.01	0.12
	Girls					
Home Visiting	0.299**	0.298**	0.300**	0.281*	0.208	0.240
	(0.117)	(0.122)	(0.139)	(0.144)	(0.155)	(0.164)
Observations	245	245	206	206	166	166
R^2	0.03	0.11	0.02	0.15	0.01	0.23
	B. Psychomotor Developmental Index (PDI)					
	Boys					
Home Visiting	0.024	-0.016	-0.023	-0.116	0.029	0.119
	(0.141)	(0.134)	(0.154)	(0.157)	(0.194)	(0.276)
Observations	223	223	169	169	113	113
R^2	0.00	0.23	0.00	0.22	0.00	0.22
	Girls					
Home Visiting	0.167	0.219*	0.172	0.060	-0.068	-0.127
	(0.120)	(0.125)	(0.140)	(0.154)	(0.159)	(0.177)
Observations	258	257	205	204	149	149
R^2	0.01	0.07	0.01	0.07	0.00	0.23

Notes: Robust standard errors in parentheses. Controls include demographic, psychological and physical baseline characteristics. The treatment effects on MDI for boys and girls are significantly different at the 10 percent level at 6 months and 24 Months.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

control variables. In the other language outcomes, the home visiting effect size is also larger for the girls than for the boys.

4.6 Discussion of the *Pro Kind* Program Effects

The first result of the analyses is that the *Pro Kind* program has an impact on cognitive development (MDI) of the children already in very young age. In contrast, the intervention did not affect the psychomotor development (PDI), and affects the language skills only in a small magnitude. One explanation for the differential effects might be that the treatment only increases investments in childeren's cognitive development because these investments are less time-costly for the mothers. For example, looking at picture books and reading or telling stories, which can be undertaken at home, are correlated with cognitive development (Baker and Milligan, 2013; Price, 2012); on the other hand, activities like going with the child to the play-

Table 9: Impact of *Pro Kind* on Language Outcomes

	Understanding Words and Sentences		Production Words and Sentences		Aver. Utterance Length	
	Basic	All Controls	Basic	All Controls	Basic	All Controls
	(1)	(2)	(3)	(4)	(5)	(6)
			Full Sample			
Home Visiting	-0.08	-0.07	0.11	0.09	-0.03	-0.06
	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)	(0.12)
Observations	334	333	268	267	269	268
R^2	0.00	0.09	0.00	0.17	0.00	0.12
			Boys			
Home visiting	-0.18	-0.18	-0.10	-0.29	-0.06	-0.16
	(0.17)	(0.21)	(0.18)	(0.23)	(0.19)	(0.21)
Observations	156	156	127	127	128	128
R^2	0.01	0.18	0.00	0.22	0.00	0.32
			Girls			
Home Visiting	-0.00	0.04	0.28*	0.25	-0.00	-0.07
	(0.14)	(0.14)	(0.16)	(0.17)	(0.16)	(0.17)
Observations	178	177	141	140	141	140
R^2	0.00	0.20	0.02	0.27	0.00	0.22

Notes: Robust standard errors in parentheses. Controls include demographic, psychological and physical baseline characteristics. The treatment effects on production of words and sentences for boys and girls are significantly different at the 10 percent level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ground, for which the mother has to leave the house, is correlated with psychomotor development. This explanation is supported by the results shown in Table 10 and Appendix M which examine if the mother undertakes cognitive activities with her child. Although there is only a significant treatment effect at 24 months for reading or telling stories, at all assessment points mothers in the treatment group undertake cognitive activities more often with their child.

The second finding of the analysis is that girls benefit more from the intervention. This gender difference might be also explained by the influence of the treatment on cognitive activities. The figures in Table 10 reveal that the treatment enhances the cognitive activities more strongly for girls than for boys at five of the six assessment points. This is especially true for reading or telling stories at 6 and 24 months, where the difference between boys and girls is significant. Again, it might be that the treatment increases the investments with the lowest costs. This leads us to assume that investment in boys is more costly than investment in girls. A recent study by Baker and Milligan (2013) supports this assumption. They document for the US, the UK, and Canada that parents spend more time with girls reading, telling stories, singing songs, drawing, and teaching new words and letters, starting as early

as nine months of age. They explain that it is less rewarding to provide inputs (like reading time) to boys than to girls because boys wiggle and squirm more than girls do.

The third finding of the *Pro Kind* analysis is that the effects on cognitive development fade-out when the home visiting frequency gets lower. This fade-out might explain the small effects on language because language skills are only measured at 24 months. The fade-out is not caused by the fact that younger mothers attrite more often. At six months, the treatment effects for the children of these mothers are not higher than those for other mothers. In addition, the attrition of mothers with other characteristics does not cause the fade out. Therefore, the lesser home visit frequency seems to explain the smaller effects at 24 months. However, the effects on cognitive outcomes also fade-out in pre-school programs like the Head Start or the Perry Pre-School Program (Currie and Thomas, 1995; Anderson, 2008), although the fade-out in these programs occurs later. Despite the rapid decrease in test score gains, studies of these interventions find dramatic improvements in long-term outcomes among program participants (Deming, 2009; Campbell et al., 2002; Belfield, 2006; Anderson, 2008). Most likely, a boost of non-cognitive skills like personality traits and preferences causes these effects (Cunha and Heckman, 2008). However, these skills are even more difficult to measure objectively than the cognitive and psychomotor development during infancy.[3]

[3]For example time preference, other-regarding preferences or risk aversion is only measured for pre-school children older than three years (Mischel et al., 1989; Fehr et al., 2008; Sloviv, 1966).

Table 10: Investments in Children

	Activities with Child					
	Looking at Picture Books			Reading or Telling Stories		
	6 Mo.	12 Mo.	24 Mo.	6 Mo.	12 Mo.	24 Mo.
	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample					
Home Visiting	0.049	0.035	0.024	0.001	0.053	0.085**
	(0.052)	(0.022)	(0.022)	(0.059)	(0.050)	(0.041)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	320	374	345	320	373	343
R^2	0.12	0.11	0.06	0.08	0.09	0.08
	Boys					
Home Visiting	0.066	0.034	0.013	-0.063	0.061	0.017
	(0.081)	(0.045)	(0.032)	(0.094)	(0.086)	(0.063)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	145	171	157	146	171	155
R^2	0.24	0.27	0.20	0.19	0.22	0.13
	Girls					
Home Visiting	0.099	0.037	0.027	0.084	0.039	0.117**
	(0.078)	(0.026)	(0.030)	(0.086)	(0.071)	(0.056)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	203	188	174	202	188
R^2	0.17	0.19	0.10	0.09	0.11	0.11

Notes: Robust standard errors in parentheses. All Data is obtained in the personal interviews. Controls include demographic, psychological and physical baseline characteristics. All dependent variables are binary. The dependent variables are 1 if the mother undertakes the activity daily, several times per week or at least once a week with the child. The dependent variables are 0 if the mother does not undertake the activity with the child.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Sensitivity Analysis

Table 11: Test Scores of Children not Participating in the Next Developmental Test

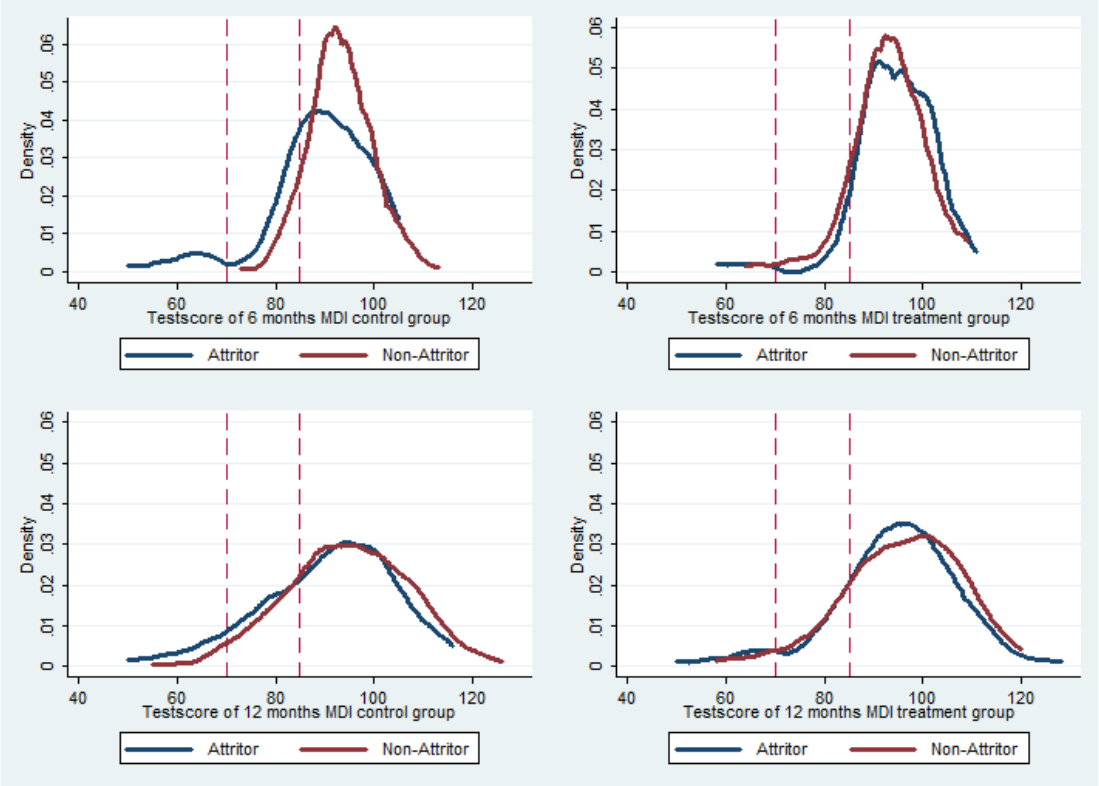
	Control Group		Treatment Group		Difference
	Test Score Attritors	n	Test Score Attritors	n	TG/CG
6-Month MDI	89.02	50	94.26	65	-5.242***
12-Month MDI	90.64	74	94.47	70	-3.836*
6-Month PDI	82.78	69	80.66	74	2.120
12-Month PDI	91.66	76	92.76	88	-1.103

P-values are based on two sided T-tests. Appendix N describes the composition of the attritors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The previous sections showed that attrition did not result in unbalanced groups with respect to baseline characteristics. Nevertheless, attrition might be selective for outcomes which are influenced by the *Pro Kind* intervention like the developmental tests. Table 11 documents that this is the case for the MDI. At six months, children of mothers who dropped out of the control group score 5.2 points lower at the MDI than children of mothers who dropped out in the treatment group.The effect is

smaller but still significant at 12 months.

Figure 1: Comparison MDI Test Scores of Attritors and Non-Attritors



This selective attrition could be caused by the procedure of the researcher sending a letter to mothers in both groups with the test results of their children. In both groups, poor results could cause frustration and skepticism towards the tests. However, mothers in the treatment group could discuss the results with their home visitors. This could reduce disaffirmation and avoid attrition. This opportunity is not given to the mothers in the control group, and, therefore, mothers of poorly performing children might drop out more often.

Figure 1 supports this hypothesis. It compares the distribution of the MDI test scores of attritors and non-attritors separately for treatment and control groups. While in the treatment group the attritors and non-attritors have an almost similar distribution, attrition in the control group is clearly focused in the range below 85 and 70 points. In this range, the letter to the mothers contains the information that their child has a developmental delay (below 85) or serious developmental delay (below 70). For developmental delay, the term *geistige Verzögerung* is used, which

is a harsh term in the German language. The attrition of all mothers in the control group of children who scored less than 70 points in the MDI at six months supports the idea that this additional information is a major reason for attrition.

If mothers with children performing poorly on the MDI attrite, one can ask why this is not the case for PDI. Mothers of children who scored low at the PDI were told in their letter that their child has movement difficulties (*Schwierigkeiten bei der Beweglichkeit*), which is a less harsh term in German language. This difference in language use in the information letters for low MDI and PDI scoring children might explain why the selective attrition does not occur for the PDI.

Table 12: Impact of Home Visiting on Children's Development in SD with imputations

	6 Months		**12 Months**		**24 Months**	
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
Home visiting	0.155	0.200**	0.259**	0.303***	0.106	0.133
	(0.096)	(0.097)	(0.101)	(0.103)	(0.111)	(0.111)
Community fixed effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	524	524	524	524	524	524
R^2	0.00	0.08	0.02	0.07	0.01	0.08

Notes: Robust standard errors in parentheses. Controls include socio-demographic, psychological and medical maternal baseline characteristics.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Impact of Home Visiting on Children's Development in SD with Imputations (Boys and Girls)

	6 Months		**12 Months**		**24 Months**	
	Basic	All controls	Basic	All controls	Basic	All controls
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
			Boys			
Home Visiting	-0.036	-0.021	0.141	0.216	-0.063	-0.033
	(0.145)	(0.148)	(0.151)	(0.160)	(0.163)	(0.170)
Observations	242	242	242	242	242	242
R^2	0.00	0.11	0.00	0.10	0.00	0.05
			Girls			
Home Visiting	0.323**	0.358***	0.362***	0.341**	0.256*	0.229
	(0.127)	(0.132)	(0.137)	(0.141)	(0.149)	(0.150)
Observations	282	282	282	282	282	282
R^2	0.02	0.10	0.04	0.12	0.02	0.16

Notes: Robust standard errors in parentheses. Controls include demographic, psychological and physical baseline characteristics. The treatment effects on MDI for boys and girls are significantly different at the 10 percent level at 6 months and 24 Months.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To correct for this selective attrition related to the MDI, I impute missing test

scores through a multiple multivariate imputation procedure (Royston, 2004). I only impute scores for children who participated in at least one test. In most cases, attrition is linear in a way that children participate in the first test, or tests and then refuse to participate. However, there also cases in which children just miss the first or second test (Appendix N). Overall, this leads to imputed values for 524 children. I include the baseline characteristics, and interactions between baseline characteristics and treatment group in the imputation regression, which is repeated 300 times. For the gender-wise analysis, the imputation is conducted separately for boys and girls.

The imputation increases the effects of MDI at all three assessment points (Table 12). As in section 4.4 the effect is highest at 12 months and reduce at 24 months. The gender difference after imputation reveals the same picture as without the imputation. The effect of the home visiting is greater for girls than for boys (see Table 13). At 12 months, girls in the treatment group score 0.36 SD higher than girls in the control group. These results demonstrate how sensitively the effect sizes react to the selective attrition of the bad performing children in the control group. Although the results in Table 12 might be exaggerated because of the imputation assumptions, it is likely that the estimates with the raw data in Table 7 give the lowest bound of the treatment effects.

6 Skill Formation Dynamics

The *Pro Kind* experiment gives the unique possibility to analyze the skill formation process in children's first two years. The *Pro Kind* data is unique in the respect that all other studies about skill formation, which I am aware of, collect data later in children's lives or less frequently in the first two years (Cunha et al., 2006). The knowledge about this very early skill formation can examine whether the dynamics in the skill formation process, as predicted by Cunha and Heckman (2007), occur already at this early stage. If not, the efficacy of the programs which try to enhance skills at such an early stage must be reconsidered. Furthermore, the insights can shed light on the mechanisms of how home visiting generates effects, and why these effects occur with girls but not with boys.

In accordance with Cunha and Heckman (2007), self-productivity as well as direct and dynamic complementarity are the components through which skills beget skills and abilities beget abilities, and therefore, they are the dynamic factors in the skill production function. Equation 3 formalizes this skill production function, where S_t denotes the vector of skills acquired at stage t.

$$S_{t+1} = f_t(h, S_t, HV) \tag{3}$$

As in Equation 2, h is defined as demographic and psychological family characteristic at baseline. Cunha and Heckman (2007) propose to include family investment in the production function. I use the home visiting variable HV as a proxy for family investment. Self-productivity in the skill formation process arises when

$$\frac{\partial S_{t+1}}{\partial S_t} = \frac{\partial f_t(h, S_t, HV)}{\partial S_t} > 0, \tag{4}$$

that is, when higher stocks of skills in one period create higher stocks of skills in the next period. In accordance with self-productivity, direct complementarity applies if one set of skills is productive for the formation of other skills in previous periods, and vice versa. The following investigation methods are based on Blomeyer et al. (2009) and Coneus et al. (2012), who also analyzed early childhood skill formation

in the German context with data of the *Mannheim Risiko Studie* (MARS)[4].

I use four stages in my approach. My basic estimation equation for all four stages is a linear representation of the skill production function described in Equation 3. In Equation 5, $S^k_{t,i}$ denotes the skill indicator in t, and $S^k_{t+1,i}$ denotes skills k acquired in the next period. At stage t_1 birth weight is the measure for S^k_i, at stages t_2,t_3, and t_4, I use MDI and PDI test scores 6, 12, and 24 months as a measure for S^k_i.

$$S^k_{t+1,i} = \gamma S^k_{t,i} + \phi HV + \eta h + \epsilon_{i,t} \tag{5}$$

My coefficients of interest are γ, indicating self-productivity or direct complementarity and ϕ, indicating the effects of the home visiting investment. All variables are standardized as explained in Chapter 4.

Table 14 summarizes the results. For the whole sample, I find self-productivity for MDI and PDI at every stage. The coefficients for self-productivity rise gradually indicating that skills get more stable with age. Direct complementarity appears only at stage 3, where MDI at 24 months increases by 0.14 SD, if PDI increases by one SD at 12 months. If I separate the sample by gender, the picture changes. For boys, I find no self-productivity for MDI at stage 2, and no self-productivity for PDI at stage 3. Instead of self-productivity, I find direct complementarity of 6-month PDI for 12-month MDI. For girls, self-productivity is sustained at all stages, with direct complementarity occurring as well. The HV coefficients report the net impact of home visiting in each stage, because the estimates are controlled for the impact of home visiting in previous stages. All net impact coefficients on PDI and MDI are smaller than estimated in Table 7 and Table 8, with the exception of the coefficient for boys for MDI at stage 2.

The results of self-productivity for the whole sample are in line with the results of Coneus et al. (2012), who find significant self-productivity for the MDI from three months to two years with a coefficient of 0.3. However, Coneus et al. (2012) found only small gender-specific differences with respect to the skill formation process. The more detailed *Pro Kind* data reveals that boys younger than 12 months do

[4]MARS is a longitudinal epidemiological cohort study following at risk infants from birth to adulthood. The initial sample contains 382 children born between February 1986 and February 1988 (Laucht et al., 1997).

not benefit in the next period by an increase of the cognitive skills in the previous period. Furthermore, the investigation with the *Pro Kind* data gives a new insight that already from the age of six months, self-productivity gradually increases and that on the other hand, direct complementarity develop when the child is older than one year.

For the effectiveness of home visiting programs, the results give answers but also raise new questions. First, the hypothesis that interventions that start prenatally or at infant age have the highest lifelong effects seems to be valid. This is shown by the size of the self-productivity coefficients, which demonstrate that the dynamic nature of skill formation already occurs at infant age. Second, as direct complementarity is low at this age, home visiting has to concentrate on each skill separately, if it wants to affect each skill. Third, the coefficient of HV indicates that the main reason for the insufficient effect for boys lies in the first six months of the home visiting. In contrast, at 12 months the net effect is comparable with that of the girls. Here, the question remains if the small effects of the intervention for boys are related to the fact that there is no skill self-productivity between 6 and 12 months for boys. This is a question for further interdisciplinary research which also has to examine whether home visiting should intervene differently for mothers of boys at this age.

Table 14: Estimates of the Skill Production Function with Two Skill Factors and Home Visiting as Investment

	Whole Sample			Boys			Girls		
	MDI t-1	PDI t-1	HV	MDI t-1	PDI t-1	HV	MDI t-1	PDI t-1	HV
t = 24 Months									
MDI	0.41***	0.13*	0.02	0.39***	0.17	0.01	0.27***	0.20**	0.08
PDI	0.09	0.34***	-0.04	0.35*	0.20	-0.23	-0.06	0.33***	-0.01
t = 12 Months									
MDI	0.28***	0.06	0.20**	0.13	0.19**	0.25	0.35***	-0.02	0.21
PDI	0.10	0.41***	0.01	-0.01	0.44***	-0.01	0.18**	0.43***	-0.08
	Birth Weight		HV	Birth Weight		HV	Birth Weight		HV
t = 6 Months									
MDI	0.20***		0.16*	0.24**		-0.07	0.13		0.30***
PDI	0.24***		0.13	0.12		-0.02	0.31***		0.24**

Notes: Robust standard errors in parentheses. Controls include demographic, psychological and physical baseline characteristics. The estimates include all observations for which data is available at two subsequent assessment points.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 Conclusion

Antenatal home visiting for disadvantaged families are shown to be effective for child's development in the long run. The analysis of the *Pro Kind* project using objective assessments instead of screening methods as measures for child development demonstrates that effects on child development also occur at infant age. The results suggest a better cognitive development at the age of 12 months. However, the program effects on cognitive development are concentrated on girls, who achieve higher test scores at 6, 12, and 24 months than their counterparts in the control group, whereas there are no differences between the groups for boys. As an explanation for the gender-specific effects, I can show that the treatment increases investments for boys and girls of differential magnitude. The program does not affect psychomotor development, and the effects on cognitive development fade-out at 24 months. The findings of gender differences in cognitive development and the fade-out of these effects are in line with reevaluations of other early childhood interventions like the Perry Preschool program, where the intervention is exclusively effective for girls and the cognitive effects fade-out over time.

The effects of *Pro Kind* on child development are robust to several specifications and increase dramatically, when missing observations were imputed by multivariate imputation methods. I estimate models with different family baseline characteristics and community fixed effects as controls. The home visiting effect is hardly influenced by any specification. The results show the advantage of a longitudinal design compared to a design with only on observation point in which either the effect at 12 months or the fade-out at 24 months would not have been found. I also investigated the dynamic nature of the skill formation process because of its importance for the interpretation of the effect sizes. I showed that self-productivity is present at all stages. I do not find direct complementarity between MDI and PDI. After estimating separate models for boys and girls, I find strong differences in the skill formation process, which could explain some of the gender differences in the effectiveness of *Pro Kind* for cognitive development.

The *Pro Kind* intervention is relatively costly (on an average €8,705, which is

approximately US-$ 11,752 per intervention (Maier-Pfeiffer et al., 2013)). Nevertheless, the *Pro Kind* effects on child mental development could still generate a positive benefit-cost ratio because of the dynamic nature of the skill formation process. For example, Coneus et al. (2012) find in a German sample of children that the cognitive development measured by the BSID at 24 months is strongly related to high school graduation. Furthermore, the meaning of the effect size is enlarged because the home visitors do not directly interact with the child; rather they interact with the mothers. Thus, it is likely that the mother uses the acquired skills in other aspects of life as well for her own or her child's health, or in the planning of her own life course. Additionally, there could be spill-over effects of the acquired skills for the second child. If these effects are considered, the *Pro Kind* effects on early cognitive development will clearly lead to a benefit-cost ratio greater than one. This conclusion is supported by benefit-cost analyses of the pre-school program, Head Start. For this program, Ludwig and Phillips (2007) estimated that effect sizes around 0.1 to 0.2 SD on cognitive development would be sufficient to pass benefit-cost tests. In another study about Head Start, Deming (2009) reveals that an effect size of 0.06 SD on cognitive development is enough to reach the break-even point with program costs of US-$ 6000 and a 3 percent discount rate.

References

Almond, D. and J. Currie (2011). Human Capital Development before Age Five. In Orley Ashenfelter and David Card (Ed.), *Handbook of Labor Economics*, pp. 1315–1486. Elsevier.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Baker, M. and K. Milligan (2013). Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries. *NBER Working Paper 18893*.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Journal of Psychology* (37), 122–147.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: W.H. Freeman.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Cooperation.

Bayley, N. (1993). *Manual for the Bayley Scales of Infant Development (2nd Edition)*. San Antonio: Psychological Corporation.

Belfield, C. R. (2006). The High/Scope Perry Preschool Program: Cost-benefit Analysis Using Data from the Age-40 Followup. *Journal of Human Resources 41*(1), 162–190.

Bergmann, K. E., R. L. Bergmann, U. Ellert, and J. W. Dudenhausen (2007). Perinatal Risk Factors for Long-Term Health. Results of the German Health Interview and Examination Survey for Children and Adolescents (KIGGS). *Bundesgesundheitsblatt 50*(5-6), 670–676.

Berlin, L. J., K. Appleyard, and K. A. Dodge (2011). Intergenerational Continuity in Child Maltreatment: Mediating Mechanisms and Implications for Prevention. *Child Development 82*(1), 162–176.

Blomeyer, D., K. Coneus, M. Laucht, and F. Pfeiffer (2009). Initial Risk Matrix, Home Resources, Ability Development, and Children. *Journal of the European Economic Association 7*(2), 638–648.

Bowlby, J. (1969). *Attachment and Loss: Vol. 1 Attachment*. New York: Basic Books.

Bronfenbrenner, U. (1992). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge and Mass: Havard University Press.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-John (2002). Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. *Applied Developmental Science 6*(1), 42–57.

Coneus, K., M. Laucht, and K. Reuß (2012). The Role of Parental Investments for Cognitive and Noncognitive Skill Formation—Evidence for the first 11 Years of Life. *Economics & Human Biology 10*(2), 189–209.

Cunha, F. and J. Heckman (2007). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. Heckman, L. Lochner, and D. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, pp. 697–804. Amsterdam and London: North-Holland.

Currie, J. and D. Thomas (1995). Does Head Start Make a Difference? *American Economic Review 85*(3), 341–364.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics 1*(3), 111–134.

Doyle, O., C. Harmon, J. Heckman, C. Logue, and S. Moon (2013). Measuring Investments in Human Capital Formation: An Experimental Analysis of Early Life Outcomes. *NBER Working Paper 19316*.

Doyle, O., C. Harmon, J. Heckman, and R. Tremblay (2009). Investing in Early Human Development: Timing and Economic Efficiency. *Economics & Human Biology 7*(1), 1–6.

Eckenrode, J., M. Campa, D. Luckey, Ch. Henderson, R. Cole, H. Kitzman, E. Anson, K. Sidora-Arcoleo, J. Powers, and D. Olds (2010). Long-term Effects of Prenatal and Infancy Nurse Home Visitation on the Life Course of Youths. *Archives of Pediatrics and Adolescent Medicine 164*(1), 9–15.

Fehr, E., H. Bernhard, and B. Rockenbach (2008). Egalitarianism in Young Children. *Nature 454*(7208), 1079–1083.

Florida Institute of Education (2004). *Birth to Three Screening and Assessment Resource Guide*. Jacksonville.

Gomby, D. S. (2005). *Home Visitation in 2005: Outcomes for Children and Parents*. Washington DC.

Heckman, J. J. (2011). The American Family in Black and White: A Post-Racial Strategy for Improving Skills to Promote Equality. *Daedalus 140*(2), 70–89.

Jungmann, T., V. Kurtz, T. Brand, S. Sierau, and K. v. Klitzing (2010). Präventionsziel Kindergesundheit im Rahmen des Modellprojektes Pro Kind: Vorläufige Befunde einer längsschnittlichen, randomisierten Kontrollgruppenstudie. *Bundesgesundheitsblatt 53*, 1180–1187.

Jungmann, T., Y. Ziert, V. Kurtz, and T. Brand (2009). Preventing adverse developmental outcomes and early onset conduct problems through prenatal and infancy home visitation: The German Pilot Project Pro Kind. *European Journal of Developmental Science 3*(3), 292–298.

Laucht, M., G. Esser, and H. Schmidt (1997). Developmental Outcomes of Infants

Born with Biological and Psychosocial Risks. *Journal of Child Psychology and Psychiatry 38*, 843–854.

Ludwig, J. and D. Phillips (2007). The Benefits and Costs of Head Start. *NBER Working Paper 12973*.

Maier-Pfeiffer, A., S. Hartmann, K. Adamaszek, R. Schneider, G. Helm, M. Refle, S. Geisler, and T. Brand (2013). Auswahl der Fachkräfte. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 61–69. Weinheim: Beltz Juventa.

McAffee, O. and D. Leong (1997). *Assessing and Guiding Young Children's Development and Learning.* (Second Edition ed.). Bosto: Allyn & Bacon.

Mischel, W., Y. Shoda, and M. I. Rodriguez (1989). Delay of Gratification in Children. *Science 244*, 933–938.

Olds, D., C. Henderson, H. Kitzman, J. C. R. Eckenrode, and Tatelbaum R.C (1999). Prenatal and Infancy Home Visitation by Nurses: Recent Findings. *The Future of Children* (9), 44–63.

Olds, D. L. (2006). The Nurse–Family Partnership: An Evidence-Based Preventive Intervention. *Infant Mental Health Journal 27*(1), 5–25.

Price, J. (2012). The Effects of Parental Time Investments: Evidence From Natural Within Family Variation. *BYU Working Paper*.

Royston, P. (2004). Multiple Imputation of Missing Values. *The Stata Journal 4*(3), 227–241.

Salkind, N. J. (2010). *Encyclopedia of Research Design.* Thousand Oaks and Calif: Sage.

Sloviv, P. (1966). Risk-Taking in Children: Age and Sex Differences. *Child Development 37*(1), 169–176.

Sweet, M. A. and M. I. Appelbaum (2004). Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families with Young Children. *Child Development 75*(5), 1435–1456.

Wu, S., C. Ma, R. Carter, M. Ariet, E. Feaver, M. Resnick, and J. Roth (2004). Risk Factors for Infant Maltreatment: A Population-Based Study. *Child Abuse & Neglect 28*(12), 1253–1264.

Appendix A: Randomization Outcomes per Municipality

Federal State	Community	CG	TG	Enrollment Period
Lower Saxony	Braunschweig	26	32	1.11.2006 - 30.4.2009
	Celle	15	25	
	Garbsen	10	12	
	Göttingen	12	13	
	Laatzen	4	4	
	Wolfsburg	11	15	
	Hannover	54	52	
Bremen	Bremen	77	83	15.4.2007 - 15.3.2009
	Bremerhaven	31	29	
Saxony	Leipzig	36	44	1.1.2008 - 31.12.2009
	Plauen	13	18	
	Muldentalkreis	16	12	
	Dresden	46	43	
	Vogtlandkreis	10	12	
$\sum$		**361**	**394**	

Appendix B: Baseline Variable Definitions - Demographic Characteristics

Variable	**Type**	**Description**	**n**
Age in Years	Metric	Participants' Age in Years at Baseline	755
Week in Pregnancy	Metric	Week in Pregnancy at Randomization	755
Underage	Binary	1 if Participant is Younger than 18 Years	755
Migration	Binary	1 if Participant is not Born in Germany or has no German Nationality	755
Monthly HH-Income (€)	Metric	Monthly Net-Income in Participants' Household	647
Debt over €3000	Binary	1 if Debt is over €3000 in Participants' Household	728
Education Risk	Binary	1 if Participant has less than 11 Years of Schooling	755
Income Risk	Binary	1 if Net-Income is below €1250 in Participants' Household	647
Employment Risk	Binary	1 if Participant has no Regular Employment	755
No Partner	Binary	1 if Participant is in a Partnership	755
Living with Parents	Binary	1 if Participant Lives in her Parents Household	751
Persons in HH	Metric	Number of Persons in Participants' Household at Baseline	737

Appendix C: Baseline Variable Definitions - Psychological and Physical Characteristics

Variable	**Type**	**Description**	**n**
Unwanted Pregnancy	Binary	1 if Participant States that Pregnancy was Unwanted	747
Daily Smoking	Binary	1 if Participant Smokes Daily	755
Isolation	Binary	1 if Participant has Infrequently Contact to Friends or Relatives	747
Foster Care Experience	Binary	1 if Participant Lived at Least Once in a Foster Family or Foster Care	735
Neglect Experience	Binary	1 if Indication of Neglect Experience during Childhood	730
Lost Experience	Binary	1 if Participant Lost an Attachment Figure due to Death or Divorce	736
Violence Experience	Binary	1 if Participant ever Experienced Violence in her Life	751
Depression	Binary	1 if Value higher 20 for Depression on the Depression Anxiety Stress Scale (DASS)	749
Anxiety	Binary	1 if Value higher 15 on Anxiety on the DASS	744
Stress	Binary	1 if Value higher 25 on Stress on the DASS	749
Aggression	Binary	1 if Value higher 10 on the *Fragebogen zur Erfassung von Aggressivitätsfaktoren (FAF)*	743
Medically Indicated Risk Preg.	Binary	1 if participant has physical problems or if participant is older than 35	724
Body-Mass-Index	Metric	Participants' $Weight/Height^2$ (Weight Before Pregnancy)	750
Sum Risk Factors	Metric	Sum of Risk Factors	755

Appendix D: *Pro Kind* Locations

Note: Orange points indicate locations in Lower Saxony, yellow points in Bremen and red points in Saxony.

Appendix E: Selective Attrition between TG and CG - Reliable MDI Tests

	Difference TG/CG for MDI		
	6 Months	12 Months	24 Months
	(1)	(2)	(3)
Demographic Characteristics			
Age in Years	0.014 (0.416)	0.047 (0.465)	0.169 (0.546)
Week in Pregnancy	-0.247 (0.539)	-0.326 (0.579)	0.129 (0.681)
Migration	-0.054* (0.033)	-0.050* (0.035)	-0.029 (0.042)
Underage	-0.002 (0.035)	0.011 (0.036)	0.031 (0.039)
Mon. HH-Inc. in €	-0.279 (0.533)	0.622 (0.560)	0.425 (0.597)
Debt over 3000 €	0.021 (0.035)	0.034 (0.040)	0.054 (0.047)
Education Risk	0.037 (0.040)	0.040 (0.045)	0.047 (0.054)
Income Risk	0.019 (0.038)	0.014 (0.043)	0.000 (0.045)
Employment Risk	-0.029 (0.037)	-0.053 (0.041)	-0.019 (0.047)
No Partner	0.015 (0.042)	0.051 (0.045)	0.002 (0.053)
Living with Parents	-0.012 (0.041)	0.026 (0.045)	-0.038 (0.049)
Persons in HH	0.060 (0.152)	0.117 (0.162)	0.024 (0.176)
Selected Psychological and Physical Characteristics			
Unwanted Pregnancy	0.014 (0.034)	0.045 (0.038)	0.011 (0.042)
Daily Smoking	-0.013 (0.043)	0.035 (0.047)	-0.007 (0.054)
Isolation	-0.020 (0.023)	0.002 (0.026)	0.014 (0.032)
Foster Care Exper.	0.017 (0.036)	0.038 (0.039)	0.084** (0.042)
Neglect Experience	-0.004 (0.045)	0.007 (0.050)	0.019 (0.057)
Lost Experience	-0.048 (0.046)	-0.081 (0.051)	-0.096 (0.058)
Violence Ever	-0.026 (0.024)	-0.028 (0.027)	-0.036 (0.030)
Depression	-0.014 (0.028)	0.019 (0.031)	0.015 (0.035)
Anxiety	0.025 (0.033)	0.025 (0.038)	0.003 (0.040)
Stress	0.037 (0.043)	0.044 (0.048)	0.034 (0.053)
Aggression	-0.065* (0.034)	-0.085** (0.037)	-0.042 (0.042)
Medic. Indic. Risk Preg.	-0.011 (0.029)	-0.012 (0.032)	-0.018 (0.035)
Body-Mass-Index	-0.024 (0.529)	0.605 (0.558)	0.583 (0.647)
Sum Risk Factors	-0.172 (0.221)	-0.107 (0.237)	-0.131 (0.264)
Observations	464	393	299

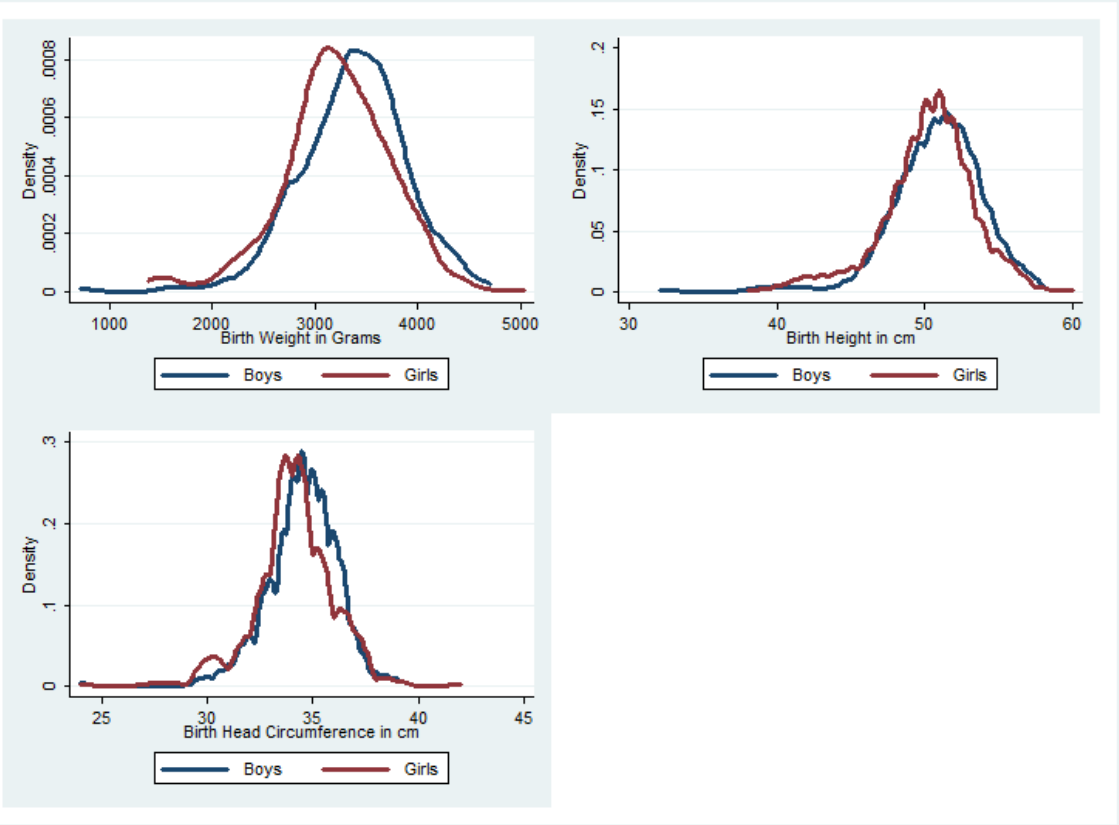
Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Appendix B and C for variable definitions.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix F: Selective Attrition between TG and CG - Reliable PDI Tests

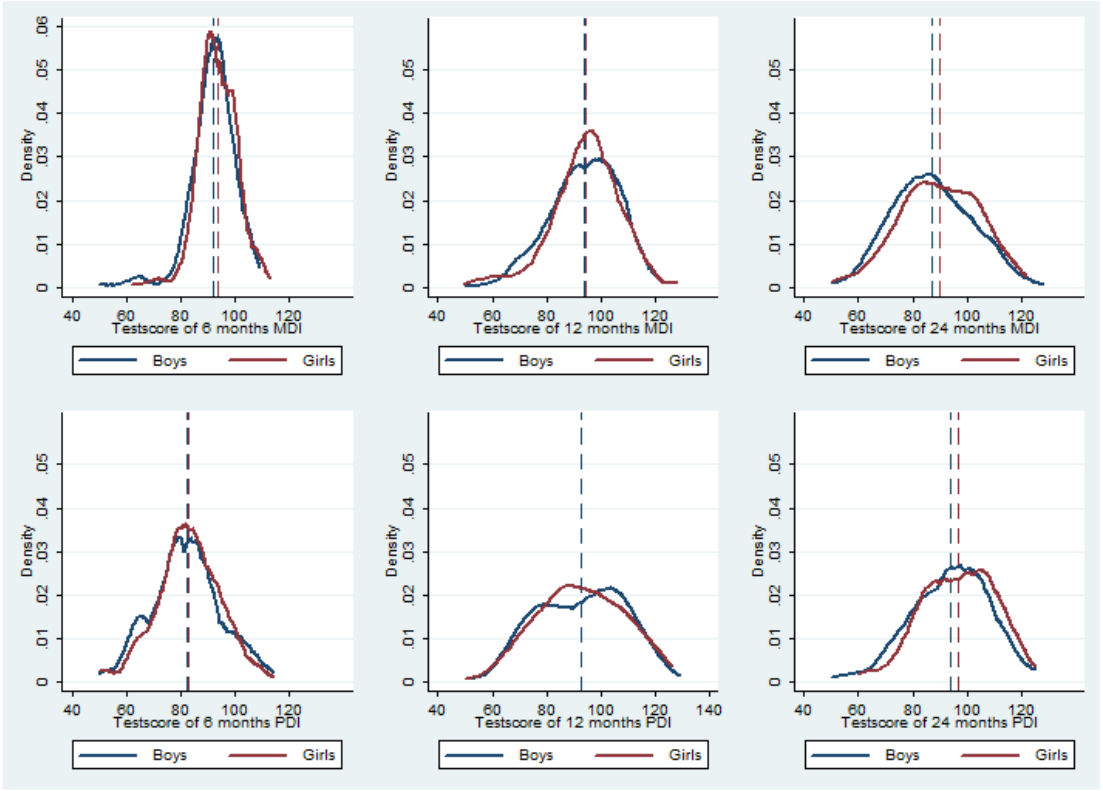
	Difference TG/CG for PDI		
	6 Months	12 Months	24 Months
	(1)	(2)	(3)
Demographic Characteristics			
Age in Years	0.016 (0.395)	-0.034 (0.471)	0.175 (0.589)
Week in Pregnancy	-0.379 (0.519)	-0.499 (0.600)	0.057 (0.708)
Migration	-0.056* (0.032)	-0.061* (0.037)	0.018 (0.043)
Underage	0.003 (0.035)	0.040 (0.037)	0.015 (0.040)
Mon. HH-Inc. in €	-0.279 (0.533)	0.181 (0.571)	-.68 (0.613)
Debt over 3000 €	0.021 (0.035)	0.043 (0.040)	0.075 (0.052)
Education Risk	0.037 (0.040)	0.032 (0.047)	0.077 (0.059)
Income Risk	0.025 (0.037)	-0.009 (0.043)	0.019 (0.049)
Employment Risk	-0.036 (0.035)	-0.044 (0.041)	0.020 (0.049)
No Partner	0.014 (0.041)	0.024 (0.047)	0.005 (0.057)
Living with Parents	-0.012 (0.040)	0.025 (0.045)	-0.001 (0.053)
Persons in HH	0.087 (0.147)	0.112 (0.161)	0.022 (0.181)
Selected Psychological and Physical Characteristics			
Unwanted Pregnancy	0.010 (0.034)	0.057 (0.038)	0.022 (0.048)
Daily Smoking	-0.018 (0.043)	0.022 (0.049)	0.001 (0.058)
Isolation	-0.015 (0.021)	-0.015 (0.028)	0.024 (0.033)
Foster Care Exper.	0.020 (0.036)	0.037 (0.040)	0.069 (0.044)
Neglect Experience	0.012 (0.045)	-0.003 (0.051)	0.032 (0.062)
Lost Experience	-0.054 (0.045)	-0.059 (0.052)	-0.106 (0.062)
Violence Ever	-0.016 (0.024)	-0.025 (0.026)	-0.047 (0.032)
Depression	-0.002 (0.027)	0.014 (0.032)	0.010 (0.037)
Anxiety	0.025 (0.033)	0.040 (0.039)	-0.009 (0.045)
Stress	0.053 (0.042)	0.037 (0.049)	0.073 (0.058)
Aggression	-0.058* (0.034)	-0.088** (0.039)	-0.052 (0.046)
Medic. Indic. Risk Preg.	-0.013 (0.029)	-0.003 (0.033)	0.013 (0.035)
Body-Mass-Index	-0.021 (0.518)	0.448 (0.593)	1.066 (0.650)
Sum Risk Factors	-0.120 (0.216)	-0.086 (0.252)	-0.031 (0.283)
Observations	481	374	262

Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Appendix B and C for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

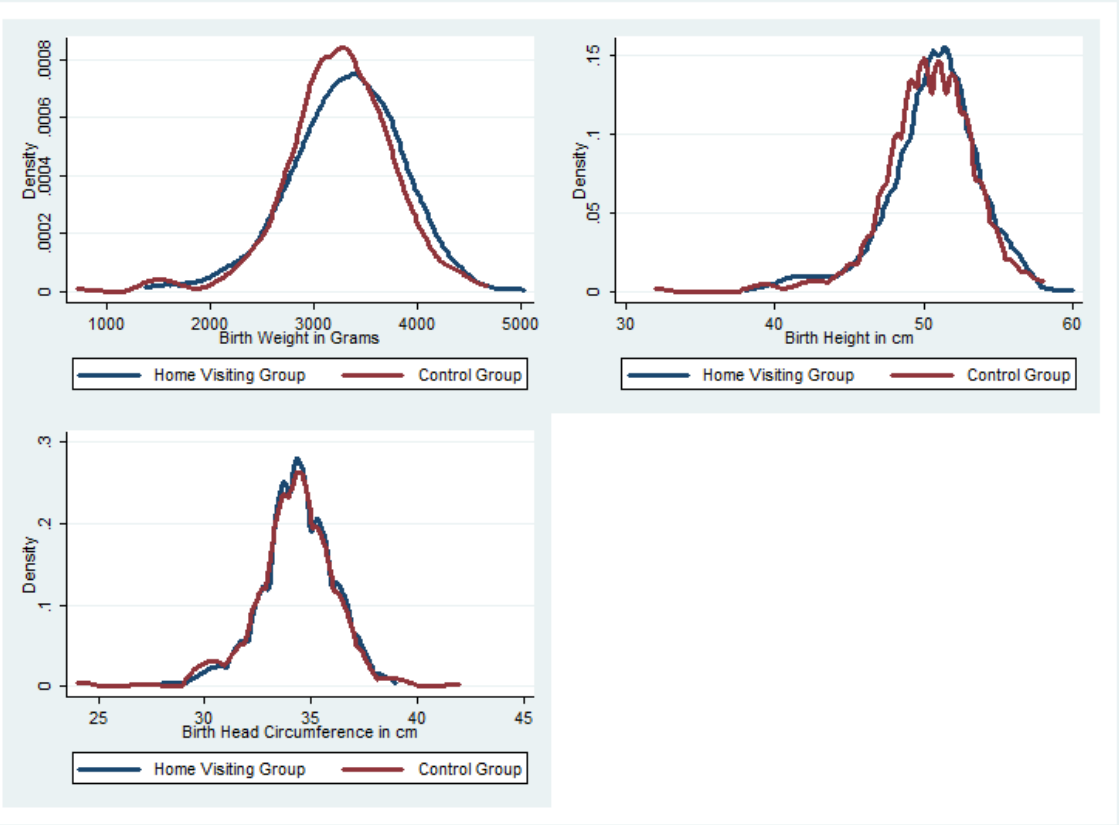
Appendix G: Birth Outcomes for Boys and Girls



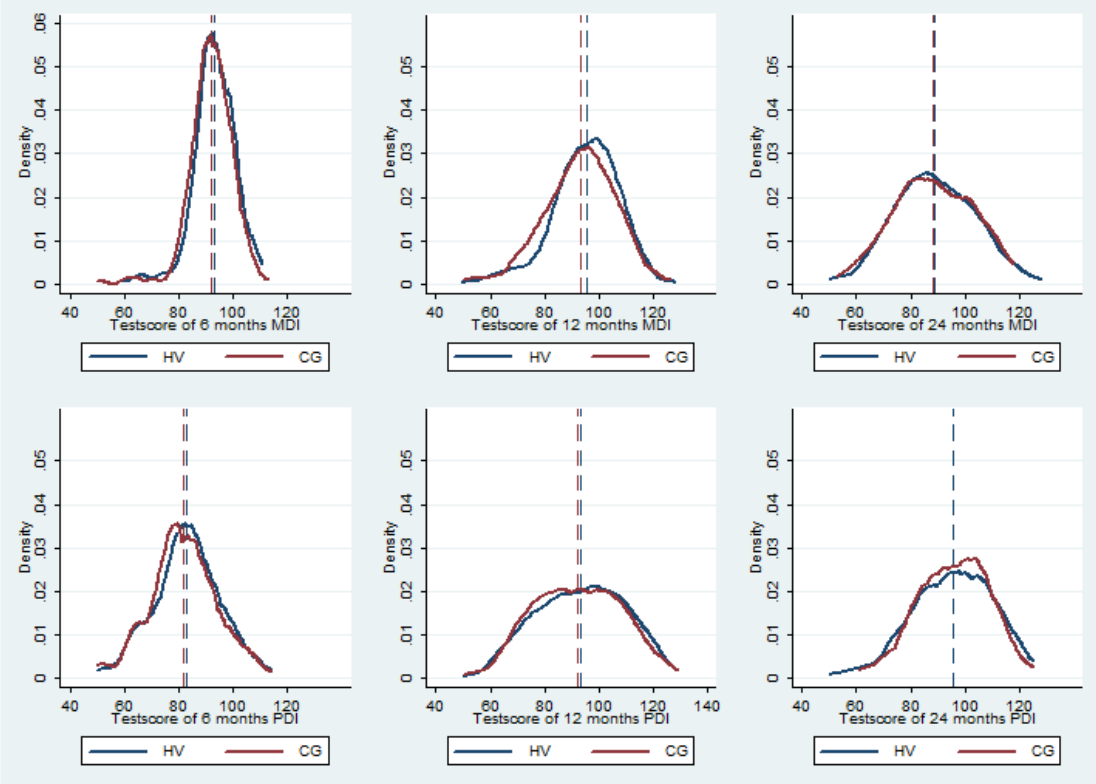
Appendix H: BSID Test Scores for Boys and Girls



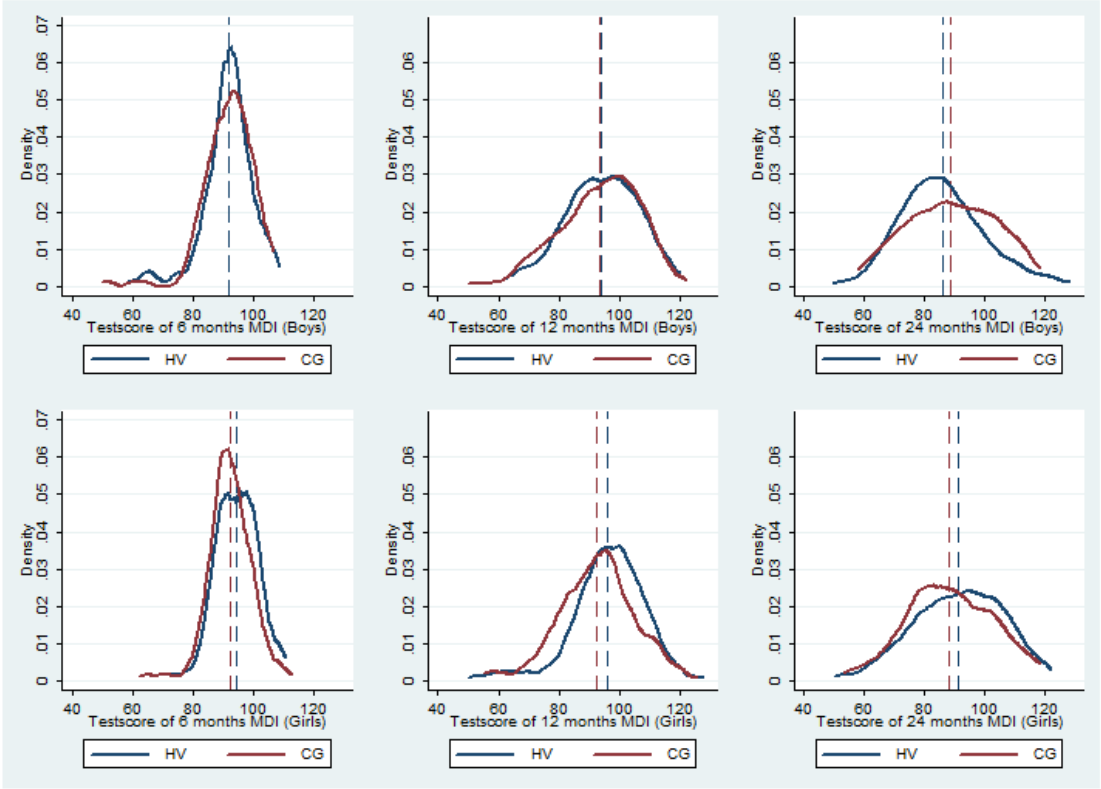
Appendix I: Birth Outcomes for Treatment and Control Group



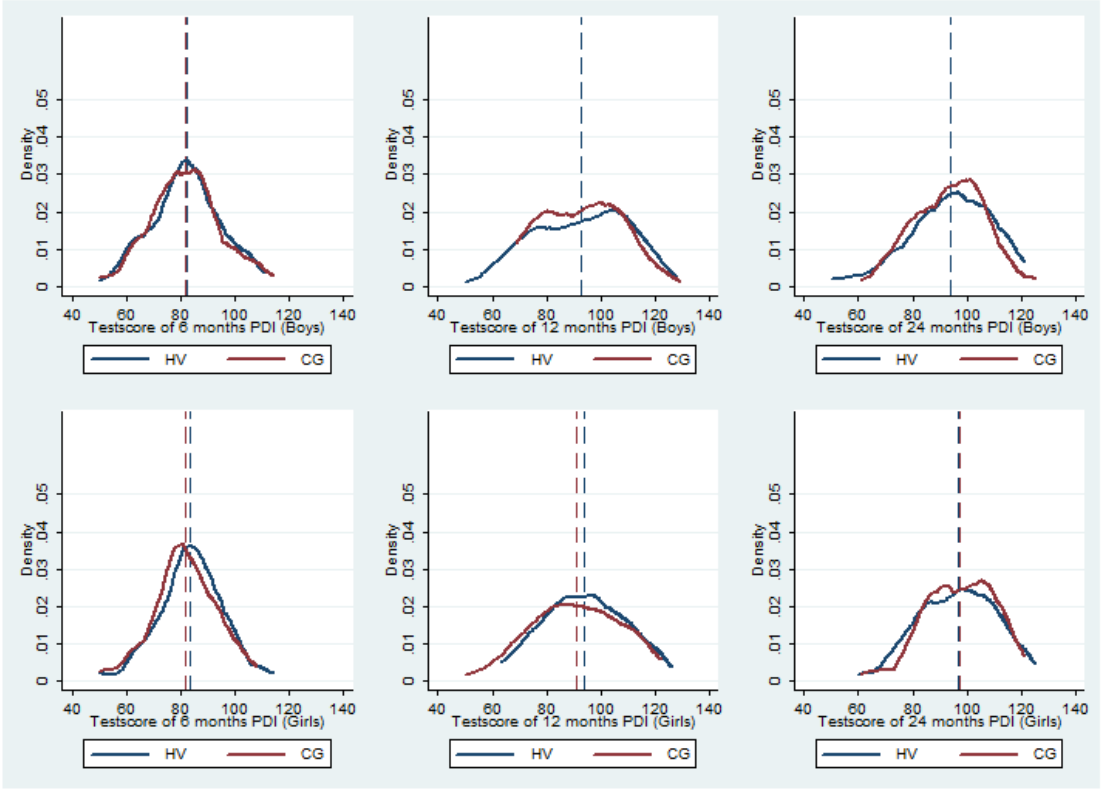
Appendix J: BSID Test Scores for Treatment and Control Groups



Appendix K: MDI Test Scores for Treatment and Control Groups (Boys and Girls)



Appendix L: PDI Test Scores for Treatment and Control Groups (Boys and Girls)



Appendix M: Investments in Children - Descriptive Statistics

	Total		**Boys**		**Girls**	
	CG	TG	CG	TG	CG	TG
At 6 Months						
Looking at Picture Books	0.70	0.74	0.77	0.73	0.63	0.74
Reading or Telling Stories	0.49	0.51	0.53	0.46	0.46	0.56
Observations	*149*	*171*	*70*	*75*	*79*	*96*
At 12 Months						
Looking at Picture Books	0.93	0.97	0.90	0.97	0.95	0.98
Reading or Telling Stories	0.69	0.75	0.69	0.76	0.69	0.75
Observations	*149*	*199*	*81*	*90*	*94*	*109*
At 24 Months						
Looking at Picture Books	0.96	0.98	0.96	0.98	0.96	0.98
Reading or Telling Stories	0.81	0.89	0.86	0.85	0.77	0.92
Observations	*168*	*177*	*75*	*81*	*92*	*96*

Notes: All data are obtained from the personal interviews. All dependent variables are binary. The figures give the rate of mothers who look at picture books with their children or read or tell stories to their children daily, several times per week or at least once a week.

Appendix N: Distribution of MDI Tests

	Number of MDI Tests at		
	6 Months	12 Months	24 Months
MDI Tests at 6, 12 and 24 Months	228	228	228
MDI Tests at 6 and 12 Months	121	121	-
MDI Tests at 6 and 24 Months	34	-	34
MDI Tests at 12 and 24 Months	-	21	21
MDI Test only at 6 Months	81	-	-
MDI Test only at 12 Months	-	23	-
MDI Test only at 24 Months	-	-	16
$\sum$	464	393	299