Economic Evaluation of an Early Childhood Intervention and a Student Mentoring Program.

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -
genehmigte Dissertation

von

Diplom Ökonom Malte Sandner
geboren am 15.7.1980 in Hannover

2013

Referent: PD Dr. Peter F. Lutz
Korreferent: Prof. Dr. Patrick Puhani
Tag der Promotion: 28.6.2013

Abstract:

This thesis adds causal evidence to the field of social and educational policies in Germany. The work investigates a home visiting program for disadvantaged first time mothers and a student mentoring program. A randomized field experiment and a quasi-experimental difference-in-difference approach is used to establish causal inference. Both programs are publicly financed and aim to increase human capital. Therefore, both programs have strong fiscal and social relevance. The results of the analysis can give policy makers evidence how to use public money efficiently.

The evaluation of the home visiting program shows that by the end of the first year of the program, children in home visited families perform significantly better than those in the control families in cognitive measures. However, the effects fade-out after 24 months. Examination of gender differences reveal that the effects are concentrated on girls. Investigating the effects of the intervention on the maternal life course reveals an increase in fertility and maternal life-satisfaction and well-being, whereas the treatment does not affect maternal employment, school attendance and child care use. These results are in contrast to previous studies from the US where home visiting programs decreased fertility.

The impact of the university mentoring program on first year study success is evaluated by a difference-in-difference framework. The mentoring is only offered to students in an economics and management program, whereas it is not offered to students in an industrial engineering program. However, students in both programs take the same classes and write the same exams in their first study year. Results show that the mentoring program significantly raised the grades and decreases the failure rates in the first semester exams.

Keywords:

Randomized Controlled Trial, Early Childhood Intervention, Student mentoring, Natural Experiment

Zusammenfassung:

Diese Dissertation liefert kausale Erkenntnisse für den Bereich der sozial- und Bildungsmaßnahmen in Deutschland. Die Arbeit untersucht ein Hausbesuchsprogramm für benachteiligt erstgebärende Mütter und ein Mentoringprogramm für Studenten anhand eines randomisierten Feldexperimentes und eines quasi-experimentellen *difference-in-difference* Ansatzes. Beide Programme sind öffentlich finanziert und sollen eine Humankapitalsteigerung bewirken. Somit weisen beide Programme eine starke fiskalische und gesellschaftliche Relevanz auf. Die Ergebnisse der Untersuchung können politischen Entscheidungsträgern helfen, öffentliche Gelder effizient zu verwenden.

Die Evaluation des Hausbesuchsprogramms zeigt, dass Kinder, deren Familien die Hausbesuche erhalten haben, am Ende des ersten Lebensjahres bei kognitiven Tests besser abschneiden als Kinder in der Kontrollgruppe. Allerdings werden die Effekte nach 24 Monaten geringer. Eine Analyse von Geschlechtsunterschieden führt zu dem Ergebnis, dass die Effekte auf Mädchen konzentriert sind. Für den mütterlichen Lebensweg bewirkt das Hausbesuchsprogramm einen Anstieg der Fertilität, der mütterlichen Lebenszufriedenheit und des Wohlbefindens. Im Gegensatz dazu lassen sich keine Auswirkungen auf die mütterliche Erwerbstätigkeit, den Schulbesuch und die Nutzung von Kinderbetreuungseinrichtungen finden. Die Ergebnisse stehen im Gegensatz zu Ergebnissen amerikanischer Studien, die zeigen, dass Hausbesuchspogramme die Fertilität der Teilnehmerinnen senken.

Der Einfluss des Mentoringprogramms auf den Studienerfolg im ersten Semester wird anhand eines *difference-in-difference* Ansatzes evaluiert. Dieser Ansatz nutzt den Umstand, dass das Programm nur für Studenten eines wirtschaftswissenschaftlichen Studiengangs und nicht für Studenten eines Wirtschaftsingenieurstudiums angeboten wurde. Allerdings besuchen im ersten Studienjahr Studenten aus beiden Studiengängen dieselben Kurse und schreiben dieselben Klausuren. Die Ergebnisse der Studie zeigen, dass das Mentoringprogramm die Noten der wirtschaftswissenschaftlichen Studenten signifikant

erhöht und die Durchfallquoten signifikant reduziert hat.

Schlagwörter:

Randomisiertes kontrolliertes Experiment, Frühkindliche Bildungsmaßnahme, Mentoring für Studenten, Natürliches Experiment

Contents

List of Tables

List of Figures

1 Introduction

Policy makers in Europe and Germany increasingly recognize the need for evidence-based social and educational policies when deciding how to allocate limited funds (e.g. Commission of the European Communities, 2007; Bundesministerium für Bildung und Forschung, 2008). These policy makers mainly call for evidence from randomised control trials (RCT) or quasi-experiments. However, until now in Europe and especially in Germany evidence-based policy evaluations are limited in the area of social and educational policies.[1] This might be explained by the need of technical complex research methods in quasi-experiments or high investments in the case of field experiment to establish causal inferences.

This thesis tries to add causal evidence to the field of social and educational policies in Germany. For this aim the work investigates a home visiting program for disadvantaged first time mothers and a student mentoring program. I use a randomized field experiment and a quasi-experimental difference-in-difference approach to establish causal inference. Both programs are publicly financed and aim to increase human capital. Therefore, both programs have strong fiscal and social relevance. The results of the analysis can give policy makers evidence how to use public money efficiently. To attain the result, this thesis procedures as follows:

Chapter 2 describes the implementation of the *Pro Kind* program and its accompanying research. *Pro Kind* is a home visiting program for disadvantaged first time mothers

[1] Arni (2012) gives an overview about RCTs in various policy areas in Europe. He shows that in German-speaking countries RTCs are only conducted to evaluate labor market policies (e.g. Krug and Stephan, 2011).

and their families. 755 mothers participate in the program in three German federal states. The mothers are randomly allocated equally to a treatment and a control group. This randomized controlled trial gives the possibility to identify the causal effects of the intervention. An implementation research documents that the intervention is implemented as aspired. Data for the evaluation is collected by development test, personal interviews and telephone interviews until the third birthday of the child. Because participation in *Pro Kind* is voluntary, some mothers do not participate in the entire research and do not get the full number of home visits. However, due to the research design, the special population and the sound implementation of the intervention, the research created a unique data set which is convenient to evaluate the intervention.

Using this data set, chapter 3 analyses the effects of the *Pro Kind* Program on child development. 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 reveal that home visited girls score 0.30 standard deviations higher than girls in the control families, whereas boys score similar in both groups. The effects fade-out after 24 months. However, sensitivity analyses present 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.

Chapter 4 investigates the effects of the *Pro Kind* intervention on the maternal life course. In focus of the analysis are maternal employment, school attendance, child care use, fertility, life-satisfaction and well-being. For the analysis I use data from the telephone interviews. The longitudinal design of these data gives the possibility to apply duration methods. I find that *Pro Kind* increases fertility and maternal life-satisfaction and well-being, whereas the treatment does not affect maternal employment, school attendance and child care use. These results are in contrast to previous studies from the US where

home visiting programs decreased fertility.

Finally, chapter 5 presents evidence from a natural-experiment which is used to evaluate the effectiveness of a student mentoring program. The mentoring includes several compulsory, scheduled, face-to-face appointments between a mentor who is graduated and employed by the institution and a student in the first study year. For the evaluation, I use the fact that the mentoring is only offered to students in an economics and management program, whereas it is not offered to students in an industrial engineering program. However, students in both programs take the same exams. Therefore, the industrial engineering students present a reliable control group for evaluating the program. I find that the mentoring program significantly decreases the failure rates in the first semester exams.

2 The Early Childhood Intervention *Pro Kind*

2.1 Introduction

In the last years early childhood has gained much interest in economic research. In 2000, there were no articles on this topic in the *Journal of Political Economy*, *Quarterly Journal of Economics*, or the *American Economic Review* (excluding the Papers and Proceedings), but there have been five or six per year in these journals since 2005 (Almond and Currie, 2011). This growing interest is caused by findings that early childhood plays an important role for lifelong human capital accumulation. Most of these findings came from field experiments with randomized controlled trials (RCT) which demonstrated that relatively small investments cause lifelong effects and high benefit-cost ratios. These experiments were able to identify causal program effects and have generally been influential, sometimes extremely so (Imbens and Wooldridge, 2009).[1]

However, economic research is mainly focused on field experiments which analyse interventions, starting when children have reached preschool age (Belfield, 2006; Puma et al., 2005) or primary school age (Krueger, 1999). Field experiments which analyse earlier starting childhood interventions received less attention from economic research, although these programs have the potential to be as efficient or even more efficient than

[1]For example the US preschool program Head Start (cost of $7 billion in 2009) was mainly initialised because of the results of the Perry Preschool experiment (Ludwig and Phillips, 2008).

preschool or school programs because of the dynamic skill formation process (**?**). Furthermore, the conducted field experiments are mainly located in the US. In Europe and especially in Germany field experiments which investigate early childhood interventions are rare or just in the beginning (Arni, 2012).

This chapter describes the implementation of a field experiment, analysing an already prenatal starting early childhood intervention, the *Pro Kind* program, in Germany. *Pro Kind* is a home visiting program for disadvantaged first time mothers and their families. An interdisciplinary research team accompanied the program in order to evaluate its effectiveness and efficiency. The main instrument of the research is a randomized controlled trial. Altogether 755 mothers participated in the program and were followed from pregnancy to their child's third birthday. The effects of the *Pro Kind* program are the subject of chapters 3 and 4.

The rest of this chapter is organized as follows: section 2.2 gives a background about home visitation and explains the *Pro Kind* intervention. Afterwards, I introduce the accompanying research, the sample composition and the available data. The last section of this chapter draws conclusions.

2.2 Background and Concept of the *Pro Kind* Intervention

Sweet and Appelbaum (2004) define home visiting as an umbrella term that describes a strategy for delivering a service, rather than a type of intervention per se. Programs vary for example in frequency, duration, visitor's profession or starting point. Despite these differences, all programs are linked by their method of delivery service at the family's own home, their goal of helping children by helping the parents and their focus on infants and toddlers. Home visiting has been popular for delivering family services in the US for many years. In Europe and especially in Germany home visiting programs have less

tradition (Wasik and Bryant, 2001). One reason for this might be the developed system of compulsive gestational and postpartum checkups in Germany which makes home visiting less necessary than in the US.

Despite the young tradition of home visiting in Germany, the *Pro Kind* program was realized as part of a broader German initiative, the National Center on Early Prevention *(Nationales Zentrum Früher Hilfen - NZFH)*. The German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth *(Bundesministerium für Familien, Senioren, Frauen und Jugend - BMFSFJ)* founded this initiative after a series of child abuse and neglect cases which received high media attention in 2007. In order to reduce cases of child abuse and neglect the initiative set up ten pilot projects in several German federal states. Most of these projects aimed to minimize child abuse due to a better connection between the medical system and the child protection service. Besides *Pro Kind* the *NZFH* supported three other programs which used home visits to reduce child abuse and neglect. All *NZFH* projects were accompanied by at least some research. In the case of the three other home visiting programs the research consisted only of a process evaluation which monitored the implementation of these programs. Furthermore, in these three programs the sample size was small and in two programs the home visits started postnatal (see NZFH, 2011, 2010, for more details about the *NZFH* and the other supported pilot projects).

In the *Pro Kind* program the home visiting starts between the 12th and 28th week of pregnancy and ends at the second birthday of the child. 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 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 topic and the aim of each home visit. Nevertheless, the home visitors are free to flexibly react to the needs of the mothers and their families.

All home visitors regularly receive feedback, encouragement, reflection and support from nurse supervisors. These supervisors have an academic qualification and they do not consult more than ten home visitors each. *Pro Kind* is an adaption of the Nurse Family Partnership (NFP) program (Olds et al., 1997). The NFP provided instruction for home visitation frequency, employee selection, teaching materials, guidebooks and supervision.

Pro Kind focuses on three major goals: (1) improvement of birth outcomes by changing health behavior, (2) improvement of child's subsequent health and development by sufficient child stimulation and avoidance of child abuse and neglect as well as (3) improvement of families' economic self-sufficiency by helping parents to develop a perspective for their future and make appropriate decisions about planning future pregnancies, finishing their education and finding employment.

One element to achieve these goals is a strong theoretical foundation of *Pro Kind*. In this foundation the *concept of self–efficacy*, the *theory of human attachment* and the *human ecology theory* are most important:

The *concept of self–efficacy* (Bandura, 1997) is rooted in the notion that people are more likely to engage in a desirable behavior if they believe that the behavior will produce a desired outcome. Therefore, the home visitors help the parents to define and to establish realistic and achievable goals that – once accomplished – should increase the parents' reservoir of successful experiences.

According to the *theory of human attachment* (Bowlby, 1969), the home visitors help parents to understand the importance of a secure and empathic relationship between parents and child for the child's development. In this context, the program explicitly promotes sensitive, responsive, and engaged care-giving in the early years of the child's life. Furthermore, the home visitors train the parents to perceive children's signals accurately and to answer them sensitively.

The *human ecology theory* (Bronfenbrenner, 1992) emphasizes that the parents' care influences children's development. In turn the parents' care is influenced by characteristics

of their family members, social networks, neighborhoods, and communities as well as the interrelation among these factors. Therefore, home visitors attempted to enhance the material and social environment of the family by involving other family members, especially the fathers, during the home visits.

The guidebooks for the home visitors are based on these theories. For example the *Pro Kind* concept on how to tackle smoking behavior during pregnancy uses the self-efficacy theory. First the mother defines a small smoking reduction and if she meets this reduction aim, this achieved goal motivates her to reach even a higher smoking reduction. The nurse supervisors demonstrate how the nurses can integrate the three theories in their daily work. Additionally, in order to be successful in sensitive topics like smoking or parental behavior the nurses receive training in psychological conversation techniques.

In addition to the three theories, the program is guided by the assumption that the mothers are "the experts on their own lives". Therefore, the home visitors accept the preferences of the mothers and give support and recommendations, but no rules or instructions. This requires that the participation in *Pro Kind* is voluntary. Hence, at any point the mother can decide to pause or to end the home visits. Furthermore, the three theories imply that even after birth of the child the mother and not the child is still in focus of the intervention. This is based on the idea that the life circumstances of the child can be improved only through a change in maternal behavior and life course, as the main influence for the child.

In order to keep pace with the fast developments and the different requirements during pregnancy and the child's first two years, the guidebooks of the teaching materials are divided in three stages: the pregnancy stage, the infant stage from birth to the first birthday and the toddler stage from the first to the second birthday. To structure the wide range of goals, the home visits are subdivided in six topical categories: *Maternal Health*, *Environmental Health*, *Life Course Development*, *Maternal and Parental Role*, *Family and Friends* as well as *Social and Health Services*. While in pregnancy stage the home visits

focus on maternal health, in infant stage they concentrate on the maternal role and during toddler stage they focus on the maternal life course. (see Jungmann et al., 2008, 2009; Adamaszek et al., 2013; Schneider et al., 2013; Nurse-Family Partnership, 2010, for more information about the structure, the supervision and the theoretical background of NFP and *Pro Kind*).

To ensure that the home visits can be conducted at all, *Pro Kind* only includes mothers with sufficient German language skills and permanent residence. Additionally, *Pro Kind* participants must obtain three characteristics in order to reach the goals of the intervention effectively:

Firstly, only mothers who are during their 12th and 28th week of pregnancy are enrolled to *Pro Kind*. Enrollment during pregnancy allows time for the client and the home visitor to establish a relationship before the birth of the child and leaves time to address prenatal health behavior which affects birth outcomes and the child's neuro-development. Additionally, NFP data show that early entry into the program is related to longer stays during the infancy phase, increases a client's exposure to the program and offers more opportunities for behavior changes (Nurse-Family Partnership, 2010).

Secondly, only first time mothers can participate in *Pro Kind*. This enables spill over effects because the mothers could use the acquired parental skills also for further offspring. Furthermore, psychological research indicates that first time mothers feel more vulnerable and therefore they are more open to help and advice than mothers who have already acquired experience in raising a child (Olds, 2006).

Thirdly, all participating mothers have to be disadvantaged. In the definition of *Pro Kind*, mothers are disadvantaged if the mother or the family shows an economic constraint and if 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 having a high amount of debt. The considered social risk factors include: low education, teenage pregnancy, isolation, experienced violence or health problems. *Pro*

Kind only affiliates disadvantaged mothers assuming that less resources are available to these mothers outside of the program and that, therefore, the marginal benefit from the program is greater for low income than for higher income mothers.

Pro Kind cooperated with project partners to get access to this accurately defined client group. Project partners are professions or public services which are in frequent contact with the clients and which are able to refer them to *Pro Kind*. Gynecologists who referred 22 percent of the participants to *Pro Kind*, acted as the most important project partners. The other project partners were pregnancy information centers (18 percent), job centers (15 percent) and youth welfare offices (14 percent). 18 percent of the participants registered to the program by themselves. Most likely these women became aware of the program by brochures or due to social contacts. The project partners and the brochures mainly told the mothers that *Pro Kind* is a research project about disadvantaged first time mothers. The home visits were mentioned only as a minor point in advertising *Pro Kind*.

The *Pro Kind* project started officially in Lower Saxony in autumn 2006. An eight months test phase with nine participating mothers initiated the official launch. In spring 2007, the federal state Bremen implemented *Pro Kind* and in January 2008 the federal state Saxony followed as the third project location. *Pro Kind* affiliated mothers in time periods of 24 and 30 months in the different federal states. The last mother joined *Pro Kind* in Saxony in December 2009. In the three federal states *Pro Kind* is conducted in 13 municipalities (Figure 2.1 in Appendix A describes the location of the municipalities in detail). Four of these municipalities are cities with more than 500.000 inhabitants (Hannover, Bremen, Leipzig and Dresden). In contrast, three municipalities can be classified as rural (Celle, Vogtlandkreis and Muldentalkreis). Due to the mixture between municipalities in West and East Germany as well as rural and urban municipalities the *Pro Kind* locations give a representative sample for entire Germany.

60 female home visitors worked for *Pro Kind* (36 midwives, 23 social pedagogues, 1

nurse), with a mean age at entrance of 40.3 years (range = 22–53). Most of them had substantial work experience (M = 15.2 years, range = 0–31) and many years of experience dealing with socially disadvantaged families (Mean = 11.1 years, Range = 0–30). Two of the midwives had a migrational background. A full-time home visitor carries a caseload of no more than 25 active clients. All home visitors received an intensive training in the guidelines and theories of *Pro Kind* of at least ten days (see Maier-Pfeiffer et al., 2013; Adamaszek et al., 2013; Refle et al., 2013, for more information about the home visitors, the project partners and the training of the home visitors).

2.3 Experimental Design and Data Collection

The *Pro Kind* Project was accompanied by an intensive research which conducted telephone interviews with the participating mothers, tested the development of their children and collected information about frequency, topic and duration of the home visits. The team which conducted this research consisted of highly interdisciplinary scientists, including economists, psychologists, sociologist and physicians in five different institutions. The *Pro Kind* research used a randomized control trial (RCT) as the main instrument to investigate the effectiveness and the efficiency of the *Pro Kind* home visiting program. The next paragraphs explain the design and the procedure of the RCT in more detail.

Pro Kind used a RCT because in non-experimental frameworks individuals who receive treatment are usually different from individuals without treatment. This is not only with regard to their treatment status, but also in other covariates that affect outcomes. If this selection occurs, a simple comparison of outcomes between treated and non-treated mothers and children would not be appropriate. Random assignment of the treatment solves this selection problem because random assignment is uncorrelated with the attributes of the mothers and therefore on average the mothers in the treatment group are similar to individuals in the control group. In other words, random assignment ensures that the assignment is independent of the mothers' and their families' characteristics

which may be correlated with the outcomes of the *Pro Kind* intervention. If this holds true, any differences in outcomes between the two groups can be causally attributed to the intervention. Therefore, randomization can be seen as the gold standard to identify causal effects of an intervention (see Imbens and Wooldridge, 2009; Angrist and Pischke, 2009, for more information about randomization in economics).

In the beginning of the randomization process all women who were referred or came forward to *Pro Kind* answered a short screening questionnaire to check if the affiliation criteria were fulfilled. Most of the time this screening questionnaire was conducted by 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 research. Afterwards, participants 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 and as little information as possible about the home visits in order to minimize the "John Henry" effect for those mothers who were later allocated to the control group.[2] After answering the baseline questionnaire, women received the results of the randomization which sorted them into a home visiting group or a control group. A computer calculated the randomization which is stratified for communities, immigration and being underage.

After randomization mothers in the control group and the home visiting group had access to the regular welfare state services. Both groups received an address list with support services and received monetary incentives for participating in the research. Additionally, the research send letters to the mothers in both groups with feedback about the development of their child. However, only the home visiting group was eligible for the *Pro Kind* home visits.

Table 2.1 reports that 1157 mothers were referred or came forward to *Pro Kind*. 263 of

[2]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).

Table 2.1: Randomization Outcomes

Eligible Invited to Participate	1157	
Active Refusals	263	
Passive Refusals	139	
Randomized	755	
	CG	**TG**
Allocated to Treatment	361	394

these mothers did not meet the affiliation criteria when the short screening questionnaire was conducted (Active Refusals). 139 mothers met the criteria, but decided against joining *Pro Kind* before filling out the baseline questionnaire or the informed consent (Passive Refusals). These mothers mainly refused the participation because they felt that they received too little information about the treatment or they did not sign the informed consent because of data privacy issues. No data was collected about the mothers who refused the participation. Hence, no statement can be made if these mothers were less or more disadvantaged than the enrolled mothers. Finally, 755 mothers participated in the randomization process with 394 mother allocated to the treatment group and 361 mothers to the control group.

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. This procedure was necessary because the home visitors should have at least one client in each community as soon as possible after the beginning of the program. Table 2.2 presents the enrollment results and enrollment periods for each federal state and each municipality. It shows that in some location more participants were allocated to the home visiting group (Celle) and in others more in the control group (Hannover). This distribution is an outcome of the state wide randomization, and it occurs although the randomization was stratified for municipality.

Table 2.2: 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**	

2.3.1 Baseline Sample Characteristics

As described in the section before, the randomization process should provide two groups with equally balanced characteristics. To prove if this condition is fulfilled, I use the following basic model:

$$h_{ic} = \beta_0 + \beta_1 HV_{ic} + \alpha_c + \epsilon_i \tag{2.1}$$

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 mothers. Additionally, I include a community fixed effect estimator α_c in equitation 2.1 because the randomization results in Table 2.2 reveal that the number of participants in treatment and control group 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 2.3. Column 2 contains β_0 which

gives the average of characteristic h_{ic} in the control group. Columns 3 and 4 present the estimated differences between the treatment and control groups for demographic characteristics and selected psychological and physical characteristics. The model in Column 3 does not include any controls, while the model in Column 4 controls for community fixed effects.

If a missing occurred in one of the base line variables, I include sample means or values from a multivariate imputation procedure for the missing. However, for most variables complete data is available. Only in the income variables the share of missings is higher than three percent (see Table 2.9 and 2.10). The results do hardly change if the missings are used instead of the sample means or imputed values. In almost all variables the missings are equally distributed between control and treatment group.

The differences in average characteristics between the control and the treatment group are all small and mostly statistically insignificant. Migration status, defined as women who do not have German citizenship or who are not born in Germany, is the only demographic characteristic which is significantly different having a higher proportion of immigrants in the control group. None of the differences in psychological or physical risk characteristics are statistically significant. 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 which does not reject that the characteristics in the treatment and control group are the same. 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 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 the

Table 2.3: Sample Balance Across Treatments

Dependent Variable	Control Mean	Treatment Difference Including No Controls	Treatment Difference Including 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 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)
Risk Pregnancy	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 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 Column (2) and (3) are one if the mother is in the treatment group. They contain estimates of the average difference in characteristics between the control and treatment students, without controls and with community fixed effects, respectively. See Table 2.9 and 2.10 for variable definitions.
$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

average household size of 2.49 persons, this average income is below the poverty line in Germany.

Table 2.4 presents a comparison between *Pro Kind* participants and first time mothers

Table 2.4: Comparison *Pro Kind* Sample and GSOEP First Time Mothers

	Pro Kind Participants		GSOEP First Time Mothers	
	N	%	N	%
Country of Birth				
Germany	568	86.4	568	88.8
Turkey	6	0.9	13	2.0
East-Europe	35	5.3	41	6.5
Others	48	7.3	18	2.7
Living Situation in the first 15 Years of Life				
With both Parents	325	45.5	333	81.0
At least one Year with just one Parent	286	40.0	67	16.3
At least one Year in Foster- Care/Parents	104	14.5	11	2.7
Family Status				
Unmarried	623	86.6	179	33.9
Married	80	11.1	329	62.3
Divorced/Widowed	16	2.2	20	3.8
School Degree				
Higher Degree (Abitur / Realschule)	280	37.1	455	80.9
Basic Degree (Hauptschule / Ausland)	255	33.9	95	16.9
No Degree (incl. Förderschule)	137	18.2	10	2.0
Still Going to School	81	10.8	2	0.4
Age in Years (Mean)	21.4		28.3	

Notes: *Pro Kind* data are derived from telephone interview during pregnancy. In the GSOEP sample answers are given from mothers whose first child is between 0 and 12 months old.

from the German Socioeconomic Panel (GSOEP) which is a longitudinal panel study representative for the German society (Wagner et al., 2007). In this study all mothers who gave birth to a child in the past 12 months were asked about their children and their life circumstances with a special questionnaire. The average *Pro Kind* mother is about seven years younger than the average GSOEP first time mother. Furthermore, in the GSOEP sample 80 percent of mothers lived in a two parent household until the age of 15 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 underline the disadvantage 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.

2.3.2 Process Evaluation of the *Pro Kind* Intervention

A process evaluation monitored the implementation of the *Pro Kind* Project (see Brand and Jungmann, 2010a, for more details of the process evaluation). The main instrument for the monitoring was a documentation of the conducted home visits. To get a complete documentation the home visitors filled in a report in which the duration and the covered topic are noted for each home visit. The *Pro Kind* implementation data can be compared with the NFP data which has been collected for 20 years to improve and to monitor the program. Therefore, the NFP gives guidelines and comparison for the implementation of the *Pro Kind* program. As a further instrument, the process evaluation interviewed the participating mothers, the nurses, the nurse supervisors and the project partners about their contentment with the program.

Mothers participate in *Pro Kind* voluntarily. Therefore, it is expectable that some mothers in the treatment group will quit the intervention before the second birthday of the child with less home visits than originally scheduled. In the end, 166 of the 394 mothers (42.2 percent) withdrew early from the program. These attrition rates are similar to the rates in the NFP (The National Center for Children Families and Communities, 2005). However, only 9 of the 394 mothers (2.3 percent) in the treatment group received no home visit. In three of these cases the mother moved away from a *Pro Kind* community before the intervention started and in four cases the participants contact details were not sufficient.

Table 2.5 reports the number of attriting mothers and the reasons which caused the attrition for each stage. The reasons for attrition can be divided into endogenous and exogenous to the program. Exogenous to the program are fetal demises, intervention by the child protection service, moving into a community where *Pro Kind* is not offered or time restriction because of maternal employment. Endogenous reasons contain attritions

Table 2.5: Attrition from the Home Visits

	Pregnancy		Infancy		Toddler		Together	
Reason for Attrition	n	%	n	%	n	%	n	%
Fetal Demises / Infant Deaths	3	5.8	1	1.1	-	-	4	2.4
Child Protection or other Services	3	5.8	13	14.9	3	11.1	19	11.4
House Moving	8	15.4	16	18.4	4	14.8	28	16.8
Employment / Time Restriction	-	-	9	10.3	1	3.7	10	6.0
Not Reachable	11	21.1	18	20.7	8	29.6	37	22.3
No further Interest in HV	27	52.9	30	34.5	11	40.7	68	40.9
Sum of Attritors	**52**		**87**		**27**		**166**	

because the mother was not interested in the home visiting anymore or the home visitor was not able to contact the participant for further visits. In both cases it is likely that the mother expected something different of the program or that the relationship to the nurse was stressed.

In the first stage of the intervention, before child birth, 52 mothers quit the home visiting. At this stage 73 percent of the participants attrite because of reasons endogenous to the program. In the next stage, before the first birthday, 87 additional mothers canceled the intervention. At this stage the percentage of mothers who didn't have interest or were not able to be contacted anymore decreased to 55 percent. In the toddler stage only 27 mothers attrite. 70 percent of these due to endogenous reasons. In the sum of all stages 105 mother (63 percent) quit because they were not satisfied with the intervention. This figure is important because it is likely that these mothers also quit participation in the *Pro Kind* research.

The analysis of the home visit reports of all 394 mothers reveals that in average a family got 32.7 home visits with a minimum of 0 and a maximum of 94 and a standard deviation of 18.6 home visits. During pregnancy the families received 8.5 home visits on average, in the infancy stage 15 and in the toddler stage 9.1. Considering only the 228 families who received the intervention per-protocol increases the average number of home visits to 45.3 with a minimum of 11 and a standard deviation of 10.7 home visits. The 342 mothers who stayed in the program during the complete pregnancy stage got on average

9.3 visits during pregnancy. For the next stage the average number increases to 20.3 if only the 255 mothers who completed this stage are considered. Altogether 12,894 home visits were conducted. The duration of an average home visit is 82 minutes. Overall, for the mothers who received the treatment per protocol, the frequency and duration of the home visits are very close to the NFP guidelines (Olds et al., 2002).

The *Pro Kind* home visits are structured in six topics. Five of these topics are also used in the NFP. Only the topic *Social and Health Services* is exclusively added to *Pro Kind* and is not part of the NFP. Table 2.6 shows the average time devoted to the topics in comparison to the average NFP time and NFP recommendations. During pregnancy *Pro Kind* focuses, similarly to NFP, on the topic *Maternal Health*. After birth *Maternal and Parental Role* is in center of the home visits with over 30 percent of the time devoted to this topic which is comparable to the NFP average and NFP recommendations. The topic *Life Course Development* becomes more important over time and is always addressed more often than in the NFP. Considering that *Pro Kind* uses one topic more than the NFP, the *Pro Kind* implementation is very close to the NFP average and its recommendations.

Furthermore, the implementation research investigates how the participating mothers and the home visitors perceived *Pro Kind*. On the child's first birthday the mothers were asked about their satisfaction with *Pro Kind* in general, about the relationship between mother and home visitor and about the structure and dynamic of the home visits. The participants gave positive ratings in all three dimensions. In particular, the general satisfaction is 3.72 on a four point likert scale where four indicates the highest satisfaction and almost all mothers would recommend *Pro Kind* to friends who are in a similar situation. Furthermore, the relationship between mother and home visitor is rated as trustful and open and the home visits are perceived as dynamic and structured. However, the positive ratings are not surprising because only mothers who did not attrite the home visits were asked. Since participation in *Pro Kind* is voluntary, a mother who is unhappy with the intervention would have attrited immediately. Additionally, the positive

Table 2.6: Topical Focus of the Home Visits in NFP and *Pro Kind*

	Pro Kind Average	**NFP-Average**	**Recommended Average by NFP**
During Pregnancy			
Maternal Health	28%	37%	35%-40%
Maternal and Parental Role	19%	23%	23%-25%
Environmental Health	10%	11%	5%-7%
Life Course Development	16%	13%	10%-15%
Family and Friends	15%	16%	10%-15%
Social and Health Services	12%	-	-
During Infancy			
Maternal Health	16%	20%	14%-20%
Maternal and Parental Role	30%	36%	45%-50%
Environmental Health	11%	14%	7%-10%
Life Course Development	17%	15%	10%-15%
Family and Friends	14%	15%	10%-15%
Social and Health Services	11%	-	-
During Toddlerhood			
Maternal Health	13%	17%	10%-15%
Maternal Role	30%	37%	40%-45%
Environmental Health	10%	14%	7%-10%
Life Course Development	22%	17%	18%-20%
Family and Friends	14%	15%	10%-15%
Social and Health Services	11%	-	-

Notes: The percentage rates give the share of the total time in the family which the home visitors devote for a certain topic (The National Center for Children Families and Communities, 2005).

ratings could be given by mothers who are afraid that the rating is not secret to the home visitor.

In order to reduce the dependence on the participants' self ratings, also the home visitors are asked about the commitment of the participants' in the home visits. In line with the participants self-ratings the home visitors rate the commitment with 3.36 on a four point likert scale where four indicates the highest value. Therefore, it seems that mothers who participate in the home visits are satisfied with them. Furthermore, the home visitors are also asked how they use and how they rate the *Pro Kind* materials and guidebooks. The best ratings are given to materials related to the maternal role. These materials are also used most regularly. This corresponds to the finding that most time is devoted to this topic in the home visits after birth.

The interviews with the project partners revealed reservations against the randomized trial (Brand and Jungmann, 2010b). Partners who referred just few or no mothers to *Pro Kind* explained their noncompliance with the concern that their clients could be allocated in the control group. This concern might be highest for mothers with many risk factors because the project partners do not want them in the control group without additional help. Therefore, it is likely that the *Pro Kind* program has not fully reached the extremely deprived mothers. This finding is important for the external validity of the *Pro Kind* outcomes. Furthermore, it could be important for other projects which also want to implement a randomized trial in Germany.

Overall, it seems that *Pro Kind* is implemented as aspired. The topics and the number of home visits are in accordance with the NFP guidelines. Furthermore, the sample consists of mothers for whom the program is designed. The general client satisfaction with the home visits is high. However, over 40 percent of the participants did not comply fully with the program. Nevertheless, since program participation is voluntary, the attrition rate was expectable and the rate is in line with the NFP experiences. Therefore, the attrition is not caused by inappropriate implementation and hence it is likely that the

program will have an impact on certain dimensions of the maternal and the child's life.

2.3.3 Evaluation Instruments

The *Pro Kind* research uses a mixture of developmental tests, personal interviews and telephone interviews. The different data collection methods are necessary to cover the diversity of possible outcomes which result from an intensive intervention like *Pro Kind*. The development tests and the interviews use a longitudinal design to capture the age from which on effects occur and if these effects are stable over time. Appendix A Figure 2.2 gives a detailed overview about the timing of the data collection. Many other field experiments (e.g. O'Neill et al., 2013) lack a longitudinal design which makes it difficult to interpret their results because they can not make statements about the persistence of program effects. This is especially important in early childhood where child development is very dynamic. The next sections explain the different tests and interviews in more detail. [3]

2.3.3.1 Developmental Tests and Personal Interviews

Developmental tests with the children are conducted at 6, 12 and 24 months after birth. At each test the Mental Developmental Index (MDI) of the Bayley Scales of Infant Development (BSID) measures the cognitive abilities (IQ) of the children (Bayley, 1969). Additionally, the fine and gross motor abilities, called the motor quotient (MQ), are tested at each assessment by the Psychomotor Developmental Index (PDI) of the Bayley Scales. A language test for two year old children (*Sprachentwicklungstest für zweijährige Kinder*, SETK-2) (Grimm, 2000) is conducted at 24 months. The BSID and the SETK-2 tests are video taped and assessed after the interview by a developmental psychologist who does not know the treatment group of the child. An important advantage of the

[3] In addition to the interview and development data, the *Pro Kind* research obtained health system use data of the General Local Health Insurance (*Allgemeine Ortskrankenkasse - AOK*) Lower Saxony, Bremen and Saxony and data of the *Kassenärztliche Vereinigung* Lower Saxony and *Kassenzahnärztliche Vereinigung* Lower Saxony and Bremen. These data will be examined in future research.

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 standard deviation (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 pedeatric assistance. If a child in the home visiting or the control group scores below these thresholds, the mother gets special information and advice, additional to the regular feedback of the research. MDI and PDI tests consist of different tasks. Sometimes the children refuse certain tasks of the full test battery. If the refusal or interruption rate of these tasks in one test exceeds 20 percent, the overall test result is not reliable anymore.

In combination with the developmental test the research conducts personal interviews with the mothers. These interviews contain scales which measure maternal mental health (Depression Anxiety Stress Scale – DASS (Lovibond and Lovibond, 1995)), self efficacy (Parent Expectations Survey – PES (Reece, 1998)) and attachment to the child (Maternal Postnatal Attachment Scale – MPAS (Condon, 1998)). Furthermore, the mothers are asked for the problem behavior of their children (Child Behavior Check List – CBCL (Arbeitsgruppe Deutsche Child Behavior Checklist, 2002)). All these scales base on maternal self-ratings which can be biased by social desirability. However, this problem might occur in both groups to the same amount. Therefore, it is unlikely that the impact estimates are biased. Finally, in addition to the interview and the developmental test the mother's record of natal care (*Mutterpass*) is copied in all interviews.

Table 2.7 reports the rates of developmental tests for treatment and control group and child gender. Although the baseline comparisons presented in Table 2.3 show that the treatment and control groups were similar at the baseline, it is possible that the two groups become incomparable if in one group more participants attrite than in the other. However, the rates of completed development tests indicate that there are statistically no significant differences between the control and treatment group and between child

Table 2.7: Sample Composition Developmental Tests

	Control	**Home Visiting**	**Total**
Allocated to Treatment	361	394	755
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*

genders. In both groups about one third of the 6 months tests are not available. The attrition rate for the 12 months test is about 45 percent of the baseline participants and for the 24 months 55 percent are missing.

Missings in the development tests occur because of the same reasons why mothers attrite from the home visits. For treatment and control group this is mainly because the mothers lack interest in the tests or the mothers were not reachable anymore. However, children also miss developmental tests because these tests are sensitive to time. If a test is not conducted within a time span of two months, the test is not reliable anymore and, therefore, this test is skipped. Hence, not all children of mothers who received the complete home visits are tested at each time point. In case of the treatment group 217 children of the 255 mothers (85.1 percent) who participated in the home visits are tested at 12 months and 77.6 percent at 24 months.

The time sensitivity of the test also leads to cases where 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 test. The power analysis, which was conducted before the program started, considered an attrition rate of 25 percent to detect with 80 percent probability effects with an effect size of 0.2 SD. If all available data is considered,

the assumptions in the calculation are almost met.

In the treatment group almost all families who participated in the development tests received at least some treatment. To be more precise, at six, twelve and 24 months only one family in which the assessments were conducted accepted no home visit. In average families who took part in the six months assessment got 18.5 (SD=5.5) home visits, those who took part in the twelve months assessment got 29.6 (SD=7.7) home visits and in average 45.4 (SD=10.6) home visits when they took part in the 24 months assessment.

2.3.3.2 Telephone Interviews

The telephone interviews start during pregnancy and continue afterwards every 6 months up to the third birthday. The interviews are computer assisted and contain questions about household, income, employment, childcare use, family planning as well as questions about service utilization by mother and child. The questionnaire includes all questions, which are recommended when using GSOEP as a reference data set (Siedler et al., 2009). Furthermore, the interviews include the GSOEP activity calendar to learn about the employment status of the participants on a monthly base. Other sources of the questionnaires are the Panel for Labour Market and Social Security (PASS) (Trappmann et al., 2010), The German Health Interview and Examination Survey for Children and Adolescents (KiGGS) (Kamtsiuris et al., 2007), the "Deutsches Jugend Institut" (DJI) family survey (Bien and Marbach, 1989) and the World Value Survey (WVS, 2009).

Most variables in the questionnaire measure time durations in monthly intervals. This gives an exact insight into the lives of the young mothers and their families. Due to the high frequency of the interviews the danger of recall bias is low. Furthermore, most questions in the questionnaires ask for objective outcomes like whether the mother is occupied or not. Research about questionnaire design concludes that for those questions the answer reliability is high (Bradburn, 2004; Groves et al., 2004) (see Lutz and Sandner, 2010, for more details about the telephone interviews).

Table 2.8: Sample Composition Telephone Interviews

Treatment Group	**Control**	**Home Visiting**	**Total**
Allocated to Treatment	361	394	755
Research Refusals	31	30	61
Fetal Demises	8	2	10
Infant Deaths	2	2	4
Completed Telephone Interviews			
34-Weeks Pregnancy	320 (88.6%)	360 (91.4%)	680
3 Months	273 (75.6%)	309 (78.4%)	582
9 Months	229 (63.4%)	257 (65.2%)	486
15 Months	204 (56.5%)	238 (60.4%)	442
21 Months	195 (54.0%)	223 (56.6%)	418
27 Months	199 (55.1%)	239 (60.7%)	438
36 Months	169 (46.8%)	205 (52.0%)	374
Complete data until second birthday	161 (44.6%)	187 (47.5%)	348
Complete data until third birthday	137 (38.0%)	159 (40.4%)	296

Table 2.8 demonstrates the sample composition for the conducted telephone interviews. During pregnancy the participation rate is very high. After birth of the child the rate declines and stabilizes at 60 percent from 9 months to 27 months. Only for the last interview the rate declines again to about 50 percent. The drop at 36 months is caused by the end of the research project in November 2012. At this point of time, some children had not reached the age of three yet. It is likely that the participation rate would reach 60 percent also in this interview if data collection would have continued. Participation in the telephone interviews is slightly higher in the treatment group, but the difference is not significant at a ten percent level at any interview, tested by a two side proportion test.

The telephone interviews tried to contact all randomized mothers at each time point. The only exception was an infant death or a fetal demise of a participating mother. In contrast to the personal interviews, the telephone interviews did not stop if the mother moved away from a *Pro Kind* location or because of an intervention by the child protection service. Furthermore, the telephone interviews were more flexible in time than

the personal interviews and therefore it was possible to consider time constraints of the mothers. Because of this, the research also conducted telephone interviews with mothers in the treatment group who were not visited anymore by the home visitors. At 15 months, 214 mothers in the program and 33 mothers who attrited were interviewed. At 27 months, these numbers were 190 interviews to 49 interviews.

Like in the personal interviews refused participation or switching mobile numbers were the main reasons why missings occur. To minimize missings, the interviewer tried to contact the participant four times within two months. If no contact could be made in this time span, the interviewer tried to contact the mother for the next scheduled interview which was four months later. If the contact could be realized for this interview, a combined interview was conducted. However, no interview covers a time period of more than 12 months in order to avoid recall bias. Therefore, similarly to the personal interviews, some participants miss only one or two telephone interviews. Finally, 296 mothers have participated in all interviews until the third birthday and 348 mothers participated in all interviews until the second birthday.

Like in the development test the number of mothers who participated in the telephone interviews but received no home visit is very small. Only one mothers who were available for all interviews until the third birthday of the child got no home visits. In the group of those mothers who participated until the second birthday there is also only one mother without home visits. In average those mothers who participated in all interviews got 42.5 (SD: 12.7) home visits and 42.7 (SD:13.2) home visits if they participated in all interviews until the second year. Looking at each interview separately reveals that seven mothers who did not received any home visit participated in the telephone interview during pregnancy. In the interview directly after pregnancy this number decreased to three and afterwards it declined to two or one.

2.3.3.3 Treatment Effects

The previous sections described on the one hand the data which is available to evaluate the outcomes of the *Pro Kind* program and on the other hand the compliance of the participants to their assigned treatment. Both information is necessary to assess which types of treatment effects are possible to calculate and how to interpret these effects. The explanations in this subsection base on Imbens and Wooldridge (2009); Angrist and Pischke (2009); Kling et al. (2007) where also a more detailed discussion about treatment effects can be found.

In an optimal setting with full compliance (all participants who are assigned to receive the *Pro Kind* treatment do comply with their assignment) and no data is missing (all outcomes of interest are measured), there are no problems in the interpretation of the effects. It is easy to calculate the Average Treatment Effect (ATE) which is $ATE = E[Y_{1i} - Y_{0i}]$, where Y_{1i} is the outcome with treatment and Y_{0i} the outcome without treatment. In a setting with full compliance and no missings the difference between the outcomes gives the average causal effect of the treatment.

In the *Pro Kind* program arise two problems which make the direct calculation of the ATE impossible: compliance is not complete and data missings occur. Examining the first problem in more detail shows that in the case of *Pro Kind* one-sided noncompliance occurs. Noncompliance means that participants who are assigned to receive the *Pro Kind* treatment do not comply with their assignment, and instead receive an alternative treatment (which is in the case of *Pro Kind* similar to the treatment in the control group). One-sided means that only units assigned to receive the active treatment can potentially circumvent their assigned treatment and receive the opposite treatment; all units assigned to receive the control treatment do, in fact, comply with this assignment. In the case of *Pro Kind* the noncompliance is one-sided because it was possible for the program staff to reject the access to the treatment for the control group members.

Traditional formal statistical analyses of randomized experiments with noncompli-

ance in general focus on the relationship between the random assignment and the outcome of interest, discarding entirely any information about the treatment actually received. Such an approach is generally referred to as an Intention-To-Treat (ITT) analysis. The main drawback of these ITT analyses is that they do not answer questions about causal effects of the receipt of treatment itself, only about causal effects of the assignment to treatment. In order to calculate the causal effects of the receipt of treatment, the offer of the *Pro Kind* treatment can be used as an instrumental variable for *Pro Kind* participation. The so estimated effect is called the local average treatment effect ($LATE$) which can be interpreted as the average causal effect of the receipt of treatment for compliers and gives the effect which tries to imitate the ATE. However, the problem of one-sided noncompliance is very small in the *Pro Kind* context. Only nine mothers who were assigned to the treatment group were not compliant with their assignment. All other mothers assigned to the treatment group received at least some treatment. Therefore, the difference between $LATE$ and ITT is very small.

Unfortunately, in the case of *Pro Kind* the problem of missing data arises, too. The child of only one mother who were not compliant to her assignment was tested in the development assessments. Without assumptions about the missing data, the missing data can not be included in the estimates by definition.[4] Using only cases in which outcome data is available leads to a compliance rate of over 99.5 percent in each assessment point. Therefore, using the offer of the *Pro Kind* treatment as an instrumental variable leads to a small difference between $LATE$ and ITT. Because of the small difference between $LATE$ and ITT and in order to improve clarity, I decided to present just the ITT estimates in the following sections.

[4]Only in section 3.4 assumptions about the missing data are made as a robustness check using a multiple imputation framework. These assumptions consider the findings presented in Tables 3.2, 4.16 and 4.17 that missings data occurs more often with more disadvantaged mothers.

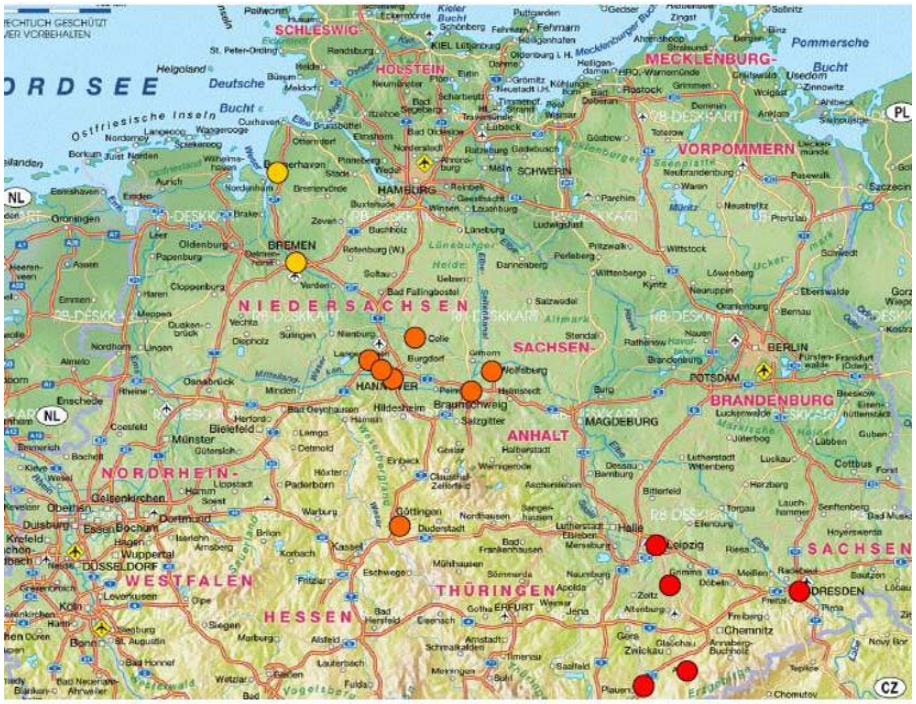
2.4 Conclusion

This chapter explained the implementation of the *Pro Kind* program. It proved that *Pro Kind* is a high quality home visiting intervention. The quality elements consist of theoretical foundation, voluntary participation of the clients, an accurately defined client group, a high visiting frequency, more than two years of visitation, trained home visitors and supervision for the home visitors. Results of the implementation research explained that these high quality elements were successfully implemented. Therefore, it is likely that the three aims of the program (improvement of maternal and child health, improvement of child development and improvement of the maternal life-course) will be achieved.

Furthermore, *Pro Kind* was successful in establishing a randomized trial with 755 participating mothers. The ambitious evaluation research consisting of development tests, personal interviews and telephone interviews was conducted as planned. However, the attrition rate in the tests and interviews is high. This can be problematic because sample size is reduced and selective attrition between treatment and control group can bias the impact estimates. However, the remaining sample size has still enough power to detect impacts resulting from the treatment and with the rich baseline characteristics available for each participant it can be tested whether the attrition is non-random.

2.5 Appendix A

Figure 2.1: *Pro Kind* Locations



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

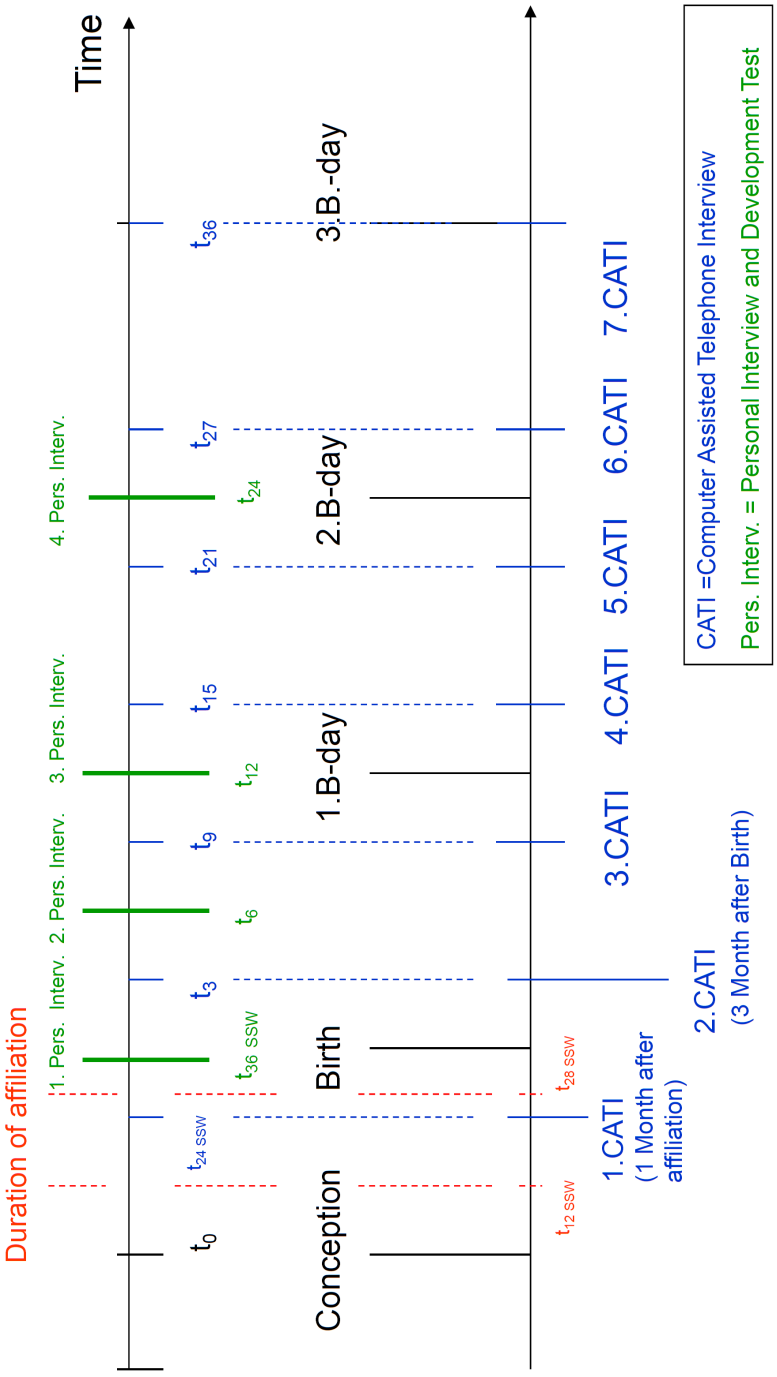
Table 2.9: Baseline Variable Definitions - Demographic Characteristics

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 in €	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 is without School Degree or Apprenticeship	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

Table 2.10: 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

Figure 2.2: *Pro Kind* Research Schedule



3 The Effects of *Pro Kind* on Child Development and Early Skill Formation[1]

3.1 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

[1]Part of this research was supported by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth *(Bundesministerium für Familien, Senioren, Frauen und Jugend - BMFSFJ)*.
The paper was presented at the Annual Conference of the Scottish Economic Society 2012, the 17th Spring Meeting of Young Economists (SMYE) 2012, the 1st Lower Saxony Workshop in Applied Economics 2012, the 26th 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 and the Ausschuss für Bildungsökonomie des VfS 2013. This paper is accepted for presentation at the Annual Conference of Royal Economic Society (RES) 2013 and the Annual Conference of the Society of Labor Economics (SOLE) 2013.

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 benefit-cost 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, documents 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 indicate 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 the effectiveness of 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 provide information about birth outcomes and investments into the child. 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., 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 distribution of treatment effects, 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, 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. As an explanation for the gender specific outcomes, I find that the treatment enhances parental investments in a different magnitude for boys and girls. 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 receive feedback about the test results. Imputing missing test scores with test scores from earlier assessments leads 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 for the whole sample.

The rest of the chapter is organized as follows: Section 3.2 discusses the impact of attrition on the internal validity and presents descriptives of the data set. Section 3.3

presents results on the impact of the home visiting program on birth outcomes, mental and psychomotoric development as well as language development. Section 3.4 conducts robustness checks and presents evidence that the main effect of the intervention might be downward biased. Section 3.5 analyses the dynamics of the skill development. Section 3.6 discusses aspects of the cost-effectiveness of the home visiting program. Section 3.7 presents conclusions.

3.2 Data

3.2.1 Attrition

To analyse the effects of *Pro Kind* on child development, birth outcomes and parental investments, I use data of the development tests, the personal interviews and data of the three month telephone interview described in section 2.3.3.1 and 2.3.3.2. Table 2.7 showed that the attrition rates for the tests and interviews were similar in treatment and control group. However, the characteristics of the attritors and 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 3.1. The 6, 12 and 24 months tests are investigated in Column 2, 3 and 4, respectively. As in section 2.3.3.1, I run regressions of mother and family characteristics from the baseline survey on treatment status, only including the mothers and families who participated in the interviews or tests.

All differences between treatment and control group 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 mothers in treatment and

control group refuse participation in the interviews and tests. Table 3.2 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 characteristics are less correlated with attrition. The differences in the demographic risk factors mainly occur between the baseline and the 3 months interview. Afterwards, the difference between attritors and non-attritors stays constant. The only characteristic which continuously decreases in the attritors group is the age. At the 24 months assessment, participating mothers are more than two years older than their 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 cumulative sum of risk factors is 5.45 in the non-attritors group in contrast to 6.08 in the attritors group.

MDI and PDI tests consist of different tasks. Sometimes the infants and toddlers refuse certain tasks of the full test battery. If the refusal or interruption rate of these tasks exceeds 20 percent in one test, the test results are not reliable. Therefore, the Bayley Scale guidelines recommend not to use these results (Bayley, 1993). Because of this, 38 (TG:18, CG:20) MDI test results at 6 months, 37 (TG:22, CG:15) MDI test results at 12 months and 48 (TG:23, CG:25) MDI test results 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). Appendix B Tables 3.13 and 3.14 demonstrate that also without these unreliable MDI and PDI tests the two groups only slightly differ regarding the baseline characteristics.

Table 3.1: 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)
Month. HH-inc. 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 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 shown in parentheses. Estimates include community fixed effects. See Table 2.9 and 2.10 for variable definitions.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Selective Attrition to Baseline - Development Tests

	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 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 shown in parentheses. Estimates include community fixed effects. See Table 2.9 and 2.10 for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

3.2.2 Descriptives

In order to allow a better interpretation of the intervention outcomes, Table 3.3 gives a combined overview of the birth outcomes and the results of the reliable tests 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 the GSOEP data (T=5.6). The gender difference in birth outcomes is similar to the GSOEP 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. At 12 months all 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, only in MDI at 6 months the difference is statically significant at a five 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. Appendix B Figures 3.2 and 3.3 present density graphs of birth outcomes and child development tests scores by gender.

3.3 Estimating Program Effects

3.3.1 Specification Model for Estimating Treatment Effects

I estimate the *Pro Kind* effects on child development by OLS-regression analysis using equation 3.1:

$$Y_{ic} = \beta_0 + \beta_1 HV_{ic} + \beta_2 h_{ic} + \alpha_c + \epsilon_{ic}, \tag{3.1}$$

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

Table 3.3: Descriptive Statistics Child Outcomes

	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 Circumfer. 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 Circumfer. 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.

family characteristics at baseline. I also include a dummy variable α_c for each community to absorb the community effects. The outcomes of interest are the standardized birth weight, birth height and birth head circumference, the standardized 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 missing values occur in the covariates, I include sample means or imputed values. However, the results are robust to including only covariates with very few missing values. For the analyses 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 early childhood interventions. I decided against clustering the standard errors at the community level, because I am concerned that clustering would produce biased standard errors as the observation sizes of the clusters is very unbalanced.[2]

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

3.3.2 Impact of *Pro Kind* on Birth Outcomes

Table 3.4: 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 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 3.4). 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

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

varies only slightly with the model specifications, which shows 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. However, this effect becomes insignificant when controls and mainly maternal smoking are included. Appendix B Figure 3.4 presents density graphs of birth outcomes in the treatment and control group.

3.3.3 Impact of *Pro Kind* on Child Development

Table 3.5: 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

My analysis of home visiting effects on cognitive abilities (MDI) or fine and gross motor abilities (PDI) begins with the whole sample (Table 3.5). 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 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 an effect size close to zero. At all assessment points the coefficients change only slightly when controls are included confirming the validity of the randomization. Appendix B Figure 3.5 shows the density graphs for MDI and PDI at 6, 12 and 24 months in the treatment and control group.

Table 3.6: 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 3.6). 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. The differences between the models with and without controls are small for girls, but larger for boys. Appendix B Figures 3.6 and 3.7 present the density graphs for MDI and PDI at 6, 12 and 24 months in the treatment and control group separated by gender.

3.3.4 Impact of *Pro Kind* on Language

Table 3.7: 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The SETK-2 results (Table 3.7) 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.

3.3.5 Discussion of the *Pro Kind* Program Effects

The first result of the analyses in the previous sections is that *Pro Kind* increased the cognitive development (MDI) of the children, whereas the intervention did not effect the psychomotoric development (PDI) and the language skills only in a small magnitude. One explanation for the different effects might be that the treatment only increases investments in child'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, is correlated with cognitive development (Baker and Milligan, 2013; Price, 2012), on the other hand activities like going with the child to the playground, for which the mother has to leave the house, is correlated with psychomotoric development. This explanation is supported by the results of Tables 3.8 and 3.15 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 by the intervention. Also this gender difference might be explained by the influence of the treatment on cognitive activities. The figures in Table 3.8 reveal that the treatment enhances the cognitive activities stronger for girls than for boys at five of the six assessment points. This is especially true for reading or telling stories at 6 months 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 assumes that investments in boys are more costly than investments in girls. A recent study by Baker and Milligan (2013) supports this assumption. They document for the US, 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 9 months of age. They explain their results that it is less rewarding to provide inputs (like reading time) to boys than to girls because, for example, boys wiggle and squirm more than girls.

The third finding of the *Pro Kind* analysis is that the effects 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 6 months the treatment effects for children of these mothers are not higher than for other mothers. Also the attrition of mothers with other characteristics does not cause the fade-out. Therefore, the lesser home visitation frequency seems to explain the smaller effects at 24 months. However, it is important to note that fade-out is also found for pre-school programs like Head Start or the Perry Pre-School Program (Currie and Thomas, 1995; Anderson, 2008) and, despite rapid fade-out of 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 difficult to measure at infant age.[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 3.8: 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.4 Sensitivity Analysis

The previous sections showed that attrition did not cause 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 3.9 documents that this is the case for the MDI. At 6 months children of attriting mothers in the control group score 5.2 points lower at the MDI than children of attriting mothers in the treatment group.[4] The effect is smaller but still significant at 12 months.

[4] Attriting means that a child participates in one test but does not participate in the subsequent test.

Table 3.9: 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 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

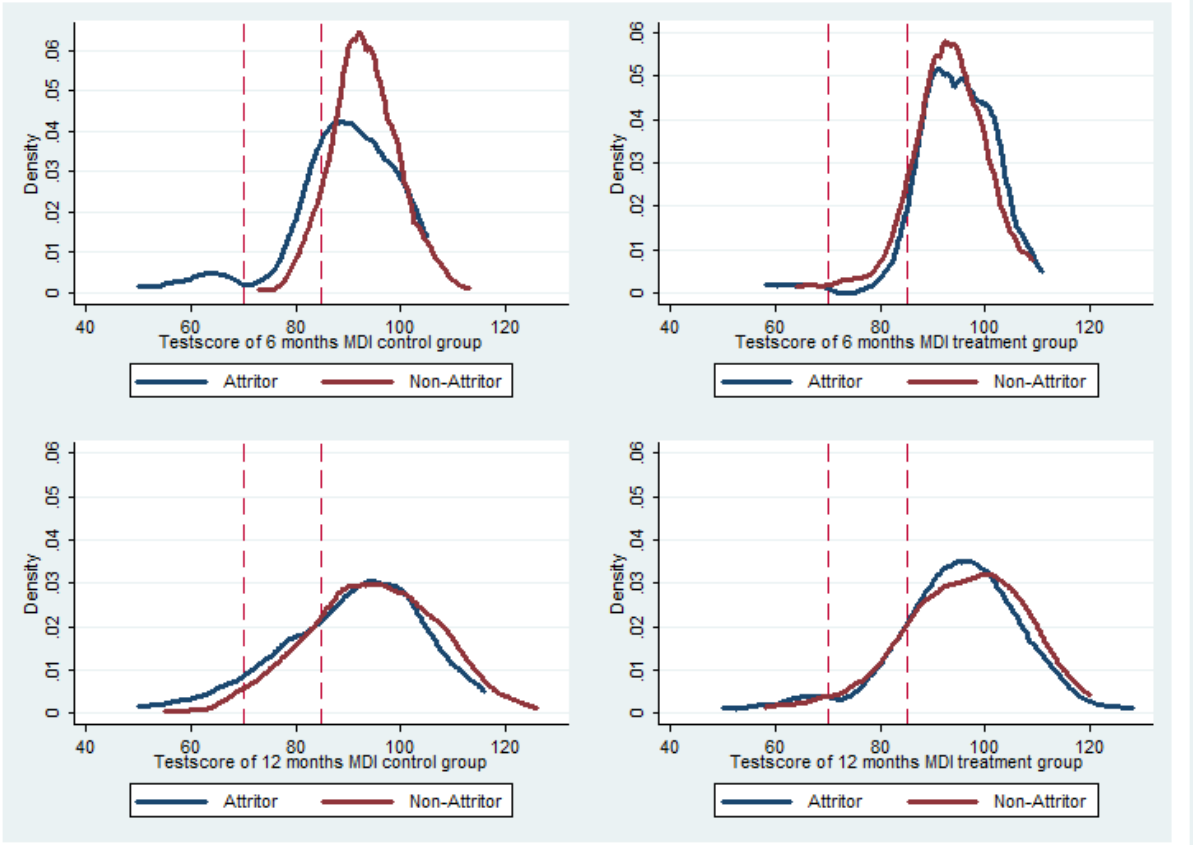
P-values base on two side T-tests. Appendix B Table 3.16 describes the composition of the attritors
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This selective attrition could be caused by the procedure that the research sends a letter to the mothers in both groups with 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 3.1 supports this hypothesis. It compares the distribution of the MDI 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 the letter to the mothers contains the information that their child has a mental delay (below 85) or serious mental delay (below 70). For mental delay the term (*geistige Verzögerung*) is used which is a quite harsh term in the German language. The attrition of all control group mothers of children which scored less than 70 points in the MDI at 6 months supports that this additional information is a major reason for attrition.

If mothers with low performing children in the MDI attrite, one can ask why this is not the case for PDI. Mothers of children who scored low at the PDI received in their letter the information that their child has movement difficulties (*Schwierigkeiten bei der*

Figure 3.1: Comparison MDI Test Scores Attritors and Non-Attritors



Beweglichkeit) which is a less harsh term in the German language. This different 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.

To correct for this selective attrition at the MDI, I impute missing test scores by 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 the way that children participate in the first test or tests and than refuse participating. However, there also cases in which children just miss the first test or the second test (Appendix B Table 3.16). 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 analysis

Table 3.10: 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	0.155	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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the imputation is conducted separately for boys and girls.

The imputation increases the effects of MDI at all three assessment points (Table 3.10). As in section 3.3.3 the effect is highest at 12 months and gets smaller at 24 months. Looking at gender difference after the imputation reveals the same picture as without the imputation. The effect of the home visiting is greater for girls than for boys (see Table 3.11). At 12 months girls in the treatment group score 0.36 SD higher than girls in the control group. These results demonstrate how sensitive the effect sizes react to the selective attrition of the bad performing children in the control group. Although the results in Table 3.10 might be exaggerated because of the imputation assumptions, it is likely that the estimates with the pure data in Table 3.5 give the lowest bound of the treatment effects.

Table 3.11: 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	-0.036	-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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.5 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 prove if the dynamics in the skill formation process, predicted by Cunha and Heckman (2007), occur already at this early stage. If this is not the case, the efficacy of programs which try to enhance skills in 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 3.2 formalises this skill production function, where S_t denotes the vector of skills acquired at stage t.

$$S_{t+1} = f_t(h, S_t, HV) \tag{3.2}$$

Like in Equation 3.1, 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

$$\frac{\partial S_{t+1}}{\partial S_t} = \frac{\partial f_t(h, S_t, HV)}{\partial S_t} > 0, \tag{3.3}$$

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 complementarity apply 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)[5].

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 3.2. In Equation 3.4 $S^k_{t,i}$ 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_i, at stage t_2,t_3 and t_4 I use 6, 12 and 24 months MDI and PDI test scores as measure for S^k_i

$$S^k_{t+1,i} = \gamma S^k_{t,i} + \phi HV + \eta h + \epsilon_{i,t} \tag{3.4}$$

My coefficients of interest are γ indicating self-productivity or direct complementarity and

[5]Chapter 3.6 gives are more detailed overview about the MARS.

ϕ indicating the effects of the home visiting investment. All variables are standardized as explained in Chapter 3.3.

Table 3.12 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 months PDI for 12 months MDI. For girls self-productivity is sustained in 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 Tables 3.5 and 3.6 with the exception of the coefficient for boys on 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 MDI from 3 months to 2 years with a coefficient of 0.3. However, Coneus et al. (2012) found only small gender specific differences regarding the skill formation process. The more detailed *Pro Kind* data reveals that boys younger than 12 months do 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, compared to previous findings, the new insight that already from age 6 months self-productivity gradually increases and that, on the other hand, direct complementarity just 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. Firstly, the hypotheses that interventions which start prenatal 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. Secondly, because direct complementarity are low at this age, home visiting has to concentrate on each skill separately if it wants to affect each skill. Thirdly, the strong correlation between later test scores and test scores from previous periods document that the imputation method in chapter 3.4 was at least partly justified. Fourthly, the coefficient of HV indicates that the main reason for the insufficient effect for boys lies in the first 6 months of the home visiting. In contrast, at 12 months the net effect is comparable with the girls' effect. Here the open question remains if the small effects for boys in the first 6 months are related to the fact that there is no self-productivity between 6 and 12 months for boys. This is a question for further interdisciplinary research which also has to examine if home visiting should intervene differently for mothers of boys at this age.

Table 3.12: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6 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 the NFP. The second one analyzes the relationship between increased cognitive ability in the first two years of life and the probability to attain the *Abitur*, the highest German school degree.

Since *Pro Kind* is an adaption of the NFP Program I use the results of NFP as a benchmark for the *Pro Kind* results. An interdisciplinary research team evaluated the NFP Program in three different trials. The first trial 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, 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. Although only in one trial effects on the MDI were found, in all three trials other program effects occurred such as lower childhood injuries or fewer subsequent pregnancies.[6]

There was follow-up research with different time horizons in all three trials. In the Elmira trial, data is available for 19 years, in Memphis for 12 years and in Denver only for 4 years after birth. 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.,

[6]Until now there have been no reevaluation of gender differences for the NFP. However, at the moment the economic department at Chicago University under the direction of James J. Heckman conducts gender reevaluations with the NFP data. (http://heckman.uchicago.edu/page/nurse-family-partnership-nfp)

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 from the Bayley Scales 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 costs about \$11,511 for each participant in the Denver trial in 2006 (Olds et al., 2010). *Pro Kind* costs approximately €8,705 in 2010 which is \$11,752 assuming an exchange rate of 1.35 €/\$ (Maier-Pfeiffer et al., 2013). 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 also effects on the children (mainly a decrease of anti-social behavior) influenced the cost-benefit relation. Since MDI was not significantly changed at age one, it seems that improvments in other domains have a stronger influence on anti social behavior. In contrast, MDI might influence school success. If this is the case, the missing effects on MDI may explain the missing effects 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 I 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 (Laucht et al., 1997). 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., 2012) 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. I use this correlation to assess the *Pro Kind* impact on MDI in a numerical calculation.

At 24 months the *Pro Kind* effect on the MDI is 0.08 SD with the pure data and 0.16 SD with the imputed data. I use an effect size of 0.15 SD for the numerical example, also accounting for the greater effects sizes at 12 months. 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, a workforce entry age of 25 and 40 years of workforce participation. Following the MARS correlations, the *Pro Kind* effect of 0.15 SD on cognitive development increases the probability to attain high school by 1.95 percent. This means, on average, a higher discounted life time income of €2,515.48 for the home visited, which is 29 percent of the intervention costs. There are no analysis for the MARS data if an higher MDI at age two is correlated with the probability of dropping out or class repetition. If I assume that not only 1.95 percent more children attain *Abitur*, but also 1.95 percent less children leave school without a degree, the effect on life earnings would be €4,713.49[7] which is 54 percent of the intervention costs.

These figures reveal that if only the income effect of an increased MDI is considered, the break-even point is not reached. However, it is likely that an improved MDI also reduces unemployment and class repetition. Furthermore, increased cognitive development can directly or indirectly lead to improvement in many other areas like health or criminality. If these effects are considered, an effect size of 0.15 SD on cognitive development will clearly lead to a benefit-cost ratio greater than one. This conclusion is supported by benefit-cost analyses of 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

[7]The calculation base on figures from (Autorengruppe Bildungsberichterstattung, 2012). It includes the increased life earning for attaining the Abitur and for attaining the German basic degree (Hauptschule) instead of no degree. The calculations include just an increase in earnings and no decrease in welfare spending.

an effect size of 0.06 SD on cognitive development is enough to reach the break-even point with program costs of $6000 and a three percent discount rate.

3.7 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 measures for child development demonstrates 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 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 in a different size. 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 also 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 size of the *Pro Kind* effect on child's cognitive development is in line with other home visiting programs Sweet and Appelbaum (2004). The programs in this study start often after birth and do not focus on the first child whereas *Pro Kind* is a cost intensive high-quality program. Therefore, it was expected that *Pro Kind* generates higher effects than other programs. 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, 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 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.

3.8 Appendix B

Table 3.13: 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 Table 2.9 and 2.10 for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

Table 3.14: 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 Table 2.9 and 2.10 for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

Figure 3.2: Birthoutcomes for Boy and Girls

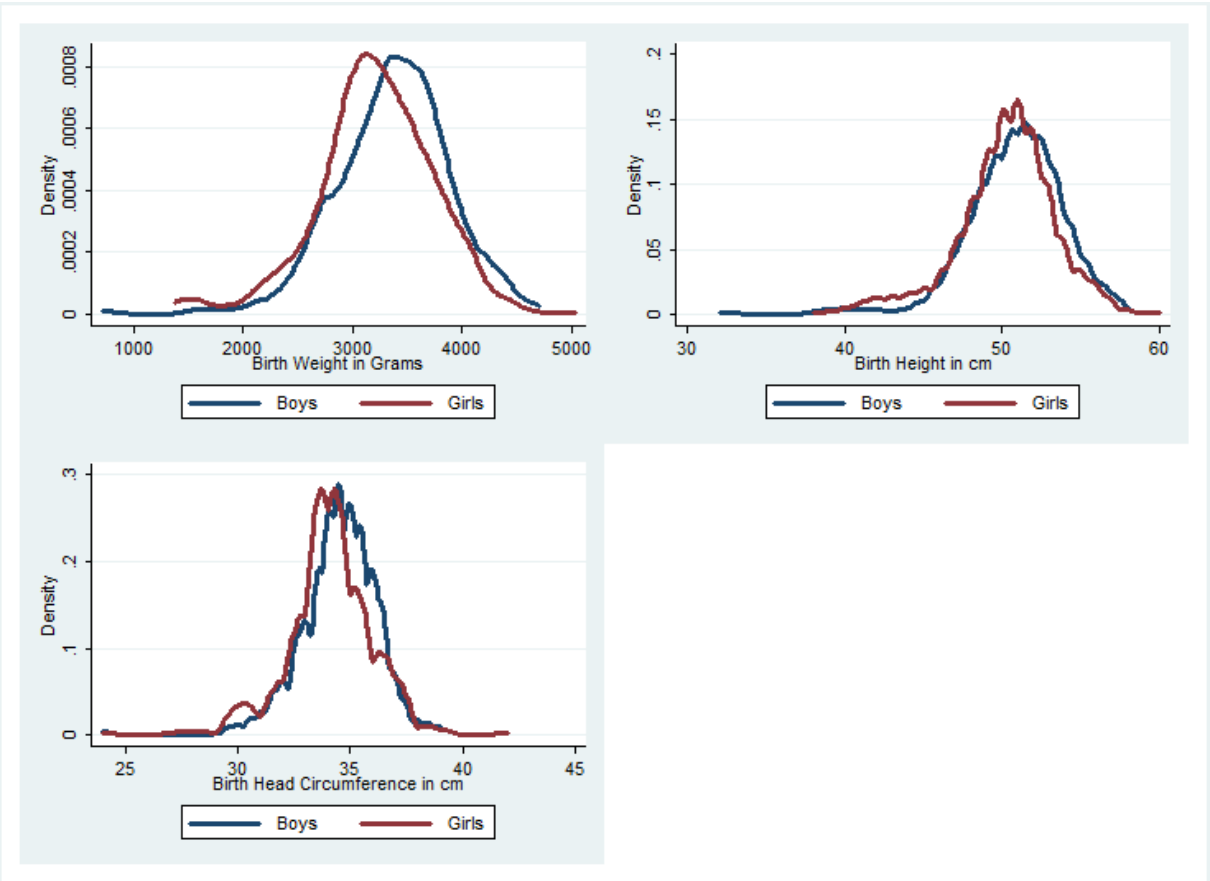


Figure 3.3: BSID Test Scores for Boy and Girls

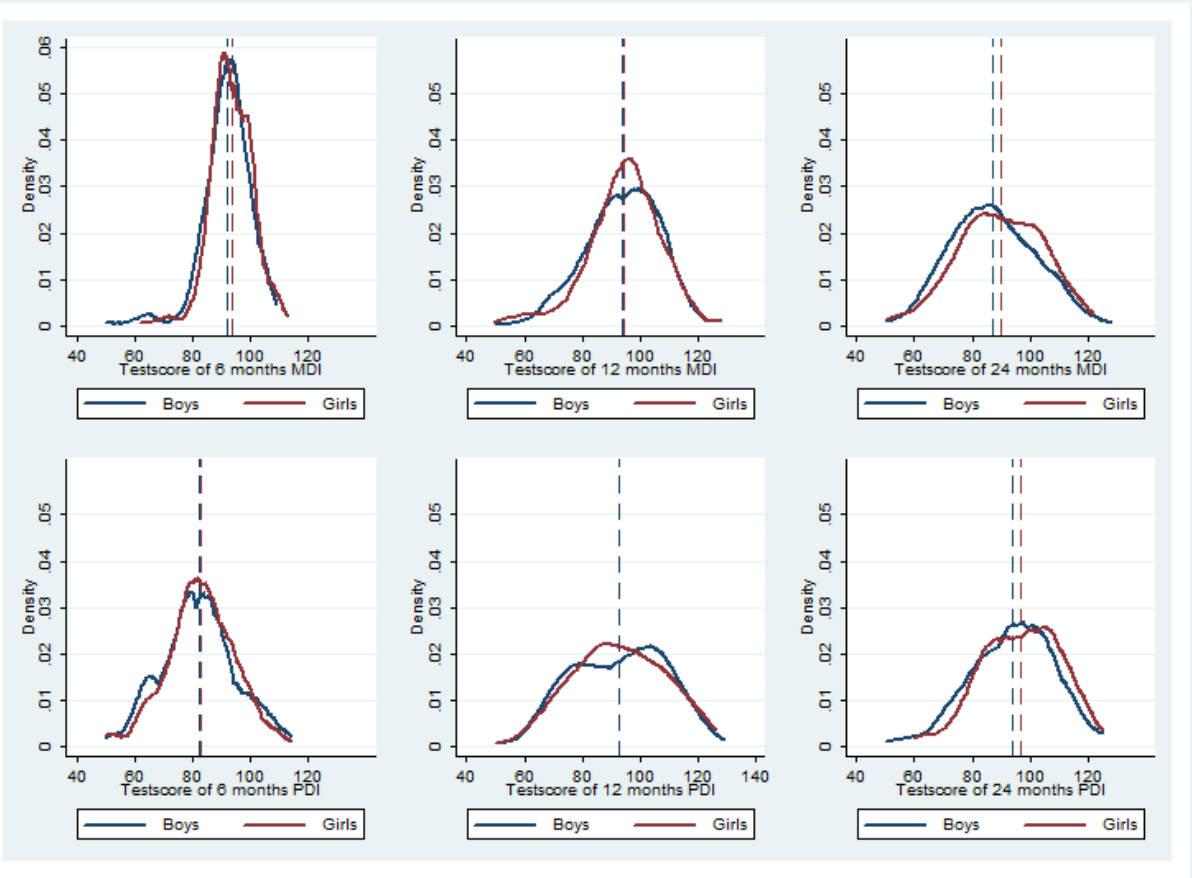


Figure 3.4: Birthoutcomes for Treatment and Control Group

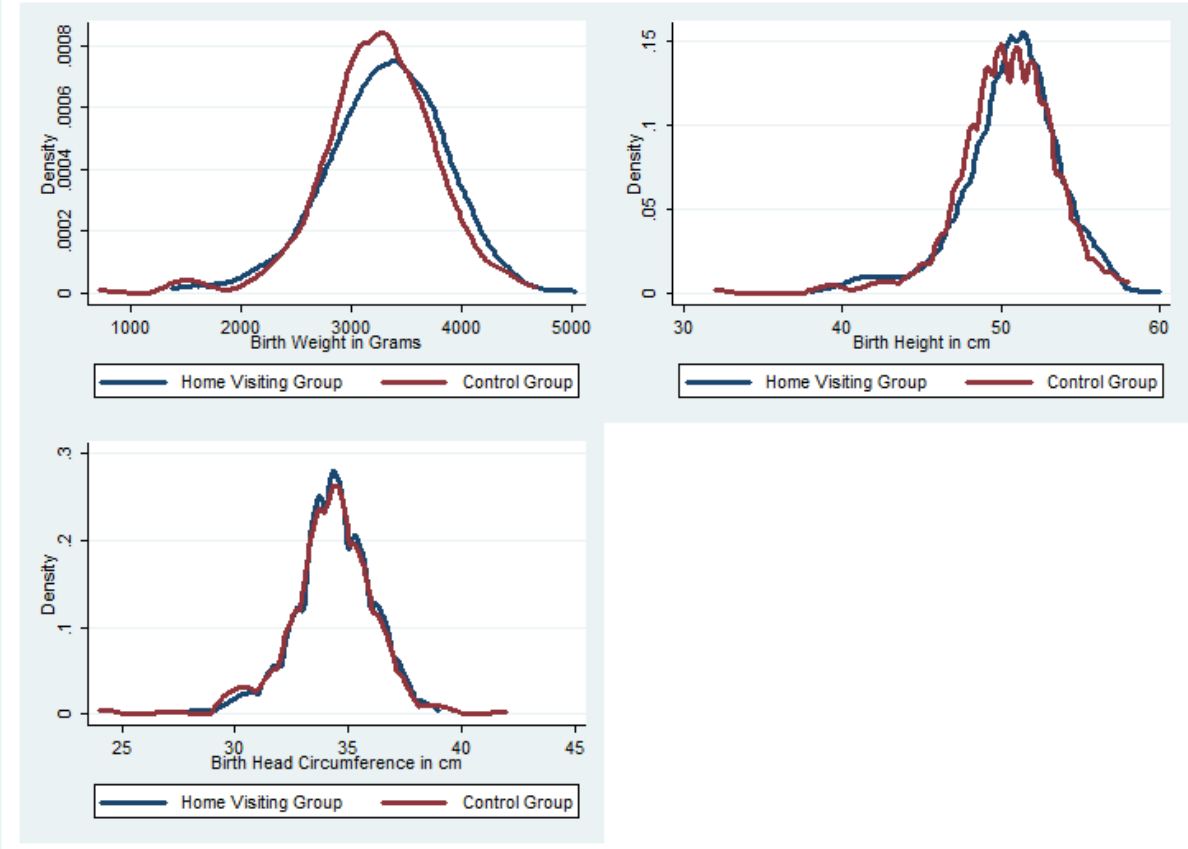


Figure 3.5: BSID Test Scores for Treatment and Control Group

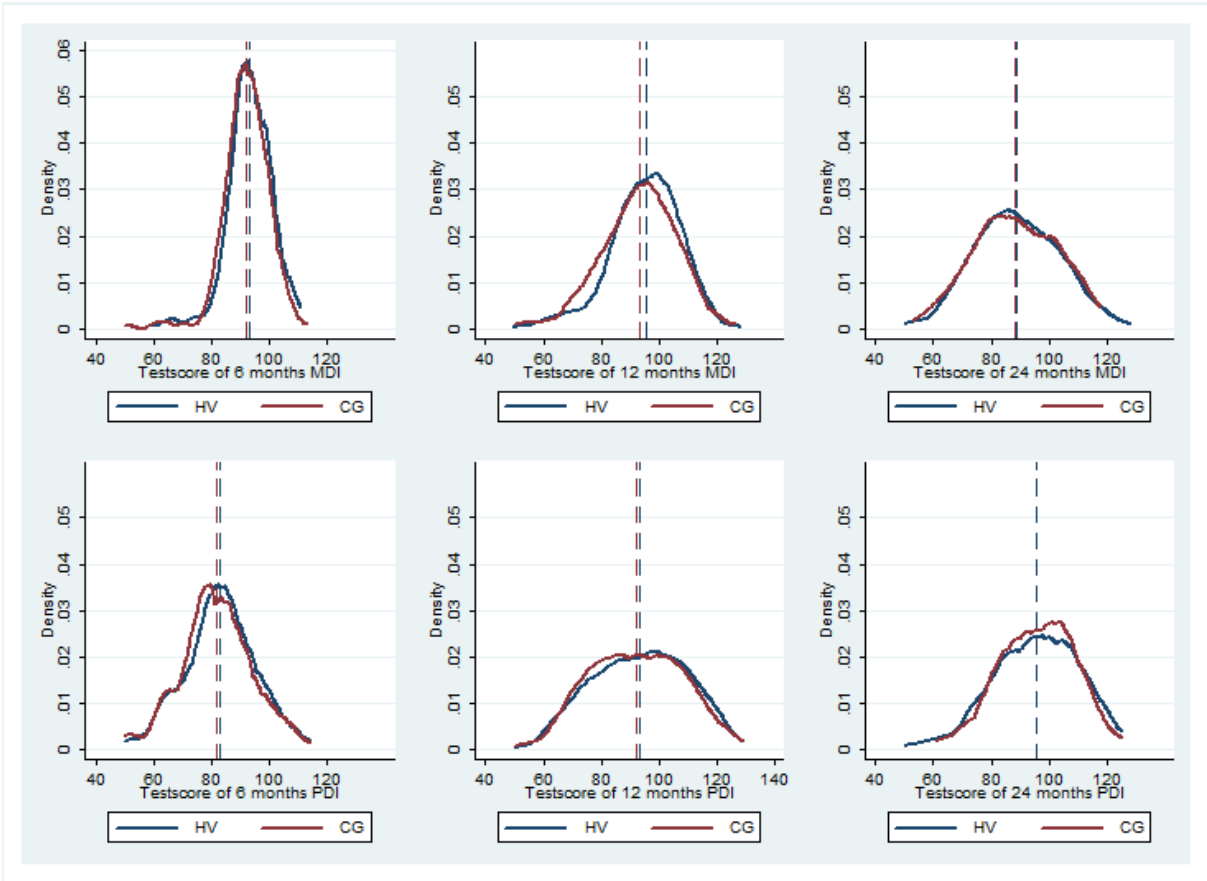


Figure 3.6: MDI Test Scores for Treatment and Control Group (Boys and Girls)

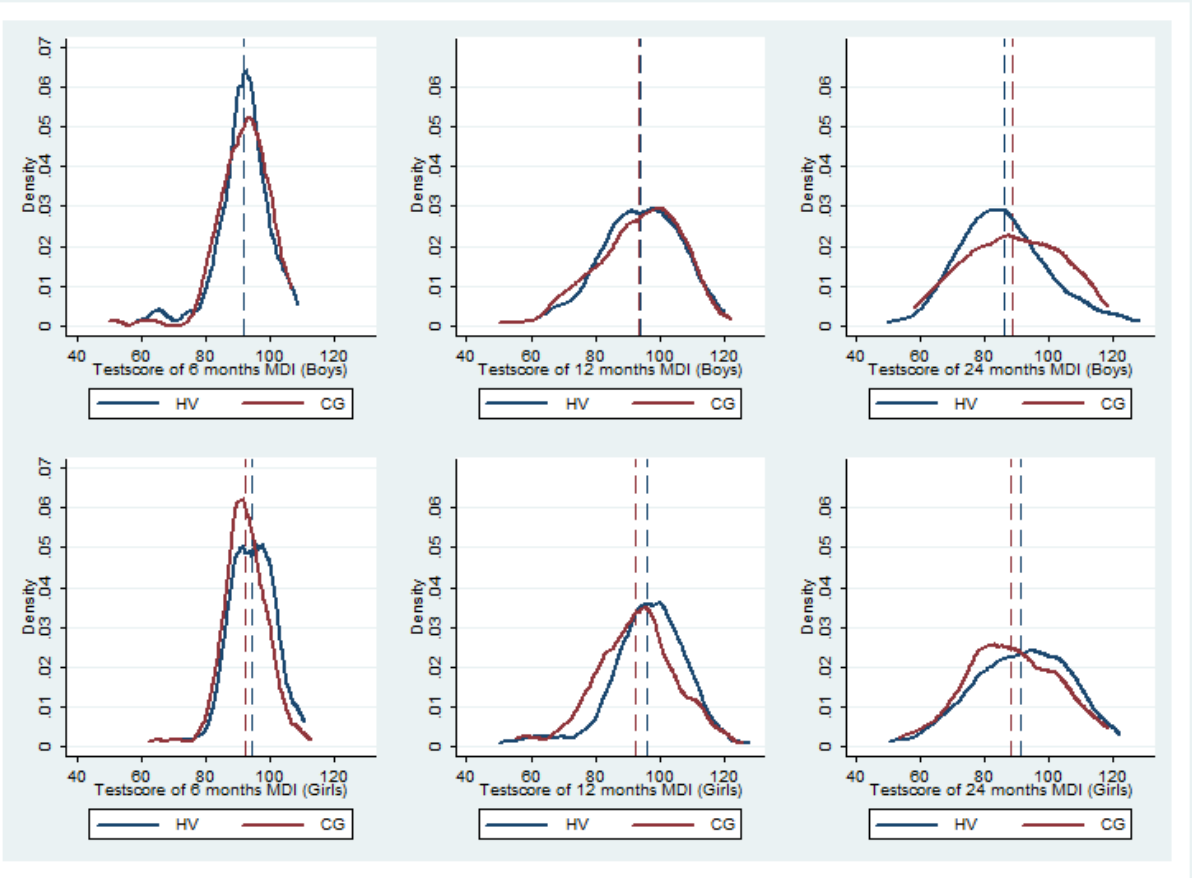


Figure 3.7: PDI Test Scores for Treatment and Control Group (Boys and Girls)

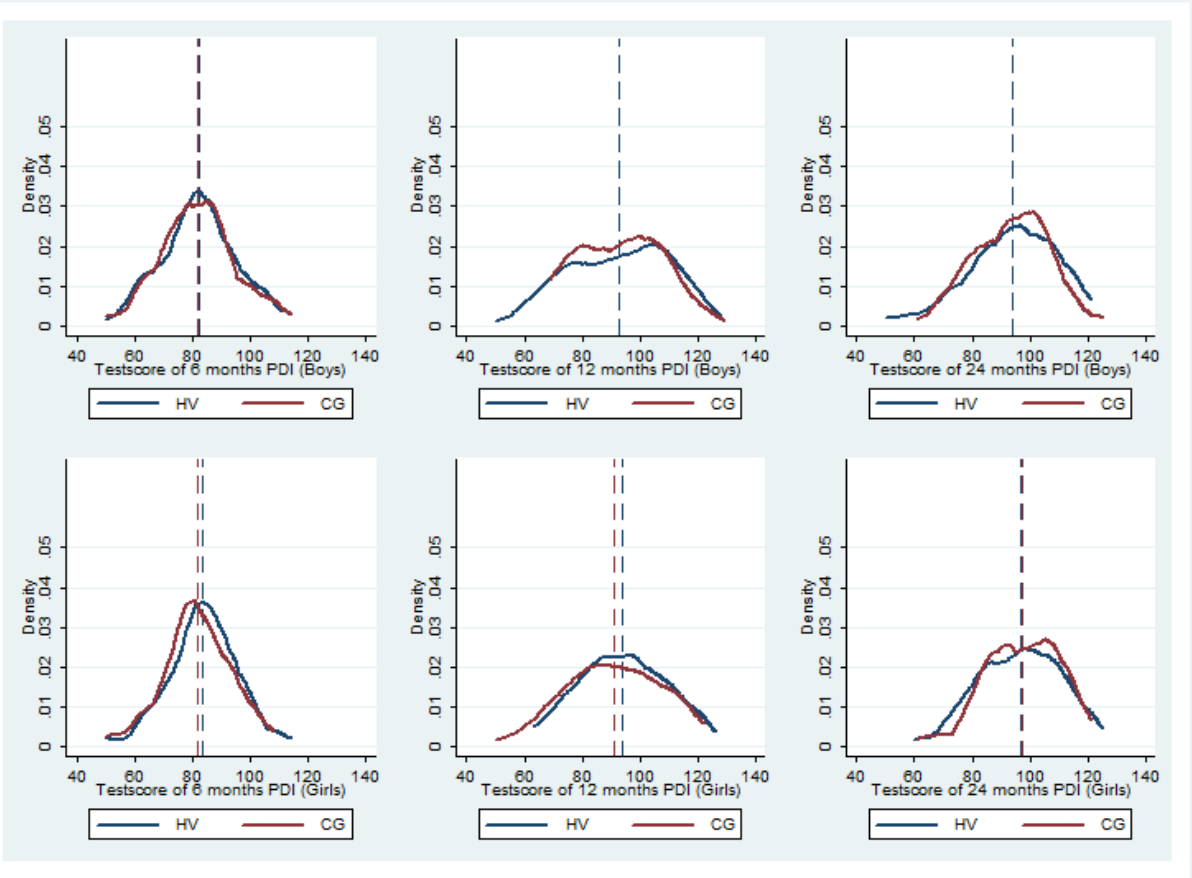


Table 3.15: Investments in Children - Descriptives

	Full Sample		**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 is obtained in the personal interviews. All dependent variables are binary. The figures give the rate of mothers who look at picture books with her child or reads or tells stories to her child daily, several times per week or at least once a week.

Table 3.16: 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

4 The Effects of *Pro Kind* on the Maternal Life Course [1]

4.1 Introduction

In recent years the outcomes of early childhood intervention programs have gained much attention in economic literature. Evidence from randomized experiments suggests that these programs improve cognitive and socioemotional abilities, as well as, the health of disadvantaged children (see Almond and Currie, 2011; Karoly et al., 2005, for a review of the literature). Because of the dynamic process of skill formation, these early investments in children can reduce later inequality and can cause high cost-benefit ratios in the long run (Cunha and Heckman, 2007; Heckman and Masterov, 2007; Belfield, 2006).

Despite these promising results for children, so far there has been little economic research on the impact of early childhood interventions on certain dimensions of the maternal life course, such as maternal employment, education, fertility, childcare use and maternal well-being. This neglect is surprising as many interventions mainly focus on the mother. Home visiting programs are probably the type of early childhood intervention with the strongest maternal focus. In these programs nurses directly address

[1]Part of this research was supported by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth *(Bundesministerium für Familien, Senioren, Frauen und Jugend - BMFSFJ)*. Earlier versions of this paper were presented at 10th Journees Louis-Andre Gerard-Varet 2012, the 25th Annual Meeting of the Associazione Italiana Economist del Lavoro 2012 and the annual Meeting of the German Society for Economic Experimental Research 2012.

disadvantaged mothers under the belief that parents mediate changes for their children. Accordingly, home visiting tries to enhance parental skills such as attachment behavior, interaction and teaching skills. Furthermore, many home visiting programs directly try to increase women's personal strengths like self-efficacy, problem-solving abilities, self-esteem or emotional functioning and the ability to tie and maintain social networks.

It is likely that these improved parental skills and personal strengths could have effects on the maternal life course. However, the direction and the size of the effects are unclear. On the one hand, the intervention could lead to higher labor market or education participation because of the improved parental skills and personal strengths. On the other hand, the intervention could increase satisfaction with the maternal role, partner stability or subjective well-being. This could lead to higher satisfaction with the maternal role and therefore to higher fertility and as a consequence to longer work force abstinence.

Until now only two home visiting programs have been systematically analyzed to reveal which of the two effects predominates. Firstly, the Nurse Family Partnership Program (NFP) presents evidence from three randomized trials that home visiting decreased maternal fertility and welfare dependency and increased maternal labor market participation (see Olds et al., 2007, 1997, for a review of the literature). Secondly, the Infant Health and Development Program also positively affected maternal employment (Brooks-Gunn et al., 1994). However, both programs are located in the US. The effects of home visiting on maternal life course could be different in a European welfare syste where incentives and support for disadvantaged young mothers systematically differ in comparison to the US.

This paper presents the first randomized experiment to evaluate the impact of one such home visiting program, the *Pro Kind* Project, on the maternal life course in an European context. The intervention starts during pregnancy and continues until the second birthday of the child. Exclusively mothers on social welfare or with low income are enrolled. Additional affiliation criteria include: being under age, poor education,

substance abuse and experiences of violence or neglect. Biannual telephone interviews with the participants give detailed information about the life outcomes in the first three years after birth. The obtained data contains information about objective outcomes like employment, fertility and child care use as well as subjective statements about well-being and life-satisfaction.

I find that the *Pro Kind* Project has significantly increased the hazard of a second birth in the intervention group. The birth rate increased, although the share of second pregnancies is equal in treatment and control group. The smaller share of pregnancies which lead to a life birth in the control group is mainly caused by more abortions in this group. There is evidence that mothers in the treatment group more often welcomed a further pregnancy and that this reduced the abortion rate. Additionally, the intervention positively influences the maternal subjective well-being and life-satisfaction which might also influence fertility decisions. There are no statistically significant effects on maternal employment, school attending, childcare use or partnership duration. However, tendency is strong that the intervention group uses institutional childcare more frequently and earlier.

The results in this study substantially differ from the results of previous studies. One interpretation for the differences might be the arrangement of the German welfare state. This welfare state is characterized by generous social assistance rules which guarantee a fixed welfare amount per child and unconstrained social assistance until the third birthday of a child. In this social environment, where mothers with small children have no work obligations and income increases with a further birth, it is likely that the interventions impact on maternal skills and life-satisfaction might lead to subsequent birth. In contrast, incentives for a further child are small in the US because of stricter budget constraints induced by the welfare regulations, especially since the mid 90s.[2] Therefore, an increased

[2]In 1996 the Temporary Assistance to Needy Families (*TANF*) eliminated the legal entitlement to cash welfare by imposing a 60-month lifetime time limit on benefit receipt and requiring individuals to leave welfare for work after 2 years. Furthermore, three of the four stated goals of *TANF* involved reducing non-marital births and encouraging marriage (Blank, 2002). Nevertheless, also the program

well-being might lead to higher work force participation in the US.

The remainder of the paper is organized as follows. The next section gives some background on home visiting programs and their potential for altering maternal life course. Section 4.3 reviews the existing literature about the effects of home visiting on maternal life course. Section 4.4 describes the data and the estimation method to identify the causal effects of the intervention. Section 4.6 presents results while section 4.7 provides concluding remarks.

4.2 Home Visiting and Maternal Life Course

As this paper concentrates on maternal life course, one could question why home visiting in general, and *Pro Kind* in particular, should produce effects in this area. This is especially crucial since the German welfare state is quite generous to mothers of infants and toddlers. For example, there are no work obligations or welfare cuts as long as a mother has no child care arrangement. Therefore, there are little incentives for maternal labor market participation. Furthermore, additionally to *Pro Kind* various services offer help and support especially for these mothers, e.g. the labor agency offers special programs for unemployed who are younger than 25 years and for single mothers.

The main answer why home visiting can still have an effect on maternal life course is the relationship which the home visitor develops with the mothers during their pregnancies and their children's early years. The strongest base for inducing and deepening this relationship is the first time experience of a new born child. Olds et al. (2010) state that through this relationship nurses could help parents gradually gain a sense of mastery in overcoming challenges and position themselves to create the kind of life they want. Furthermore, mothers with newborns are often open-minded about guidance during this

which proceeded *TANF*, Aid to Families with Dependent Children (*AFDC*), can be seen as more strict than the welfare regime in Germany today. Only single mothers were eligible to rather low cash benefits (benefits for a single-parent family with two children and no income ranged from $120 in Mississippi to $597 in Vermont). Additionally AFDC strongly used in-kind transfers like food stamps and strong work obligations (Moffitt, 1998; Gebhardt and Jacobs, 1997).

fundamental life transition, as they make important choices that shape the subsequent trajectories of their life and those of their children. Thus, building relationships and meeting open minded clients are the strongest advantages of home visiting compared to other programs and can lead to changes in maternal life.

Another question is the fiscal relevance of possible changes in maternal life course in Western European welfare states. In Germany, 11.1 percent of all transfer households include children aged below three, and 21.5 percent of all households with children aged below three receive social benefit transfers. In addition 75 percent of single parents are eligible to welfare payments at least one month in the first three years after birth. Overall, households with children under age three receive around 4.7 billion Euro of social benefits per annum (Bundesagentur für Arbeit, 2009; Statistisches Bundesamt, 2009). These figures express that changing welfare dependency of households with children would have strong short time fiscal impact. The fiscal relevance of the maternal life course is further underlined by two studies of the Nurse Family Partnership program, where the programs achieved a positive cost-benefit ratio just by changes in maternal life course (Olds et al., 1993, 2010).

The importance of maternal life course in the *Pro Kind* program is illustrated by the time which the home visitors spent for this topic. Table 2.6 illustrates that at all developmental stages the home visitors invests 40 percent of their time in the family for domains related to maternal life course. Additionally, *Pro Kind* spent more time on these domains than the NFP average and the recommended average of the NFP. These figures show that life course is a fundamental part in the implementation of the *Pro Kind* Project in which effects are likely and which therefore requires investigation.

4.3 Previous Literature

A large body of literature, mainly medical scientists and psychologists, discusses the various outcomes of home visiting programs. In order to present an overview how this

literature considers the maternal life course and which effects the literature finds, this section initially explains meta-analyses concerning home visiting programs. Afterwards, I take a closer look at the effects of the Nurse-Family-Partnership (NFP) program, which is the prototype of the *Pro Kind* intervention and the intervention where effects on the maternal life course are most systematically investigated.

Gomby (2005) reviews twelve meta-analyses regarding home visiting. These meta-analyses are not limited to studies which use an experimental or quasi-experimental design. In fact, they include studies with ex-ante/ex-post comparisons, small sample sizes and short follow-up periods. Furthermore, the interventions in many studies offer characteristics of low-quality home visiting, e.g. they start after pregnancy and they are not structured or focused either on disadvantaged or first time mothers. Nevertheless, it is interesting that only four of these twelve meta-analyses investigate maternal life course, showing the low attention for this potential outcome. Looking at the results reveals that three of these four (Geeraert et al., 2004; Layzer et al., 2001; Sweet and Appelbaum, 2004) find significant positive effects at least in one domain of maternal life course. In one study (Elkan et al., 2000) the results concerning maternal life course are inconclusive. The studies, which do not include any outcome of maternal life course, focus mainly on children's cognitive development or abuse and neglect.

The meta-analysis by Sweet and Appelbaum (2004) is the one with the highest methodological rigorousness. They include exclusively 60 home visiting programs conducted in the United States since 1965. All included studies are experiments or quasi-experiments. However, also home visiting programs with low-quality characteristics are included. Education, employment, and reliance on public assistance define maternal life course outcomes, whereas fertility is not included as an outcome. As a result home visiting has a significant impact on maternal education with a weighted mean standardized effect size of .134 and with the largest effect on teenage mothers. Sweet and Appelbaum (2004) find no significant effect on employment or reliance on public assistance. Only

five of the 60 studies included in the meta-analysis present results on maternal education, seven on employment and three on public assistance. In contrast, 41 studies investigate child cognitive development and 37 studies search for effects on parenting child rearing.

In summary, most studies, and meta-analyses, of home visiting focus dominantly on the outcomes for children. Studies located in the United States dominate the literature. Although the existing meta-analyses document overall an impact on maternal life course, the impact seems to be small in size. One reason for small or missing effects could be that life course is given too little attention and therefore relevant outcomes are not measured. This hypothesis is supported by the fact that studies are mainly conducted by medical scientists and psychologists for whom maternal life course may not lay in their core interest. Meta-analyses and reviews on early childhood interventions by economists (Nores and Barnett, 2010; Baker-Henningham and Lopez Boo, 2010; Barnett, 2006) are just at the beginning and they neither focus on home visiting nor on maternal life course.

Gomby (2005) concludes in her review: *"Evidence is insufficient for effects of home visiting on maternal life course. The exception is the Nurse-Family-Partnership Program, which has the largest effects when included in the presented meta-analysis."* As the *Pro Kind* Project is conceptually similar to NFP, this subsection presents a closer look at NFP and its results. NFP is a program of prenatal and infancy home visiting for low income, first-time mothers and their families. The nurses start visiting families as early as possible during pregnancy and continue the visits until the child's second birthday. NFP is evaluated in three different trials by randomized experiments. The first evaluation was conducted in Elmira, New York, starting 1980 with mainly white first-time mothers participating. The next evaluation started in Memphis, Tennessee, in 1990 enrolling mainly black, low income, single, first-time mothers. In 1995 the third evaluation was initiated in Denver, Colorado. Participants were mainly Hispanic low income, single, first-time mothers. In all three trials maternal life course was always of core interest beside child outcomes. Depending on the start of the trial, outcomes for different endpoints are

Table 4.1: NFP Results Elmira (Olds et al., 1988, 1997).

Outcome	Observation Period		
	6 Months	**4 Years**	**15 Years**
School:	More School Enrollment of School Dropouts		
Employ.:		More Employment (15.54 Months vs. 8.64 Months)	By trend more Employment (95 months vs. 80 Months)
Fertility:		Fewer Subsequent Pregnancies (0.58 vs. 1.02)	Fewer Subsequent Births (1.3 vs. 1.6) Longer Interval Between First and Subsequent Birth (65 Months vs. 37 Months)
Transfer:			Less Months Eligible to Transfer (60 Months vs. 90 Months)

Notes: If not indicated differently, all treatment effects are significant at a five percent level. Employ. = Employment

available. Follow-up data is available between four years in Memphis and 15 years in Elmira. Table 4.1 to 4.3 presents results concerning maternal life course for the three trials.

Overall, the literature shows that NFP reduces the rates of subsequent pregnancies and births and increases the intervals between first and second pregnancies and births in all three trials within the first four years. In two trials an increase in maternal employment is found. Women's use of welfare is reduced in all three trials. Mainly more stable partnerships and reduced subsequent births explain these effects. Long-term follow-ups reveal that effects on maternal life course do not fade-out over the years. The intervention does not affect school graduation in any trial, although higher school attendance is recognized in Elmira.

In the Elmira and the Memphis trial NFP reaches the fiscal break-even point through the presented changes in maternal life course. In Elmira the program cost of $3.133 face discounted savings of $3.246 expressed in 1980 US-Dollars by child age four. Higher

Table 4.2: NFP Results Memphis (Kitzman et al., 1997; Olds et al., 2004, 2007, 2010).

Outcome	Observation Period			
	2 Years	**6 Years**	**9 Years**	**12 Years**
Employ.:		By trend more Employment ($p<0.1$)	By trend more Employment ($p<0.1$)	By trend more Employment ($p<0.1$)
Fertility:	Fewer Subsequent Pregnancies (0.36 vs. 0.47)	Fewer Subsequent Pregnancies (1.16 vs. 1.38)	Fewer Cumulative Subsequent Births per Year (0.81 vs. 0.93)	
Transfer:		Less Months Eligible to Transfer per Year (7.21 Months vs. 8.96 Months)	Less Months Eligible to Transfer per Year (5.21 Months vs. 5.92 Months)	

Notes see Table 4.1

Table 4.3: NFP Results Denver (Olds, 2002; Olds et al., 2004).

Outcome	Observation Period	
	2 Years	**4 Years**
Employ.:	More Employment (6.83 Months vs. 5.65 Months)	More Employment (15.13 Months vs. 13.38 Months)
Fertility:	Fewer Subsequent Births (0.12 vs. 0.19)	Longer Interval Between First and Subsequent Birth (24.51 Months vs. 20.39 Months)

Notes see Table 4.1

maternal employment, shown in Table 4.1, is the main reason for savings. In Memphis the NFP causes $12.300 in discounted savings compared with a program cost of $11.511, both expressed in 2006 US-Dollars by child age twelve. Higher maternal employment and less government spending on food stamps, Medicaid, and AFDC and TANF, presented in Table 4.2, generate the savings.

Additionally, a recent study examines the effects of NFP on time to second pregnancy within two years of the first infant's birth with a quasi-experimental research design (Rubin et al., 2010). The investigation takes place after statewide NFP implementation in the Commonwealth of Pennsylvania. Therefore, it is possible to examine whether the effects of earlier NFP trials sustain after dissemination in state level. Rubin et al. (2010) find that the program effects on pregnancy planning emerge after an implementation period of three years and the effects are particularly strong among younger mothers.

4.4 Data and Identification Strategy

4.4.1 Sample Attrition

The biannual telephone interviews are used as the main data source to examine the impact of *Pro Kind* on maternal life course. In some analysis I also use data from the personal interviews which were conducted jointly with the development tests. The data collecting process and the questionnaires are explained in detail in chapter 2.3.3. As expressed in Table 2.8 less than 50 percent of the randomized mothers participated in all telephone interviews until the third birthday of the child. These attrition rates do not bias the randomization outcome if they are not selective regarding the baseline characteristics between treatment and control group. I use the basic model from equation 2.1 to prove if there is any selective attrition between the two groups. Table 4.4 presents the differences in the baseline demographic characteristics between treatment and control group for each interview. Table 4.14 in Appendix C shows the differences of the psychological characteristics

and Table 4.15 in Appendix C presents the differences in the baseline characteristics only for these mothers who participated in all interviews until the second and third birthday.

The results reveal that the equal distribution of the baseline characteristics is only slightly reduced by the attrition. Only the difference in the share of mothers with migrational background which is already existent at baseline stays significant in almost all interviews. The psychological characteristics show some smaller differences. Similarly to the personal interviews mothers with risk of aggression attrite more often in the control group. However, mothers with foster care experience participate more often in the treatment group. For the mothers who participated in all interviews until the second or third birthday the picture is similar indicating that mothers with similar baseline characteristic attrite in treatment and control group.

However, there could be selective attrition compared to the baseline between attritors and participants. Again, I use the model in equation 2.1 to estimate the difference between attritors and non-attritors. As seen in Appendix C Table 4.16 and 4.17 some characteristics and risk factors are different between the attriting population and the participants in the follow- up interviews. Mainly the participating mothers are older and have less cumulative risk factors.[3] The psychological characteristics are less correlated with attrition but almost all coefficients have a positive sign which indicates higher risk rates for the attritors. However, as in the personal interviews the mothers who participate still belong to a highly disadvantaged sub-population.

4.5 Estimation Methods

I use objective and subjective outcomes to measure the impact of *Pro Kind* on maternal life course. The objective data contains labor market participation, fertility, childcare use and partnership stability. Most of the objective data is measured in time durations. The

[3]Most mothers who attrite are between 19 and 21 years old.

Table 4.4: Selective Attrition between TG and CG Demographic Characteristics - Telephone Interviews

	Difference TG/CG						
	Pregnancy	3 months	9 months	15 months	21 months	27 months	36 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demographic Characteristics							
Age in Years	-0.164	0.004	0.107	-0.043	0.114	-0.184	0.236
	(0.332)	(0.366)	(0.411)	(0.433)	(0.447)	(0.434)	(0.481)
Week in Pregnancy	-0.491	-0.579	-0.151	-0.235	-0.362	-0.318	-0.017
	(0.441)	(0.472	(0.528)	(0.553)	(0.557)	(0.552)	(0.596)
Migration	-0.063**	-0.068**	-0.045	-0.059*	-0.065*	-0.064*	-0.059
	(0.027)	(0.029	(0.032)	(0.035)	(0.035)	(0.036)	(0.040)
Underage	0.036	0.021	0.016	0.038	0.032	0.044	0.033
	(0.029)	(0.032)	(0.033)	(0.035)	(0.035)	(0.034)	(0.037)
Mon. HH-Inc. in €	11.01	18.73	-18.35	-17.39	-0.52	-11.44	-22.77
	(43.38)	(47.24)	(50.34)	(53.45)	(52.69)	(51.69)	(57.09)
Debt over 3000 €	0.032	0.031	0.033	0.039	0.043	0.017	0.041
	(0.030)	(0.033)	(0.037)	(0.038)	(0.039)	(0.038)	(0.042)
Education Risk	0.036	0.021	0.021	0.035	0.020	0.043	0.020
	(0.033)	(0.036)	(0.041)	(0.043)	(0.044)	(0.043)	(0.048)
Income Risk	0.016	0.012	0.025	0.031	0.020	0.022	0.033
	(0.030)	(0.033)	(0.036)	(0.039)	(0.040)	(0.040)	(0.044)
Employment Risk	-0.036	-0.038	-0.046	-0.037	-0.055	-0.051	-0.051
	(0.029)	(0.032)	(0.035)	(0.038)	(0.039)	(0.037	(0.041)
No Partner	0.011	0.024	0.046	0.033	0.020	0.033	0.032
	(0.035)	(0.037)	(0.041)	(0.043)	(0.045)	(0.044	(0.049)
Living with Parents	0.013	-0.001	-0.024	-0.020	-0.016	-0.016	-0.048
	(0.034)	(0.037)	(0.039)	(0.041)	(0.043)	(0.043	(0.047)
Persons in HH	0.100	0.116	0.034	0.022	-0.018	-0.054	-0.072
	(0.124)	(0.136)	(0.144)	(0.154)	(0.148)	(0.143)	(0.157)
Observations	680	582	486	442	418	438	374

Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Table 2.9 and 2.10 for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

months after birth of the treatment child is the analysis time.[4] Therefore, an analysis time of 36 months is available for mothers who participated in all telephone interviews. The subjective data includes maternal well-being and life-satisfaction, obtained at the 27 months telephone interview. I examine intent-to-treat (ITT) effects because in the balanced panel compliance to the randomization outcome is almost complete, see sub-

[4]Throughout this chapter, the treatment child indicates the first child of the mother who was in focus of the intervention.

section 2.3.3.2. Because data is only available for a very small number of non-complying mothers (those who did not receive any home visit) the results do hardly change if the randomitzation outcome is used as an instrument for the treatment received.

To give a descriptive overview of the outcomes, I start my analysis with comparing the means of the objective outcomes. Afterwards, I estimate equation 4.1:

$$Y_{ic} = \beta_0 + \beta_1 HV_{ic} + \beta_2 h_{ic} + \alpha_c + \epsilon_{ic}, \tag{4.1}$$

where Y_{ic} denotes an outcome variable (employment, fertility, childcare use and partnership stability) for mother i from community c. HV_{ic} is a dummy variable that takes value one if the mother receives the home visits. h_{ic} is a vector of demographic and psychological family characteristics at baseline; α_c are community dummies; and ϵ_{ic} is the error term. β_1 measures the difference between treatment and control group in outcome Y. In the descriptive overview and in the estimation only mothers who participated in all interviews are included to avoid biases due to right censoring of the mothers who attrite. All outcomes are binary coded and take the value one if the outcome, e.g. employment or second pregnancy, occurs. Therefore, I estimate linear probability models and report the marginal effect of HV_{ic} on outcome Y_{ic}.

Next, I examine the mothers' probability of "surviving" beyond a certain point of time t, where t is measured in months since birth of the intervention child. In context of maternal life course "surviving" means not bearing a second child, not starting work or school after birth or not using institutional childcare. To capture the nature of the duration data, I apply statistical methods within a hazard rate framework (Kleinbaum and Klein, 2005; Cleves, 2010). Such techniques have the huge advantages of accounting for censoring and of taking into account the precise duration until the event, which causes "failing", occurs. Accounting for right-censoring is necessary, because some participants may not experience the event of "failing" within the observation period of 36 months and for some participants the complete data is not available for the whole observation period.

Surviving is reported by the survivor function $S(t)$ (Equation 4.2), with $0 \leq S(t) \leq 1$ and with T as a non-negative random variable that denotes the time of the event.

As $S(t)$ is estimated by the nonparametric Kaplan-Meier estimator (Equation 4.3), where n_j is the number of participants at risk at time t_j and d_j the number of events at t_j there is no matter at which point of time censoring occurs.

$$S(t) = 1 - F(t) = P(T > t) \tag{4.2}$$

$$\widehat{S}(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \tag{4.3}$$

Therefore, I can test the equality of survivor functions in intervention and control group. If the test of equality is rejected, an impact of the intervention can be assumed.

In a next step I use Cox proportional hazards regression models for covariate analysis to improve the precision of the coefficients. The Cox regression asserts that the hazard rate for the jth subject in the data is

$$h(t|x_j) = h_0(t)exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \tag{4.4}$$

where the regression coefficients, β_x are to be estimated from the data. The hazard rate $h(t)$ can be calculated by the hazard function

$$h(t) = lim_{\Delta t \to 0} \frac{P(t + \Delta t > T > t | T > t)}{\Delta t} \tag{4.5}$$

The baseline hazard $h_0(t)$ in Equation 4.4 is given no particular parametrization and, in fact, can be left unestimated.

In order to guarantee the duration structure of the data I use the following procedure: data for childcare use is collected in each interview beginning at the 9 months interview. Participants are asked whether their child attends institutional childcare and if so, they are

asked for the starting month. If the participant states that her child attends institutional childcare but does not know the exact starting point, the date of the interview is used as starting point. For subsequent pregnancy and birth I follow the same procedure with the only difference that mothers are asked for fertility at the 15 months interview the first time. Employment and school attendance are surveyed in each interview on the base of a monthly activity calendar. I only consider the first status switch after birth as a fail, therefore, it is not recognized how long the participants stay in this status. For example, in the case of school enrollment, analysis considers if a participant starts school but not if she continues school a month later.

Finally, the subjective ratings of the mothers are analyzed. For this I also compare the ratings in control and treatment group with equation 4.1. Additionally, I estimated ordered probit models for those dependent variables which are measured on a scale from 0 to 10. However, the results do not bring you new insights in comparison to the OLS estimates. Furthermore, I compare the self-ratings of well-being and satisfaction with ratings from GSOEP first time mothers.

4.6 Results

4.6.1 Objective Outcomes

4.6.1.1 Comparison of Means and Multivariate Analysis

Table 4.5 reports the percentage of mothers who had a second pregnancy, a second birth or used child care until the second or third birthday of the treatment child. The variable partner in household indicates whether the mother lives with her partner in the same household at the child's second or third birthday. The table only includes mothers who either participated in all interviews until the end of the intervention, at the second birthday of the treatment child or in all interviews until the third birthday.

24 months after birth of the treatment child, the rates of second pregnancies and

Table 4.5: Fertility, Child Care Use and Partnership in Treatment and Control Group

	Only Mothers with complete 24 Months Data		
	Control	Treatment	P-value
24 Months after first Birth (Means)			
Second Pregnancy	0.23	0.27	0.358
Second Birth	0.07	0.12	0.117
Child Care Use	0.35	0.41	0.316
Partner in HH	0.50	0.48	0.773
Observations	161	187	
	Only Mothers with complete 36 Months Data		
	Control	Treatment	P-value
24 Months after first Birth (Means)			
Second Pregnancy	0.23	0.26	0.451
Second Birth	0.08	0.13	0.202
Child Care Use	0.32	0.40	0.180
Partner in HH	0.50	0.47	0.597
36 Months after first Birth (Means)			
Second Pregnancy	0.32	0.35	0.574
Second Birth	0.18	0.28	0.038**
Child Care Use	0.58	0.67	0.142
Partner in HH	0.48	0.46	0.735
Observations	137	159	

Notes: P-values base on z-statistic of a two-group test of proportions. Observations in *Only Mothers with 24 Months Data* are based on mothers who participated in all interviews until the **second** birthday of their child. Observations in *Only Mothers with 36 Months Data* are based on mothers who participated in all interviews until the **third** birthday of their child. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

births are 4-5 percentage points higher in the treatment group than in the control group. Child care use is slightly higher in the treatment group. In both groups around 50 percent of the mothers live with their partner in the same household which is equal to the rate at baseline.

At 36 months the group difference in the cumulated share of pregnancies is smaller than at 24 months. However, in the treatment group ten percentage points more mothers state that they gave birth to a second child. Comparing the pregnancies at 24 months and the births at 36 months reveals that almost all pregnancies in the treatment group lead to a life birth while this is not the case in the control group (I discuss this fact in

Table 4.6: Maternal Work Participation in Treatment and Control Group

	(1)	(2)	(3)	(4)	(5)	(6)
	Only Mothers with complete 24 Months Data					
	Control Group		Treatment Group			
	at least one Month	Average Months	at least one Month	Average Months	P-Value Diff.	P-Value Diff.
24 Months after Birth (Means)						
Apprenticeship	0.16	1.14	0.18	1.46	0.597	0.448
Minijob	0.21	1.56	0.14	0.94	0.076*	0.093*
Part/Fulltime employed	0.20	1.17	0.16	0.83	0.225	0.241
Any occupation	0.45	3.87	0.38	3.23	0.202	0.290
School	0.10	0.91	0.09	0.91	0.788	0.997
Welfare	0.90	19.71	0.96	20.28	0.020**	0.997
Observations	161		187			
	Only Mothers with complete 36 Months Data					
	Control Group		Treatment Group			
	at least one Month	Average Months	at least one Month	Average Months	P-Value Diff.	P-Value Diff.
24 Months after Birth (Means)						
Apprenticeship	0.15	1.21	0.16	1.35	0.810	0.775
Minijob	0.20	1.55	0.15	1.04	0.228	0.217
Part/Fulltime employed	0.21	1.11	0.17	0.92	0.359	0.563
Any occupation	0.43	3.87	0.39	3.31	0.477	0.404
School	0.09	0.90	0.08	0.83	0.691	0.872
Welfare	0.90	19.57	0.96	20.27	0.052*	0.406
36 Months after Birth (Means)						
Apprenticeship	0.26	2.67	0.22	3.11	0.475	0.559
Minijob	0.25	2.53	0.23	1.86	0.756	0.265
Part/Fulltime employed	0.30	2.36	0.29	1.84	0.851	0.331
Any occupation	0.55	7.57	0.55	6.82	0.896	0.481
School	0.10	0.93	0.09	1.01	0.679	0.880
Welfare	0.91	26.51	0.96	27.78	0.127	0.298
Observations	137		159			

Notes: P-values base on results from two-sample mean-comparison tests and from two-group test of proportions. P-values in column 5 base on comparison between the means in column 1 and 3. P-values in column 6 base on comparison between the means in column 1 and 3. Minijob is an employment with less than 400 Euros wage per month. In *24 Months* all mothers are included who participated in the interviews until the **second** birthday of their first child. In *36 Months* all mothers are included who participated in the interviews until the **third** birthday of their first child. Welfare includes households which receive *Arbeitslosengeld II (ALGII)*, *Hilfe zum Lebenunterhalt (HLU)* or *Sozialhilfe.*
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

more detail in chapter 4.6.1.3). In the treatment group two thirds of the children are in institutional childcare at 36 months. This rate is nine percentage points smaller in the control group. The share of mothers who live with their partner in the same household decreases slightly in both groups.

The higher birth rates at 36 months could be caused by selective attrition of mothers who are pregnant at 24 months because these mothers might be more time constraint and therefore have less time for the telephone interviews. However, 24 months after birth of the treatment child the outcome rates look similar between the mothers with complete 24 months data and complete 36 months data. If disproportionate or selective attrition would be the case, 24 months after birth the rate of mothers with a second pregnancy or birth would have been higher in the sample with the complete 24 months data compared to the sample with the complete 36 months data. Therefore, the similar results 24 months after birth confirm that women who are pregnant or gave birth do not leave the sample disproportionate or selective in treatment or control group.

Next, I examine the effects of *Pro Kind* on occupation and public assistance. The first rows in column 1 and 3 of Table 4.6 present the percentage of mothers who worked at least one months in apprenticeship, *Mini Job* or part/full time employment. The variable "any occupation" is one if the mother worked at least one month in one of the three kinds of occupation independently whether she worked in more than one kind of occupation. The last two rows show if a mother went to school or lived in a household which received public assistance. Column 2 and 4 present the average sum of months a mother spent in a certain occupation, attended school or received public assistance. A mother can have only one kind of occupation in each month. However, she can be employed but also receive public assistance. Just as Table 4.5, Table 4.6 only includes mothers who participated in all interviews until the second birthday and until the third birthday, respectively.

24 months after birth the average sum of months in occupation or at school is very low. This is not surprising since the mothers have an infant at home. In contrast, the

percentage of mothers who started any occupation (45 percent) is high compared to the average sum of months in occupation. This indicates a high job fluctuation and short employment periods. The differences between the treatment and control group are small. Nevertheless, in the control group more mothers worked in a *Mini-Job*.

Ten percent of the control group mothers live in households who have not received any public assistance since birth of the treatment child versus four percent in the treatment group. This is surprising since receiving public assistance is an affiliation criteria and there were no differences between this characteristic at baseline. However, the average months with public assistance did not differ. In both groups the household received welfare in average almost 20 months. Therefore, it is likely that in the treatment group some households receive public assistance just for one or two months and that similar household are completely without public assistance in the control group. Again, no difference occurs for occupation or for public assistance between women with complete 36 and 24 months data.

36 months after birth the share of mothers who are employed or do an apprenticeship increases. The average sum of months almost doubles indicating that from the second to the third birthday the mothers participate more in the workforce than in the first two years. The sum of months in apprenticeship increases most strongly and is the most frequent occupation after 36 months. This is not surprising considering that the *Pro Kind* sample consists of many young women without a complete apprenticeship. The differences between the two groups are smaller than 24 months after birth. This is surprising since more second births which might reduce employment occur in the treatment group. Hence, analyzing only mothers without a second child reveals a higher employment rate in the treatment group. However, the difference is not significant. The average sum of months on public welfare increases. Nevertheless, the months without public welfare increase stronger. This indicates a greater economic self-sufficiency when the child gets older. The group difference between the households who received any public assistance vanishes.

Table 4.7: Maternal Life Course Multivariate Analysis

	(1) Second Pregnancy	(2) Second Birth	(3) Child Care Use	(4) School	(5) Any Employment
			ITT		
Home visiting	0.017	0.104***	0.064	0.019	-0.042
	(0.046)	(0.025)	(0.054)	(0.029)	(0.046)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	296	296	296	296	296
R^2	0.12	0.14	0.15	0.20	0.05

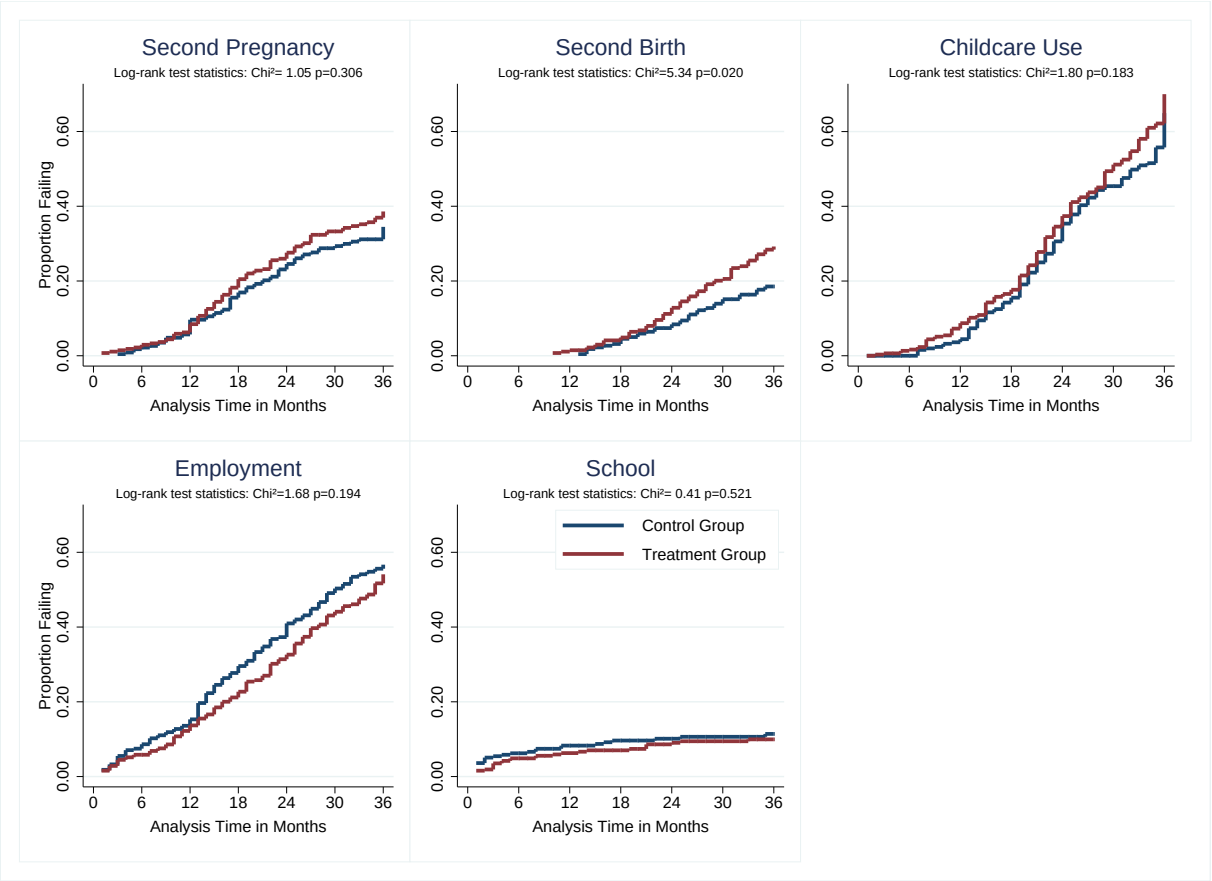
Notes: Standard errors (in parentheses). All outcome variables are binary. Therefore, all estimates are linear probability models and coefficients report the influence of home visiting on the outcome in percentage points. The models include all mothers who participated in the interviews until the **third** birthday of their first child. All models use extended baseline control variables ans community fixed effects. ITT=Intent-to-Treat;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the descriptive statistics show that *Pro Kind* affected maternal fertility, whereas other life course outcomes are not significantly affected. Next, I examine whether these results remain in a multivariate model which includes baseline controls. Table 4.7 presents the results 36 months after birth for fertility, child care, school attendance and any employment. All dependent variables are binary and take the value one if the mother was in the status until the third birthday of her first child. The multivariate analysis confirms the descriptive results. In the treatment group the rate of second births is 10.4 percentage points higher than in the control group which means that the second births are increased by more than 50 percent in the treatment group. There are no significant effects on the other outcomes.

4.6.1.2 Duration Analysis

This section considers the duration characteristics of the data. Duration analyses are robust when censoring occurs at some time other than an observed failure time. Therefore, all observations can be included, independent whether data is available only from one or from more interviews. Figure 4.1 provides the failure graphs of the investigated outcomes

Figure 4.1: Kaplan Meier Survival Functions - All Participants



for the treatment and the control group. In line with the previous results the graphs for second birth significantly diverge between treatment and control group.

The duration analysis gives the possibility to investigate the timing of the second birth. This is important because rapid repeat birth (RRB) which is defined as a birth occurring within 24 months after a previous birth has been identified as a risk factor for adverse perinatal outcomes (Klerman et al., 1998; Zhu et al., 1999). Appendix H shows that the hazards for a second birth diverge mainly after 18 months and the divergence pikes around 30 months. Therefore, RRB occurs more often in the treatment group. However, because there are only small differences in the hazard rate before 18 months it seems that RRB is only a slightly greater problem in the treatment group compared to the control group.

Looking at the other outcomes reveals that child care use and employment increase well correlated. However, some mothers use childcare without being in any employment. As in the analyses above, employment does not drop behind in the treatment group although more second births occur in this group. Most mothers who start to attend school do this close after the birth of the treatment child indicating that mostly mothers who went to school before the birth of the treatment child continue their school education. Appendix I presents the failure graphs which only include the data of mothers who participated in all interviews until the third birthday. The results hardly change compared to Figure 4.1.

Table 4.8 presents the results of the multivariate Cox regression including baseline characteristics as controls. Being in the treatment group increases the hazard of having a second child by 68 percent. The effect is significant at a 1 percent level. In contrast the hazard of having a second pregnancy is not increased by the treatment. The results hardly change in an estimation without covariates (not shown in a table). Overall, these multivariate duration analyses confirm the results of the previous sections.

Table 4.8: Maternal Life Course Cox Regression Analysis

Change to	(1) Second Pregnancy	(2) Second Birth	(3) Child Care Use	(4) School	(5) Any Employment
			ITT		
Home visiting	1.190	1.684***	1.183	0.945	0.858
	(0.151)	(0.222)	(0.134)	(0.259)	(0.098)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	500	499	598	594	594
Number of failure	164	101	293	56	250

Notes: Standard errors (in parentheses). All coefficients present hazard rates. Observation time is 36 months. The models include all mothers who participated in an interview after birth of the treatment child. All models use extended baseline control variables and community fixed effects. ITT=Intent-to-Treat

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

4.6.1.3 Effects on Pregnancy Outcomes

The results of the former sections illustrated that the *Pro Kind* intervention affected second births but not second pregnancies. Table 4.5 gave first evidence that attrition of pregnant participants does not cause these results. Therefore, it is likely that more pregnancies in the control group did not lead to a live birth. Table 4.9 shows the pregnancy outcomes of all 122 mothers who stated to be pregnant until the second birthday of the treatment child. I only include pregnancies until the second birthday of the treatment child to ensure that the pregnancy outcome is within the observation period. Pregnancy outcomes could be either loss to follow-up, live birth, abortion or miscarriage.

Along with the results of the previous sections Table 4.9 reveals that the percentage of pregnancies which lead to a live birth is significantly higher in the treatment group. Beyond that the table demonstrates that abortions and miscarriages are significantly higher in the control group. However, the rate of pregnant women who are loss to follow-up is only slightly higher in the treatment group. This confirms that selective attrition does not cause the effects on fertility, but reduced abortions and miscarriages in the treatment group.

Table 4.9: Second Pregnancy Outcomes in Treatment and Control Group

	Control	Treatment
Pregnancies at 24 Months	53	69
Pregnancy Outcome in %		
Abortion	0.28	0.12
Miscarriage	0.12	0.06
Life Birth	0.54	0.72
Loss to follow-up	0.06	0.10
Diff. TG-CG: p=0,032**;		

Notes: P-Values base on χ^2 test statistics. The data includes all mothers who stated to be pregnant until 24 months after first birth.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The program goal of *Pro Kind* is not to decrease or increase fertility but to enhance appropriate decisions about fertility. In this context appropriate decisions means that

only mothers who want a second child and who are able to deal with the challenges of a further child get pregnant. After demonstrating that in the treatment group a lower percentage of pregnancies ended in an abortion, it is still unclear whether this is the result of appropriate decisions about planning future pregnancies and births. To investigate this question I analyse the life situation and the attitudes towards second pregnancies of the mothers who gave birth to a second child or were pregnant a second time.

Table 4.10: Life Situation of Mothers who Gave Birth to a Second Child

	Control		Treatment		
	n	%	n	%	P-value
Before Birth of sec. Child					
Wish for sec. preg. at 6 Mo.	31	0.23	48	0.21	0.854
Wish for sec. preg. at 12 Mo.	29	0.28	44	0.25	0.805
After Birth of sec. Child					
Unplanned Preg.	37	0.57	65	0.62	0.636
Father Does not Live In HH	37	0.27	63	0.38	0.259
No Other Care Apart From Mother	37	0.30	64	0.50	0.047**
Mother has no Partner	35	0.06	61	0.16	0.128
Age of the Sec. Child in Mo.	34	8.03	65	6.62	0.398
Age of the Moth. at Births in Years	35	23.4	65	23.9	0.594

Notes: P-values base on z-statistic of a two-group test of proportions. The presented data contains all second children for who data is available. ITT=Intent-to-Treat; C=Control Group.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.10 only includes mothers who gave birth to a second child. The first two rows in the table present statements before occurrence of the pregnancy which lead to a second birth. In treatment and control group only around 25 percent of the second time mothers wished a second pregnancy 6 and 12 months after birth of their first child. The next rows present answers to questions whether the child was unplanned or whether the mother is without a partner. These questions are asked after birth of the second child. If the mothers makes appropriate decisions about family planning, one can expect that these characteristics are uncommon with second time mothers. However, 62 percent of the mothers in the treatment group state that their second child was unplanned. In the control group this rate is 57 percent. Also the other characteristics, like “no partner” or “father does not live in the household” occur more often in the treatment group. The

difference in "no other care giver apart from the mother" is even significant.

These results could indicate that mothers with less resources got pregnant in the treatment group and that these mothers are less responsible in their family planning. However, these group differences are difficult to interpret because more mothers abort their pregnancies in the control group. Analyzing the mothers who abort their pregnancies reveals that two third of the mother have no partner and that these mothers often call a potential further pregnancy catastrophic before the pregnancy occurs. Therefore, the higher rate of abortions in the control group supposes that the attitudes regarding a further pregnancy are more positive in the treatment than in the control group before a pregnancy occurs.

To proof this, Table 4.11 analyses the use of contraception and the attitude towards a second pregnancy of those mothers who got pregnant a second time. The information come from questions which are asked before the second pregnancy occurred. In the treatment group significantly more mothers are happy or have no worries about a further pregnancy. Additionally, there are less mothers who do not want a further child. These statements indicate that the mothers who give birth to a second child in the treatment group were more positive about a further pregnancy which could explain the lower rate of abortions in the treatment group.

Table 4.11: Attitudes Towards a Second Pregnancy

	Control Group		Treatment Group		
	n	%	n	%	P-value
Happy or no Worries about Further Pregnancy	31	0.35	41	0.56	0.083*
No Further Child Wanted	29	0.17	39	0.10	0.401
No or Unregular Use of Contraception	31	0.39	42	0.33	0.619

Notes: Notes: P-values base on z-statistic of a two-group test of proportions. The presented data contains statements of mothers before their second pregnancy which lead to a second birth.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, less abortions and less miscarriages mainly cause the effect of the *Pro Kind*

project on fertility. It seems that at least partly more appropriate decisions about family planning in the treatment group causes this effect since the treatment group has more positive attitudes towards a second child and, considering the higher abortion rate in the control group, less unplanned pregnancies. One explanation why *Pro Kind* increases the rate of mothers who want a second child might be higher maternal satisfaction. This satisfaction might be caused by more positive experiences with their first child and higher personal strengths. Another explanation why more mothers in the control group decide for an abortion might be depression and low well-being. Both is strongly correlated with abortion (Suri et al., 2004; Aavitsland, 2009). These mental health problems might be reduced due to the intervention. If these explanations are valid, they must be confirmed by measures of maternal subjective life-satisfaction and well-being which are analysed in the next section.

.

4.6.2 Subjective Outcomes

This section investigates whether the *Pro Kind* intervention influences maternal subjective life-satisfaction and well-being. These outcomes are obtained at the interview 27 months after birth of the treatment child. The questionnaires use measures which are also used by the GSOEP and which are intensively tested on reliability and validity (Krueger and Schkade, 2008; Bertrand and Mullainathan, 2001). It is important to investigate these subjective outcomes because previous research showed that a higher maternal life-satisfaction and well-being can positively influence child outcomes (Berger and Spiess, 2011). Furthermore, the investigation of the maternal life-satisfaction can help to explain the reason why *Pro Kind* affects fertility.

The telephone questionnaire contains 13 items concerning subjective life-satisfaction and well-being, eight of these 13 items measure various satisfaction dimensions, one item measures general life-satisfaction and four measure well-being. Appendix J gives a de-

scriptive overview about the outcomes in the treatment and control group and for GSOEP first time mothers. In eight of the nine satisfaction dimensions the mothers in the treatment group state to be more satisfied than the mothers in control group. The picture is similar in the four questions regarding well-being. The mothers in the treatment group feel less often sad, angry, worried and more often happy. Comparison with the GSOEP mothers shows that these mothers are less often sad, more often happy and in most categories more satisfied than the *Pro Kind* mothers. Only in the category housework the *Pro Kind* mothers are more satisfied which could be arise from lower opportunities in the labor market and therefore higher satisfaction with home production.

Table 4.12: Well-Being in the last Four Weeks and Satisfaction with Life in General

	(1)	(2)	(3)	(4)	(5)
	In the last four weeks				Satisfaction with
	Angry	Worried	Happy	Sad	Life in General
			ITT		
Home Visiting	-0.106	-0.289***	0.088	-0.191**	0.147**
	(0.062)	(0.086)	(0.085)	(0.083)	(0.062)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	429	427	427	427	427

Notes: Standard errors (in parentheses). The dependent variables in columns 1-4 are measured in a five point likert scale. The dependent variables in column 5 is measured in a eleven point likert scale. All dependent variables are standardized with mean of zero. All models include extended baseline control variables, community fixed effects and age of the treatment child. Measurement is in average at 28 months after birth of the treatment child.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12 and 4.13 show that the difference between control and treatment group is significant in seven of the 13 items at a ten percent level after including controls. Furthermore, the non significant coefficients are all positive indicating higher satisfaction and well-being in the treatment group. The standardized effect sizes are meaningful with values around 0.2 SD.

After showing that *Pro Kind* increased the maternal life-satisfaction and well-being, I continue investigating if these subjective measures are related to the fertility decisions.

In line with the literature that unhappier women more often tend to an abortion, these mothers have a general satisfaction value of 5.74. Although it is not clear if low life-satisfaction caused the abortion or the abortion the low life-satisfaction, it gives a first hint that low life-satisfaction is correlated with abortions.

Table 4.13: Life-Satisfaction in Different Areas

	(1) Health	(2) Housework	(3) Household Income	(4) Personal Income	(5) Place of Dwelling	(6) Free Time	(7) Child Care Availability	(8) Family Life
				ITT				
Home Visiting	0.096*	0.175	0.253***	0.156*	0.014	0.148*	-0.047	0.067
	(0.054)	(0.132)	(0.056)	(0.073)	(0.065)	(0.081)	(0.085)	(0.061)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	424	419	423	419	424	424	415	424

Notes: Standard errors (in parentheses). The dependent variables are measured in an eleven point likert scale. All dependent variables are standardized with mean of zero. All models include extended baseline control variables, community fixed effects and age of the treatment child. Measurement is in average at 28 months after birth of the first child. ITT=Intent-to-Treat;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Another hint that the higher life-satisfaction in the treatment group is related to fertility comes from a comparison between mother who gave birth to a second child in the treatment and control group. The life-satisfaction differs significantly with a value of 7.61 in the treatment group and 6.42 in the control group (T=-3.06; nTG=60; nCG=33). It is possible that the birth of the second child caused this happiness increase. However, it is unlikely that the higher life-satisfaction is not influenced by better experiences with the first child and that, therefore, the mothers are already happier before their second pregnancies. If this is the case, this higher happiness could give an explanation for the smaller rate of abortions in the treatment group.

4.7 Conclusion

Home visiting programs are a popular type of early childhood interventions to support disadvantaged families. While many studies investigated how these programs affect child

outcomes, this study has explored by a randomized experiment the much less investigated question how home visiting programs affect maternal life course. The few previous studies which have investigated this topic have found positive effects on maternal employment and negative effects on fertility. In contrast, the analysis of the *Pro Kind* Project reveals that the intervention has no effects on employment but strong positive effects on fertility. The effects on fertility are mainly driven by less abortions in the treatment group. Furthermore, *Pro Kind* increased the life-satisfaction and well-being of the participating mothers.

The previous studies which examined the effects of home visiting on the maternal life course were located in the US, whereas the *Pro Kind* program is located in Germany. Therefore, the different welfare state systems might explain much of the variation between the outcomes in the previous studies and the *Pro Kind* study. In the US welfare state, mothers who receive welfare have less incentives to give birth to a second child than in Germany. In this European country each additional child increases the amount of welfare and there are no work obligations or benefit cuts until the third birthday of the child. Therefore, an increase in maternal skills and life-satisfaction due to the intervention could lead to further birth in Germany, whereas in the US these improved skills might be used for higher labor market participation.

A randomized experiment is used to evaluate the effects of *Pro Kind* on the maternal life course. Therefore, the effects can be causally linked to the intervention. However, around half of the randomized mothers did not participated in all interviews until the third birthday of the treatment child. Nevertheless, this attrition is not selective between treatment and control group in respect to the baseline characteristics and there is no indication that mothers who are pregnant or employed leave the sample selectively. Therefore, it is unlikely that the sample attrition causes problems to the validity of the results.

The results of this study can help to better understand the mechanisms through which early childhood interventions work. It is likely that improved maternal life-satisfaction and

well-being can explain partly why these programs improve various child outcomes. Furthermore, the results give new insights how welfare regimes influence fertility. Although the literature presents inconclusive results whether welfare affects fertility (Moffitt, 1998; Kearney and Levine, 2012) the picture could be different if the welfare regime interacts with an early childhood intervention. Considering these results might be helpful for other policies from the US which will be implemented in Europe in the future.

4.8 Appendix C

Table 4.14: Selective Attrition TG and CG Psychological Characteristics - Telephone Interviews

	Difference TG/CG						
	Pregnancy	3 Months	9 Months	15 Months	21 Months	27 Months	36 Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Selected Psychological and Physical Characteristics							
Unwanted Pregnancy	0.016	0.025	0.023	0.007	0.010	0.000	0.000
	(0.029)	(0.031)	(0.032)	(0.034)	(0.034)	(0.035)	(0.037)
Daily Smoking	0.006	-0.008	-0.014	-0.001	-0.010	-0.020	-0.027
	(0.036	(0.039)	(0.042)	(0.044)	(0.046)	(0.045)	(0.049)
Isolation	-0.021	-0.022	-0.013	-0.006	0.001	0.006	0.025
	(0.019	(0.021)	(0.023)	(0.024)	(0.024)	(0.025)	(0.029)
Foster Care Exper.	0.053*	0.057*	0.051	0.068*	0.066*	0.060*	0.051
	(0.030	(0.032)	(0.035)	(0.036)	(0.036)	(0.035)	(0.038)
Neglect Experience	-0.007	-0.003	0.006	0.006	0.009	0.001	-0.003
	(0.037	(0.040)	(0.044)	(0.046)	(0.048)	(0.046)	(0.051)
Lost Experience	-0.041	-0.074*	-0.050	-0.037	-0.077	-0.043	-0.004
	(0.038	(0.041)	(0.045)	(0.048)	(0.049)	(0.048	(0.052)
Violence Ever	-0.008	-0.009	-0.031	-0.032	-0.020	-0.009	-0.022
	(0.020	(0.021)	(0.023)	(0.025)	(0.026)	(0.026)	(0.029)
Depression	-0.022	-0.001	0.006	0.008	0.010	0.013	0.006
	(0.024	(0.025)	(0.028)	(0.028)	(0.030)	(0.029)	(0.032)
Anxiety	-0.008	0.007	0.008	0.001	-0.010	0.001	0.002
	(0.029	(0.031)	(0.033)	(0.036)	(0.037)	(0.035)	(0.038)
Stress	0.031	0.033	0.037	0.037	0.024	0.000	-0.006
	(0.036	(0.038)	(0.042)	(0.044)	(0.046)	(0.044)	(0.049)
Aggression	-0.048*	-0.051*	-0.066*	-0.084**	-0.064*	-0.050	-0.041
	(0.029	(0.030)	(0.033)	(0.036)	(0.036)	(0.035)	(0.037)
Medic. Indic. Risk Preg.	0.008	0.003	-0.020	-0.034	-0.017	-0.002	0.004
	(0.024	(0.025)	(0.028)	(0.029)	(0.031)	(0.030)	(0.033)
BMI	-0.293	-0.276	-0.140	-0.063	-0.161	-0.125	-0.125
	(0.419	(0.458)	(0.506)	(0.537)	(0.527)	(0.550)	(0.614)
Sum Risk Factors	-0.076	-0.125	-0.099	-0.108	-0.160	-0.078	-0.054
	(0.181	(0.193)	(0.209)	(0.221)	(0.226)	(0.226)	(0.250)
Observations	680	582	486	442	418	438	374

Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Table 2.9 and 2.10 for variable definitions.
* p < 0.1, ** p < 0.05, *** p < 0.01

Table 4.15: Sample Balance Across Participants with all Interviews until the Second and Third Birthday

	All Interviews 36		All Interviews 24	
Demographic Characteristics				
Age in Years	0.413	(0.562)	-0.057	(0.503)
Week in Pregnancy	(0.071)	0.663	-0.291	(0.620)
Migration	-0.066	(0.045)	-0.074*	(0.040)
Underage	0.028	(0.040)	0.038	(0.037)
Mon. HH-inc. in €	7.67	(64.45)	17.34	(58.41)
Debt over 3000 €	0.064	(0.047)	0.037	(0.042)
Education Risk	0.028	(0.054)	0.059	(0.049)
Income Risk	0.026	(0.050)	0.012	(0.045)
Employment Risk	-0.072	(0.049)	-0.057	(0.044)
No Partner	0.029	(0.056)	0.021	(0.050)
Living with Parents	-0.051	(0.051)	-0.021	(0.046)
Persons in HH	-0.143	(0.182)	-0.096	(0.166)
Selected Psychological and Physical Characteristics				
Unwanted Pregnancy	-0.018	(0.041)	-0.015	(0.038)
Daily Smoking	-0.022	(0.055)	-0.012	(0.050)
Isolation	0.026	(0.033)	0.002	(0.029)
Foster Care Exper.	0.060	(0.042)	0.081*	(0.039)
Neglect Experience	0.034	(0.058)	0.038	(0.053)
Lost Experience	-0.004	(0.058)	-0.031	(0.054)
Violence Ever	-0.043	(0.034)	-0.031	(0.030)
Depression	0.013	(0.036)	0.013	(0.033)
Anxiety	0.003	(0.045)	-0.005	(0.040)
Stress	0.010	(0.055)	0.007	(0.050)
Aggression	-0.086**	(0.043)	-0.078*	(0.040)
BMI	0.087	(0.662)	0.191	(0.593)
Medic. Indic. Risk Preg.	-0.023	(0.036)	-0.029	(0.033)
Sum Risk Factors	-0.097	(0.279)	-0.082	(0.254)
Observations	296		348	

Notes: Robust standard errors shown in parentheses. See Table 2.9 and 2.10 for variable definitions.
p < 0.1 ** p < 0.05, *** p < 0.01

Table 4.16: Selective Attrition to Baseline Demographic Characteristics - Telephone Interviews

	Difference Attritors Non-Attritors						
	Pregnancy	3 Months	9 Months	15 Months	21 Months	27 Months	36 Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demographic Characteristics							
Age in Years	-1.363***	-1.183***	-1.718***	-1.612***	-1.858***	-1.946***	-1.848***
	(0.455)	(0.335)	(0.301)	(0.313)	(0.301)	(0.297)	(0.308)
Week in Pregnancy	-1.929***	-1.267**	-1.007**	-0.925**	-0.523	-1.217***	-1.425***
	(0.740)	(0.511)	(0.443)	(0.430)	(0.427)	(0.428)	0.426)
Migration	-0.057	-0.035	-0.004	-0.028	-0.023	-0.061**	-0.058**
	(0.038)	(0.028)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
Underage	0.057	0.049	0.075**	0.059*	0.099***	0.110***	0.088***
	(0.049)	(0.035)	(0.031)	(0.030)	(0.029)	(0.029)	(0.029)
Mon. HH-Inc. in €	-211.82***	-198.50***	-108.16**	-146.50***	-125.92***	-154.13***	-171.91***
	(56.04)	(43.73)	(44.86)	(44.02)	(42.89)	(43.48)	(42.59)
Debt over 3000 €	-0.072*	-0.043	-0.054*	-0.024	-0.044	-0.036	-0.045
	(0.040	(0.032	(0.029	(0.029	(0.028	(0.028	(0.028)
Education Risk	0.179***	0.130***	0.158***	0.149***	0.158***	0.161***	0.167***
	(0.038	(0.032	(0.029	(0.030	(0.030	(0.030	(0.031)
Income Risk	0.090**	0.080**	0.067***	0.091***	0.080***	0.089***	0.106***
	(0.039)	(0.031)	(0.029)	(0.028)	(0.028)	(0.028)	(0.029)
Employment Risk	0.106***	0.060**	0.075***	0.093***	0.099***	0.084***	0.094***
	(0.034)	(0.029)	(0.027)	(0.026)	(0.026)	(0.027)	(0.027)
No Partner	-0.009	-0.003	0.007	-0.010	-0.005	-0.046	-0.055*
	(0.056)	(0.040)	(0.035)	(0.034)	(0.034)	(0.034)	(0.034)
Living with Parents	-0.057	0.011	0.048	0.044	0.031	0.007	-0.014
	(0.053)	(0.040)	(0.035)	(0.034)	(0.033)	(0.034)	(0.033)
Persons in HH	0.070	0.106	0.219*	0.133	0.183*	0.224	0.159
	(0.234)	(0.145)	(0.131)	(0.126)	(0.123)	(0.125)	(0.121)
Observations	755	755	755	755	755	755	755

Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Table 2.9 and 2.10 for variable definitions.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.17: Selective Attrition to Baseline Psychological Characteristics - Telephone Interviews

	Difference Attritors Non-Attritors						
	Pregnancy	3 Months	9 Months	15 Months	21 Months	27 Months	36 Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Selected Psychological and Physical Characteristics							
Unwanted Pregnancy	-0.038	0.002	0.069**	0.073**	0.067**	0.034	0.027
	(0.044)	(0.034)	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)
Daily Smoking	0.048	0.093**	0.039	0.061*	0.067*	0.045	0.040
	(0.058)	(0.042)	(0.037)	(0.036)	(0.035)	(0.035)	(0.035)
Isolation	0.050	0.020	0.008	0.000	0.014	-0.005	-0.016
	(0.040)	(0.025)	(0.021)	(0.020)	(0.020)	(0.019)	(0.019)
Foster Care Exper.	0.180***	0.111***	0.081**	0.101***	0.114***	0.121***	0.112***
	(0.058)	(0.039)	(0.033)	(0.032)	(0.030)	(0.031)	(0.030)
Neglect Experience	0.101	0.103**	0.049	0.062*	0.039	0.081**	0.069*
	(0.062)	(0.043)	(0.038)	(0.037)	(0.036)	(0.036)	(0.036)
Lost Experience	-0.035	-0.008	0.004	0.043	0.037	0.016	0.039
	(0.062)	(0.044)	(0.039)	(0.038)	(0.037)	(0.037)	(0.037)
Violence Ever	0.133***	0.075**	0.068***	0.050**	0.035	0.029	0.021
	(0.047)	(0.029)	(0.024)	(0.022)	(0.021)	(0.021)	(0.021)
Depression	0.090*	0.049	0.029	0.041*	0.027	0.037	0.024
	(0.048)	(0.032)	(0.025)	(0.025)	(0.024)	(0.025)	(0.024)
Anxiety	0.079	0.057	0.034	0.028	0.012	0.050*	0.047*
	(0.051)	(0.035)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)
Stress	0.036	0.016	0.001	0.017	0.009	0.032	0.025
	(0.058)	(0.041)	(0.035)	(0.035)	(0.034)	(0.035)	(0.035)
Aggression	0.035	0.041	0.031	0.030	0.032	0.053*	0.063**
	(0.049)	(0.034)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)
BMI	-0.839	-0.503	-0.751*	-0.764*	-0.710*	-1.391***	-1.208***
	(0.640)	(0.452)	(0.409)	(0.399)	(0.402)	(0.388)	(0.399)
Medic. Indic. Risk Preg.	0.024	0.033	0.009	0.007	-0.008	0.009	-0.004
	(0.044)	(0.029)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)
Sum Risk Factors	1.110***	0.735***	0.705***	0.760***	0.742***	0.824***	0.818***
	(0.336)	(0.229)	(0.190)	(0.184)	(0.180)	(0.182)	(0.178)
Observations	755	755	755	755	755	755	755

Robust standard errors shown in parentheses. Estimates include community fixed effects.
See Table 2.9 and 2.10 for variable definitions.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 4.2: Kaplan Meier Survival Functions - Only Participants with Complete 36 Months Data

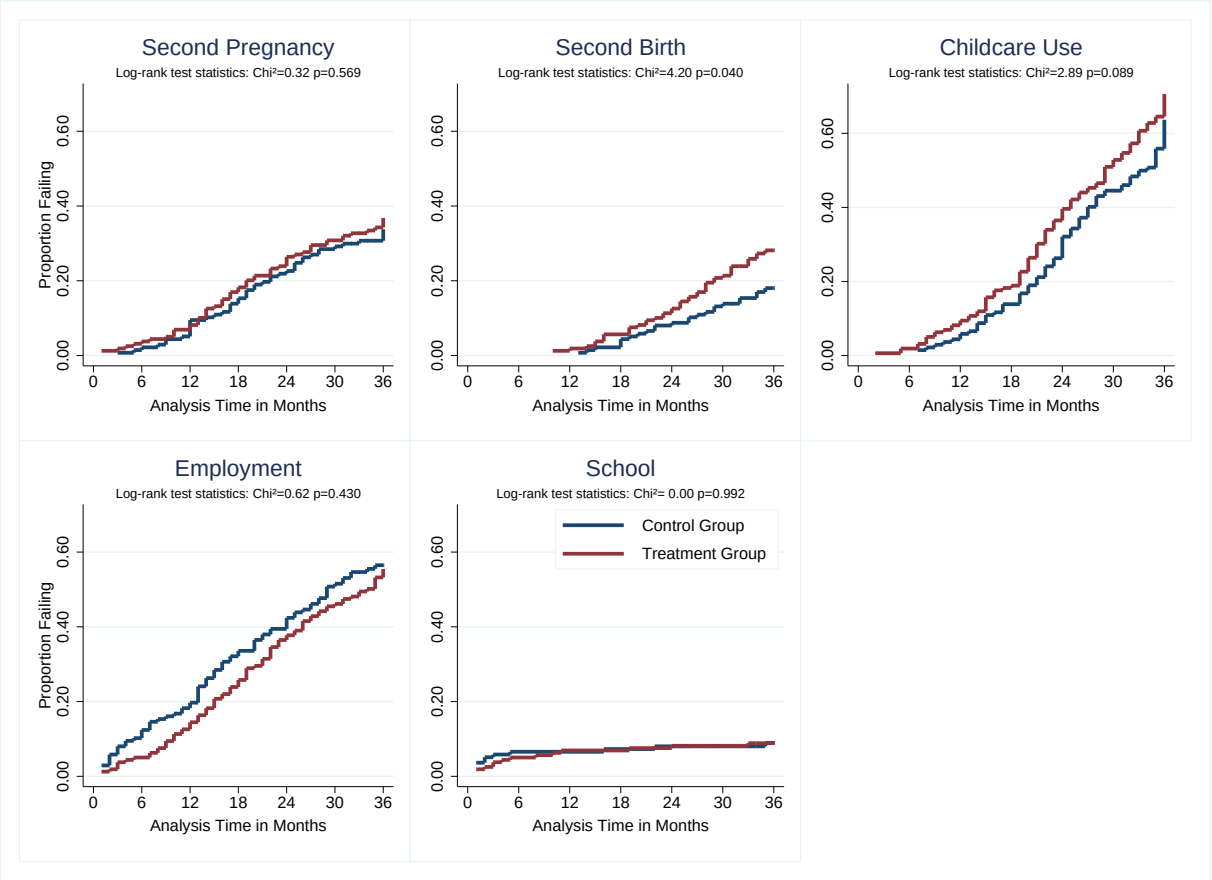


Figure 4.3: Hazard Rates for Second Pregnancy and Birth

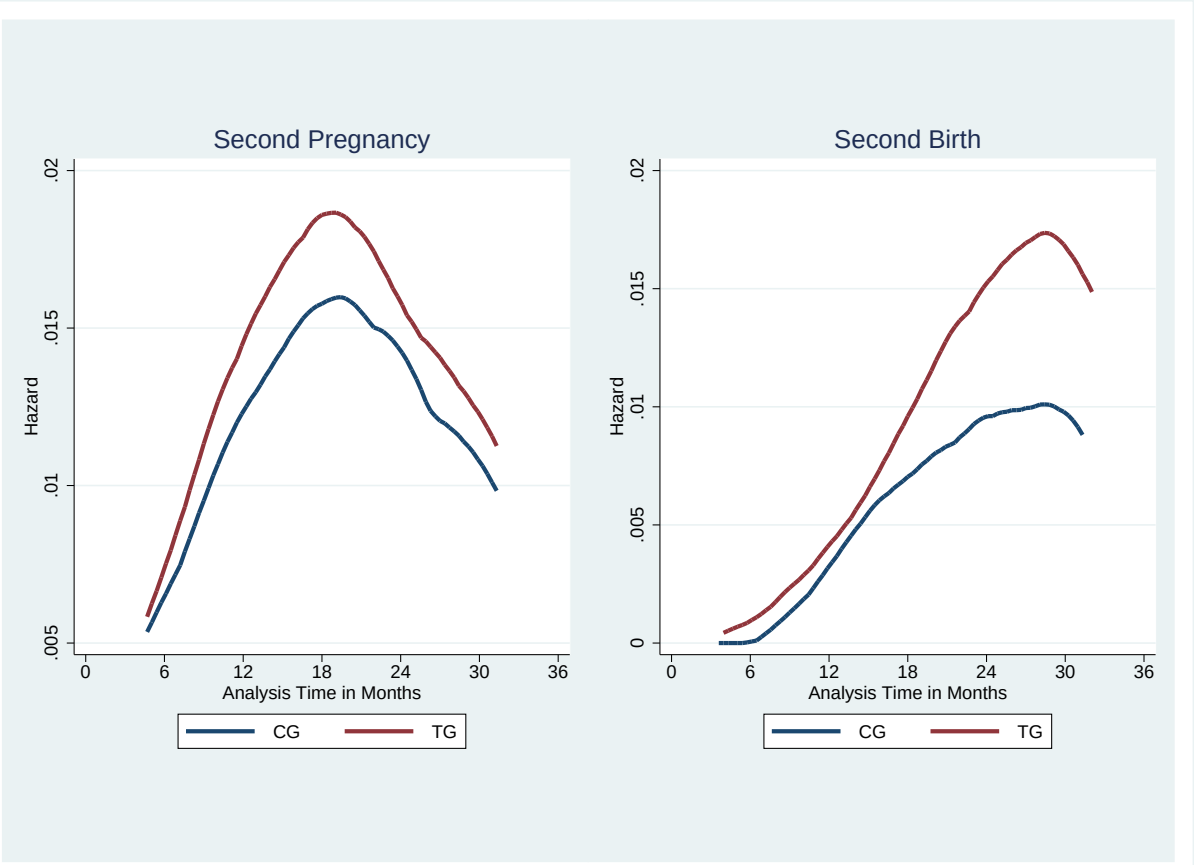


Table 4.18: Descriptive Statistics for Well-Being and Life-Satisfaction

	Control Group			Treatment Group			GSOEP		
	Mean	sd	n	Mean	sd	n	Mean	sd	n
How Often or Seldom Have You Experienced this Feeling in the Last Four Weeks?									
Angry	3.05	1.00	195	2.91	1.09	239	3.09	0.89	394
Worried	2.09	1.04	194	1.77	0.94	238	1.99	0.91	393
Happy	3.66	0.90	195	3.76	0.88	237	3.90	0.78	394
Sad	2.71	1.07	195	2.49	1.03	237	2.40	0.98	394
How Satisfied are you Today with the Following Areas of Your Life?									
Health	6.55	2.97	194	6.83	2.88	235	7.38	1.89	601
Housework	6.92	2.33	193	7.37	2.32	231	6.39	2.19	579
Household Income	4.92	2.70	193	5.58	2.89	235	5.47	2.77	578
Personal Income	4.14	2.90	191	4.57	3.05	233	6.42	2.82	582
Place of Dwelling	6.56	3.16	194	6.63	3.12	235	6.83	2.34	599
Free Time	5.67	2.91	195	6.23	2.87	234	6.77	2.51	563
Child Care Availability	6.73	3.01	192	6.68	3.33	228	7.36	2.18	590
Family Life	7.46	2.35	195	7.63	2.52	234	7.43	2.19	509
Life in General	7.13	2.10	195	7.44	1.91	237	7.41	1.56	601

Notes: For the outcomes in the first four rows the scale is: 1=Very Rarely, 2=Rarely, 3=Occasionally, 4=Often, 5=Very Often. For the other outcomes the scale is: 0=totally unhappy to 10=totally happy. GSOEP includes mothers whose first child has an age between two and three years. The average age of the first child in the *Pro Kind* sample is 30.06 months. sd=standard deviation.

5 Quasi-Experimental Evaluation of a Student Mentoring Program [1]

5.1 Introduction

Student attrition is an issue in higher education worldwide. In OECD countries 31 percent of all university students who enroll fail to earn any degree (OECD, 2011). Attrition causes high economic costs because a college wage premium is only paid when a degree is attained. Especially in subjects with high expected returns like engineering, science or economics, 40 to 50 percent of the students fail or do not obtain a degree. It is well documented that up to 60 percent of these students leave the university in the first two semesters (e.g. Heublein et al., 2010). In these semesters students often fail because they lack non-academic skills like time management, self-organization or identification with the subject.

One popular way to enhance non-academic skills is student mentoring where a more experienced member (mentor) of a university maintains a relationship with a less experienced, often new member (mentee) to the university. The mentor provides information, support and guidance to enhance the mentee's chances of success in the university and beyond (Campbell and Campbell, 1997). Topics and forms of mentoring vary widely, in

[1]The author would like to thank Nicole May and Christian Heidrich of the Economics and Management Faculty and Michael Flechtner of the Leibniz University of Hanover Controlling for support.
This article is accepted for presentation at the 18th Spring Meeting for Young Economists (SMYE) 2013 and the the 27th Conference of the European Society of Population Economics (ESPE) 2013.

most cases it is not focused on academic skills, but rather on time management, motivation, self-organization and knowledge about the institution (Crisp and Cruz, 2009). Although heterogeneous forms of mentoring do exist, most higher institutions, especially in the US and UK, offer some kind of mentoring.

Despite the frequent use of mentoring in higher education institutions, only in recent years a few studies have evaluated the effectiveness of mentoring programs with an experimental or quasi-experimental design. (Bettinger and Baker, 2011; Scrivener and Weiss, 2009; Angrist et al., 2009). Overall, these studies found only small effects of mentoring programs. However, previous research only investigated low quality mentoring, in which undergraduate students delivered the mentoring or a high mentee-mentor ratio was present. So far, there are no studies of high quality mentoring which examine the promising combination of graduated mentors who are part of the faculty and a low mentee-mentor ratio.

This paper studies the effects of a high quality student mentoring program on first year study success at Leibniz University Hanover (LUH), Germany. All mentors are graduated and employed at the university. Most of them have a degree in management or economics which is often obtained at LUH. This gives the mentors a rich knowledge about the institution which can be passed to the mentees. The mentoring program is offered to first year Bachelor students who are enrolled in an Economics and Management (EM) program. The focus of the mentoring is on the first months of their studies in which the students have several mandatory appointments with their mentor. None of the mentors consults more than 20 students. Topics of the appointments include: time management, transition from high school to college or general advice.

I use a difference-in-differences approach to evaluate the effects of the mentoring program. For this purpose, I use the fact that LUH does not offer the mentoring to students in an industrial engineering (IE) program. However, in certain classes these IE students take the same exams and attend the same lectures as the EM students. Therefore,

the IE students are a reliable control group to examine the effectiveness of the mentoring treatment. For the analyses, I include five student cohorts from Winter 2006 to Winter 2010. The mentoring was not offered for any degree in the first two student cohorts. For the next three student cohorts the mentoring program was just offered for EM students. The introduction of the mentoring is the only change in the study situation between EM and IE students. Hence, changes in the difference of exam failure, grades and absent rates between EM and IE students are causally linked to the mentoring.

I find that the mentoring program significantly decreased the failure rate and increased the grades for several courses. The effects are highest in the first semester exams where the failure rate decreases between 9 and 15 percentage points which is a reduction of the failure rate by 27 to 37 percent. The effects get smaller in the second term, when the appointment frequency between mentor and mentee fades out. Investigating subsamples reveals that female students benefit most from the mentoring program at least in some exams. The results hardly change when controls for age, gender, nationality, high school GPA, as well as place and type of high school are included. Despite the strong effects on the first semester exams, the effects on the second year retention are small and similar to the previous studies.

Participation in the mentoring was compulsory for all EM students. Nevertheless, refusals were not penalized. This suggests that not all students complied in the mentoring and, therefore, all estimates measure intention-to-treat effects. However, a student questionnaire reveals that almost all students took part in the mentoring. Female students used the mentoring program more often than male students which can explain the higher effects for these students. In contrast, students with a foreign citizenship participate less often and if they participated, they had less appointments with their mentor.

The findings of this article contribute to a broader literature which examines policies to enhance college success. This literature mostly focuses on financial rewards for students or remedial classes. Angrist et al. (2009) as well as Leuven et al. (2010) and Paola et al.

(2012) investigate the results of financial incentives on college outcomes and retention in the US and Europe. Positive results occur only for women or well-performing students. Bettinger and Bridget (2009) examine college remediation by a quasi-experimental research design in the US and show that remediation reduces attrition. On the opposite, Di Pietro (2012) finds no effects of remedial courses in the UK in a quasi-experimental study.

However, financial reward interventions assume that the naturally existing incentives are not strong enough for students, whereas remediation interventions assume that students lack academic skills. In contrast to these interventions, mentoring is based on the assumption that behavioral and non-academic factors lead to failure and attrition and that these factors can be improved by a small nudge. The results of my study in which a small intervention has a big effect support this hypothesis.

The rest of the paper is organized as follows: Section 5.2 provides details of the Faculty of Economics and Management at the Leibniz University Hanover and the Hanover mentoring program. Section 5.3 describes the data and provides summary statistics. It also reports results from a questionnaire which investigates the student's utilization and the perception of the Hanover mentoring program. Section 5.4 shows the econometric framework and presents results. Section 5.5 gives robustness checks. Section 5.6 summarizes and concludes.

5.2 The Hanover Mentoring Program

5.2.1 School Background

Leibniz University Hanover (LUH) is a university with 23.000 students and 90 degree programs. Major fields of the university are science and engineering. In these fields, LUH reaches international excellence. Despite this focus on science and engineering the Faculty of Economics and Management is ranked among the top ten in Germany. As it is

typical also for well performing German universities, most undergraduate, graduate and PhD-students at LUH come from the surrounding areas. The Faculty of Economics and Management offers two undergraduate bachelor degree programs: one in economics and management (EM) and the other in industrial engineering (IE). Both programs have been offered since 2006 after the Bologna reform. Before this reform students could only attain diploma degrees at the Faculty of Economics and Management.

The EM program has a nominal duration of four years. Each year is divided into a winter and a summer semester. In the winter semester, lectures start in October and end in January and in the summer semester the lectures start in April and end in July. In the first academic year, which always begins in the winter semester, all EM students follow exactly the same program of four compulsory courses in the first semester and four compulsory courses in the second. It is important to note that it is not possible for the students to switch or postpone any class since they are automatically enrolled for the exams. The exams take place every semester once the courses are finished and the re-take exams (for students who either fail or have been absent from the regular test) are organized in June and December. It is only after the fifth semester that students choose different packages of courses to specialize either in economics or in business.

The IE program has a nominal duration of three and a half years. It is a mixture of economics and management and engineering. In economics and management the IE students have to take the same courses with the same lectures and exams as the EM students. However, the IE students attend some courses in a later semester than the EM students. Beyond that, IE students take different mathematics courses, they do not take any statistics class and they attend two classes of mechanical engineering and two classes of electrical engineering in the first year. Table 5.1 gives an overview in which semester IE and EM students have to take the first year economics and management courses.

In both programs the failure rates in the first semester exams are as high as 50 percent, particularly in mathematics for EM students and engineering for the IE students.

Table 5.1: Overview of Economic and Management Classes in the EM and IE Programs

Course	EM	IE
Business Studies I (Accounting + Business Informatics)	1. Sem.	1. Sem.
Business Studies II (Marketing + Business Management)	1. Sem.	3. Sem.
Economics I (Introduction to Economics)	1. Sem.	3. Sem.
Mathematics I	1. Sem.	Own Course
Business Studies III (Production Management + HRM)	2. Sem.	2. Sem.
Economics II (Microeconomics)	2. Sem.	4. Sem.
Mathematics II	2. Sem.	Own Course
Statistics I	2. Sem.	-

These rates are comparable to economics and business programs at other German faculties. However, the examination regulation is rather strict in the Hanover faculty. If a student fails an exam, there is just one more attempt to re-take the exam. If this student also fails the second attempt, the faculty exmatriculates the student automatically. This rule applies only for the economics and management courses but for both EM and IE students. In the engineering courses, the IE students are allowed to re-take the exams two times. The result of any exam counts for the final grade of the bachelor degree. Therefore, the first semester is on the one hand important for the final grade of the degree and on the other hand a fail leads to the re-take exam which decides on being exmatriculated or not.

5.2.2 Design of the Mentoring Program

The Hanover mentoring program was implemented after the introduction of tuition fees in Lower Saxony in the Winter term of 2006. Since then each student pays 500 Euro tuition per semester. As the government has not been allowed to cut its spending on the universities, this tuition is a top up revenue for the universities. After discussion with the students, the Faculty of Economics and Management decided to spend part of the additional money on a mentoring program. The faculty offered the revenue earned by the tuition to the departments conditional on the participation of at least one department

member in the mentoring. If the department agrees it can use the money for hiring additional staff. Afterwards, the department determines which member joins the mentoring program. The IE students are not included in the mentoring program because their tuition fees are not allocated to the Faculty of Economics and Management. However, the mentoring program is the only additional service at the Faculty of Economics and Management, which is not available for IE students because the newly hired staff, apart from the mentoring, exclusively offers additional lectures and office hours. Both the lectures and the office hours are open to IE and EM students and therefore improving the study conditions for both groups.

The Hanover Mentoring Program pursues three goals: Improving the student performance, increasing the study satisfaction and decreasing the failure and attrition rates. To reach these three goals, the mentoring covers four topics which are supposed to have a significant influence on first semester study success. The first topic is the smooth transition from high school to college. Often first semester students realize too late that the amount of class content is much greater at the university level than in high school. Therefore, the students start too late in the semester with the preparation for the exams. Furthermore, especially in a strict system like in Hanover, the students are not aware of the consequences if they fail an exam. On the contrary, some first semester students are overwhelmed and scared by the challenges of a university study. In both cases the mentor can adjust these valuations and set the student on the right track. The second topic is the reduction of separation and exclusion for the first semester students. Here the study satisfaction of the students could get improved because the students have a personal contact person and they get the feeling that the institution is interested in them. Additionally, the mentor directly brings the student to other students to establish a study group. The third topic is advice in general topics, like financing, part-time jobs or housing. Advice in these topics could save energy and time which the student could use for preparation instead. Finally, the mentor directly tackles non-academic skills such as giving advice in

how to benefit from lectures.

To achieve the goals and to cover the mentioned topics most effectively, the mentors exhibit certain characteristics. One common characteristic of all mentors is to be employed by the faculty itself. This gives the mentor the institutional knowledge for practical advices. Additionally, the mentors are probably in higher risk of sanctions than an external mentor when the commitment for the mentoring is low. Another characteristic is that all of the mentors are graduates. Most of them hold an economics or management degree which is often earned at the Hanover faculty. This makes the mentors reliable when they give advice to the mentees' because the mentor has already achieved the goal for which the mentee aims, and the mentors are able to understand the mentees problems because they can go back to experience from their own studies. Finally, most mentors are PhD-students (in Germany PhD-students are mostly employed by the departments in the faculty) which leads to a low average age of the mentors resulting in a possibility of mutual trust and identification between mentor and mentee. Table 5.2 gives an overview of the mentor characteristics in every year since implementation.

Table 5.2: Overview of Mentor Characteristics

	Cohort		
	2008	2009	2010
Number of Mentors	25	30	30
Mentees per Mentor	15.8	14.5	14.6
Average Age of the Mentor	29.7	30.4	32.2
Female Mentors in %	45.5	34.5	34.1
Mentors with Economic or Management Degree in %	100	95.2	100
Mentors Graduated at Leibniz University Hanover in %	66.5	67.3	68.4

In addition to the mentors' characteristics, the mentoring program also possesses certain design features to achieve the intended goals. One feature is that at least three appointments between mentor and mentee are compulsory, scheduled and face-to-face. The appointments are face-to-face to enable a personal relationship between mentor and mentee. The appointments are scheduled to guarantee that they take place in critical

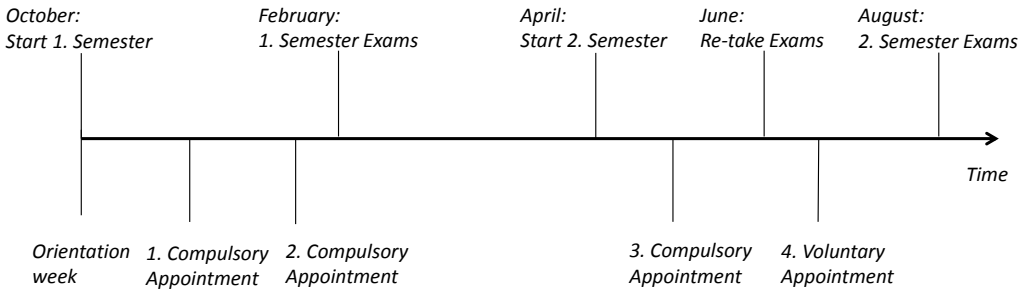
phases of the first semester and the first year. The appointments are compulsory to ensure that all students participate in the mentoring. Each of the appointments has a duration of approximately 30 minutes. Beyond these appointments the mentor encourages the mentee to hold contact via e-mail or telephone or to use the regular office hours of the mentor. Another feature of the mentoring program is a low mentor-to-mentee ratio of 1 to 20 or 1 to 10 depending on the mentor's labor contract. Furthermore, the mentors received training in negotiation and continuous feedback from supervision.

The mentoring starts in October with the beginning of the semester. Each student takes part in an orientation week which gives the students the possibility to get to know each other. The orientation week is organized in groups with a size of 40 students. Each group is accompanied by a graduated faculty employee. These accompanying employees are mentors as well. Therefore, the mentors can already build up a relationship with their mentees during this week. Additionally, during the orientation week, the mentors announce that participation in the mentoring is compulsory and that students get penalized if they do not participate. However, penalties have not been executed for non-participating students in any year of the mentoring, yet. Because orientation week group sizes are too high for an aspired mentor mentee ratio, half of the students are randomly allocated to a second mentor. This second mentor participates one day in the orientation week.

The major focus of the mentoring program is on the first 6 months of the university study. Therefore, the first compulsory meeting takes place four weeks after the start of the semester. In this meeting the focus is on the transition between high school and college. The second meeting is scheduled after Christmas four weeks before the final exams of the first semester. The focus of this second meeting is on exam preparation. This date has been chosen because there is still time for the students to adjust their learning behavior. The third meeting is compulsory for all students, but the focus is on those students who have failed an exam. The intention of the third meeting is the preparation

for the re-take exams. A fourth appointment is voluntary. Figure 5.1 summarizes the schedule for the appointments in the first year. All students get the information about the scheduled appointments during the orientation week. This makes the mentoring program transparent and reliable for the students.

Figure 5.1: Time Schedule of the Hanover Mentoring



5.3 Data and Descriptives

In order to estimate the effect of the mentoring program on the exam results, it would not be suitable just to compare the exam outcomes of the EM students before and after implementation of the mentoring. Another approach is necessary because changes in the exam outcomes could also be ascribed to time variant factors like the difficulty of the exam or the quality of the lecturers. To solve this problem, I apply a difference-in-differences approach in which I compare the difference between the exam outcomes of the

EM students to that of the IE students before and after the introduction of the mentoring program. Due to the fact that IE and EM students take the same exams and attend the same lectures a change in the difference of the exam outcomes between EM and IE students can be explained by the introduction of the Hanover Mentoring Program.[2]

I use two data sources to realize the approach. Firstly, the controlling of the LUH provided results and characteristics of the exam-takers for all first year exams for the cohorts 2006 until 2010. The data contains overall 10.000 exams taken by over 2.500 students in five classes. Secondly, 459 EM students from the mentoring cohorts filled in a questionnaire asking the students if they have participated in the mentoring and how they perceived their mentors and the program. In the next two subsections I present the descriptive statistics of these two data sets.

5.3.1 Student Characteristics and Outcomes

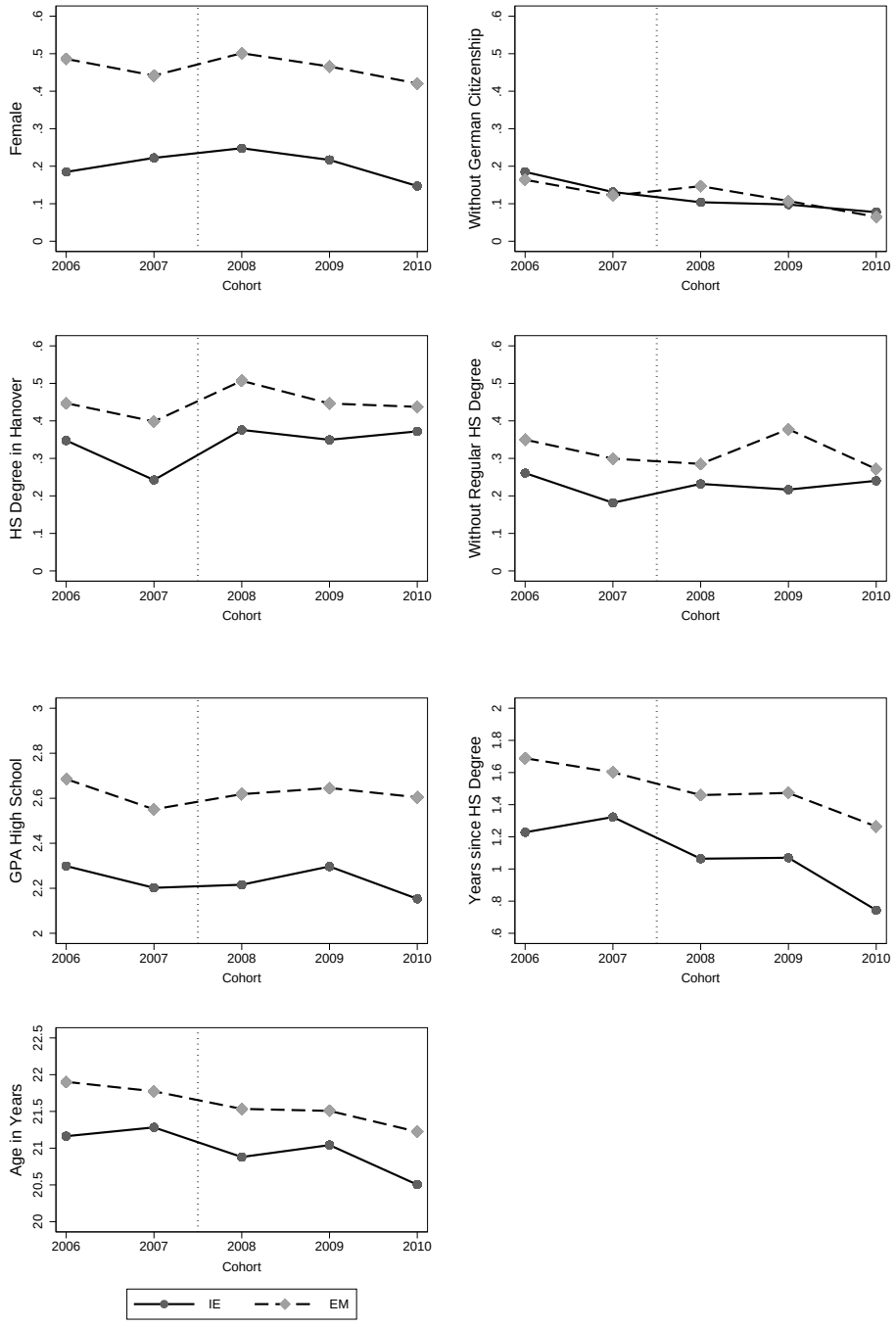
For evaluating the effect of the mentoring on study success I consider three outcomes of interest: The average exam grade, the failure rate and the rate of students who have a certified medical absence from the exam. In Germany the exam grade 1.0 is the best possible result and is equivalent to an A. The next grade is 2.0 which is equivalent to a B. Between 1.0 and 2.0 there are intervals of 1.3 and 1.7 with the same intervals for all other grades. The grade 4.0 for which a student needs at least 50 percent of the available exam points is the lowest grade necessary to pass the exam. All exams with less than 50 percent of the points are failing and marked with a grade of 5.0.[3] If an exam is failed, the student continues with the second attempt in the re-take exam. Since the students are automatically signed up for the exams, getting a medical certification is the only possible way to postpone an exam and continue with the first attempt.

[2] Beltz et al. (2012) use an approach which is methodologically related to mine. They examine the differences change between a business administration and business education program to evaluate how program and course policies affect the effort and performance of students.

[3] The administrative data from the LUH only contains the grades and not the achieved points as exam results.

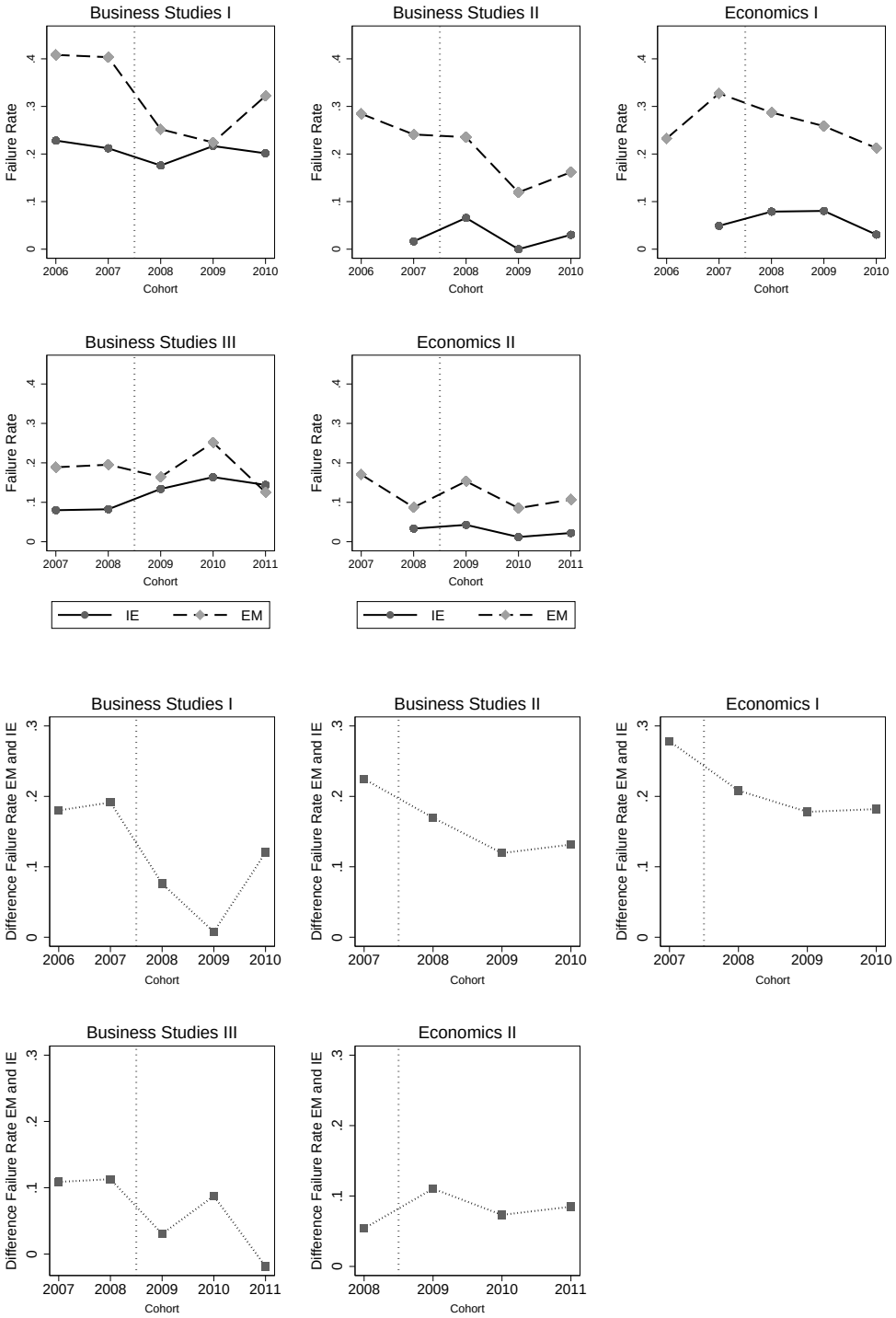
In addition to the exam outcomes the data contains information about the characteristics of each student who took the exam. This information includes gender, citizenship, age and the time between high school graduation of the student and the exam. It also gives information about the grade of his or her secondary education certificate. The secondary school GPA ranges like the exam grades, from 1.0 ("A") to 4.0 ("D") (Between these grades decimal intervals are possible). Finally, the data contains information in which municipality and at which type of school the student achieved her secondary education certificate. Besides a regular "Gymnasium" (Grammerschool) students can also attain their secondary education certificate by a "Gesamtschule" (Comprehensive School), at a job training program or due to a special examination.

Figure 5.2: First Semester Student Characteristics



Note: All figures report the means of first semester students in IE (solid line) and EM (dashed line). The vertical line indicates the introduction of the mentoring program. High School GPA ranges from 1.0 to 4.0 where 1.0 is the best grade. HS=High School.

Figure 5.3: Failure Rates



Note: The figures in the first two rows report the failure rates of IE (solid line) and EM (dashed line) students. The figures in the last two rows report the difference of the failure rates between IE (solid line) and EM (dashed line) students. The vertical line indicates the introduction of the mentoring program.

Figure 5.2 presents the first semester students' characteristics for each cohort. The comparison of IE and EM students reveals differences in several characteristics. The most obvious differences are the lower percentage of female students in IE and the better high school GPA of IE students. The first might be related to the higher technical demands of the IE program which might be less attractive for women. The HS-GPA is higher in the IE program because places are more limited in this program. Many students in both IE and EM earned their high school diploma in Hanover which is typical for a local German university. The rate is lower for the IE students which shows a higher supra-regional attractiveness of this program. In the EM program the percentage of students with a regular high school degree is lower, whereas age and duration since HS-degree is higher. These factors indicate that more career changers and students with heterogeneous career paths choose EM rather than IE. In both programs the rate of students without German citizenship is low and decreases over the years. One reason could be the strict examination regulations at the EM faculty. This regulation might cause international students to choose other universities than Hanover because these students might be more flexible in their location preference.

Overall, the graphs demonstrate that the characteristic differences between EM and IE students just slightly varies by cohort. Additionally to the impression of Figure 5.2, Table 5.10, 5.11, 5.12, 5.13 and 5.14 in the Appendix D give a more detailed descriptive overview of the characteristic means for each class. The last column of each table presents the characteristic difference of the differences before and after introduction of the mentoring. The results of the tables confirm that changes in the characteristic differences are very small and negligible.

The figures in the first two rows in Figure 5.3 report a descriptive overview about the failure rates of EM and IE students in all five classes. The figure includes all first attempt exams by a student in the regular semester of the exam. For example, exams of students who postponed the regular exam and the re-exam by a medical absence certificate are

not included. The IE students have a lower fail rate in 20 of the 22 exams. This is not surprising considering the characteristics of the IE students which are supposed to be correlated with higher study success. In Business Studies I the failure rate for both the IE and the EM students in any exam is always higher than 20 percent. This shows that failing the exam is not only a serious problem for the EM students, but for the better performing IE students as well. Also in Business Studies III at least eight percent of the IE students fail. In the Economics I and Business Studies II exams the difference in the results between IE and EM students is higher than in Business Studies I. This might be caused in large part by the fact that IE students write these exams in their third semester.

Figure 5.6 in the Appendix D reveals that IE students have better grades than the EM students in all 22 exams. Like in the failure difference, the grade difference is highest for the Economics I and Business Studies II exams. Figure 5.7 illustrates that EM students are more often medically excused than IE students. This is not surprising because EM students are in higher risk to fail an exam and therefore these students may benefit more from postponing an exam. Especially for EM students, the rate is higher in the second semester exams than in the first semester exams which might be a lagged behavior modification to the strict examination rule.

The figures in the last two rows in Figure 5.3, 5.6 and 5.7 show the development of the differences between EM and IE students over time. Analysis of the failure rate reveals that the difference has been smaller in all first semester exams after introduction of the mentoring. This is also true for the Business Studies III exam in the second semester. These results are suggestive of a positive impact of the mentoring on the failure rate. However, in the next section I use statistical analyses to establish whether this relationship is significant and holds up to the inclusion of control variables.

5.3.2 Utilization of the Mentoring

In addition to the data of the LUH-controlling, I use data from a questionnaire in which EM students were asked how they used and how they perceived the mentoring program. The questionnaires were given to students in the beginning of several randomly chosen lectures from June to August 2012. More than 400 EM students from the 2009 to 2011 cohorts filled in the questionnaire. However, this sample may not be representative of the students who started EM and received the mentoring in these cohorts because only students who are still enrolled and attend the lectures were reached by the described data collecting procedure.

Table 5.3 reports the descriptive outcomes of the survey. The questionnaire contains four groups of questions: The first set of questions investigates the number of appointments, the relation to the mentor and the perceived effectiveness of the mentoring program. Secondly, the questionnaire contains questions about the student characteristics, including the characteristics which are available from LUH-Controlling. Thirdly, the students are asked for their family's educational background. In the last set of questions, the students give information about their study performance and failed exams in the first year.

The first column in Table 5.3 shows the results for the students of all cohorts. The next columns report the results separated for the cohorts 2008, 2009 and 2010. The first row describes that in each year about 90 percent of the students had at least one appointment with their mentor and participated in the mentoring program. On average the number of appointments was approximately two in each year. This is less than expected since the mentoring calendar schedules three compulsory appointments. The explanation for this gap might be that the questionnaire only asks for personal appointments. It is likely that a third appointment is substituted by an e-mail contact or an informal meeting and therefore it is not documented in the questionnaire. Analyzing the students' perception of the mentoring program reveals that the students judge the characteristics of the

mentor best. In contrast, they are most critical with its usefulness.

A comparison between the survey sample and the first year EM students reveals that the survey sample contains fewer characteristics which are supposed to be related to failure and bad performance. It is likely that students with these characteristics are exmatriculated or they do not attend the lectures. The percentage of female students is similar between the data sets despite some differences between the cohorts. In addition to the characteristics which are available in the exam data set, I obtained information about the educational background of the student's family. This information reveals that the majority of students have a non-academic family background. These students especially might get less advice from their family than necessary for a successful university study. Mentoring could fill this gap. In the survey sample the failure rates in any exam are lower than in the actual first year exams. This lower failure rate also shows both, the selectivity of the sample and that students who fail an exam in the first semester are very likely to drop out or abstain from the lectures. Unfortunately, it was not possible to obtain data of students who dropped out or do not attend the lectures.

Table 5.3: Descriptives of the Mentoring Questionnaire

			Cohort					
	Overall		2008		2009		2010	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mentoring Utilization								
At Least one Appointment	0.91	0.3	0.90	0.4	0.89	0.3	0.92	0.3
Number of Appointments	2.08	0.8	1.87	0.7	2.01	0.8	2.28	0.7
Mentoring Perception								
Relation to Mentor	2.72	1.0	2.66	0.9	2.68	0.9	2.77	1.0
Usefulness of Mentoring	4.11	0.9	4.01	0.9	4.15	0.9	4.13	0.9
Overall Benefit	3.60	1.2	3.47	1.2	3.62	1.2	3.61	1.2
Overall Grade	3.56	1.2	3.42	1.2	3.52	1.2	3.62	1.2
Mentor Characteristics								
Mentor Female	0.35	0.5	0.41	0.5	0.33	0.5	0.33	0.5
Mentor Orientation Week Adviser	0.62	0.6	0.63	0.6	0.67	0.7	0.59	0.6
Student Characteristics								
Female	0.51	0.5	0.49	0.5	0.56	0.5	0.48	0.5
HS Degree in Hannover	0.46	0.5	0.50	0.5	0.48	0.5	0.44	0.5
Regular HS Degree	0.80	0.4	0.85	0.4	0.77	0.4	0.80	0.4
Foreign Citizenship	0.06	0.2	0.09	0.3	0.05	0.2	0.04	0.2
HS GPA	2.47	0.5	2.41	0.5	2.44	0.5	2.53	0.5
HS GPA lowest Decil	0.09	0.3	0.07	0.2	0.07	0.3	0.12	0.3
Age in Years	23.11	1.5	23.99	1.2	23.26	1.3	22.46	1.5
Family Characteristics								
Father College Degree	0.38	0.5	0.45	0.5	0.33	0.5	0.38	0.5
Mother College Degree	0.26	0.4	0.34	0.5	0.21	0.4	0.27	0.4
No Siblings at College	0.55	0.5	0.51	0.5	0.53	0.5	0.59	0.5
Failure Rate								
Business Studies I	0.08	0.3	0.08	0.3	0.07	0.2	0.09	0.3
Business Studies II	0.02	0.1	0.02	0.1	0.01	0.1	0.02	0.1
Economics I	0.06	0.2	0.05	0.2	0.08	0.3	0.06	0.2
Business Studies III	0.06	0.2	0.06	0.2	0.11	0.3	0.03	0.2
Economics II	0.03	0.2	0.05	0.2	0.03	0.2	0.03	0.2
At least one Fail	0.20	0.4	0.22	0.4	0.25	0.4	0.14	0.4
Observations	459		111		152		187	

Notes: *Relation to Mentor* is an index of four questions concerning identification, mutual trust, competency and cooperativeness. The scale reaches from 1 to 5 with 1 indicating the best relation. *Usefulness of Mentoring* is an index of four questions which ask if the mentoring motivated to study, improved grades, reduced isolation and helped to pass the exams. The scale reaches from 1 to 5 with 1 indicating most useful. The scale for *Overall Benefit* reaches from 1 to 5 with 1 indicating the greatest benefit. The scale for *Overall Grade* reaches from 1 to 6 with 1=very good and 6=insufficient. High School GPA ranges from 1.0 to 4.0 where 1.0 is the best grade.

Table 5.4 presents the results of multivariate regression analysis which investigate factors influencing the use and perception of the mentoring program. In the first column the dependent variable indicates whether the student had at least one appointment with the mentor. The analysis reveals that foreign citizenship and a failed exam are significantly related to refusing participation in the program, whereas a female mentor increases the participation rate.

In the second column the dependent variable is the number of appointments between mentor and mentee. Female students use the mentoring more often than their male counterparts. This is in line with results of Angrist et al. (2009) who even find a higher gender difference. Nevertheless, in the Hanover Mentoring the difference is also relevant with about 10 percent more appointments for women. Another finding is that foreigners do not only refuse participation more often, but also take a second or third appointment less often if they participate. Furthermore, students from the 2010 cohort use the mentoring more often than the previous cohorts. Since there is no change in the structure of the program, the more frequent appointments could be partly explained by a lower recall bias. Therefore, it might be that the real number of appointments in the years before was also higher. Finally, students with a lower high school GPA use the mentoring less often and the number of appointments increase if the mentor was the orientation week adviser.

Although the statements about the number of appointments are not provable, there is no reason to expect that female or foreign students give more or less precise answers than male or local students. Therefore, the differences between the genders and the nationalities are trustworthy, although it is not clear if the total amount of visits is correct. Statements by the mentors about the appointment frequency are weak because there is no report system for the mentoring. However, the comparison with the record keeping of several mentors reveals the same picture. Another validity problem is that students who have failed an exam are underrepresented in the survey compared to failure rates in the first year exams. The same is true for foreign students. As both characteristics

Table 5.4: Factors which Influence Utilization and Perception of the Mentoring

	(1) At least one Appointment	(2) Number of Appointments	(3) Relation to Mentor	(4) Usefulness of Mentoring
Cohort 2009	0.009	0.146	0.009	0.087
	(0.038)	(0.103)	(0.127)	(0.124)
Cohort 2010	0.026	0.447***	0.091	0.062
	(0.038)	(0.103)	(0.131)	(0.127)
Female	0.043	0.175**	0.018	0.045
	(0.027)	(0.076)	(0.099)	(0.096)
HS Degree in Hannover	-0.009	-0.107	-0.067	0.029
	(0.026)	(0.076)	(0.098)	(0.091)
Regular HS Degree	-0.016	-0.070	0.019	0.150
	(0.031)	(0.101)	(0.122)	(0.124)
Age in Years	-0.001	0.002	-0.035	-0.005
	(0.007)	(0.022)	(0.028)	(0.025)
Foreign Citizenship	-0.165*	-0.596***	-0.090	-0.222
	(0.094)	(0.205)	(0.275)	(0.316)
HS GPA	0.042	-0.166*	-0.175	-0.055
	(0.035)	(0.091)	(0.118)	(0.113)
HS GPA lowest Decil	-0.072	0.256	0.244	0.131
	(0.071)	(0.182)	(0.213)	(0.203)
At least one Failure	-0.078*	0.150	0.033	-0.213*
	(0.043)	(0.109)	(0.137)	(0.125)
No Academic Background	-0.016	0.110	-0.054	-0.066
	(0.029)	(0.079)	(0.100)	(0.099)
Mentor Female	0.054**	0.079	0.015	-0.058
	(0.027)	(0.077)	(0.104)	(0.100)
Mentor Orientation Week Adviser	-0.039	0.113*	-0.144*	-0.141*
	(0.027)	(0.063)	(0.084)	(0.079)
Observations	426	388	395	390
R^2	0.06	0.12	0.03	0.03

Notes: All models report OLS estimates. Robust standard errors in parenthesis. *At least one appointment* is a binary variable which gets 1 if at least one appointment between mentor and mentee was conducted. Models 2-4 only include students with at least one appointment. Relation to Mentor is an index of four questions concerning identification, mutual trust, competency, cooperativeness. The scale reaches from 1 to 5 with 1 indicating the best relation. Usefulness of Mentoring is an index of four questions which ask whether the mentoring motivated to study, improved grades, reduced isolation and helped to pass the exams. The scale reaches from 1 to 5 with 1 indicating most useful. No Academic Background is 1 if parents and siblings without college education. High School GPA ranges from 1.0 to 4.0 where 1.0 is the best grade.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

are correlated to non participation, it might be that the actual participation rate of all enrolled first semester students is smaller than 90 percent.

The perceived usefulness of the mentoring and the relation to the mentor are the dependent variables in columns three and four. Like the number of appointments, both increase if the mentor is the orientation week adviser. This is in line with the expectations that a longer and more intensive relationship between mentee and mentor has a positive effect. Therefore, it is likely that also the other features of the mentoring like compulsory appointments, a low mentor-to-mentee ratio and graduate mentors are recognized and positively valued by the students. Furthermore, students who have failed an exam rate the mentoring more useful. This indicates that bad performing students are more reliant on advice and, therefore, these students benefit most of the mentoring. However, it also indicates that the overall value of the usefulness of the mentoring might be downward biased because of the sample selectivity.

5.4 Effects of the Mentoring Program on Exam Outcomes

5.4.1 Evaluation Framework

I use the model in Equation 5.1 to test the first descriptive impression from the previous sections more formally:

$$Y_{it} = \alpha + \beta_1 Mentoring_t + \beta_2 EM_i + \tau(Mentoring_t \times EM_i) + \gamma X_{it} + \epsilon_{it} \tag{5.1}$$

Y_{it} is student i's exam outcome in year t (t = 2006, 2007, 2008, 2009, 2010), while $Mentoring_t$ and EM_i are dummy variables indicating the year of examination (1, if the exam was taken after the implementation of the mentoring in 2008 or later) and the students' program (1, if EM), X_{it} denotes a set of control variables, and ϵ_{it} an idiosyncratic error term. In this model, β_1 captures the time-specific variation affecting students in both programs, β_2 the time invariant group-specific effects ,τ, the effect of the reform,

and the vector γ, the effects of the control variables. Besides year dummies, I use the above mentioned variables, gender, HS-degree in Hanover, HS-degree at a "Gymnasium", citizenship, HS-GPA, duration since HS-degree and age in years in order to account for further differences between the groups and to improve the precision of my estimates.[4]

Considering the design of the program, I expect that the mentoring has the biggest effect on the failure rate. In line with the failure rate the grades will improve, too. However, it is also possible that students in the higher range of the grade distribution benefit from the mentoring. In addition, the mentoring could reduce the rate of students who are absent from the exams because the mentors could relax students who are scared of the exams. Nevertheless, it might also be that the mentees have additional information about examination regulations due to the mentoring and therefore they use the possibility of a medical certificate more often.

5.4.2 Results Main Exams

I start my analysis with the failure rate in the first semester exams. Here the main focus lies on the Business Studies I exam which is written by both the EM and IE students in the first semester. Table 5.5 shows the estimates of Equation 5.1 for this exam. The model in Column 1 only includes year dummies as controls, whereas the model in Column 2 includes the whole set of controls. The coefficient in the first row of Column 1 shows that the difference in the failure rate between EM and IE students decreases by 12.0 percentage points after the implementation of the mentoring. Including control variables increase this effect to 15.4 percentage points. The estimations for the other two first semester exams (Economics I and Business Studies II) confirm the results. However, the magnitude of the effects are smaller than in Business Studies I. The difference in the

[4]For the outcomes *Failure* and *Absence* I estimate linear probability models. I replicated the analysis using a Probit model and find very similar results. For the outcome variable *Grades* I use OLS models. I replicated the analysis using a Tobit model since the grade 5.0 is the corner for all exams with less than 50 percent points. The Tobit estimates lead to larger coefficients, but do not change significance levels.

failure rate decreases by nine percentage points after implementation of the mentoring in both classes. However, in these exams only one pre-mentoring period is available because IE students take these exams in their third semester. This makes the estimations less reliable than in the Business Studies I exam.

Table 5.5: Mentoring Effects on First Semester Failure Rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Failure		Failure		Failure	
	Business Studies I		Economics I		Business Studies II	
Mentoring × EM	-0.120***	-0.154***	-0.090**	-0.097**	-0.086***	-0.087***
	(0.042)	(0.041)	(0.041)	(0.043)	(0.031)	(0.027)
EM	0.186***	0.165***	0.278***	0.212***	0.225***	0.220***
	(0.035)	(0.035)	(0.036)	(0.040)	(0.027)	(0.025)
Female		0.008		0.080***		-0.045***
		(0.019)		(0.020)		(0.017)
HS Degree in Han.		-0.007		0.017		0.009
		(0.018)		(0.019)		(0.017)
Regular HS Degree		-0.053**		-0.073***		-0.028
		(0.023)		(0.024)		(0.020)
Foreign Citizenship		0.150***		0.141***		0.158***
		(0.037)		(0.039)		(0.036)
HS GPA		0.105***		0.111***		0.041**
		(0.019)		(0.021)		(0.019)
Durat. since HS Deg.		-0.027***		-0.003		0.000
		(0.009)		(0.009)		(0.009)
Age in Years		0.026***		0.008		0.015**
		(0.008)		(0.008)		(0.008)
R^2	0.03	0.07	0.04	0.09	0.04	0.09
Observations	2529	2379	1902	1814	1905	1816

Notes: Robust standard errors in parentheses. All models include year dummies. Han.=Hannover, HS=High School; Durat.= Duration, Deg.=Degree
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The program variable and the other control variables show the expected influence on the failure rate. The coefficient of the program variable in the second row reveals that IE students fail less often than EM students. The failure rate is positively correlated with a higher HS-GPA, with the age of the student and with foreign citizenship. In contrast,

a HS-degree at a "Gymnasium" and a short period between HS-degree and study start decreases the probability to fail the exam. Gender of the student is the only variable which does not influence the failure rate in the same direction for all exams. Female students fail more often in Economics I and less often in Business Studies II.

Table 5.6: Mentoring Effects on Second Semester Failure Rates

	(1)	(2)	(3)	(4)
	Failure Business Studies III		Failure Economics II	
Mentoring × EM	-0.077**	-0.065*	0.034	0.026
	(0.035)	(0.035)	(0.032)	(0.033)
EM	0.111***	0.083***	0.054*	0.046
	(0.027)	(0.028)	(0.028)	(0.032)
Female		-0.082***		-0.025
		(0.017)		(0.016)
HS Degree in Hannover		-0.002		0.005
		(0.017)		(0.015)
Regular HS Degree		-0.067***		-0.035*
		(0.022)		(0.019)
Foreign Citizenship		-0.003		-0.035
		(0.032)		(0.026)
HS GPA		0.101***		0.061***
		(0.018)		(0.016)
Duration since HS Degree		-0.031***		-0.018**
		(0.009)		(0.008)
Age in Years		0.031***		0.022***
		(0.007)		(0.008)
R^2	0.01	0.08	0.02	0.05
Observations	2026	1920	1555	1496

Notes: Robust standard errors in parentheses. All models include year dummies. HS=High School
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, I examine the effect of the mentoring on the failure rate in the second semester final exams. In these exams the mentoring could have an effect because fewer students fail in the first semester and therefore less students have to take the re-take exam in the second semester. Thus, the students have more time to prepare for the second semester final exams. Additionally, the acquired non-academic skills could have an effect on the results. However, after introduction of the mentoring program students with lower abilities

could reach the second semester exams and therefore the results could be worse after implementation of the mentoring. However, the estimations in Table 5.6 reveal that the mentoring also decreases the failure rate in the Business Studies III by 6.5 percent points. In the Economics II exam there is no effect on failure. The controls have similar signs as in the first semester exams, despite the fact that in both exams female students fail less often.

The Tables 5.15 and 5.16 in the Appendix D examine the effects on grades. In Business Studies I the difference in the grades also decreases significantly after introduction of the mentoring. Again, including the control variables increases this effect. The grades of the other first semester exams improve, however not significantly. The grades are mainly influenced by the same factors as the failure rate. The grades in Business Studies III are also positively affected by the mentoring, although the effect gets statistically insignificant if controls are included. The grades in Economics II decrease significantly after introduction of the mentoring. However, the comparison group are fourth semester IE students and only one cohort before introduction of the mentoring is available.

Finally, the rate of absent students is not affected by the introduction of the mentoring (see Tables 5.17 and 5.18 in Appendix D). Absence because of illness is supposed to be a random event and therefore absent rates should be slightly correlated with any characteristic. However, foreign citizenship and HS-degree in Hanover are strongly positively correlated with being absent. For foreign students, postponing exams could be rational if they have language difficulties which they hope to overcome fast. For students with a HS-degree in Hanover the rate could be higher because they already know doctors who easily testify a medical certificate. In the Business Studies III exam the mentoring also slightly increases the probability of students to be absent.

Overall, in four out of five exams the mentoring decreased the failure rate difference between EM and IE students. Effects are strongest in first semester exams when the mentoring takes place. Especially in the Business Studies I, exam the mentoring almost

equalizes the difference in the failure rate between EM and IE students. The effect in percentage points is not that strong for the other two first semester exams. However, looking at the percentage failure reduction reveals that in all exams the decrease range is between 27 and 37 percent. In line with the expectations the results are more mixed in the second semester. Nevertheless, there is still a significant decrease of the failure rate in the Business Studies III exam which is the exam in the focus of investigation at the second semester. However, the failure rate in Economics II is not positive and the grades even decline.

5.4.3 Heterogeneous Effects

This section analyzes whether certain subgroups benefit more from the mentoring program. The main focus lies on the question whether the program has different effects on men and women. A stronger effect for women would be in line with findings in other studies (Angrist et al., 2009) as well as with the higher use of the mentoring by female students. Additionally, I investigate the effects on students with a HS-GPA in the lowest quartile because poor performance at high school may be related to a non-academic family background for which the mentoring is expected to be beneficial. Finally, I investigate the effect on students without a German citizenship. For these students, I expect smaller effects of the mentoring because of their lower participation rates.

Table 5.7: Heterogeneous Mentoring Effects on Failure Rate

	(1)	(2)	(3)	(4)	(5)
	Failure Rate				
	Busin. St. I	Busin. St. II	Econ. I	Busin. St. III	Econ. II
	Heterogeneous Effects: Gender				
Mentoring × EM	-0.140***	-0.046	-0.103**	-0.020	0.028
	(0.048)	(0.036)	(0.049)	(0.044)	(0.037)
Mentoring × EM	-0.040	-0.063	0.061	-0.126*	0.050
× Female	(0.101)	(0.056)	(0.122)	(0.074)	(0.108)
Observations	2379	1816	1814	1920	1496
	Heterogeneous Effects: Lowest Quartile				
Mentoring × EM	-0.095**	-0.076***	-0.052	-0.042	0.025
	(0.047)	(0.029)	(0.050)	(0.037)	(0.035)
Mentoring × EM	-0.201**	-0.034	-0.144	-0.086	-0.002
× Lowest Quartile	(0.095)	(0.067)	(0.095)	(0.090)	(0.093)
Observations	2379	1816	1814	1920	1496
	Heterogeneous Effects: Foreign Citizenship				
Mentoring × EM	-0.171***	-0.090***	-0.121***	-0.046	0.023
	(0.042)	(0.027)	(0.042)	(0.036)	(0.037)
Mentoring × EM	0.200	0.124	0.330*	-0.244*	0.065
× Foreign Citizenship	(0.174)	(0.108)	(0.182)	(0.125)	(0.060)
Observations	2379	1816	1814	1920	1496

Notes: Robust standard errors in parentheses. All models include year dummies, all available controls and interactions of the subgroup variable with the program and the mentoring variable. All Models use failure rate as dependent variable. Busin. St. = Business Studies, Econ.= Economics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.7 shows the mentoring effects on heterogeneous subgroups. The table presents only the effects on the failure rate. The effects on the other exam outcomes are comparable to the presented effects. These results are available on request. The estimates include all available covariates. However, Table 5.7 only reports the interaction term between EM and the period in which the mentoring was offered and additionally the interaction of this term with gender, low HS-GPA or foreign citizenship.

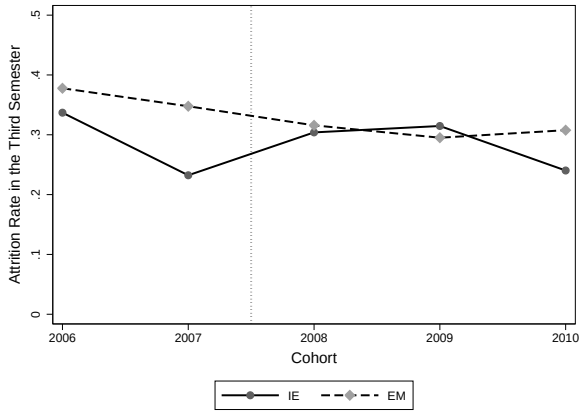
For gender the estimates reveal that only for Business Studies III the effect for women are significantly higher than for men. Nevertheless, in the two business exams in the first semester the effect is also stronger for women but not significantly. It seems that at least for the business exams the women benefit more from the mentoring. However, the gender effects are lower than expected. For students with a HS-GPA in the lowest quartile the picture is different. Although a significantly stronger effect only occurs at Business Studies I, the coefficients for all other exams show a decrease, too. For students with a foreign citizenship the results are as expected. In four of the five exams the effect of the mentoring is lower for these students. However, the effect is only statistically significant in Economics I. Only in Business Studies III in the second semester the effect of the mentoring is higher for students without a German citizenship.

Overall, the coefficients in the three subgroups have mostly the expected signs. Probably most interesting is the fact that in line with the lower participation in the mentoring the foreign students benefit less of the mentoring. This is another evidence for effectiveness of the mentoring if students participate. Nevertheless, foreign students are a small subgroup leading to large standard errors of the estimates and although the size of the coefficients is large they lack statistical significance.

5.4.4 Effects on Attrition

This section analyzes the effects of the mentoring on students' attrition. I consider attrition rates for the second (nine months after the start of the first semester) and third semester (15 months after the start of the first semester) exams. In order to calculate the attrition rate, the first semester exam gives the number of students who start their degree. Because all students are automatically signed up for the exams, the number of drop out students results from the difference between the students who are signed for the exams in the first and in the later semesters. Like in the other analyses, the difference in the differences between EM and IE students since the introduction of the mentoring gives

Figure 5.4: Attrition Third Semester



Note: The figure reports the rate of first semester students which are not enrolled for the exams in the third semester.

the treatment effect. However, in this estimation I can not control for the characteristics because I do not have any information about the students who drop out, but only about the total number.

Table 5.8: Mentoring Effects on Attrition

	(1) Attrition second Sem.	(2) Attrition third Sem.
Mentoring $\times EM$	0.003	-0.061
	(0.037)	(0.045)
EM	0.086***	0.080**
	(0.031)	(0.037)
Constant	0.168***	0.306***
	(0.031)	(0.037)
Observations	2529	2529
R^2	0.01	0.01

Notes: Robust standard errors in parentheses. All models include year dummies. The dependent variable is the rate of first semester student who are not enrolled for the exams in the second or third semester.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 5.4 illustrates the share of students who drop out before the third semester exam. It reveals that the attrition rate is between 25 percent and 40 percent in all cohorts

and in both degrees. Table 5.8 shows the results of the mentoring on attrition using Equation 5.1. In this estimation the dependent variable is 1 if a student dropped out. The coefficient in the first row presents the mentoring effect. For the second semester exams, the difference in the attrition rate between EM and IE remains constant after implementation of the mentoring. However, the difference decreases by 6.1 percentage points in the third semester. Nevertheless, this effect is not statistically significant. The second row shows that EM students drop out more often than IE students which is also visible in Figure 5.4.

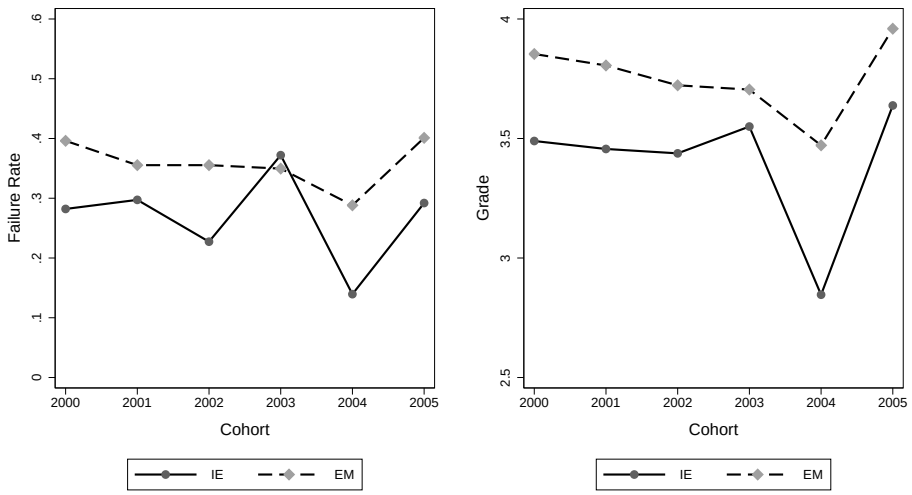
It is surprising that the mentoring does not affect attrition rates although it reduces the failure rates. However, the reduction in the third semester could be an indicator that an effect on attrition develops with time. A reason for this might be that students with low interest in EM drop out in the first year independent whether or not they have failed exams in the first year. In contrast, students who drop out in later semesters do this because they have failed exams in the first year and the pressure of re-take and final exams get too high. This pressure could be reduced due to the lower fail rates in the first year and therefore effects on attrition start to occur in the second year. Interestingly, for the third semester the results are as high as in Bettinger and Baker (2011) where the mentoring effects on attrition also lie in a range between five and seven percentage points.

5.5 Robustness Checks

To interpret the estimated effects as causal, certain assumptions must hold. Firstly, I assume that two cohorts before the introduction of the mentoring and three cohorts after the introduction are enough time periods to show that a change between the exam outcomes of IE and EM students is not random. To proof that a decrease in the difference between IE and EM students is unusual, the results of the diploma programs can give further insights. Table 5.5 indicates that the differences between EM and IE are quite stable over the years in the diploma exam outcomes. Only in 2003 the difference in

the failure rate and in the grades decreases strongly. However, in this year the EM students did not suddenly improve, rather the IE students did worse. This is a different picture than after introduction of the mentoring when the difference decreases because of an improvement of the EM students. Therefore, the diploma results present another evidence that the equalization in the exam outcomes between IE and EM students is not random, but the mentoring caused the change. Table 5.19 gives the numerical values of the diploma exam outcomes.

Figure 5.5: Differences between EM and IE in Diploma Program



Note: The figure reports the failure rate and the grades for IE and EM diploma student.

Secondly, I assume that the other classes in the syllabus of the IE and EM degree have not changed before and after the introduction of the mentoring. If this was the case, the students could reallocate their time resources for the classes and this could cause the equalization in the exam outcomes. This assumption concerns Mathematics I and II for the EM students and the engineering classes for the IE students. In none of the exams I find a change in the curriculum, the structure of the exam or in the lecturers. Therefore,

it is unlikely that the IE students put more effort in the engineering classes and, because of this, less effort in an economics and management class. The same holds for the EM students in mathematics.

Thirdly, I assume no further support than the mentoring for the EM students and no decrease in the study situation for the IE students. As I described above, all additional policies which the faculty introduced with the money of the tuition are open for all students. It is unlikely that the study situation decreased at the engineering faculty because budget allocation to the faculty has not changed. Furthermore, one could argue that the exam correctors favor EM students in the economics and management classes. It is unlikely that this is the case because the correction is almost anonymous. In addition, it is most likely, if the correctors have preferences for EM students, that these preferences also persisted before the introduction of the mentoring.

Fourthly, differences should not diverge as long as the mentoring is offered to EM students. Furthermore, each cohort after introduction of the mentoring should benefit by the mentoring in the same size. Looking at the descriptive statistics, this is most arguable for the Business Studies I exam in the 2010 cohort. In this year the difference in the failure rate between IE and EM students increases in comparison to the 2008 and 2009 cohort. Therefore, I conduct a placebo tests to investigate if the change in the difference is statistically significant. The first two columns in Table 5.9 compare the difference between the pre-mentoring cohorts 2006 and 2007 with 2010. In Column 1 the effect is not significant. However, including controls (column 2) changes the picture and the difference reduces significantly. Column 3 and column 4 investigate the difference between the 2008 and 2009 cohorts and the 2010 cohort. The estimates reveal that the change in the difference is not significant and goes towards zero if controls are included (column 4). The results indicate that a change in the students characteristics, most likely the high school GPA (2.15 for IE students in 2010 whereas the average for all years is 2.26), causes the smaller mentoring effects for the 2010 cohort.

Table 5.9: Placebo Test Business Studies I

	(1)	(2)	(3)	(4)
	Failure Rate		Failure Rate	
2010 $\times EM$	-0.065	-0.135**	0.081	0.035
	(0.055)	(0.054)	(0.051)	(0.051)
EM	0.186***	0.158***	0.040	0.007
	(0.035)	(0.036)	(0.029)	(0.031)
Controls	No	Yes	No	Yes
Observations	1490	1364	1584	1522
R^2	0.03	0.07	0.01	0.05

Notes: Column (1) compares the cohort 2010 with the pre-mentoring cohorts 2006 and 2007. Column (2) includes the available controls. Both models include year dummies for 2007 and 2010. Column (3) compares the cohort 2010 with the post-mentoring cohorts 2008 and 2009. Column (4) includes available controls. Both models include year dummies for 2009 and 2010. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, I assume that the unobserved differences in the characteristics between IE and EM students are constant over the cohorts. I cannot prove this with the available data which does not include, for example, family background. Nevertheless, the difference of the characteristics which are observed in the data do not change significantly. Therefore, there is no reason for suggesting that this is the case for characteristics which are not observed.

5.6 Conclusion

This paper reports on a natural experiment that investigated the effects of a high quality mentoring program on students' first year study success. The mentoring consists of scheduled, compulsory, face-to-face appointments between the student and a graduated member of the faculty. For the evaluation of the effectiveness, I used the fact that students in an economics and management degree and students in an industrial engineering degree take the same exam, whereas only the economics and management students received the

mentoring. Student questionnaires reveal that the participation rate in the mentoring is very high, although the students take less appointments than scheduled by the structure of the program.

I find that the effects on the first semester exam outcomes are large and significant, while the effects are smaller in the second semester. These effects occur, although the students judge the usefulness of the mentoring as low. Female students take more appointments and they benefit more from the mentoring in some exams. Furthermore, the mentoring lowers the attrition rate after 12 months. However, this effect is small and not statistically significant. This shows that many students drop out, although the mentoring caused more students to pass the first semester exams.

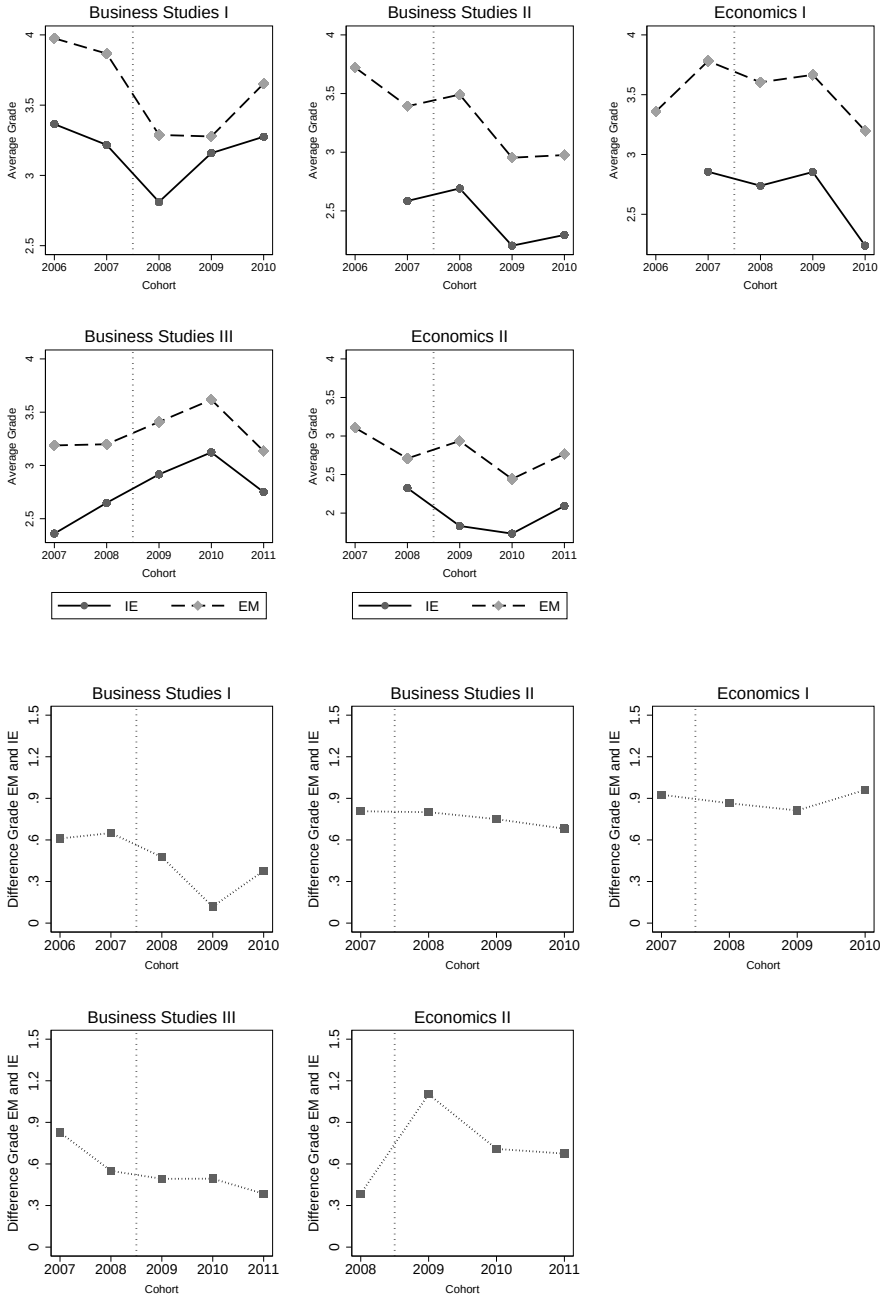
One interpretation of the findings is that students' drop out is independent of whether or not they have passed the first semester exams; rather they realize EM is not the right subject for them. In this case the mentoring is useful as well because students who quit have passed more exams. This could lead to shorter study duration if these students change the subject or they will receive higher salaries if they apply for a job. The strong effects on the failure rate and the small effect on the attrition rate could also be explained due to EM students who failed the exam in the first attempt and then passed the re-take exam before the introduction of the mentoring. After introduction of the mentoring these students pass the exams in the first attempt. Therefore, before the introduction of the mentoring the failure was the nudge which the students needed. However, if this is the case, the mentoring should lead to shorter study durations and better results in later semesters because less time is claimed for the re-take exams and therefore more time is available for the preparation for the final exams.

Overall, it seems that the mentoring can improve the study results, and because the intervention is not costly, it seems more efficient than other policies like financial rewards or class size reductions. Therefore, one could argue that more resources should be allocated from lectures to mentoring programs. However, the analyzed mentoring

program obtained certain high-quality characteristics which are not common to usual mentoring programs. Therefore, further research should investigate which characteristics are most beneficial and which mentoring dosage is optimal.

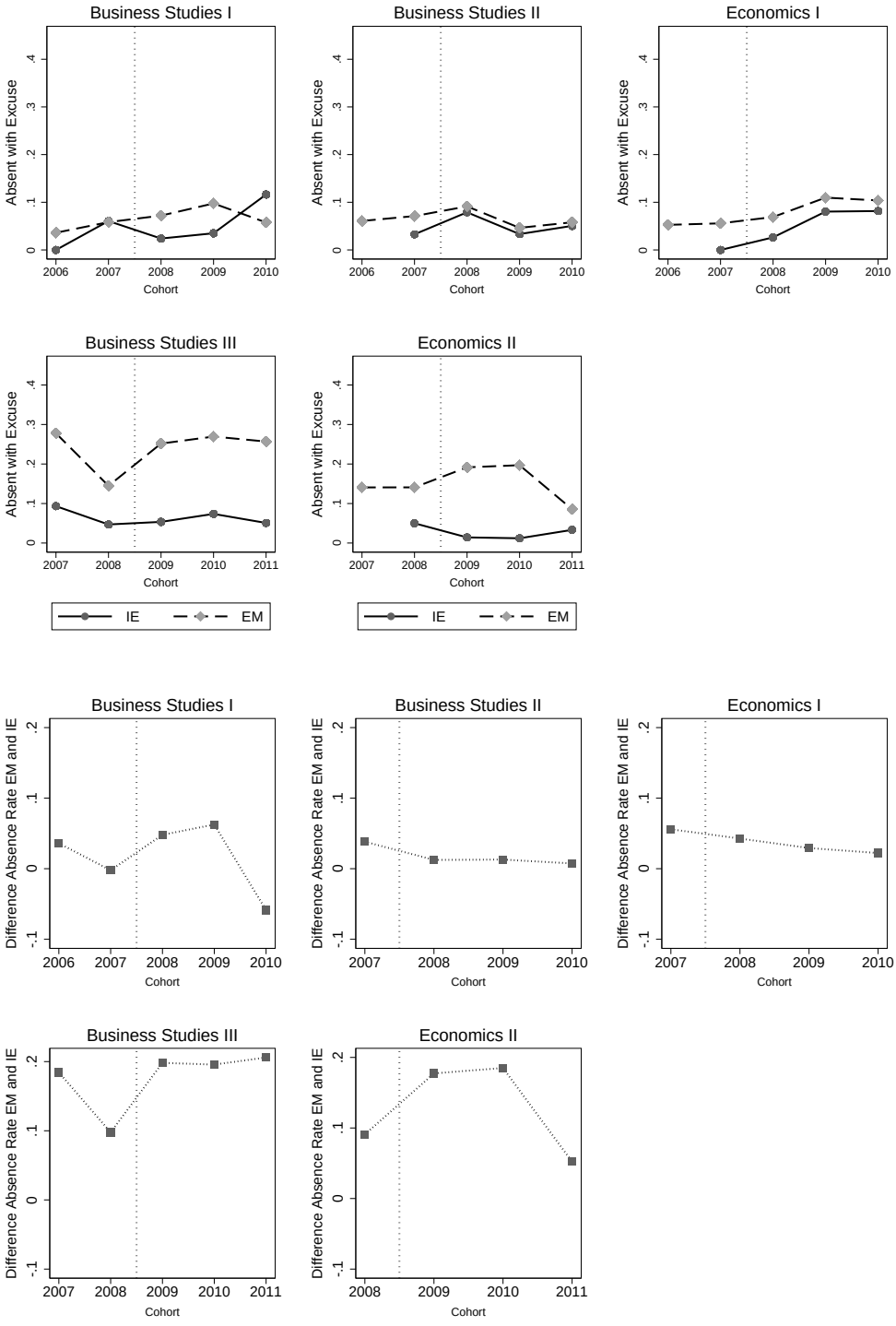
5.7 Appendix D

Figure 5.6: Grades



Note: The figures in the first two rows report the average grades of IE (solid line) and EM (dashed line) students. The figures in the last two rows report the difference of the average grades between IE (solid line) and EM (dashed line) students. The vertical line indicates the introduction of the mentoring program. Grades range from 1.0 to 5.0 where 1.0 is the best grade and 4.0 is lowest grade to pass the exam.

Figure 5.7: Absent with excuse



Note: The figures in the first two rows report the absence rate of IE (solid line) and EM (dashed line) students. The figures in the last two rows report the difference of the absence rate between IE (solid line) and EM (dashed line) students. The vertical line indicates the introduction of the mentoring program.

Table 5.10: Student Characteristics Business Studies I

	Pre-Mentoring				Post-Mentoring										
	2006		2007		2008		2009		2010		Total Pre		Total Post		
	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	Diff.
Exam Results															
Percent Fail	0.23	0.41	0.21	0.40	0.18	0.25	0.22	0.22	0.20	0.32	0.22	0.41	0.20	0.27	-0.12
	(0.42)	(0.49)	(0.41)	(0.49)	(0.38)	(0.43)	(0.41)	(0.42)	(0.40)	(0.47)	(0.42)	(0.49)	(0.40)	(0.44)	
Absent with Excuse	0.00	0.04	0.06	0.06	0.02	0.07	0.03	0.10	0.12	0.06	0.03	0.05	0.06	0.08	0.01
	(0.00)	(0.19)	(0.24)	(0.23)	(0.15)	(0.26)	(0.18)	(0.30)	(0.32)	(0.23)	(0.17)	(0.21)	(0.23)	(0.26)	
Grade	3.37	3.98	3.22	3.87	2.81	3.29	3.16	3.28	3.28	3.65	3.29	3.92	3.08	3.41	-0.30
	(1.05)	(1.03)	(1.19)	(1.16)	(1.26)	(1.27)	(1.20)	(1.20)	(1.17)	(1.15)	(1.12)	(1.10)	(1.22)	(1.22)	
Characteristics															
Female	0.18	0.49	0.22	0.44	0.25	0.50	0.22	0.47	0.15	0.42	0.20	0.46	0.20	0.46	0.00
	(0.39)	(0.50)	(0.42)	(0.50)	(0.43)	(0.50)	(0.41)	(0.50)	(0.36)	(0.49)	(0.40)	(0.50)	(0.40)	(0.50)	
HS Degree in Han.	0.35	0.45	0.24	0.40	0.38	0.51	0.35	0.45	0.37	0.44	0.29	0.42	0.37	0.46	-0.04
	(0.48)	(0.50)	(0.43)	(0.49)	(0.49)	(0.50)	(0.48)	(0.50)	(0.49)	(0.50)	(0.46)	(0.49)	(0.48)	(0.50)	
Regular HS Degree	0.74	0.65	0.82	0.70	0.77	0.71	0.78	0.62	0.76	0.73	0.78	0.68	0.77	0.69	0.02
	(0.44)	(0.48)	(0.39)	(0.46)	(0.42)	(0.45)	(0.41)	(0.49)	(0.43)	(0.45)	(0.42)	(0.47)	(0.42)	(0.46)	
Foreign Citizenship	0.18	0.16	0.13	0.12	0.10	0.15	0.10	0.11	0.08	0.06	0.16	0.14	0.09	0.10	0.03
	(0.39)	(0.37)	(0.34)	(0.33)	(0.31)	(0.35)	(0.30)	(0.31)	(0.27)	(0.25)	(0.36)	(0.35)	(0.29)	(0.31)	
HS GPA	2.30	2.69	2.20	2.55	2.22	2.62	2.30	2.65	2.15	2.60	2.25	2.61	2.22	2.62	0.04
	(0.48)	(0.47)	(0.41)	(0.50)	(0.51)	(0.48)	(0.49)	(0.48)	(0.46)	(0.50)	(0.45)	(0.49)	(0.49)	(0.49)	
Durat. since HS Degree	1.23	1.69	1.32	1.60	1.06	1.46	1.07	1.47	0.74	1.26	1.28	1.64	0.96	1.40	0.08
	(2.27)	(2.81)	(1.97)	(2.80)	(1.78)	(2.13)	(2.33)	(1.91)	(1.09)	(1.88)	(2.11)	(2.80)	(1.83)	(1.97)	
Age in Years	21.16	21.90	21.28	21.77	20.88	21.53	21.04	21.51	20.50	21.22	21.23	21.84	20.82	21.42	-0.01
	(2.78)	(3.14)	(2.15)	(3.45)	(2.03)	(2.48)	(2.61)	(2.17)	(1.70)	(2.31)	(2.47)	(3.31)	(2.17)	(2.32)	
N	92	360	99	394	125	361	143	410	129	416	191	754	397	1187	

Notes: Mean coefficients; standard deviations in parentheses; HS=High School. *Diff.* indicates the difference of the differences between IE and EM in Pre-Mentoring and Post-Mentoring. Grades range from 1.0 to 5.0 where 1.0 is the best grade and 4.0 is lowest grade to pass the exam. High School GPA ranges from 1.0 to 4.0 where 1.0 is the best grade.

Table 5.11: Student Characteristics Business Studies III

	Pre-Mentoring				Post-Mentoring										
	2007		2008		2009		2010		2011		Total Pre		Total Post		
	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	Diff.
Exam Results															
Percent Fail	0.08	0.19	0.08	0.19	0.13	0.16	0.16	0.25	0.15	0.13	0.08	0.19	0.15	0.18	-0.08
	(0.27)	(0.39)	(0.28)	(0.40)	(0.34)	(0.37)	(0.37)	(0.43)	(0.36)	(0.33)	(0.27)	(0.39)	(0.36)	(0.38)	
Absent with Excuse	0.09	0.28	0.05	0.16	0.07	0.26	0.08	0.28	0.07	0.27	0.07	0.21	0.07	0.27	0.06
	(0.29)	(0.45)	(0.21)	(0.36)	(0.26)	(0.44)	(0.27)	(0.45)	(0.25)	(0.44)	(0.25)	(0.41)	(0.26)	(0.44)	
Grade	2.36	3.19	2.65	3.21	2.92	3.41	3.12	3.62	2.78	3.14	2.52	3.20	2.94	3.39	-0.23
	(1.12)	(1.26)	(1.07)	(1.18)	(1.13)	(1.04)	(1.12)	(1.22)	(1.14)	(1.07)	(1.10)	(1.21)	(1.14)	(1.14)	
Characteristics															
Female	0.16	0.48	0.21	0.45	0.26	0.51	0.21	0.46	0.17	0.45	0.19	0.46	0.21	0.47	-0.01
	(0.37)	(0.50)	(0.41)	(0.50)	(0.44)	(0.50)	(0.41)	(0.50)	(0.38)	(0.50)	(0.39)	(0.50)	(0.41)	(0.50)	
HS Degree in Han.	0.37	0.46	0.25	0.41	0.34	0.50	0.34	0.46	0.37	0.44	0.31	0.43	0.35	0.47	0.00
	(0.49)	(0.50)	(0.43)	(0.49)	(0.48)	(0.50)	(0.48)	(0.50)	(0.48)	(0.50)	(0.46)	(0.50)	(0.48)	(0.50)	
Regular HS Degree	0.81	0.66	0.82	0.70	0.77	0.73	0.77	0.63	0.75	0.75	0.82	0.68	0.77	0.70	0.07
	(0.39)	(0.47)	(0.38)	(0.46)	(0.42)	(0.44)	(0.42)	(0.48)	(0.43)	(0.43)	(0.39)	(0.47)	(0.42)	(0.46)	
Foreign Citizenship	0.15	0.16	0.13	0.11	0.11	0.13	0.10	0.09	0.11	0.07	0.14	0.14	0.10	0.10	0.00
	(0.36)	(0.37)	(0.34)	(0.32)	(0.31)	(0.34)	(0.30)	(0.28)	(0.31)	(0.25)	(0.35)	(0.34)	(0.30)	(0.29)	
HS GPA	2.30	2.65	2.18	2.56	2.18	2.60	2.27	2.62	2.15	2.59	2.24	2.60	2.20	2.61	0.05
	(0.50)	(0.49)	(0.42)	(0.46)	(0.49)	(0.48)	(0.50)	(0.47)	(0.48)	(0.50)	(0.46)	(0.47)	(0.49)	(0.48)	
Durat. since HS Degree	1.05	1.51	1.32	1.49	1.17	1.45	1.01	1.37	0.86	1.29	1.19	1.50	1.01	1.37	0.08
	(1.93)	(2.06)	(1.88)	(2.25)	(1.89)	(2.29)	(1.34)	(1.70)	(1.24)	(1.80)	(1.90)	(2.16)	(1.51)	(1.93)	
Age in Years	21.32	22.19	21.60	22.05	21.28	21.88	21.33	21.96	21.08	21.69	21.47	22.12	21.23	21.84	-0.04
	(2.22)	(2.49)	(1.78)	(2.81)	(1.97)	(2.37)	(1.70)	(2.39)	(1.74)	(2.27)	(2.00)	(2.67)	(1.80)	(2.34)	
N	75	270	85	297	112	286	122	334	118	327	160	567	352	947	

Notes: see Table 5.10.

Table 5.12: Student Characteristics Economics I

	Pre-Mentoring		Post-Mentoring										
	2007		2008		2009		2010		Total Pre		Total Post		
	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	Diff.
Exam Results													
Percent Fail	0.05	0.33	0.08	0.29	0.08	0.26	0.03	0.21	0.05	0.33	0.06	0.25	-0.09
	(0.22)	(0.47)	(0.27)	(0.45)	(0.27)	(0.44)	(0.17)	(0.41)	(0.22)	(0.47)	(0.24)	(0.43)	
Absent with Excuse	0.00	0.06	0.03	0.07	0.08	0.11	0.08	0.10	0.00	0.06	0.07	0.10	-0.03
	(0.00)	(0.23)	(0.16)	(0.25)	(0.27)	(0.31)	(0.28)	(0.31)	(0.00)	(0.23)	(0.25)	(0.30)	
Grade	2.86	3.78	2.74	3.60	2.85	3.67	2.24	3.20	2.86	3.78	2.58	3.49	-0.01
	(0.97)	(1.08)	(0.98)	(1.12)	(1.02)	(1.06)	(0.95)	(1.24)	(0.97)	(1.08)	(1.02)	(1.16)	
Characteristics													
Female	0.20	0.44	0.20	0.50	0.29	0.47	0.17	0.42	0.20	0.44	0.22	0.46	0.00
	(0.40)	(0.50)	(0.40)	(0.50)	(0.46)	(0.50)	(0.38)	(0.49)	(0.40)	(0.50)	(0.41)	(0.50)	
HS Degree in Han.	0.41	0.40	0.24	0.51	0.33	0.45	0.37	0.44	0.41	0.40	0.32	0.46	0.15
	(0.50)	(0.49)	(0.43)	(0.50)	(0.47)	(0.50)	(0.48)	(0.50)	(0.50)	(0.49)	(0.47)	(0.50)	
Regular HS Degree	0.82	0.70	0.83	0.71	0.79	0.62	0.80	0.73	0.82	0.70	0.79	0.68	0.01
	(0.39)	(0.46)	(0.38)	(0.45)	(0.41)	(0.49)	(0.41)	(0.45)	(0.39)	(0.46)	(0.40)	(0.47)	
Foreign Citizenship	0.16	0.12	0.13	0.15	0.08	0.11	0.09	0.07	0.16	0.12	0.10	0.11	0.05
	(0.37)	(0.33)	(0.34)	(0.36)	(0.27)	(0.31)	(0.29)	(0.25)	(0.37)	(0.33)	(0.30)	(0.31)	
HS GPA	2.25	2.55	2.18	2.61	2.15	2.65	2.28	2.60	2.25	2.55	2.20	2.62	0.12
	(0.48)	(0.50)	(0.43)	(0.49)	(0.48)	(0.48)	(0.44)	(0.49)	(0.48)	(0.50)	(0.45)	(0.49)	
Durat. since HS Degree	0.66	1.60	1.33	1.50	1.15	1.47	0.92	1.25	0.66	1.60	1.15	1.44	-0.65
	(1.25)	(2.80)	(1.89)	(2.24)	(2.04)	(1.91)	(1.17)	(1.87)	(1.25)	(2.80)	(1.75)	(2.04)	
Age in Years	21.44	21.81	22.18	21.60	21.77	21.55	21.84	21.24	21.44	21.81	21.95	21.52	-0.80
	(1.42)	(3.45)	(1.81)	(2.53)	(2.18)	(2.18)	(1.69)	(2.31)	(1.42)	(3.45)	(1.95)	(2.45)	
Observations	61	394	76	362	87	410	98	414	61	394	261	1186	

Notes: see Table 5.10.

Table 5.13: Student Characteristics Business Studies II

	Pre-Mentoring		Post-Mentoring										
	2007		2008		2009		2010		Total Pre		Total Post		
	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	Diff.
Exam Results													
Percent Fail	0.02	0.24	0.07	0.24	0.00	0.12	0.03	0.16	0.02	0.24	0.03	0.17	-0.08
	(0.13)	(0.43)	(0.25)	(0.42)	(0.00)	(0.32)	(0.17)	(0.37)	(0.13)	(0.43)	(0.18)	(0.38)	
Absent with Excuse	0.03	0.07	0.08	0.09	0.03	0.05	0.05	0.06	0.03	0.07	0.07	0.07	-0.04
	(0.18)	(0.26)	(0.27)	(0.29)	(0.18)	(0.21)	(0.22)	(0.23)	(0.18)	(0.26)	(0.26)	(0.25)	
Grade	2.58	3.39	2.69	3.49	2.20	2.95	2.29	2.98	2.58	3.39	2.38	3.13	-0.06
	(0.80)	(1.18)	(0.91)	(1.09)	(0.72)	(0.99)	(0.93)	(1.20)	(0.80)	(1.18)	(0.89)	(1.13)	
Characteristics													
Female	0.20	0.44	0.20	0.50	0.28	0.47	0.17	0.42	0.20	0.44	0.21	0.46	0.00
	(0.40)	(0.50)	(0.40)	(0.50)	(0.45)	(0.50)	(0.38)	(0.49)	(0.40)	(0.50)	(0.41)	(0.50)	
HS Degree in Han.	0.41	0.40	0.24	0.51	0.34	0.45	0.37	0.44	0.41	0.40	0.32	0.46	0.15
	(0.50)	(0.49)	(0.43)	(0.50)	(0.48)	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)	(0.47)	(0.50)	
Regular HS Degree	0.82	0.70	0.83	0.71	0.79	0.62	0.80	0.73	0.82	0.70	0.78	0.68	0.02
	(0.39)	(0.46)	(0.38)	(0.45)	(0.41)	(0.49)	(0.40)	(0.45)	(0.39)	(0.46)	(0.41)	(0.47)	
Foreign Citizenship	0.16	0.12	0.13	0.15	0.09	0.11	0.09	0.07	0.16	0.12	0.12	0.11	0.03
	(0.37)	(0.33)	(0.34)	(0.35)	(0.29)	(0.31)	(0.29)	(0.25)	(0.37)	(0.33)	(0.33)	(0.31)	
HS GPA	2.25	2.55	2.18	2.62	2.17	2.65	2.27	2.60	2.25	2.55	2.21	2.62	0.11
	(0.48)	(0.50)	(0.43)	(0.48)	(0.50)	(0.48)	(0.45)	(0.49)	(0.48)	(0.50)	(0.46)	(0.49)	
Durat. since HS Degree	0.66	1.60	1.33	1.46	1.23	1.47	0.91	1.25	0.66	1.60	1.24	1.42	-0.76
	(1.25)	(2.80)	(1.89)	(2.13)	(2.10)	(1.91)	(1.16)	(1.87)	(1.25)	(2.80)	(1.84)	(2.00)	
Age in Years	21.44	21.80	22.16	21.56	21.82	21.54	21.82	21.24	21.44	21.80	22.01	21.50	-0.87
	(1.42)	(3.45)	(1.82)	(2.47)	(2.20)	(2.18)	(1.68)	(2.31)	(1.42)	(3.45)	(1.99)	(2.42)	
Observations	61	394	76	361	90	410	99	414	61	394	265	1185	

Notes: see Table 5.10.

Table 5.14: Student Characteristics Economics II

	Pre-Mentoring		Post-Mentoring										
	2008		2009		2010		2011		Total Pre		Total Post		
	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	IE	EM	Diff.
Exam Results													
Percent Fail	0.05	0.09	0.04	0.16	0.01	0.08	0.02	0.11	0.05	0.09	0.02	0.11	0.05
	(0.22)	(0.28)	(0.20)	(0.36)	(0.11)	(0.28)	(0.15)	(0.31)	(0.22)	(0.28)	(0.15)	(0.32)	
Absent with Excuse	0.95	0.85	0.99	0.81	0.99	0.79	0.96	0.90	0.95	0.85	0.98	0.84	0.04
	(0.22)	(0.36)	(0.12)	(0.39)	(0.11)	(0.41)	(0.20)	(0.30)	(0.22)	(0.36)	(0.15)	(0.37)	
Grade	2.37	2.72	1.84	2.95	1.72	2.44	2.11	2.77	2.37	2.72	1.89	2.71	0.47
	(0.88)	(1.01)	(1.03)	(1.29)	(0.71)	(1.20)	(0.90)	(1.17)	(0.88)	(1.01)	(0.89)	(1.23)	
Characteristics													
Female	0.20	0.44	0.21	0.51	0.24	0.46	0.18	0.45	0.20	0.44	0.21	0.47	0.02
	(0.40)	(0.50)	(0.41)	(0.50)	(0.43)	(0.50)	(0.39)	(0.50)	(0.40)	(0.50)	(0.41)	(0.50)	
HS Degree in Han.	0.41	0.41	0.24	0.50	0.35	0.45	0.38	0.44	0.41	0.41	0.33	0.46	0.13
	(0.50)	(0.49)	(0.43)	(0.50)	(0.48)	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)	(0.47)	(0.50)	
Regular HS Degree	0.84	0.69	0.86	0.73	0.85	0.63	0.83	0.75	0.84	0.69	0.84	0.70	0.01
	(0.37)	(0.46)	(0.35)	(0.45)	(0.36)	(0.48)	(0.38)	(0.43)	(0.37)	(0.46)	(0.36)	(0.46)	
Foreign Citizenship	0.16	0.12	0.10	0.14	0.06	0.09	0.09	0.07	0.16	0.12	0.08	0.10	0.06
	(0.37)	(0.33)	(0.30)	(0.34)	(0.24)	(0.29)	(0.28)	(0.26)	(0.37)	(0.33)	(0.27)	(0.30)	
HS GPA	2.23	2.57	2.18	2.61	2.13	2.61	2.25	2.58	2.23	2.57	2.19	2.60	0.07
	(0.47)	(0.47)	(0.43)	(0.49)	(0.46)	(0.48)	(0.43)	(0.50)	(0.47)	(0.47)	(0.44)	(0.49)	
Durat. since HS Degree	0.61	1.49	1.21	1.51	0.99	1.39	0.84	1.27	0.61	1.49	1.00	1.38	-0.50
	(1.26)	(2.28)	(1.64)	(2.49)	(1.67)	(1.72)	(1.09)	(1.78)	(1.26)	(2.28)	(1.47)	(2.01)	
Age in Years	21.79	22.04	22.37	21.90	21.99	21.92	22.17	21.64	21.79	22.04	22.16	21.82	-0.59
	(1.36)	(2.80)	(1.56)	(2.49)	(1.70)	(2.28)	(1.67)	(2.25)	(1.36)	(2.80)	(1.65)	(2.34)	
N	60	298	70	287	83	340	90	327	60	298	243	954	

Notes: see Table 5.10.

Table 5.15: Mentoring Effects on First Semester Grades

	(1)	(2)	(3)	(4)	(5)	(6)
	Grade Business St. I		Grade Economics I		Grade Business St. II	
Mentoring × EM	-0.315***	-0.394***	-0.044	-0.162	-0.069	-0.165
	(0.117)	(0.110)	(0.153)	(0.137)	(0.136)	(0.123)
EM	0.630***	0.458***	0.926***	0.730***	0.807***	0.797***
	(0.092)	(0.091)	(0.135)	(0.124)	(0.121)	(0.115)
Female		0.069		0.221***		-0.322***
		(0.051)		(0.052)		(0.050)
HS Degree in Han.		-0.003		0.031		0.011
		(0.048)		(0.050)		(0.049)
Regular HS Degree		-0.203***		-0.304***		-0.253***
		(0.058)		(0.060)		(0.058)
Foreign Citizenship		0.622***		0.648***		0.670***
		(0.092)		(0.097)		(0.100)
HS GPA		0.549***		0.660***		0.428***
		(0.054)		(0.058)		(0.057)
Durat. since HS Degree		-0.119***		-0.075***		-0.020
		(0.024)		(0.025)		(0.028)
Age in Years		0.082***		0.055***		0.052**
		(0.019)		(0.020)		(0.026)
R^2	0.07	0.16	0.13	0.26	0.11	0.23
Observations	2374	2245	1750	1677	1785	1708

Notes: Standard errors in parentheses. All models include year dummies. Only students who were not absent are included.

Table 5.16: Mentoring Effects on Second Semester Grades

	(1)	(2)	(3)	(4)
	Grade Business Studies III		Grade Economics II	
Mentoring × EM	-0.219*	-0.184	0.425***	0.347**
	(0.130)	(0.123)	(0.144)	(0.140)
EM	0.675***	0.475***	0.385***	0.238*
	(0.106)	(0.105)	(0.124)	(0.125)
Female		-0.279***		-0.069
		(0.057)		(0.064)
HS Degree in Hannover		0.057		0.026
		(0.055)		(0.060)
Regular HS Degree		-0.287***		-0.311***
		(0.071)		(0.073)
Foreign Citizenship		0.410***		-0.188
		(0.120)		(0.117)
HS GPA		0.667***		0.583***
		(0.061)		(0.066)
Duration since HS Degree		-0.129***		-0.054**
		(0.028)		(0.027)
Age in Years		0.098***		0.070***
		(0.018)		(0.021)
R^2	0.07	0.21	0.09	0.18
Observations	1630	1559	1355	1306

Notes: Robust standard errors in parentheses. All models include year dummies. Only students who were not absent are included.

Table 5.17: Mentoring Effects on First Semester Absent Rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Absent Business St. I		Absent Economics I		Absent Business St. II	
Mentoring × EM	0.002	0.000	-0.025	-0.020	-0.027	-0.020
	(0.020)	(0.018)	(0.021)	(0.021)	(0.030)	(0.032)
EM	0.016	0.005	0.056***	0.061***	0.038	0.043
	(0.015)	(0.013)	(0.012)	(0.016)	(0.026)	(0.030)
Female		0.017*		-0.009		-0.001
		(0.010)		(0.013)		(0.012)
HS Degree in Han.		0.030***		0.021*		0.003
		(0.010)		(0.013)		(0.011)
Regular HS Degree		-0.001		-0.012		-0.014
		(0.011)		(0.016)		(0.013)
Foreign Citizenship		0.089***		0.132***		0.116***
		(0.025)		(0.033)		(0.030)
HS GPA		0.017		0.023		-0.005
		(0.012)		(0.014)		(0.013)
Durat. since HS Degree		0.000		-0.007		-0.002
		(0.005)		(0.008)		(0.007)
Age in Years		0.006		0.012**		0.009*
		(0.004)		(0.006)		(0.005)
R^2	0.01	0.03	0.01	0.05	0.01	0.04
Observations	2529	2379	1902	1814	1905	1816

Notes: Robust standard errors in parentheses. All models include year dummies.

Table 5.18: Mentoring Effects on Second Semester Absent Rates

	(1)	(2)	(3)	(4)
	Absent Business Studies III		Absent Economics II	
Mentoring × EM	0.061*	0.058*	0.043	0.041
	(0.032)	(0.031)	(0.038)	(0.040)
EM	0.139***	0.104***	0.091***	0.054
	(0.026)	(0.027)	(0.035)	(0.039)
Female		0.022		0.025
		(0.019)		(0.019)
HS Degree in Hannover		0.024		0.034*
		(0.018)		(0.018)
Regular HS Degree		-0.063***		-0.024
		(0.023)		(0.023)
Foreign Citizenship		0.221***		0.120***
		(0.041)		(0.040)
HS GPA		0.057***		0.074***
		(0.020)		(0.020)
Duration since HS Degree		0.003		0.003
		(0.009)		(0.009)
Age in Years		0.002		0.005
R^2	0.05	0.09	0.03	0.07
Observations	2026	1920	1555	1496

Notes: Robust standard errors in parentheses. All models include year dummies.

Table 5.19: Differences between EM and IE in Diploma Program

Year	N		Failure Rate			Grade		
	EM	IE	EM	IE	Diff.	EM	IE	Diff.
2000	457	39	0.40	0.28	0.12	3.85	3.49	0.36
2001	537	74	0.36	0.30	0.06	3.81	3.46	0.35
2002	574	66	0.36	0.23	0.13	3.72	3.44	0.28
2003	552	86	0.35	0.37	-0.02	3.70	3.55	0.15
2004	583	86	0.29	0.14	0.15	3.47	2.85	0.62
2005	536	113	0.40	0.29	0.11	3.96	3.64	0.32

Note: Diff. reports the difference between EM and IE students.

Figure 5.8: Mentoring Questionnaire

MUSTER

EvaSys	Mentoring Befragung [Copy]	Electric Paper

Markieren Sie so: ☐ ☒ ☐ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
Korrektur: ☐ ■ ☐ ☒ ☐ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

1. An der wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover wird seit dem Jahr 2008 ein Mentoringprogramm für Erstsemester-Studenten angeboten. Zur Verbesserung des Programms würden wir Ihnen hierzu gerne einige Frage stellen. Die Umfrage ist selbstverständlich anonym und freiwillig.

Zunächst möchten wir Ihnen gerne ein paar Fragen zu Ihrer Person stellen:

1.1 Bitte geben Sie ihr Geschlecht an:
☐ Weiblich ☐ Männlich

1.2 Wie alt sind Sie heute in Jahren?

1.3 Besitzen Sie die deutsche Staatsbürgerschaft?
☐ Ja ☐ Nein

1.4 Bitte beantworten Sie folgende Frage: Schon während der Schulzeit war ich mir sicher, dass ich Wiwi studieren möchte? Sehr sicher ☐ ☐ ☐ ☐ ☐ Sehr unsicher

1.5 In welchem Jahr haben Sie ihr Wiwi-Studium begonnen?
☐ 2010 ☐ 2009 ☐ 2008
☐ Früher

1.6 Wo sehen Sie Ihren Studienschwerpunkt?
☐ Eher VWL ☐ Eher BWL

1.7 Wie ist Ihre Hochschulzugangsberechtigungsnote (z.B. Abi-Note, Abschlussnote beim Studienkolleg)?

1.8 Haben Sie Ihre Hochschulzugangsberechtigung an einem regulären Gymnasium in Deutschland erlangt (Ohne Fachgymnasium, Gesamtschule oder Studienkolleg)?
☐ Ja ☐ Nein

1.9 Haben Sie ihre Hochschulzugangsberechtigung in Hannover (Region+Stadt) erlangt?
☐ Ja ☐ Nein

1.10 Haben Sie vor dem Studium eine Ausbildung absolviert?
☐ Ja ☐ Nein

1.11 War das Wiwi-Studium ihr Wunschstudium?
☐ Ja ☐ Nein

1.12 Hat Ihr Vater oder Ihre Mutter die allgemeine Hochschulreife (Abitur) erlangt?
☐ Ja ☐ Nein

1.13 Hat Ihr Vater einen Universitäts- oder Fachhochschulabschluss erlangt?
☐ Ja ☐ Nein

MUSTER

MUSTER

EvaSys	Mentoring Befragung [Copy]	Electric Paper

1. An der wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover wird seit dem Jahr 2008 ein Mentoringprogramm für Erstsemester-Studenten angeboten. Zur Verbesserung des Programms würden wir Ihnen hierzu gerne einige Frage stellen. Die Umfrage ist selbstverständlich anonym und freiwillig. [Fortsetzung]

1.14 Hat Ihre Mutter einen Universitäts- oder Fachhochschulabschluss erlangt?
☐ Ja ☐ Nein

1.15 Besucht eine/r Ihrer Brüder oder Schwestern eine Hochschule oder hat eine/r Ihrer Brüder oder Schwestern eine Hochschule besucht (Universität oder Fachhochschule)?
☐ Ja, studiert noch ☐ Ja, Studium erfolgreich abgeschlossen ☐ Ja, Studium abgebrochen
☐ Nein ☐ Ich habe keine Geschwister

2. Studienergebnisse

2.1 Haben Sie im ersten Semester folgende Klausuren NICHT bestanden (Mehrfachnennungen möglich)?
☐ BWL I ☐ BWL II ☐ VWL I
☐ Mathe I

2.2 Haben Sie im zweiten Semester folgende Klausuren NICHT bestanden (Mehrfachnennungen möglich)?
☐ BWL III ☐ VWL II ☐ Mathe II
☐ Statistik I

2.3 Wie ist Ihr Notendurschnitt heute?
☐ <1,5 ☐ 1,5 - 2,0 ☐ 2,1 - 2,3
☐ 2,4 - 2,5 ☐ 2,6 - 2,9 ☐ 3,0 - 3,5
☐ > 3,5

3. Charakteristika des Mentorings

Nun würden wir gerne wissen, wie Sie das Mentoringprogramm wahrgenommen haben:

3.1 Geschlecht Ihrer/s Mentors/in?
☐ Weiblich ☐ Männlich

3.2 War Ihre Mentor/in Betreuer/in Ihrer O-Phasengruppe?
☐ Ja ☐ Nein ☐ Weiß nicht mehr

3.3 Ist mindestens ein Gesprächstermin zwischen Ihnen und Ihrem Mentor zustandegekommen?
☐ Ja ☐ Nein

3.4 Wie viele Treffen gab es zwischen Ihnen und Ihrem Mentor?
☐ 0 ☐ 1 ☐ 2
☐ 3 ☐ 4 ☐ 5 oder mehr
☐ Weiß ich nicht mehr

3.5 Fand eines der Treffen nach der ersten Klausurenphase statt?
☐ Ja ☐ Nein

3.6 Was wurde bei den Gesprächen thematisiert (Mehrfachnennungen möglich)?
☐ Zeiteinteilung ☐ Ziele des Studiums ☐ Nebenjob
☐ Klausurvorbereitung ☐ Vorklausuren ☐ Tipps für das Studium
☐ Auslandssemester ☐ Lerntechniken

MUSTER

MUSTER

EvaSys	Mentoring Befragung [Copy]	Electric Paper

3. Charakteristika des Mentorings [Fortsetzung]

3.7 Was wurde außerdem bei den Gesprächen thematisiert?

3.8 Was hätte bei den Gesprächen außerdem thematisiert werden sollen?

3.9	War der/die Mentor/in hilfsbereit?	Sehr hilfsbereit	☐	☐	☐	☐	☐	Nicht hilfsbereit
3.10	War der/die Mentor/in kompetent?	Sehr kompetent	☐	☐	☐	☐	☐	Nicht kompetent
3.11	Wie bewerten Sie das Vertrauensverhältnis zu Ihrem/r Mentor/in?	Starkes Vertrauens-verhältnis	☐	☐	☐	☐	☐	Kein Vertrauens-verhältnis
3.12	Wie gut konnten Sie sich mit Ihrem/r Mentor/in identifizieren?	Hohe Identifikation	☐	☐	☐	☐	☐	Niedrige Identifikation
3.13	Glauben Sie, dass das Mentoring dazu beigetragen hat die Klausuren zu bestehen?	Stark	☐	☐	☐	☐	☐	Gar Nicht
3.14	Glauben Sie, dass das Mentoring dazu beigetragen hat Ihre Noten zu verbessern?	Stark	☐	☐	☐	☐	☐	Gar Nicht
3.15	Haben Sie sich durch das Mentoring weniger isoliert gefühlt?	Stark	☐	☐	☐	☐	☐	Gar Nicht
3.16	Hat Sie das Mentoring zusätzlich für Ihr Studium motiviert?	Stark	☐	☐	☐	☐	☐	Gar Nicht
3.17	Hat Ihnen das Mentoring ingesamt etwas gebracht?	Sehr Viel	☐	☐	☐	☐	☐	Gar Nichts

3.18 Wenn Sie bei der letzten Frage "Gar Nichts" oder den Kasten davor angekreuzt haben, warum nicht?

3.19	Wie würden Sie das Mentoring-Programm insgesamt bewerten?	Sehr gut	☐	☐	☐	☐	☐	☐ Ungenügend

VIELEN DANK FÜR IHRE MITARBEIT!

MUSTER

Bibliography

Aavitsland, P. (2009). Abortion and Depression: Cause or Confounding? *Scandinavian Journal of Public Health 37*, 327–328.

Adamaszek, K., R. Schneider, and T. Brand (2013). Strukturiertes Vorgehen in der Familienbegleitung: Das Konzept der leitfadengestützten Begleitung im Modellprojekt Pro Kind. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 36–48. Weinheim: Beltz Juventa.

Adamaszek, K., R. Schneider, M. Refle, G. Helm, and T. Brand (2013). Zugangswege zu sozial benachteiligten Familien. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 50–60. Weinheim: Beltz Juventa.

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.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and Services for College Achievement. *American Economic Journal: Applied Economics* (1), 136–163.

Angrist, J. and J. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arbeitsgruppe Deutsche Child Behavior Checklist (2002). *Elternfragebogen für Klein- und Vorschulkinder (CBCL/1 1/2-5)*. Köln.

Arni, P. (2012). Kausale Evaluation von Pilotprojekten: Die Nutzung von Randomisierung in der Praxis. *LeGes - Gesetzgebung und Evaluation* (3), 355–386.

Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012: Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: Bertelmann.

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

Baker-Henningham, H. and F. Lopez Boo (2010). Early Childhood Stimulation Interventions in Developing Countries: A Comprehensive Literature Review. *IZA Discussion Paper 5282*.

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

Barnett, W. S. (2006). What is the Value of Early Childhood Education for our Society: Maximizing Returns from Prekindergarten Education? In J. J. van Kuyk (Ed.), *The quality of Early Childhood Education*, pp. 57–72. Arnhem.

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.

Beltz, P., S. Link, and A. Ostermaier (2012). Incentives for Students: Evidence from Two Natural Experiments. *Ifo Working Papers 133*.

Berger, E. M. and C. K. Spiess (2011). Maternal Life Satisfaction and Child Outcomes: Are They Related? *Journal of Economic Psychology 32*(1), 142–158.

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.

Bertrand, M. and S. Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review 91*(2), 67–72.

Bettinger, E. and R. Baker (2011). The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring. *NBER Working Paper 16881*.

Bettinger, E. and T. Bridget (2009). Addressing the Needs of Underprepared Students in Higher Education: Does College Remediation Work? *Journal of Human Resources 44*(3), 736–771.

Bien, W. and J. Marbach (1989). *Die Untersuchung familiarer Lebensformen im Familiensurvey: Arbeitsbericht*. München.

Blank, R. M. (2002). Evaluating Welfare Reform in the United States. *Journal of Economic Literature 40*(4), 1105–1166.

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.

Bradburn, N. M. (2004). Understanding the Question-Answer Process. *Survey Methodology 30*(1), 5–15.

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

Brand, T. and T. Jungmann (2010b). Zugang zu sozial benachteiligten Familien: Ergebnisse einer Multiplikatorenbefragung im Rahmen des Modellprojekts Pro Kind. *Prävention und Gesundheitsförderung 5*(2), 109–114.

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

Brooks-Gunn, J., M. McCormick, S. Shapiro, A. A. Benasich, and G. Black (1994). The Effects of Early Education Intervention on Maternal Employment, Public Assistance, and Health Insurance: The Infant Health and Development Program. *American Journal of Public Health* (84), 924–931.

Bundesagentur für Arbeit (2009). Analyse der Grundsicherung für Arbeitssuchende.

Bundesministerium für Bildung und Forschung (2008). *Wissen für Handeln - Forschungsstrategien für eine evidenzbasierte Bildungspolitik*. Berlin.

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.

Campbell, T. and D. Campbell (1997). Faculty/Student Mentor Program: Effects on Academic Performance and Retention. *Research in Higher Education 38*(6), 727–742.

Cleves, M. A. (2010). *An Introduction to Survival Analysis Using Stata* (3 ed.). College Station and Tex: Stata Press.

Commission of the European Communities (2007). *Towards More Knowledge-Based Policy and Practice in Education and Training.* Commission Staff Working Document SEC. Brussels.

Condon, J. T. &. C. C. J. (1998). The Assessment of Parent-to-Infant Attachment: Development of a Self-Report Questionnaire Instrument. *Journal of Reproductive and Infant Psychology 16*, 57–76.

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.

Crisp, G. and I. Cruz (2009). Mentoring College Students: A Critical Review of the Literature Between 1990 and 2007. *Research in Higher Education 50*(6), 525–545.

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.

Di Pietro, G. (2012). The Short-Term Effectiveness of a Remedial Mathematics Course: Evidence from a UK University. *IZA Discussion Paper 6358*.

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.

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

Gebhardt, T. and H. Jacobs (1997). American conditions? - Social Assistance in the U.S.A. and Germany : a Comparison From a Historical, Institutional and Legal Perspective: Amerikanische Verhältnisse? - Sozialhilfe in den USA und Deutschland : ein Vergleich aus historischer, institutioneller und rechtlicher Perspektive. *Zeitschrift für Sozialreform 43*(8), 597–634.

Geeraert, L., W. van den Noortgate, H. Grietens, and P. Onghena (2004). The Effects of Early Prevention Programs for Families with Young Children at Risk for Physical Child Abuse and Neglect: A Meta-Analysis. *Child Maltreatment 9*(3), 277–291.

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

Grimm, H. (2000). *SETK-2 : Sprachentwicklungstest für zweijährige Kinder. Diagnose rezeptiver und produktiver Sprachverarbeitungsfähigkeiten.* Göttingen: Hogrefe.

Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer, and T. R. (2004). *Survey Methodology.* Hoboken and New Jersey: John Wiley & Sons.

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.

Heckman, J. J. and D. V. Masterov (2007). The Productivity Argument for Investing in Young Children. *Review of Agricultural Economics 29*(3), 446–493.

Heublein, U., C. Hutzsch, J. Schreiber, D. Sommer, and G. Besuch (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen: Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08*. HIS: Forum Hochschule.

Imbens, G. and J. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jungmann, T., V. Kurtz, and T. Brand (2008). Das Modellprojekt Pro Kind und seine Verortung in der Landschaft früher Hilfen in Deutschland. *Frühförderung Interdisziplinär 27*(2), 67–78.

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.

Kamtsiuris, P., M. Lange, and R. Schaffrath (2007). Der Kinder- und Jugendgesundheitssurvey (KIGGS): Stichprobendesign, Response und Nonresponse-Analyse. *Bundesgesundheitsblatt 50*, 547–556.

Karoly, L., M. R. Kilburn, and J. Cannon (2005). *Early Childhood Interventions: Proven Results, Future Promise*. Labor and Population. Santa Monica: RAND Corporation.

Kearney, M. S. and P. B. Levine (2012). Why is the Teen Birth Rate in the United States So High and Why Does It Matter? *Journal of Economic Perspectives 26*(2), 141–166.

Kitzman, H., D. Olds, C. Henderson, C. Hanks, R. Cole, R. Tantelbaum, and et al (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.

Kleinbaum, D. G. and M. Klein (2005). *Survival Analysis: A Self-Learning Text* (2 ed.). New York and NY: Springer.

Klerman, L., S. Cliver, and R. Goldenberg (1998). The Impact of Short Interpregnancy Intervals on Pregnancy Outcomes in a Low-Income Population. *American Journal of Public Health 88*, 1182–1185.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics 114*, 497–532.

Krueger, A. B. and D. A. Schkade (2008). The Reliability of Subjective Well-Being Measures. *Journal of Public Economics 92*(8-9), 1833–1845.

Krug, G. and G. Stephan (2011). Is Contracting-out Intensified Placement Services More Effective than In-house Production? Evidence from a Randomized Field Experiment. *LASER Discussion Papers 51*.

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.

Layzer, J. I., B. D. Goodson, L. Bernstein, and C. Price (2001). *National Evaluation of Family Support Programs. Final Report Volume A: The Meta-Analysis.* Cambridge: Abt Associates, Inc.

Leuven, E., H. Oosterbeek, and B. van der Klaauw (2010). The Effect of Financial Rewards on Students' Achievment: Evidence from a Randomized Experiment. *Journal of the European Economic Association 8*(6), 1243–1265.

Lovibond, P. and S. Lovibond (1995). *Manual for the Depression Anxiety Stress Scales* (2 ed.). Sydney: Psychology Foundation of Australia.

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

Ludwig, J. and D. A. Phillips (2008). Long-Term Effects of Head Start on Low-Income Children. *Annals of the New York Academy of Sciences 1136*(1), 257–268.

Lutz, P. F. and M. Sandner (2010). Zur Effizienz Früher Hilfen: Forschungsdesign und erste Ergebnisse eines randomisierten kontrollierten Experiments. *DIW-Vierteljahrshefte zur Wirtschaftsforschung 79*(3), 79–97.

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.

Maier-Pfeiffer, A., K. Kutz, and M. Sandner (2013). Kosten, Nutzen und Finanzierung in den Frühen Hilfen. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 111–124. Weinheim: Beltz Juventa.

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

Moffitt, R. (Ed.) (1998). *Welfare, the Family and Reproductive Behaviour: Research Perspectives*. Washington D.C.: National Academy Press.

Nores, M. and W. S. Barnett (2010). Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young. *Economics of Education Review 29*(2), 271–282.

Nurse-Family Partnership (2010). *Implementation Overview & Planning: A guide for prospective Nurse-Family Partnership Implementing Agencies*.

NZFH (2010). *Nationales Zentrum Frühe Hilfen*. Köln.

NZFH (2011). *Pilotprojects in the German Federal States: Summary of Results*. Cologne.

OECD (2011). *Education at a Glance*. OECD Publishing.

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., J. Eckenrode, C. Henderson, H. Kitzman, J. Powers, R. Cole, and et al (1997). Long-Term Effects of Home Visitation on Maternal Life Course and Child Abuse and Neglect: 15-Year Follow-Up of a Randomized Trial. *The Journal of the American Medical Association 278*(8), 637–643.

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., C. Henderson, C. Phelps, and et. al (1993). Effects of Prenatal and Infancy Nurse Home Visitation on Government Spending. *Med Care* (31), 155–174.

Olds, D., C. Henderson, R. Tatelbaum, and R. Chamberlin (1988). Improving the Life-Course Development of Socially Disadvantaged Mothers: A Randomized Trial of Nurse Home Visitation. *American Journal of Public Health 78*(11), 1436–1445.

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. *Archives of Pediatrics and Adolescent Medicine 164*(5), 419–424.

Olds, D., H. Kitzman, R. Cole, and J. Robinson (1997). Theoretical Formulations of a Program of Home Visitation for Pregnant Women and Parents of Young Children. *Journal of Community Psychology 25*, 9–26.

Olds, D., J. Robinson, L. Pettitt, D. Luckey, J. Holmberg, R. Ng, and et al (2004). Effects of Home Visits by Paraprofessionals and by Nurses: Age 4 Follow-Up Results of a Randomized Trial. *Pediatrics 114*(6), 1560–1568.

Olds, D. L. (2002). Prenatal and Infancy Home Visiting by Nurses: From Randomized Trails to Community Replication. *Prevention Science 3*(3), 153–172.

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

Olds, D. L., H. Kitzman, C. Hanks, R. Cole, E. Anson, K. Sidora-Arcoleo, D. W. Luckey, C. R. Henderson, J. Holmberg, R. A. Tutt, A. J. Stevenson, and J. Bondy (2007). Effects of Nurse Home Visiting on Maternal and Child Functioning: Age-9 Follow-up of a Randomized Trial. *Pediatrics 120*(4), e832–e845.

Olds, D. L., J. Robinson, R. O'Brien, D. W. Luckey, L. M. Pettitt, C. R. Henderson, R. K. Ng, K. L. 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.

O'Neill, D., S. McGilloway, M. Donnelly, T. Bywater, and P. Kelly (2013). A Cost-Effectiveness Analysis of the Incredible Years Parenting Programme in Reducing Childhood Health Inequalities. *The European Journal of Health Economics 14*(1), 85–94.

Paola, M. d., V. Scoppa, and R. Nisticó (2012). Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment. *Journal of Human Capital 6*(1), 56–85.

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

Puma, M., St. Bell, R. Cook, C. Heid, and M. Lopez (2005). *Head Start Impact Study: First Year Findings. U.S. Department of Health and Human Services.* Westat.

Reece, S. M. (1998). Parent Expectations Survey. In B. K. Redman (Ed.), *Measurement Tolls in Patient Education*, pp. 225–232. New York: Springer Pub Co.

Refle, M., S. Geisler, G. Helm, K. Adamaszek, R. Schneider, A. Maier-Pfeiffer, and S. Hartmann (2013). Fortbildung der Fachkräfte. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 70–84. Weinheim: Beltz Juventa.

Rubin, D., A. O'Reilly, X. Luan, D. Dai, A. Localio, and C. Christian (2010). Variation in Pregnancy Outcomes Following Statewide Implementation of a Prenatal Home Visitation Program. *Archives of Pediatrics and Adolescent Medicine 165*(3), 198–204.

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

Schneider, R., K. Adamaszek, A. Maier-Pfeiffer, S. Hartmann, G. Helm, M. Refle, and S. Geisler (2013). Fachberatung im Modellprojekt Pro Kind. In T. Brand and T. Jungmann (Eds.), *Kinder schützen, Familien stärken*, pp. 76–84. Weinheim: Beltz Juventa.

Scrivener, S. and M. Weiss (2009). *More Guidance, Better Results? Three-Year Effects of an Enhanced Student Services Program at Two Community Colleges*. New York.

Siedler, T., J. Schupp, C. K. Spieß, and G. G. Wagner (2009). The German Socio-Economic Panel as Reference Data Set. *Schmollers Jahrbuch 129*(2), 367–374.

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

Statistisches Bundesamt (2009). *Genesis-Datenbank: Kindertagesbetreuung*. Wiesbaden.

Suri, R., L. Altshuler, and J. Mintz (2004). Depression and the Decision to Abort. *American Journal of Psychiatry 161*(8), 1502.

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.

The National Center for Children Families and Communities (2005). Sample NFP: Sample County Health Department - Sample Report IV Inition(August 1999) through December 31, 2003.

Trappmann, M., B. Christoph, J. Achatz, W. Wenzig, G. Müller, and D. Gebhardt (2010). Design and Stratification of PASS: A New Panel Study for Research on Long Term Unemployment. *IAB Discussion Paper 5*.

Wagner, G., J. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch 127*(1), 139–169.

Wasik, B. H. and D. M. Bryant (2001). *Home Visiting: Procedures for Helping Families* (2 ed.). Newbury Park: Sage.

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.

WVS (2009). *World Value Survey 1981-2008 Official Aggregate v.20090902, 2009.* Madrid.

Zhu, B., R. Rolfs, B. Nangle, and J. Horan (1999). Effect of the Interval Between Pregnancies on Perinatal Outcomes. *New Englan Journal of Medicine 340*, 589–594.