The effects of early childhood intervention on child development and early skill formation. Evidence from a randomized experiment.

Malte Sandner *

Leibniz Universität Hannover
Wirtschaftswissenschaftliche Fakultät
October 25th, 2012

Do not cite or circulate without permission.

Abstract

This paper presents results from a randomized evaluation of a home visiting program for disadvantaged first-time mothers and their families implemented in three German federal states. By the end of the first year of the program, children in home visited families perform significantly better than those in the control families by 0.18 standard deviations in the Mental Developmental Index. Examination of gender differences revealed that home visited girls scored 0.30 standard deviations higher than girls in the control families, whereas boys scored similar in both groups. The effects faded out after 24 months. However, sensitivity analyses show strong evidence that the estimated effects are downward biased by additional treatment for the control families. Analyzing the infant skill formation process reveals self productivity of skills but in different magnitude for boys and girls. Furthermore, I analyze possible monetary returns of the program.

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.

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: Firstly, 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). Secondly, 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). Thirdly, 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 prenatal, and which alter parenting behavior are supposed to have a lasting effect on children's life outcomes and can produce high cost-benefit ratios.

Home visiting is a type of early intervention which can fulfill these requirements. In the high quality versions of home visiting, trained midwifes, nurses or social pedagogues visit disadvantaged families at their own home starting already during pregnancy. These home visitors typically interact with the parents to encourage and train them how to raise their children. Evidence from meta-analyses including all varieties of home visiting, e.g. programs which start after birth, shows that home visiting has a modest effect on improving child development. (Sweet and Appelbaum, 2004). High quality home visiting, concentrating on disadvantaged families and starting during pregnancy, appears to be more effective for child development (Olds et al., 1999; Gomby, 2005). The few existing studies on long-term effects show that the results on child development are stable over time (Eckenrode et al., 2010).

However, up until now only medical scientists or psychologists have investigated this promising type of early childhood intervention; whereas economic research has so far neglected this topic. Therefore, previous research fails to consider efficiency questions and to investigate the influence of home visiting on skill formation dynamics. Furthermore, the previous research on high quality home visiting mainly refers to the US or developing countries. The outcomes could be different in continental European countries due to a higher degree of health insurance coverage, higher welfare payments and a system of mandatory doctor visits during pregnancy.

This paper provides an econometric analysis of the first randomized experiment on high quality home visiting conducted in Germany, the *Pro Kind* Project. The *Pro Kind* Project is a longitudinal study in which disadvantaged first-time mothers in three federal states are randomly assigned to either a treatment group with home visits both during pregnancy and the following two years, or a control group. The home visits are conducted by midwives, nurses or social pedagogues. The frequency of the home visits varies between weekly and bi-weekly. 755 mothers are involved

in the project. All of the mothers receive welfare benefits or have other financial restrictions and they additionally possess a psychological risk characteristic. Trained research assistants conducted reliable video-controlled mental and psychomotoric child development tests at the age of 6, 12 and 24 months and a language test at 24 months. Personal interviews and hospital data provide information about birth outcomes. The obtained data is unique in the respect that all other studies of early childhood interventions assess cognitive development later in childhood or less frequently. Therefore, the data does not only give the possibility 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 child development psychologists before. This analysis found that children in home visited families tend to have better birth outcomes and achieve higher mental development test scores (Jungmann et al., 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 distribution of treatment effects and the efficiency of home visiting 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, bias 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 were missing in at least one developmental test. These limitations of the experiment 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) in 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 psychomotoric skills, the birth outcomes or 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 around 0.30 SD at 6 and at 12 months and 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 by treatment in any of these outcomes. Investigating the skill formation process in the first 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 group and treatment group concerning baseline characteristics. However, in the control group the test scores of the children who quit participating in the research are lower than in the treatment group. This might be caused by the fact that mothers in both groups get feedback about test result. Imputing missing test scores with test points from earlier assessments lead to much higher treatment effects. After the imputation the mental development is increased significantly at all three assessment points in a range between 0.2 and 0.3 SD.

The rest of the paper is organized as follows: Chapter 2 provides a description of the *Pro Kind* Project. Section 3 describes the experimental design and data collection, while Section 4 discusses the randomization results and the impact of attrition on the internal validity. Section 5 presents results on the impact of the home visiting program on birth outcomes, mental and psychomotoric development as well as language development. Section 6 conducts robustness checks and presents evidence that the main effect of the intervention might be downward biased. Section 7 analyses the dynamics of the skill development. Section 8 discusses aspects of the cost effectiveness of the home visiting program. Section 9 presents 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 starts between the 12th and 28th week of pregnancy and ends at the second birthday of the child. The program runs in three German federal states, two in West- and one in East Germany. Families were affiliated between November 2006 and December 2009. Midwives, nurses or social pedagogues conduct the home visits alone or in a team. The frequency of the home visits varies between weekly, bi-weekly and monthly with highest frequency directly before and after birth. Home visitors use teaching materials and a guidebook to structure the topic and the aim of each home visit. Nevertheless, the home visitors are free to react flexibly to the demands of the mothers and their families. All home visitors receive feedback, encouragement and support from specially trained supervisors regularly. *Pro Kind* is an adaption of the Nurse Family Partnership (NFP) program, which provided instruction for home visitation frequency, employee selection, teaching materials and guidebooks.

Improving birth outcomes and child development are major goals of *Pro Kind*. For birth outcomes, the personal health of the mother during pregnancy is vital while for child development parental skills, i.e. that parents understand signals of their child, play an important role. Therefore, the home visits cover issues like smoking and a balanced diet to generate a healthier environment. To enhance parental skills 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* reverts to different psychological theories like the ecologic theory, the attachment theory and the self-efficacy theory (Bronfenbrenner, 1992; Bowlby, 1969; Bandura, 1982, 1997).

Affiliation criteria for participating in *Pro Kind* include an economic constraint on the family and that the mother has at least one social risk factor. The economic constraint is defined as receiving social welfare, unemployment benefits, an income, that is as low as social welfare or has a high amount of debt. The considered social risk factors include: Low education, teenage pregnancy, isolation, experienced violence or health problems. Project partners, like gynecologists, job centers and youth welfare offices referred three quarters of the participants to *Pro Kind*. About one quarter of the participants registered to the program by themselves.

A process evaluation monitored the implementation of the *Pro Kind* program. For this reason, home visitors fill out a report for each home visit in which the duration and the covered topic is recorded. The process evaluation reveals that in average a family got 33 home visits with a minimum of 0 and a maximum of 69 with a standard deviation of 19 home visits. During pregnancy the families received 9 home visits on average. I include the participants with 0 home visits because in the analyses below I estimate the intention to treat effects. Considering only families where the intervention is conducted per protocol increases the average number of home visits to 47 with a minimum of 31. The duration of an average home visit is 82 minutes. 28% of the home visits are devoted to the topic maternal health, 20% to parental skills and 10% to healthy child environment (see Brand and Jungmann, 2010, for details of the process evaluation).

3 Experimental Design and Data Collection

All women who were referred or came forward to *Pro Kind* filled in a short screening questionnaire to check if the affiliation criteria were fulfilled. If the requirements were met, participants, or if they were underage their parents, signed an informed consent for participating in research. Afterwards, participants answered a baseline questionnaire to obtain socio-demographic and psychological characteristics and risk factors. After answering the baseline questionnaire women got the results of the randomization which allocates them into a home visiting group (394 women) and a control group (361 women). A computer calculated the randomization, which is stratified for communities, immigration and being underage. After randomization the control group and the home visiting group have access to the regular welfare state services. Both groups get an address list with support services and they get the results of the developmental and language tests continuously. Only the home visiting group is eligible for the *Pro Kind* intervention. The home visiting group includes slightly more participants than the control group because the first woman

in each community was automatically allocated to the home visiting group.

After the randomization, home visitation begins for the intervention group. For both groups research starts with a telephone interview and a personal interview during pregnancy. Telephone interviews continue in an interval of six months 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) are 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), are also assessed at each personal interview by the Psychomotor Developmental Index (PDI) of the Bayley Scales. Additionally at 24 months, a language test for two years old children (*Sprachentwicklungstest für zweijährige Kinder*, SETK-2) is conducted. The BSID and the SETK-2 tests are video taped and after the interview evaluated by developmental psychologist who do not knoe the treatmentgroup 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 hundred with a 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 the need of medical assistance. If a child in the home visiting or the control group scores below these thresholds the mother get special information and advise additional to the regular feedback of the research. For my regression analysis I standardized the test scores and birth outcomes with a mean of 0 and a SD of 1. The standardization allows to compare effects on birth outcomes and test scores and facilitates the comparison to other home visiting interventions. MDI and PDI tests consist of different tasks. If the refusal or interruption rate of these tasks in one test exceeds 20 percent the reliability of the test becomes problematic. Therefore, I exclude tests with a refusal or interruption rate higher than 20 percent and include them only for robustness checks of the results. Birth data outcomes are collected at two times by the telephone interviews and the personal interviews. The data is only used when the mothers give the identical information in both interviews. Additionally, part of the birth outcomes are checked by medical records which revealed a high reliability of the self statements.

4 Baseline Comparison and Attrition

4.1 Baseline Comparison

Randomly assigning families to the home visitation program ensures that the assignment is independent of the mothers' and their families' characteristics that may

be correlated with birth outcomes and child development. If this holds true, any differences in outcomes between the two groups post-intervention can be causally attributed to the intervention. To check that the mother and family characteristics were indeed similar between the two groups, I run regressions of baseline mother and family characteristics on treatment status, and then verify that changes in the sample due to attrition are also uncorrelated with treatment status.

I present the comparison of mother and family characteristics at baseline in Table 1. Column 1 contains the average characteristics for the control group. Columns 2 and 3 present the estimated differences between the treatment and control groups for demographic characteristics and selected psychological risk characteristics. The results in column 2 do not include any controls, while those in column 3 control for community fixed effects, because randomization was conducted at state level.

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	20.3	-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 in €	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 risk 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 pregnancy	0.113	0.000 (0.023)	-0.005 (0.023)
Body-Mass-Index	25.31	-0.088 (0.394)	-0.160 (0.394)
Sum risk factors	5.864	-0.131 (0.178)	0.035 (0.028)
Observations	361	755	755

Robust standard errors shown in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The differences in average characteristics between the control and treatment groups are all practically small and mostly statistically insignificant. Migration status, defined as women who have no German citizenship or who are not born in Germany, is the only demographic characteristic which is significantly different with

a higher proportion of immigrants in the control group. None of the differences in psychological risk characteristics are statistically significant. Thus, 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 life time and over half of the women lost an important person during childhood. Both are related to attachment problems with their own child (Olafson, 2004). Appendix A shows a comparison between *Pro Kind* participants and first time mothers from the German Socioeconomic Panel (GSOEP) which is a longitudinal panel study representative for the German society. In this study all new mothers were asked about their children and life circumstances with a special questionnaire. The average *Pro Kind* mother is around 7 years younger than the average GSOEP first time mother. Furthermore, in the GSOEP sample 80 percent of mothers lived their first 15 years in a two parent household compared to less than 40 percent in the *Pro Kind* sample. Age and family situation during childhood are just two examples of many characteristics which prove the disadvantage of the *Pro Kind* participants. Therefore, *Pro Kind* was successful in acquiring high burdened women and families who are the target population of the intervention.

4.2 Attrition

Table 2: Sample Composition

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

Although, the baseline comparisons presented in Table 1 show that the treatment and control groups were similar at the baseline, it is possible that nonrandom attrition from the two groups between the baseline and follow up surveys may have rendered the two groups incomparable. Table 2 shows the attrition rates for both groups and child genders. There are statistically no significant differences between

the attrition rates for the control and treatment groups nor between genders. In both groups about 20 percent of the birth outcomes are not available. The attrition rate for the 6 months test is around one third of the baseline participants, for the twelve months test 45 percent and for the 24 months 55 percent are missing.

Attrition happens mainly because the participants change their mobile number or move away from their old address. In both cases the research team tries to find out new contact data. However, in many cases this is time consuming or not successful. The time consuming cases often lead to a missing in just one test. Therefore, 71% of the randomized families and their children participated in at least one test. The power analysis considered an attrition rate of 25% to detect with 80% probability effect with an effect size of 0.2 SD. If all available data is considered the assumptions in the calculation is almost met.

Table 3: Selective Attrition

	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 in €	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 risk 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 shown in parentheses. Estimates include community fixed effects.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Even though the attrition rates were similar for both groups, the characteristics of the attritors and non attritors still could have differed between the two groups. I check this possibility in Table 3 for the three months telephone interview in column 1, the six, twelve and 24 months tests in column 2, 3 and 4, respectively. Again

I run regressions of mother and family characteristics from the baseline survey on treatment status just with the mothers and families who did not attrite. All of the differences are statistically insignificant with the exception of the proportion of mothers with risk of aggression and lost experience at 24 months. The difference in mothers with immigration background becomes insignificant just at the 24 months interview which shows that even this unbalance in the randomization process sustains almost stable. I, therefore, conclude that the comparability of the control and home visited families has been sustained throughout the follow up tests.

Nevertheless, it might be that more or less disadvantaged mother in treatment and control group refuse participation in the interviews and tests. Table 4 compares maternal baseline characteristics of attritors with non attritors. The results reveal that younger mothers and mothers with demographic risk factors like low education or income have a higher risk to refuse participation in the research. Psychological risk characteristics are less correlated with attrition. Most of the risk mothers attrite before the 3 months interview because the higher percentage of risk factors stays constant. However, age sharply decreases from 6 to 12 months in the attritors group. After 24 months mothers who still participate are more than two years older than attriting 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 and 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 cumulated sum of risk factors is 5.45 in the non attriors group in contrast to 6.08 in the attriors group.

5 Estimating Program Effects

5.1 Descriptive Data

In order to allow a better interpretation of the intervention outcomes, Table 4 gives a combined overview of the birth outcomes and test results for treatment and control group members. A comparison of the *Pro Kind* birth outcomes with the first-borns from the GSOEP reveals that birth weight and height are similar in both samples. Nevertheless, head circumference is statistically smaller in the *Pro Kind* sample than in GSOEP data (T=5.6). The gender difference in birth outcomes is similar to the average population. Looking at the developmental test scores reveals that the *Pro Kind* average is below the population norm of 100 points in all tests. As expected the *Pro Kind* eligibility criteria seem to be negatively related with test score results. After 12 months all test scores are closer to the norm of 100 points. However, after 24 months the mean of MDI declines again. Girls score in almost all tests better than boys. However, only in MDI at 6 months the difference is statically significant at a five percent level (T=2.1). The fact that girls score higher is well documented also in other studies (Quelle). Using the Levene-Test, variance of test scores is not

Table 4: Selective Attrition to baseline

	Difference 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)
Mon. HH-income in €	-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 risk 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 shown in parentheses. Estimates include community fixed effects.
* p < 0.1, ** p < 0.05, *** p < 0.01

significantly different between the genders at any point. Additionally, appendices B and C show density graphs of birth outcomes and child development tests scores by gender.

5.2 Specification Model for Estimating Treatment Effects

I estimate the *Pro Kind* effects on child development by OLS-regression analysis using equation 1:

$$Y_{ic} = \beta_0 + \beta_1 HV_{ic} + \beta_2 h_{ic} + \alpha_c + \epsilon_{ic}, \tag{1}$$

where Y_{ic} is the outcome of child i in community c. HV_{ic} is a dummy variable indicating whether the child's family is home visited. h_{ic} is a vector of demographic and psychological family characteristics at base line. I also include a dummy variable α_c for each community to absorb the community effects. The outcomes of interest are the normalized birth weight, birth height and birth head circumference, the normalized MDI and PDI test scores at six, 12 and 24 months, as well as, the results of the SETK-2. The coefficient of interest is β_1, which indicates the size of the causal effect of *Pro Kind*. The first model in each analysis includes no controls.

Table 5: Descriptive statistic

	Whole Sample		Boys		Girls	
	Mean	N	Mean	N	Mean	N
Birth Outcomes Pro Kind						
Weight in grams	3283 (540.7)	603	3370 (526.2)	280	3210 (544.3)	321
Height in cm	50.49 (3.17)	602	50.83 (3.15)	280	50.20 (3.18)	320
Head Circumference in cm	34.28 (1.85)	588	34.51 (1.71)	272	34.10 (1.94)	314
Birth Outcomes GSOEP						
Weight in grams	3253 (597.3)	825	3303 (613.7)	417	3203 (576.4)	408
Height in cm	50.86 (3.21)	824	51.20 (2.81)	417	50.51 (2.81)	407
Head Circumference in cm	35.11 (3.22)	765	35.26 (3.28)	386	34.95 (3.14)	379
6 Months 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 Months 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 Months 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

The second model is estimated with community fixed effects and controls for most available baseline characteristics. In those cases where there are missing values for the covariates, I include sample means or imputed values. The results are also robust for including more control variables or only using control variables with very few missings.

I run separate regressions for boys and girls because gender is a child characteristic which is unlikely to be correlated to any family characteristic. Therefore, different intervention effects between boys and girls can be fully attributed to gender. Furthermore, reevaluations of preschool programs suggest these programs benefit girls but not boys (Anderson, 2008). Such gender reevaluations are absent for home visiting programs so far.

5.3 Impact of *Pro Kind* on Birth Outcomes

Table 6: Impact of Home Visiting on Children's 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 in parentheses. Controls include socio-demographic, psychological and medical maternal 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. Nevertheless, the home visiting coefficient has a positive sign for all outcomes and close to The size of the coefficients only slightly varies with the model specifications, which shows that control variables are independent of the home visiting variable (Table 5). Additionally, Appendix D presents density graphs of birth outcomes in the treatment and control groups. Analyzing the effects separated for for boys and girls reveals that boys in the home visiting group have a significantly higher birth weight. However, this effect becomes insignificant when controls and mainly maternal smoking are included.

5.4 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 six months all coefficients are positive and MDI get significant when controls are included. The coefficients have similar sizes for MDI and PDI. At 12 months the coefficient for MDI increases and becomes significant also without controls. The effect for PDI is smaller than at 6 months. At 24 months the effect sizes for both MDI and PDI decline. While the effect for MDI is still positive the effect for PDI gets negative with a size very close to zero. At all assessment points the coefficients change only slightly when controls are included confirming the validity of the randomization. Appendix E shows the density graphs for MDI and PDI at six, 12 and 24 months in the treatment and control group.

Table 7: Impact of Home Visiting on Children's Development in SD

	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	0.070
	(0.093)	(0.094)	(0.101)	(0.100)	(0.116)	(0.116)
Community fixed effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	464	464	393	393	299	298
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	0.066	-0.022	-0.017
	(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	373	262	261
R^2	0.00	0.10	0.00	0.07	0.00	0.13

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

Splitting the sample by gender reveals that the coefficient of home visiting for boys are close to zero or small for MDI and PDI at any assessment point (table

Table 8: Impact of Home Visiting on Children's Development in SD (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	0.226
	(0.117)	(0.122)	(0.139)	(0.144)	(0.155)	(0.163)
Observations	245	245	206	206	166	165
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	-0.130
	(0.120)	(0.125)	(0.140)	(0.154)	(0.159)	(0.177)
Observations	258	257	205	204	149	148
R^2	0.01	0.07	0.01	0.07	0.00	0.22

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

8). In contrast, girls benefit strongly for MDI by 0.3 SD at 6 and 12 months and 0.2 SD at 24 months. The PDI effect for girls is significant with controls at 6 months but vanishes after 12 months and gets negative after 24 months. (Table 8). The coefficient for PDI increases but does not reach significance. Appendix F and G shows the density graphs for MDI and PDI at six, 12 and 24 months in the treatment and control group separated by gender.

Appendices H-J show estimates in which I include the tests with a task refusal or interruption rate higher than 20 percent. The newly included observations increase the sample size by 36 and 37 at six months and at 12 months, respectively. The coefficients of home visiting show little change, indicating that refusals or interruptions are not related to treatment. The home visiting coefficients are slightly smaller for girls alone when including the additional observations.

5.5 Impact of *Pro Kind* on Language

The SETK-2 results reveal no effects of the home visiting on the language development for the whole sample. The coefficients are both positive and negative but always below 0.10 SD. However, in the category production of words and sentences

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

	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 socio-demographic, psychological and medical maternal baseline characteristics. The treatment effects on production of words and sentences for boys and girls is significantly different at the 10 percent level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Sensitivity Analysis

The previous sections showed that attrition did not caused unbalanced groups with respect to baseline characteristics. Nevertheless, also outcomes which are influenced by the intervention could lead to selective attrition. In the case of *Pro Kind* it might be that the results of the developmental tests causes selective attrition. Table 9 shows that this is the case. At six months children of attriting mothers in the treatment group score 5.2 points higher at the MDI than children of attriting mothers in the control group. The effect is smaller but still significant at 12 months.

This selective attrition could be caused by the procedure that mothers in both groups get information about the test results of their children. In both groups bad results could cause frustration and skepticism towards the tests. However, mothers in the treatment group could discuss the results with their home visitor. This could reduce disaffirmation and avoid attrition. This opportunity is not given to the mothers in the control group and therefore mothers of bad performing children might attrite more often. Figure 1 supports this hypothesis. It compares the distribution of test scores of attritors and non-attritors separately for treatment and control group. 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 points and 70 points. In this range mothers get additional information that their child has developmental delay. At six months all control group mothers of child which scored less than 70 points at MDI did not continue the research.

To correct for this selective attrition I impute missings at with the standardized available test score at another assessment point. In most cases I use data from earlier assessment points to impute them in missings scores of later assessments. However, in some cases I also use later scores to impute earlier missings. Therefore, the data includes for all assessment points the same number of observations. This procedure is based on the assumption that the test scores are stable over the time. Section 7 will prove this assumption. Furthermore, there is no reason to suppose that the control group children catch up more than the treatment group children. Therefore, the imputation procedure seems even quite defensive.

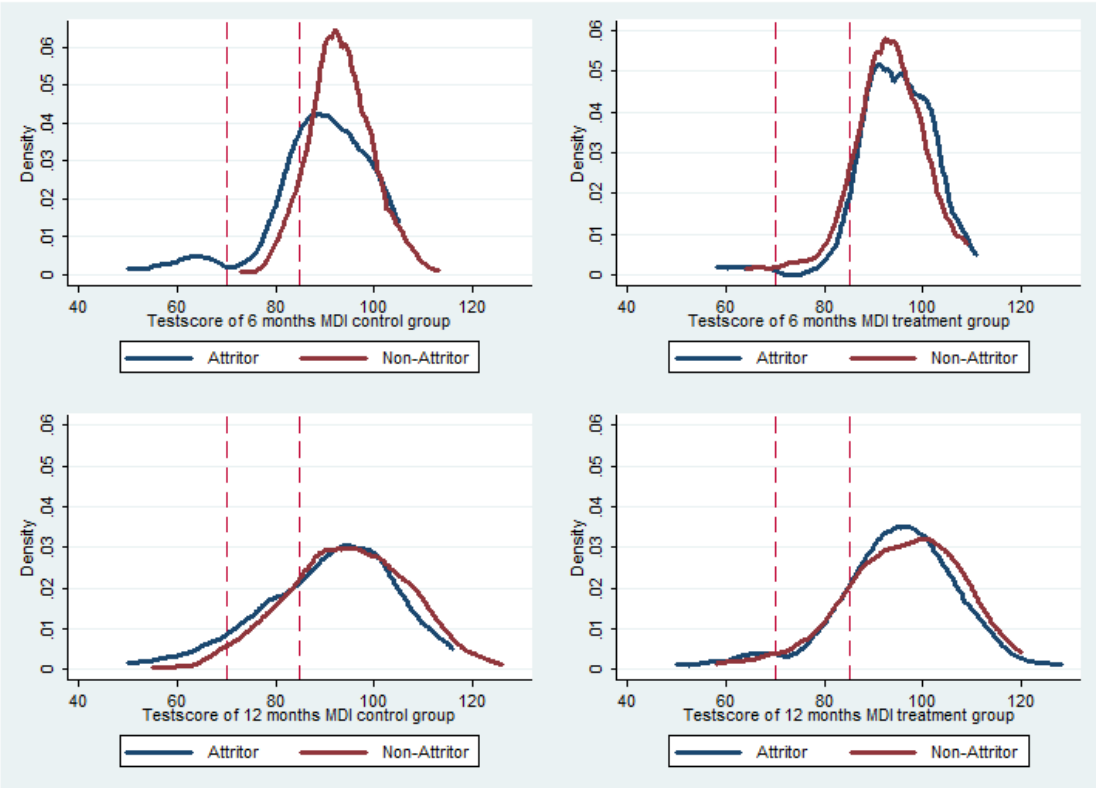
Table 9: Test scores of children who do not participate in the next developmental test

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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This imputation leads to significant effects of MDI in all three assessment points (Table 10). Again the effect is highest at 12 months and gets smaller at 24 months but is still significant at a 5 percent level. The PDI score does not increase by

Figure 1: Comparison MDI Test scores Attritors and Non-Attritors



the imputation. Looking at gender difference after the imputation reveals the same picture than before. The effect of the home visiting is grater for girls than for boys. Girls score at the MDI always over 0.25 SD significant at 5 percent level or higher.

Table 10: 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	0.144	0.187**	0.279***	0.334***	0.190**	0.222**
	(0.092)	(0.093)	(0.094)	(0.093)	(0.094)	(0.093)
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
	B. Psychomotor Developmental Index (PDI)					
Home visiting	0.085	0.124	0.032	0.044	-0.051	-0.033
	(0.088)	(0.089)	(0.088)	(0.090)	(0.088)	(0.091)
Community fixed effects	No	Yes	No	Yes	No	Yes
Household Controls	No	Yes	No	Yes	No	Yes
Observations	529	529	529	529	529	529
R^2	0.00	0.08	0.00	0.05	0.00	0.06

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 11: 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	-0.040	-0.005	0.126	0.223	0.024	0.084
	(0.141)	(0.151)	(0.145)	(0.147)	(0.147)	(0.149)
Observations	242	242	242	242	242	242
R^2	0.00	0.12	0.00	0.10	0.00	0.08
			Girls			
Home visiting	0.308***	0.278**	0.413***	0.380***	0.339***	0.263**
	(0.119)	(0.124)	(0.121)	(0.120)	(0.120)	(0.122)
Observations	282	282	282	282	282	282
R^2	0.02	0.10	0.04	0.12	0.03	0.16
B. Psychomotor Developmental Index (PDI)						
			Boys			
Home visiting	0.021	-0.040	0.007	-0.074	-0.039	-0.100
	(0.136)	(0.134)	(0.134)	(0.130)	(0.135)	(0.136)
Observations	242	242	242	242	242	242
R^2	0.00	0.17	0.00	0.15	0.00	0.14
			Girls			
Home visiting	0.141	0.207*	0.051	0.039	-0.056	-0.052
	(0.116)	(0.121)	(0.118)	(0.126)	(0.114)	(0.126)
Observations	287	287	287	287	287	287
R^2	0.01	0.07	0.00	0.04	0.00	0.06

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

7 Skill Formation Dynamics

The *Pro Kind* experiment gives the unique possibility to analyze the skill formation process in children's first two years. The 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. The knowledge we get about early skill formation can shed light on the mechanisms of how home visiting generates effects and why these effects occur with girls but not with boys. Furthermore, investigating early skill formation can show if it is possible that the effects of home visiting are sustained later in life.

In accordance with Cunha and Heckman (2007) skill formation is a dynamic process in which self productivity as well as direct and dynamic complementaries are important factors. Equation 2 presents the skill production function, where S_t denotes the vector of skills acquired at stage t.

$$S_{t+1} = f_t(h, S_t, HV) \tag{2}$$

Like in Equation 1, h is defined as demographic and psychological family character-

istic at baseline. Cunha and Heckman (2007) propose to include family investment in the production function. We use the home visiting variable HV as a proxy for family investment. Self productivity in the skill formation process arises when

$$\frac{\partial S_{t+1}}{\partial S_t} = \frac{\partial f_t(h, S_t, HV)}{\partial S_t} > 0, \tag{3}$$

i.e., when higher stocks of skills in one period create higher stocks of skills in the next period. In accordance with self-productivity, direct complementarities apply if one set of skills is productive for the formation of other skills in previous periods and vice versa. The investigation methods are based on Blomeyer et al. (2009) and Coneus et al. (2011), who also analyzed early childhood skill formation in the German context.

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 2. In Equation 4 $S^k_{t,i}$ denotes the skill indicator in t, $S^k_{t+1,i}$ denotes skills k acquired in a next period. At stage t_1 birth weight is the measure for $S^k_{t,i}$, at stage t_2,t_3 and t_4 we use 6, 12 and 24 months MDI and PDI test scores as measure for $S^k_{t,i}$

$$S^k_{t+1,i} = \gamma S^k_{t,i} + \phi HV + \eta h + \epsilon_{i,t} \tag{4}$$

My coefficients of interest are γ indicating self productivity or direct complementaries and ϕ indicating the effects of the home visiting investment. All variables are standardized as explained in Chapter 3.

For the whole sample we 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 complementaries appear only at stage 3, where IQ, at 24 months, increases by 0.14 SD if MQ increases by one SD at 12 months. If we separate the sample by gender, the picture changes. For boys we find no self productivity for IQ at stage 2 and no self productivity for MQ at stage 3. Instead of self productivity we find direct complementarity of 6 months MQ for 12 months IQ. For girls self productivity is sustained in all stages, with direct complementaries occurring as well. The HV coefficients show 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 MQ and IQ are smaller than estimated in tables 7 to 9 with the exception of the coefficient for boys on IQ at stage 2. This coefficient is larger because HV has a negative value in stage 1.

The results of self productivity are in line with the expectations. Coneus et al. (2011) find significant self productivity for IQ from 3 months to 2 years with a coefficient of 0.3 with a slightly higher value for boys than for girls. The increase of self productivity is also in line with previous studies finding values of 0.9 at the age of 8 years which represent a high stability of skills (Cunha and Heckman, 2008). The size of the self productivity coefficients demonstrates the relevance of early

interventions. On the one hand, the skills in each stage are related to the skills from the previous stage. On the other hand, skills seem to be malleable because self-productivity is not close to one. The direct complementarities are surprisingly low. Further research has to reveal the reasons and the consequences for home visiting. The coefficient of HV shows that the main reason for the insufficient effect for boys lies in the first six months of home visiting. At 12 months the net effect is comparable with the girls' effect. In further steps we investigate, in more detail, dynamic complementaries, if home visiting effects self productivity and why home visiting has no effect on boys especially in the first six months.

Table 12: Estimates of the skill production function with two skill factors and home visiting as investment

	Whole Sample			Boys			Girls		
	IQ t-1	MQ t-1	HV	IQ t-1	MQ t-1	HV	IQ t-1	MQ t-1	HV
t = 24 Months									
IQ	0.41***	0.13*	0.02	0.39***	0.17	0.01	0.27***	0.20**	0.08
MQ	0.09	0.34***	-0.04	0.35*	0.20	-0.23	-0.06	0.33***	-0.01
t = 12 Months									
IQ	0.28***	0.06	0.20**	0.13	0.19**	0.25	0.35***	-0.02	0.21
MQ	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									
IQ	0.20***		0.16*	0.24**		-0.07	0.13		0.30***
MQ	0.24***		0.13	0.12		-0.02	0.31***		0.24**

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All estimates conducted with robust standard errors and include community fixed effects

Notes: Robust standard errors in parentheses. Controls include socio-demographic, psychological and medical maternal baseline characteristics.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Cost-Benefit Analysis

I use two approaches to value the impact of *Pro Kind* on child development. The first approach links the impact of *Pro Kind* directly to NFP. The second one analyzes the relationship between increased cognitive ability in the first two years of life and the probability to attain the highest German school degree.

Since *Pro Kind* is an adaption of the NFP Program I use the results of NFP as a benchmark. An interdisciplinary research team evaluated the NFP Program in three different trials. The first trials started in Elmira in the early 1980's, the second in Memphis in 1990 and the third in Denver in 1995. All three trials used the Mental Developmental Index (MDI) of the Bayley Scales of Infant Development to assess child development. However, the tests in Memphis and Denver were conducted only at the age of 24 months and in Elmira only at the age of 12 months. Neither in the Elmira trial nor in the Memphis trial home visited infants score significantly better at MDI (Olds et al., 1986; Kitzman et al., 1997). Only the Denver trial revealed an impact of NFP on the child mental development (Olds et al., 2002). Home visited children scored 4 points higher on a scale with the population mean of 100, which is higher than the *Pro Kind* effect. Nevertheless, in all three trials other program effects occurred such as lower childhood injuries or fewer subsequent pregnancies.

The time period for follow up research is different for each trial. In the Elmira trial, data is available for 19 years, in Memphis for 12 years and in Denver only for 4 years later. In Elmira home visiting reduces reported serious antisocial behavior and emergent use of substances for the home visited adolescents at age of 15 and 19. (Olds et al., 1998; Eckenrode et al., 2010). The only measure for school success was high school graduation at age 19, where the intervention caused no effect (Eckenrode et al., 2010). In contrast, the program in Memphis not only reduced antisocial behavior but also improved the academic achievement of children at age 12. The four year follow up in Denver showed that home visited children scored better in a series of cognitive tasks focusing primarily on the children's capacity for sustained attention and inhibitory control (Olds et al., 2010). Therefore, it appears that the effect on child development lasts for at least a time span of four years.

When comparing the costs of the *Pro Kind* and NFP interventions one can see they are within a similar range. The NFP program in the Denver trial cost about \$11,511 in 2006 (Olds et al., 2010) and *Pro Kind* cost approximately €8,790 in 2010 which is \$11,866 assuming an exchange rate of 1.35 €/\$. The monetary benefits caused by the Elmira and Memphis trial are higher than the program costs where the major part of the benefits occurred by changes in maternal life course. Only in Elmira did the decrease of anti social behavior of the home visited children played a role as well. Since MDI was not significantly changed at age one it seems that other domains have a stronger influence on anti social behavior. In contrast, MDI could strongly influence school success. Then, the missing change in MDI may explain

the weak results on school performance in the Elmira trial. To estimate the impact of MDI on school success and to consider the German setting of *Pro Kind*, another data base is needed.

Therefore, in the second approach we use data from the German Mannheim Risk Study (MARS). 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. The MDI of the Bayley Scales was used to assess children's cognitive development at an age of three months and 24 months. This gives the unique possibility to analyze the relationship between early cognitive development and later school success in a German context. An analysis of the MARS data (Coneus et al., 2011) shows that an increase in cognitive development by one SD at 24 months increases the probability of attaining a high school degree by 13 percentage points. To assess the economic relevance of the increase we continue with a numerical calculation.

In Germany the life earning premium for attaining a high school degree is €230,548 (Fritschi and Oesch, 2008). The Net Present Value of this amount is €118,837 assuming a discount rate of 1.5 percent, an entry age of 25 and 40 years of workforce participation. Furthermore, I use the MDI result (0.22 SD) of the sensitivity analysis at 24 months. Then the *Pro Kind* effect on cognitive development increases the probability to attain high school by 2.86 percent. This means, on average, a higher life time income of €3,398.74 for the home visited, which is 39 percent of the intervention costs. Until now I have no data about how increased cognitive development is related to the probability of dropping out or class repetition. These adverse school events could have a high relevance in the *Pro Kind* sample. Hence, they are the focus of our next investigations.

9 Conclusion

Home visiting for disadvantaged families is proven to be effective for child development in the US. The analysis of the *Pro Kind* project using the Bailey scales of infant development and birth outcomes as measure for child development shows that this is also true for continental Europe. The results suggest a better cognitive development at the age of 12 months. However, program effects on cognitive development are concentrated on girls. Girls in the treatment group achieve higher test scores at six, 12 and 24 months than their counterparts in the control group, whereas boys show no difference. My findings of gender differences in cognitive development are in line with reevaluations of other early childhood interventions like the Perry Preschool program, where the intervention is also exclusively effective for girls.

The effects of *Pro Kind* on child development are robust to several specifications and increases dramatically when missing observations were imputed by the value of the proceeding test. I use community fixed effects estimations because randomiza-

tion was done on a state level. Furthermore, I estimate models with different family baseline characteristics as controls. The home visiting effect is hardly influenced by any specification. Since there was more than 20% refusals or interruptions of tasks within one test affect reliability of the overall test result I excluded these tests from the main estimations. If I include these tests, effects of home visiting on child development are still robust with marginal changes.

I also investigated the dynamic nature of the skill formation process because of its importance for the interpretation of the effect size. I showed that self productivity is present at all stages. We do not find direct complementarities between MDI and PDI. After estimating separate models for boys and girls we 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.

Considering the question if the size of the *Pro Kind* effect on child development is meaningful, psychological and behavioral literature would claim that effect sizes below 0.2 SD are small. Nevertheless, the *Pro Kind* effects on child mental development could have large lifelong impact because of the dynamic nature of the skill formation process. For example, the cognitive development at 24 months is strongly related to high school graduation, which is a strong indicator for life income. Especially for girls the effect size, which is classified as moderate, could have a lifelong impact. The meaning of the effect size is also enlarged because the home visitors do not directly interact with the child rather with the mothers. Thus, it is likely that the mother uses the acquired skills in other aspects of life as well; in respect to her own or her child's health or in the planning of her own life course. Furthermore, there could be spill-over effects of the acquired skills for the second child because *Pro Kind* just affiliates first time mothers.

References

Almond, D. and J. Currie (2011). *Human capital development before age five*, Volume 4b of *Handbook of Labor Economics*. 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.

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.

Blomeyer, D., K. Coneus, M. Laucht, and F. Pfeiffer (2009). Initial risk matrix, home resources, ability development, and children. *European Economic Association: Journal of the European Economic Association 7*(2), 638–648.

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

Brand, T. and T. Jungmann (2010). Abschlussbericht der Implementationsforschung zum Modellprojekt Pro Kind.

Bronfenbrenner (1992). *The ecology of human development: Experiments by nature and design*. Cambridge and Mass.: Havard University Press.

Coneus, K., M. Laucht, and K. Reuß (2011). The role of parental investments for cognitive and noncognitive skill formation—Evidence for the first 11 years of life. *Economics & Human Biology*, in press.

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. and J. J. Heckman (2007). The technology of skill formation. *The American Economic Review 97*(2), 31–47.

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.

Fritschi, T. and T. Oesch (2008). *Volkswirtschaftlicher Nutzen von frühkindlicher Bildung in Deutschland: Eine ökonomische Bewertung langfristiger Bildungseffekte bei Krippenkindern.*

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*, Volume 5495 of *IZA Discussion Paper.* Bonn.

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

Kitzman, H., D. Olds, C. Henderson, C. Hanks, R. Cole, and R. Tantelbaum (1997). Effects of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: A randomized controlled trial. *The Journal of the American Medical Association* (278), 644–652.

Olds, D., Ch. Henderson, R. Cole, J. Eckenrode, H. Kitzman, D. Luckey, L. Pettitt, K. Sidora, P. Morris, and J. Powers (1998). Long-term Effects of Nurse Home Visitation on Children's Criminal and Antisocial Behavior: 15-Year Follow-up of a Randomized Controlled Trial. *Journal of the American Medical Association 280*(14), 1238–1244.

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., Henderson C., Chamberlin R., and Tatelbaum R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics 78*, 65–78.

Olds, D., H. Kitzman, R. Cole, C. Hanks, and K. Arcoleo (2010). Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Maternal Life Course and Government Spending: Follow-up of a Randomized Trial Among Children at Age 12 Years. *Arch Pediatr Adolesc Med. 2010;164(5):419-424 164*(5), 419–424.

Olds, D., J. Robinson, R. O'Brien, D. Luckey, L. Pettitt, C. Henderson, R. Ng, K. Sheff, J. Korfmacher, S. Hiatt, and A. Talmi (2002). Home Visiting by Paraprofessionals and by Nurses: A Randomized, Controlled Trial. *Pediatrics 110*(3), 486–496.

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.

A Comparison between *Pro Kind* Mothers and SOEP first time mothers

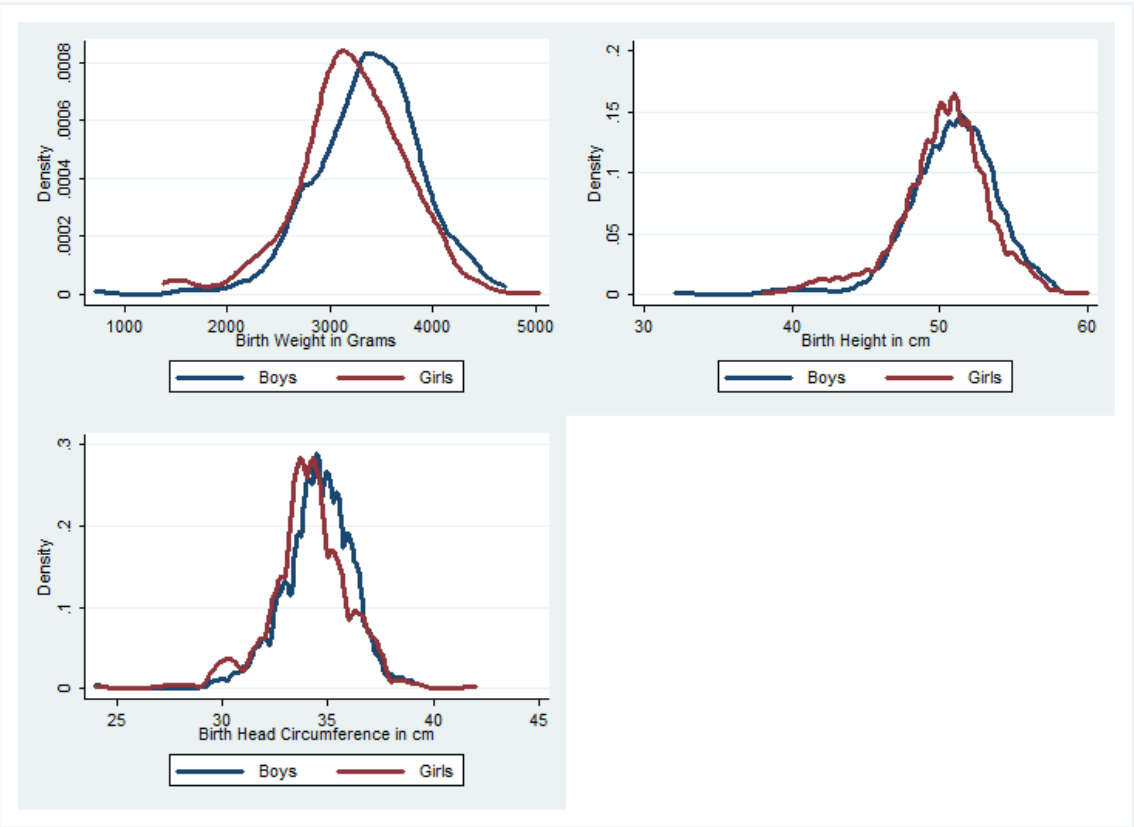
Figure 2: Comparison between *Pro Kind* Mothers and SOEP first time mothers

	TG		CG		SOEP	
	n	%	n	%	n	%
Economic sitituation during pregnancy:						
- Alg II-Transfer	246	74,8%	223	74,6%		
- Alg-Transfer	8	2,4%	6	2,0%	68[1]	10,2%
- HLU-Transfer	16	4,9%	11	3,7%		
- Overindebtedness	5	1,5%	6	2,0%	9	1,3%
- Little income	22	6,7%	18	6,0%	14	1,8%
- Participant in training	18	5,5%	26	8,7%	34	5,1%
- No economic risk factor noticed	14	4,3%	9	3,0%	553	81,4%
- ***Total***	***329***		***299***		***664***	
Age:						
- in years, at start of the intervention	21,48 (4,36)		21,57 (4,39)		28,3 (5,50)	
Total	***343***		***314***		***662***	
Country of birth						
- Germany	307	89,5%	261	83,1%	568	88,8%
- Turkey	4	1,2%	2	0,6%	13	2,0%
- East-Europe	12	4,1%	23	7,3%	41	6,5%
- Others	20	5,2%	28	9,0%	18	2,7%
Total	***343***		***314***		***640***	
Living situation within the first 15 years of life						
- With both parents	131	38,4%	130	41,8%	333	81,0%
- At least one year with just one parent	160	46,9%	142	45,7%	67	16,3%
- At least one year in foster care or with foster parents	50	14,7%	39	12,5%	11	2,7%
Total	***341***		***311***		***411***	
Family status						
- Unmarried	294	85,7%	273	87,5%	179	33,9%
- Married	41	12,0%	33	10,6%	329	62,3%
- Divorced/widowed	8	2,3%	6	1,9%	20	3,8%
Total	***343***		***312***		***528***	
School graduation in the year of pregnancy						
- Upper-track secondary qualifying for university entry (Fachhochschulreife/Abitur)	19	7,5%	23	10,3%	238	42,2%
- Intermediate-track secondary (Realschule)	79	31,2%	63	28,1%	217	38,7%
- Lower-track secondary (Hauptschule)	94	37,2%	88	39,3%	66	11,8%
- School for special needs/ other graduation	15	5,9%	9	4,0%	14	2,5%
- Left school without graduating	26	10,3%	20	8,9%	3	0,5%
- Graduation in a foreign country	10	4,0%	11	4,9%	22	3,9%
- Still attending school	10	4,0%	10	4,5%	2	0,4%
Total	***253***		***224***		***561***	

TG=Treatment group, CG= Control group

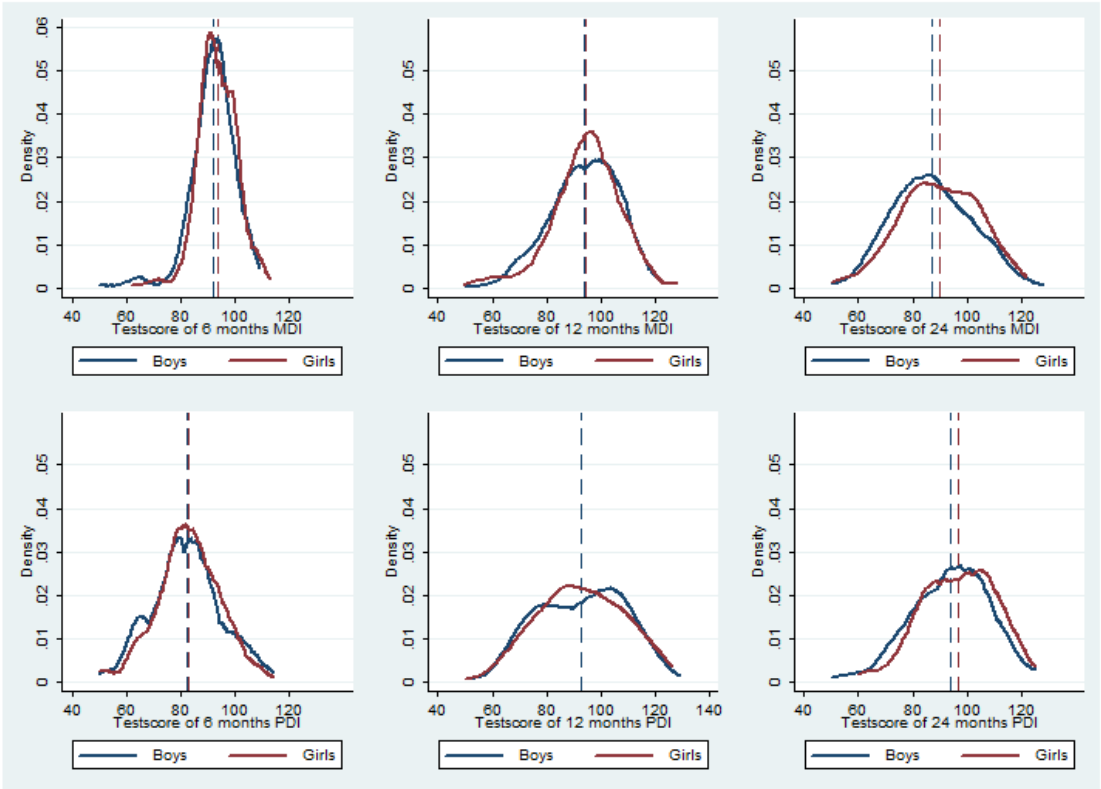
B Density Birth Outcomes for Boys and Girls

Figure 3



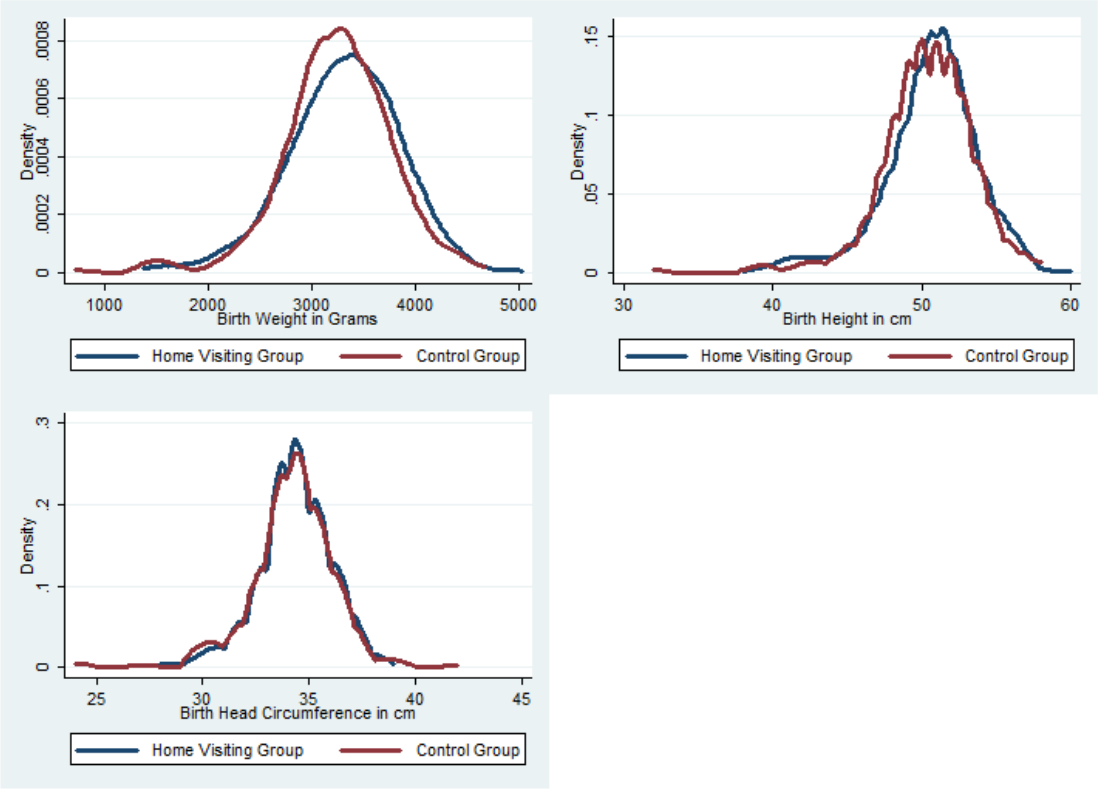
C Density Bayley Scales for Boys and Girls

Figure 4



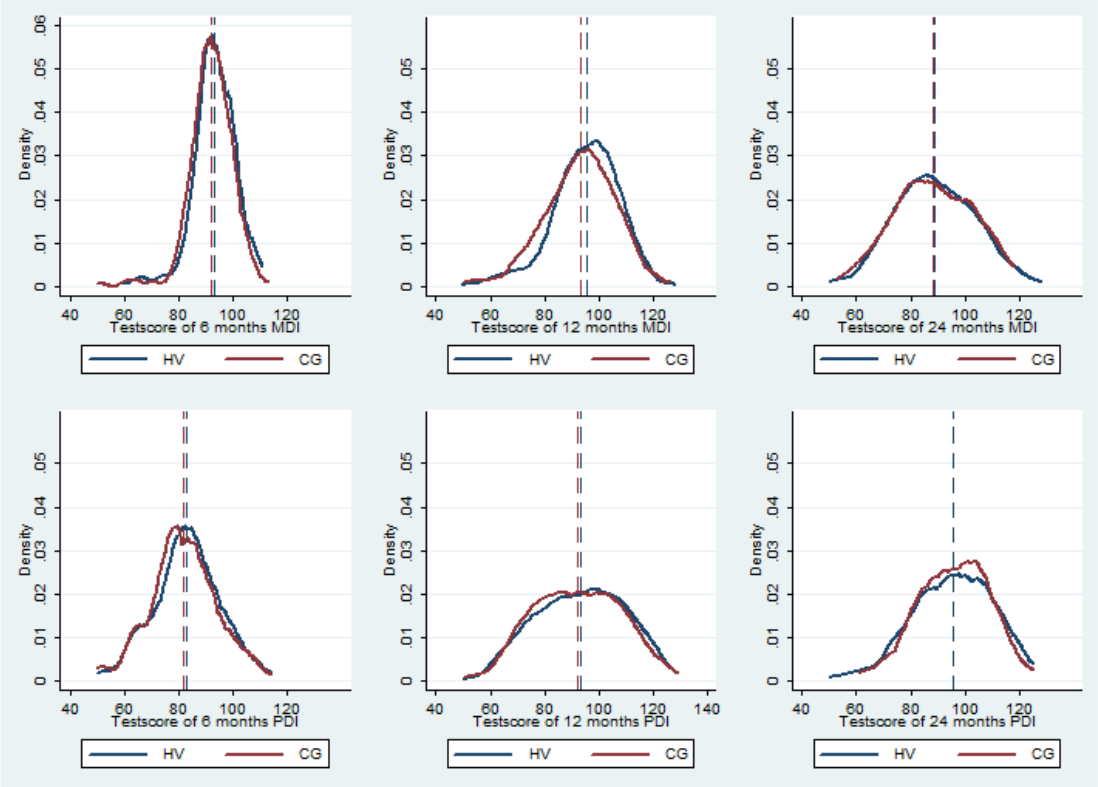
D Density Birth Outcomes for Treatment Group and Control Group

Figure 5



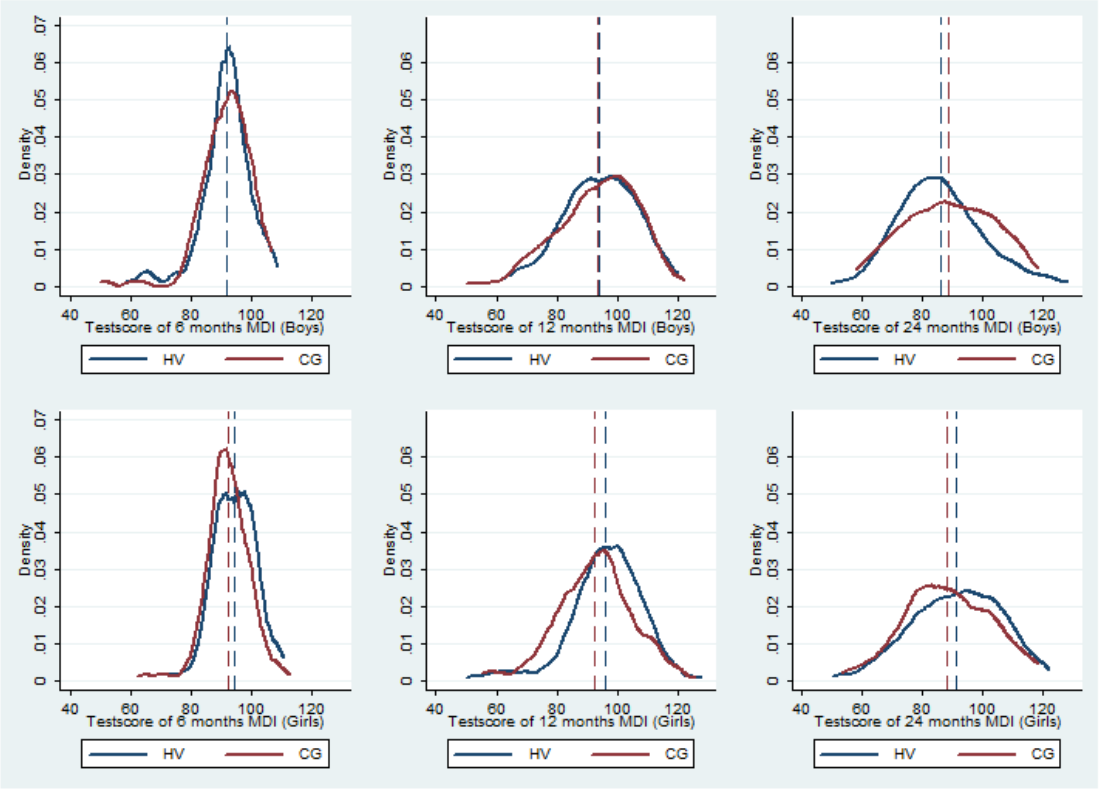
E Density MDI and PDI for Treatment Group and Control Group

Figure 6



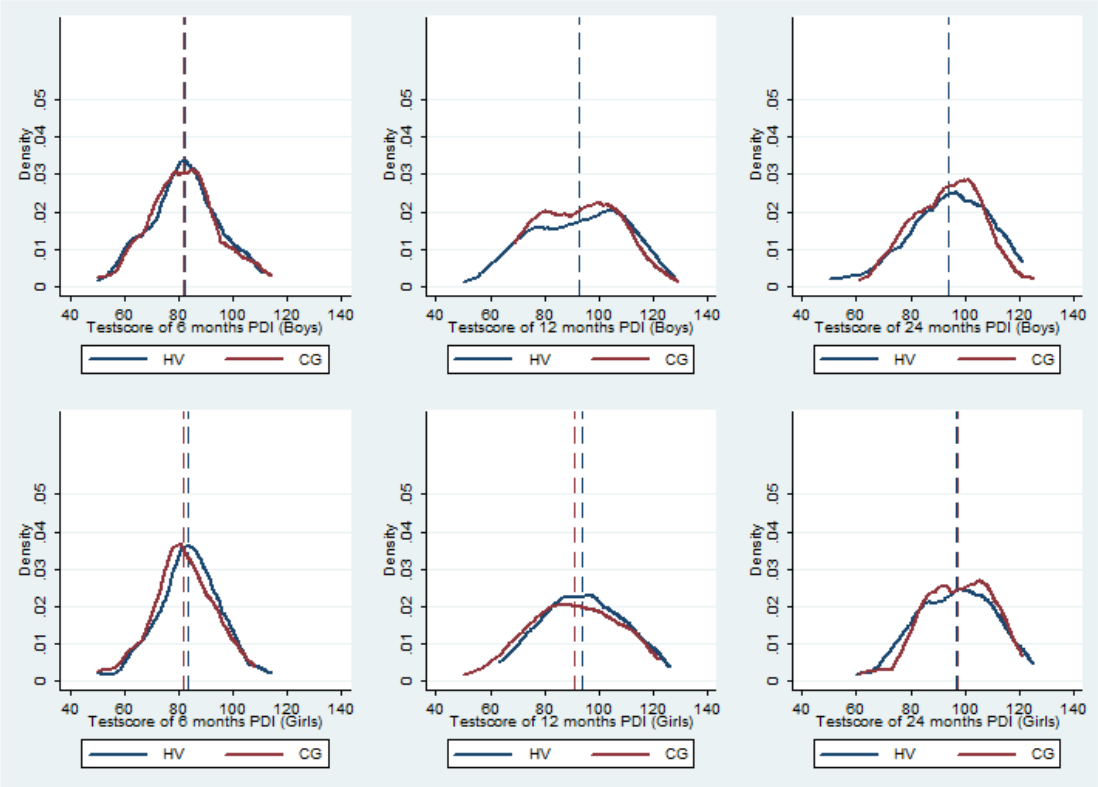
F Density MDI Treatment Group and Control Group Boys and Girls

Figure 7



G Density PDI for Treatment Group and Control Group Boys and Girls

Figure 8



H Child development including all tests

Table 13: Impact of Home Visiting on Children's Development in SD with all Tests

	6 Months			**12 Months**		
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
Home visiting	0.116	0.111	0.130	0.188*	0.200**	0.225**
	(0.089)	(0.088)	(0.090)	(0.097)	(0.094)	(0.096)
Community fixed effects	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
Observations	502	502	502	430	430	430
R^2	0.00	0.00	0.01	0.01	0.01	0.02
	B. Psychomotor Developmental Index (PDI)					
Home visiting	0.109	0.118	0.123	0.0589	0.0621	0.0767
	(0.090)	(0.089)	(0.091)	(0.096)	(0.096)	(0.097)
Community fixed effects	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
Observations	499	499	499	431	431	431
R^2	0.00	0.00	0.03	0.00	0.00	0.04

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I Child development including all tests

Table 14: Impact of Home Visiting on Children's Development by Gender at Six Months in SD with all Tests

	6 Months					
	Boys			Girls		
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
Home visiting	0.0268	0.0183	0.0150	0.196*	0.167	0.167
	(0.138)	(0.133)	(0.139)	(0.116)	(0.116)	(0.114)
Community fixed effects	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
Observations	235	235	235	267	267	267
R^2	0.00	0.00	0.04	0.01	0.01	0.03
	B. Psychomotor Developmental Index (PDI)					
Home visiting	0.0459	0.0211	0.000340	0.165	0.191	0.200
	(0.137)	(0.134)	(0.138)	(0.118)	(0.120)	(0.122)
Community fixed effects	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
Observations	233	233	233	266	266	266
R^2	0.00	0.00	0.05	0.01	0.01	0.04

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

J Child development including all tests

Table 15: Impact of Home Visiting on Children's Development by Gender at twelve Months in SD with all Tests

	12 Months					
	Boys			**Girls**		
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mental Developmental Index (MDI)					
Home visiting	0.111	0.147	0.201	0.256**	0.277**	0.288**
	(0.147)	(0.145)	(0.147)	(0.127)	(0.126)	(0.125)
Household Controls	No	No	Yes	No	No	Yes
N	199	199	199	231	231	231
R^2	0.00	0.01	0.05	0.02	0.02	0.03
	B. Psychomotor Developmental Index (PDI)					
Home visiting	0.0274	0.0262	0.00733	0.0850	0.0372	0.0721
	(0.143)	(0.141)	(0.143)	(0.131)	(0.130)	(0.131)
Household Controls	No	No	Yes	No	No	Yes
N	199	199	199	232	232	232
R^2	0.00	0.00	0.07	0.00	0.00	0.03

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$