DEPARTMENT OF ECONOMICS

Why Sisters are Better than Brothers - The Effect of Sibling Gender on Attitudes and other Essays in Gender and Education Economics

Johanna Luise Reuter

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 24 February 2022

European University Institute
Department of Economics

Why Sisters are Better than Brothers - The Effect of Sibling Gender on Attitudes and other Essays in Gender and Education Economics

Johanna Luise Reuter

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Economics
of the European University Institute

Examining Board

Prof. Michèle Belot, Cornell University (Supervisor)
Prof. Andrea Ichino, EUI (co-supervisor)
Prof. Almudena Sevilla, UCL
Prof. Arnaud Chevalier, Royal Holloway University of London

Researcher declaration to accompany the submission of written work

I Johanna Luise Reuter certify that I am the author of the work Why Sisters are Better than Brothers – The Effect of Sibling Gender on Attitudes and Other Essays in Gender and Education Economics I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297)).

Statement of inclusion of previous work (if applicable):

I confirm that chapter 1 was jointly co-authored with Mr. Martin Habets and I contributed 70% of the work.

I confirm that chapter 3 was jointly co-authored with Ms. Nurfatima Jandarova, PhD and I contributed 50% of the work.

Signature and Date:

13.01.2022

Abstract

This thesis is composed of three independent essays in applied microeconomics. The first contributes to the field of gender and family economics and analyzes the effect of the gender of the second-born sibling on first-born individuals' attitudes. The second chapter speaks to the health economics literature, evaluating the unintended consequences of a liberalization of the morning after pill. The topic of the final chapter lies within the economics of education, proposing a way to differentiate between degrees depending on the type of higher education institution. Even though the three chapters seem separate, all of them share my interest in gender and education economics, as well as causal estimation.

In Chapter 1, joint with Martin Habets, we analyze the causal effect of sibling gender on attitudes and preferences. Comparing first-born women with a next-born sister to first-born women with a next-born brother allows us to estimate the causal effect of sibling gender. In particular, we find that a next-born sister leads first-born women to have less stereotypically female preferences in education. We also explore how the gender of the next-born sibling influences parental involvement. Our findings indicate that parents are more involved in the education of their first-born daughter if their next-born sibling is also a girl. These results shed light on how sibling gender influences preferences and attitudes, specifically those for education choices that are gender role conforming. To further explore the role of sibling gender in shaping attitudes, we have designed an online survey – currently in progress – to measure gender roles more precisely.

In Capter 2, I analyze the causal effects of liberalizing access to emergency hormonal contraception (EHC), also known as the morning after pill, on young adults' reproductive behavior in England. The liberalization, which changed the prescription status from "on doctor's prescription only" to "available without prescription in pharmacies", created easier and more timely access to EHC for all women aged 16 years or older. In a theoretical model of individual behavior I find that EHC, which can be seen as

insurance against pregnancies, acts both as a substitute for regular contraception, as well as a substitute for abortions. This creates the need for analyzing the issue empirically since overall effects on outcomes such as births and abortions are unclear. Using a difference-in-differences approach, I find that easier access to EHC increases births only among 20-24 year olds. I find no effects on abortions or sexually transmitted infections.

Chapter 3, attempts to differentiate the degree attainment in the UK by type of higher education institutions. Historically higher education in the UK has been shaped by a dual system: elite universities on the one hand and polytechnics and other higher education institutions on the other. Despite the formal equivalence of both degrees, the two institution types faced different financing, target populations, admission procedures and subjects taught. Nevertheless, in survey data they are often indistinguishable. We overcome this problem using a multiple imputation technique in the UKHLS and BHPS data sets. We examine the validity of inference based on imputed values using Monte Carlo simulations. We also verify that the imputed values are consistent with university graduation rates computed using the universe of undergraduate students in the UK.

To Paola

Acknowledgements

This thesis is the final step in a long journey: the last five and a half years of PhD have been the best years of my life, even though they were incredibly challenging. This endeavour has brought me so much joy while studying, learning to do research, but also growing on a personal level. I want to thank everyone who was part of these wonderful years and helped me get to where I am today.

First, I want to thank my advisors Michele Belot and Andrea Ichino for having been there with me on this journey. Michele has been an incredible advisor. I am hugely indebted to her for being there for each step of the way: for each step forwards and also for the many steps backwards that I took. Not only has she always asked the right questions, which made me think hard, but she has always had my back and supported me through the difficult moments. Without her I would have quit this endeavour a long time ago. Thank you for making me believe in my research and myself. I could not have asked for a better advisor. I want to thank Andrea for always challenging me to do better and for being a great teacher especially when working together. I knew I could always come to ask for an opinion and advice. I am also grateful to Andrea for the wonderful bike rides around Florence. Another big thank you goes to Joe Doyle for hosting me at MIT and for being an amazing advisor during my time there. I am greatly indebted to David Paton for providing me with data for the second chapter of this thesis. Furthermore, I want to thank my coauthors Nurfatima Jandarova and Martin Habets for having been brave enough to work with me. Finally, I want to thank Arnaud Chevalier and Almudena Sevilla for having agreed to be on my thesis committee.

However, also on a more personal level there are many people who have contributed to this voyage. The PhD years have been very special to me - incredibly difficult and so much fun at the same time. I have never cried as much nor have I had greater times than during my PhD. And hence, I want to thank everyone who made these years so exceptional. The boys: Philipp and Felix. They made even first year fun and every problem set was time well spent if we worked on it together. There is no-one I would have rather spent my days with. Fatima and Flavia, who are the best friends one could ask for in a PhD. They give the best advice on research and on life. They've always been here for me, ready to put everything into perspective with a long hug or a big cup of tea or emergency Stata help. Birgit and Henrike. Birgit was the first one who made Florence feel like home and I could not have asked for anyone better to confide in. Henrike completed our lives in Via Sercambi and made them even better. Speding time with the two of you in our kitchen made life easy, even when it wasn't.

Furthermore, I want to thank everyone who made Villa la Fonte the amazing place it is. Everyone who listened and did so much more specifically Sarah, Lucia, Antonella and Lori, and the Reps, but also all of my wonderful cohort. A special thank you also goes to Andi, Clemi and Maxi and the IHS, who prepared me for the PhD, understand all the struggles and have supported me from far away. And to everyone I went biking with, in particular Philipp, Sebastian, and Matthias. No place can feel like home without a bike. To everyone who came over for dinner, or invited me for dinner: comfort food is wonderful but is so much better when shared with the right people. I also want to thank Alice and Colby for making Boston feel like home. And to everyone else who made my life in Florence so special and who made me look forward to coming to VLF every day.

I want to thank my mother and father and my sisters Magdalena and Ophelia, who made me feel at home even when I was far away. Thank you for your never ending support, I could not have done any of this without you.

Last but not least I want to thank my husband Giovanni, with whom I've shared everything. Meeting him has been by far the best side effect of the PhD. Thank you for always being there, no matter how hard it got.

And to Paola, who stormed into our lives at the end of this PhD journey and who puts all of this into perspective.

Table of contents

Chapter 1

Why Sisters are Better than Brothers - The Effect of Sibling Gender on Attitudes

joint with Martin Habets

1.1 Introduction

Peer effects have received a vast amount of attention in economics over at least the last 20 years. In particular, this fast growing literature has focused on the role of peer effects for educational attainment and ultimately labor market outcomes. Hence, the education system has been the focal point of most studies. This environment revolves around classmates, classmates' families and college roommates[1]. However, a fundamental group of peers that has surprisingly received little attention within the peer effects literature is that of siblings. There are compelling reasons to think of siblings as very influential peers. Siblings usually live in the same household, spend a large amount of time together and share crucial experiences. Furthermore, given that the time spent together is concentrated in the early years of life, the argument in favor of considering siblings as peers becomes even stronger. This is particularly true in light of the literature that shows how these years matter to personal development and educational attainment (see *e.g.* Heckman, Pinto, and Savelyev 2013).

One aspect of the literature on peers which has been emphasized in the literature is the gender composition of peer groups, often classrooms (see *e.g.* Brenøe and Zölitz 2020; Schneeweis and Zweimüller 2012). The natural extension of this question to the siblings literature concerns the gender composition of siblings. More specifically, this implies investigating the influence of the presence of a sister or a brother – which might matter for one's attitudes or labor market outcomes. This precise consideration of the causal effects of the gender composition of siblings has not received sufficient focus in the literature.

To credibly investigate the causal effects of the gender composition of siblings, we need to overcome several challenges. A central obstacle is due to the endogeneity of parents' fertility choices. In particular, fertility choices are influenced by the gender of the existing children. Dahl and Moretti (2008) show that, in the United States, parents tend to prefer to have sons. They are accordingly more likely to stop having children once they have a son. As a consequence, parents of a first-born son who decide to have a second child might be very different from parents of a first-born daughter who decide to have a second child. These differences between parents – concerning for example the stance on gender equality or gender roles, but even family wealth or income – could then affect the attitudes, preferences and even labor market outcomes

1. A seminal contribution for this nowadays rich literature was Sacerdote (2001): for more information please refer to the literature review, section 1.2.

of their children. Therefore, by comparing for example a woman who has an older brother to a woman who has an older sister, one would capture an effect only partially driven by the gender of the siblings. Indeed, this effect would be entangled with parents' unobserved characteristics. In addition, Ichino, Lindström, and Viviano (2014) point out that the gender of the first-born child leads to different rates of separation and divorce. Ignoring these effects would lead to biased estimates of the effect of sibling gender. These challenges severely complicate the investigation of the causal effect of sibling gender composition, be it on educational and labor market outcomes, or on the formation of preferences and attitudes.

To overcome the endogeneity of parents' fertility choices, our analysis considers only the effect of the second-born's gender on the first-born child's outcomes. Hence, we compare first-born women who have a next-born brother to first-born women who have a next-born sister. Then we proceed to do the same for first-born men. This strategy leverages on the fact that conditioning on the gender of the first child, the gender of the second child is as good as random. This approach, which has only been used few times in the literature[2], allows us to to overcome the issue of endogeneity and hence to address our research question in a causal manner.

Using this approach, the question we want to answer with this project is how and why having an opposite-sex sibling versus a same-sex sibling influences attitudes. Specifically, the first point we investigate is whether the gender of the next-born sibling influences the relationship with parents and parental involvement in education. Second, we want to shed light on how sibling gender influences attitudes towards education, gender roles and political preferences, which to the best of our knowledge has not been done before. Third, we also explore the effects of sibling gender on earnings, in order to connect to the small existing literature on the the causal effects of sibling gender on earnings and occupations[3].

To address these questions we use two sets of panel data from the UK, namely the Next Steps Data and the British Cohort Study. These data sets respectively follow individuals born in 1989/90 and 1970 throughout their lives. These data sets give us the opportunity to analyze the same individuals in terms of their outcomes such as parental involvement and attitudes when they are of school age, and focus on labor market outcomes when they are of working age.

2. See Brenøe (2021), Golsteyn and Magnée (2020) and Peter et al. (2018)

3. Due to our small sample size, we however never expected significant results regarding earnings or occupations.

From the analysis of the above-mentioned panel data sets, we find that the effect of sibling gender is in itself gender specific: whereas men appear not to be affected by the gender of their next-born sibling, the effects on women are significant in several dimensions. To start with, having a next-born sister rather than a next-born brother increases parental involvement in the education of young women. Furthermore, a sister also significantly influences educational attitudes by increasing preferences for STEM subjects in school, improving attitudes towards mathematics and even improving the overall attitude towards school[4]. Even though we find an effect of sibling gender on women's attitudes towards school and future plans for remaining in school, we fail to do so for the highest qualification obtained. All of these results are reflected both in the Next Steps data as well as in the British Cohort study and hence hold true for different cohorts of women. Concerning the effects of sibling gender on women's earnings, we find no significant results[5].

The fact that we find effects only for women is in line with recent literature. Both Cools and Patacchini (2019) and Brenøe (2021) also find that for earnings, the effect of sibling gender only matters for women. However, our results also show that the sibling effect influences underlying preferences. We therefore deem that sibling gender might be influential for many more outcomes than just earnings.

Accordingly, in the Next Steps data, we look at more nuanced outcomes in order to better understand how sibling gender influences preferences. We explore sexual behavior, specifically if individuals have ever engaged in sexual activities without using contraception. This can be interpreted as a proxy for risk preferences. We find that sibling gender has no effect on the probability of engaging in such risky sexual behavior. Interestingly, we find that having a sister (relative to a brother) makes it easier for young men to talk about sexual matters with their parents. We next study career goals and find that, already at age 25, women with a sister place a higher importance on having a career. We show that women with a next-born sister also experience a higher general life satisfaction at age 25 than women with a brother.

The data from the Next Steps and British Cohort Study are, to some extent, limited. They do not allow us to explore attitudes towards gender roles, preferences towards job characteristics or political preferences in detail. Since the existing literature has

4. This overall attitude towards school captures how happy an individual is in school and with the school, how useful she or he thinks school is and how much effort he or she puts into school.

5. We do not find significant effects on women's earnings in the British Cohort Study. In the Next Steps data, we deem individuals to be too young to be able to draw any meaningful conclusions about their earnings.

highlighted gender roles as potentially the main mechanism behind the effects of sibling gender, we aim to study gender roles in more detail. In an attempt to measure these attitudes more precisely, we have designed our own online survey, which is ongoing in a population of young adults in the UK.

Summarizing, our paper aims to provide a four-fold contribution to the literature. First, we explore the causal effect of sibling gender in existing panel data – the Next Steps Data and the British Cohort Study – on attitudes and preferences, specifically at a young age. Second, we consider labor market outcomes in these data sets in order to better connect to the existing literature. Third, we design our own survey in order to measure adherence to gender roles and whether this is influenced by sibling gender. Fourth, we test whether sibling gender has an effect on political preferences using our survey. We are motivated by a strand of the economics literature which has highlighted the effect of having a daughter (rather than a son) on fathers' political preferences and attitudes[6].

The remainder of the paper is structured as follows: Section 1.2 summarizes the literature and our contribution to it. Section 1.3 outlines the identification strategy. Details on the data used can be found in Section 1.4. Section 1.5 shows the results obtained from the Next Steps Data and the British Cohort Study. Section 1.6 outlines our ongoing survey. Finally Section 1.7 concludes.

6. See Section 1.2 for more details.

1.2 Literature

With this project, we broadly speak to the literature on siblings in economics. This literature has mostly focused on the effects of the number of siblings, birth order, and birth spacing. Angrist and Evans (1996) were one of the first to use the gender composition of siblings to analyze the effects of childbirth on the labor market participation of mothers. Many then followed Angrist and Evans (1996) in using the gender of the first two children as an instrument for the number of children. Black, Devereux, and Salvanes (2005) study both the effect of family size and birth order on children's education and hereby started a branch of literature which focusses on birth order effects. More recently, also contributing to the literature on birth order effects Breining et al. (2020) find that birth order plays a role in delinquency outcomes. In general, this literature finds that second-borns typically perform worse in terms of a broad range of dimensions, including educational measures. Within the economics literature on the composition of siblings, the effect of sibling gender has received relatively little attention, potentially due to the fact that uncovering causal effects poses a challenge. Still, a growing number of studies investigate the effect of sibling gender on various outcomes, and their underlying mechanisms.

First, there are a range of papers studying the effects of sibling gender on educational attainment, starting with Butcher and Case (1994). However most of this research estimates a composite effect of sibling gender and parental types or preferences, due to their estimation strategy. Nonetheless, we briefly outline some of the more recent publications from this branch of the literature: Anelli and Peri (2014), for example, study how the gender composition of siblings affects the choice of college major in Italy. They find that individuals from families with same sex siblings are encouraged to make less gender stereotypical choices. Relatedly, Dossi et al. (2021) use sibling gender composition as an instrument for parental gender attitudes and analyze how these affect educational outcomes. They find that girls from more conservative households in terms of gender attitudes do worse on standardized math tests at age 10. Again, one major drawback of these findings are related to their identification strategy. Both papers are not able to identify is the causal effect of sibling gender composition – rather, they estimate a composite effect. Thus, they cannot disentangle the sibling gender effect from the effects of parental types of preferences. In a very recent working paper Collins (2021) provides causal evidence on the effects of sibling gender on education in less developed countries. On average, he finds small yet statistically significant negative effects on education of having a brother relative to a sister. This result masks significant heterogeneity and highlights the importance of customs and traditions in how parents' make decisions around their children's education.

Second, a strand of the literature on the gender composition of siblings looks at its effects on labor earnings. This recent literature adopts a similar identification strategy to ours to study causal effects – they study whether first-borns are affected by the gender of their next-born sibling. Peter et al. (2018) find using Swedish data that a same-sex sibling increases men's earnings, while for women, although similar, their results are less robust. In an attempt to learn more about the underlying channels of this earnings effect, they study several additional outcomes. They provide suggestive evidence that the positive effect on men's income could be largely driven by competition between brothers, while the effect on women's earnings could come from lower unemployment – through a more efficient use of a sister's job search network. However, the authors state that there might other potential explanations that they could not explore with their data. Cools and Patacchini (2019) focus on a cohort of US women. They find that the presence of a next-born brother (relative to a next-born sister) lowers women's earnings in their late 20s and early 30s by approximately 7%. They do not find statistically significant results for men. Then, they consider the mechanisms through which the presence of a brother may affect women's earnings. Their evidence seems to indicate that this earning penalty may primarily stem from (i) lower parental expectations and (ii) by leading women to adopt more traditional attitudes and behaviors toward gender roles. However, their data do not provide them with questions about preferences for many gender-specific tasks (such as housework).

Third, some articles study the effect of sibling gender on personality. Detlefsen et al. (2018) show in an experimental design that sibling gender composition has a significant impact on trust and risk preferences. They find, for example, that second-born children are only more risk taking when all siblings are of the same gender. Cyron, Schwerdt, and Viarengo (2017) study the effect of opposite sex siblings on cognitive and non-cognitive skills in early childhood. They find that boys with a sister exhibit significantly higher math and reading skills in kindergarten and better learning skills and self-control than boys with a brother, while the overall effect on girls is insignificant. In a paper closely related to ours, Golsteyn and Magnée (2020) study whether sibling gender causally affects personality traits. Using the 1970 British Cohort Study, they study the causal effect of the gender of the second-born sibling on personality traits of the oldest sibling in the household. Their identification strategy is thus the same as ours. They find that the gender of the sibling has implications for personality traits. They show that boys are more agreeable if they have a next-born younger sister.

Fourth, Brenøe (2021) studies the effect of sibling gender on women's gender conformity with Danish administrative data. She finds that having a next-born brother (relative to a sister) increases first-born women's conformity to traditional gender roles – as measured through their choice of occupation and partner. To look at the mechanisms behind this effect, she examines mothers' and fathers' quality time investment in their first-born daughter during childhood. To do so, she matched her administrative data with a detailed time use survey. She shows that parents of mixed-sex children invest their time more gender-specifically than parents of same-sex children.

All in all, this literature remains inconclusive regarding the effect of sibling gender. Moreover, a common factor of these works is their intent to uncover the mechanisms behind the effects highlighted. Most encourage further research to explore those possible mechanisms. In particular, Cools and Patacchini (2019) suggest to study the interactions with parents and interactions between siblings in adolescence. According to Golsteyn and Magnée (2020), gender norms should be explored to further explain the relationship between sibling gender and personality traits. This is particularly relevant since Brenøe (2021) states that girls' development of gender conformity by adolescence has important consequences for their later-life educational and labor market outcomes. Hence, we contribute to this new branch of literature on causal sibling effects by focusing on attitudes specifically at a young age of individuals. Our contribution is threefold: First, we use existing panel data, the Next Steps Data and the British Cohort Study, to explore the causal sibling effect on attitudes and preferences specifically at a young age. Second, we consider labor market outcomes in these data sets in order to better connect to the existing literature. Third, we design our own survey in order to hopefully measure more precisely one of the most important mechanism behind the effects which have been studied in the literature, namely adherence to gender roles.

Most of the above cited papers primarily focus on the effect on women of having a brother rather than a sister. However, the economics literature has highlighted the effect of having a daughter rather than a son on fathers and specifically their political preferences and attitudes. We suspect that there could be a similar effect on men of having a sister rather than a brother. In particular, since we look at the effect of younger on older siblings – for which the older brother could feel responsible – this literature motivates our research and the inclusion of a liberalism section in the survey questions. In a very recent paper, Ronchi (2021) studies the role of managers' gender attitudes in shaping gender inequality within the workplace. She finds that following the birth of their manager's first daughter, women's relative earnings increase by 4.4% while relative employment increases by 2.9%. Washington (2008) studies the attitudinal shift that arises from parenting daughters. She finds that children can influence parental behavior. She demonstrates that conditional on total number of children, each daughter increases a congressperson's propensity to vote liberally, particularly on reproductive rights issues. While those papers highlight the relevance of child-to-parent behavioral influence, it seems reasonable to assume that similar effects might appear in the sister-to-brother influence. We further contribute to the broader literature on family interactions and effects of these on political preferences by exploring such preferences in our own survey.

1.2.1 Review of the mechanisms behind sibling gender effects

Economics, however is clearly not the only discipline to have studied sibling gender effects. A branch of the psychology literature has studied the influence of siblings' gender composition[7], hereby focussing on behavioral outcomes. This literature is highly relevant for our research. On the one hand, this literature finds that individuals evolving in mixed-gender households are more likely to endorse gender stereotypes. This increase in conformity to traditional gender roles could be attributed to different channels. A first channel – advanced by McHale et al. (2000) – is that of parenting. They find that mixed-gender households lead to more gender specific parenting *i.e.* mothers spend more time with their daughters and fathers spend more time with their sons. This increases the propensity of children to imitate their same-gender parent, which in turn, leads to a stronger endorsement of gender stereotypes. A second channel though which mixed-gender households might lead to stronger adherence to gender roles has ben tested by Hirnstein, Andrews, and Hausmann (2014). They show with a biopsychosocial approach that mixed gender settings activate gender stereotypes. As a result, individuals from mixed gender households could be more prone to act in gender stereotypical ways. On the other hand, Conley (2000) suggests that individuals generally compete more with siblings of the same gender. This higher level of competition within the family is said to be beneficial for children, especially regarding their educational attainment. The economics literature has attempted to shed light on some of these mechanisms – connecting psychology and economics – by testing some of the hypothesis put forward by the psychology literature. While this is what our paper also does, we further aim to

7. For a review on gender in the family see Endendijk, Groeneveld, and Mesman (2018).

explore behavioral outcomes that have, to the best of our knowledge, not been objects of study in the economics literature.

1.3 Identification and Estimation Strategy

As mentioned in the literature review, the basic strategy of all projects concerned with the effects of sibling gender relies on comparing individuals who have a same sex sibling, to those who have an opposite sex sibling. However, one needs to be cautious in attempting to uncover a causal effect. Challenges arise because parents' fertility choices are influenced by the gender of their children, as shown by for example Dahl and Moretti (2008). In particular, Dahl and Moretti (2008) have shown that in the United States married parents tend to prefer a son and are more likely to stop having children once they have a son. On the contrary, Ichino, Lindström, and Viviano (2014) have shown that in the US, the UK, Italy and Sweden a first-born son increases total fertility. As a consequence, parents who decide to have another child after having a first-born son might be very different from parents who decide to have a second child after having a first-born daughter. Hence, a girl who has an older brother might grow up in very different circumstances and with considerably different parents than a girl who has an older sister. Parental characteristics which could differ across these two families are not only gender attitudes or preference for boys of parents, but also wealth or the presence of the father. All those elements can be seen as potential confounders when one is interested in the effect of sibling gender on individuals. Since we are specifically interested in how having an opposite-gender sibling might influence the gender norms of an individual, it is crucial to take care of these confounders.

The way forward, which has now been used a few times in the literature (see *e.g.* Peter et al. 2018; Brenøe 2021) is to only consider first-born siblings: once one conditions on the gender of the first-born child, the gender of the second child is deemed as good as random, meaning that parental gender preferences do not influence the gender of the second child. Hence, considering only first-born children allows us to estimate the causal effect of having an opposite gender sibling.

Therefore, in order to estimate the causal effect of sibling gender, we compare first-born women who have a next-born younger brother to first-born women who have a next-born younger sister, and likewise for men. This comparison allows us to get at the underlying causal effect since – in the absence of sex-selective abortions – the gender of a child is random.

$$Y_i = f(sister_i, X_i, \varepsilon_i) \tag{1.1}$$

In a regression analysis, we investigate whether the gender of the next-born sibling has an effect on several different outcome variables, which we explain in more detail below. The most general form for this regression is described in (1.1). The outcome Y_i is a function of $sister_i$, an indicator function that identifies whether individual i's next-born sibling is female (*i.e.* whether the next-born sibling is a sister, as opposed to a brother), a set of covariates X_i, and an error term ε_i. The set of covariates X_i contains parental and individual characteristics. Parental characteristics include variables such as parental education and age at birth. Personal characteristics include controls such as birth weight, but vary across specifications. We use robust standard errors for the estimation. Furthermore, we use cross-sectional survey weights when estimating these regressions using data from the British Cohort Study or the Next Steps Data. The function $f(\cdot)$ in (1.1) can take on different forms, depending on the outcome variable. In the case of a linear $f(\cdot)$, Equation (1.1) becomes Equation (1.2), a simple OLS regression:

$$Y_i = \alpha + \beta sister_i + \delta X_i + \varepsilon_i. \tag{1.2}$$

Function $f(\cdot)$ can also take the form of a logit or a type ordered logit, depending on the outcome variable under study. These regressions are always estimated separately for men and women, because the effect of having a younger sister is potentially different for women than it is for men. The main coefficient of interest is β, which is the coefficient on the indicator function for having a next-born sister rather than a next-born brother.

1.4 Data and Sample

As explained above, our identification strategy consists in comparing first-born individuals with a next-born sister to those with a next-born brother. In order to select these individuals in the data, we need information on their siblings composition, such as birth order and gender. Even though survey data often includes information on the number of siblings (and in some cases siblings' gender), siblings' birth order is rarely available. Hence, to pursue our identification strategy, one needs very detailed data, which is not easy to come across.

The data we use for this project is from the UK. We use the British Cohort Study (University College London, UCL Institute of Education, Centre for Longitudinal Studies 2021a) as well as the Next Steps Study (University College London, UCL Institute of Education, Centre for Longitudinal Studies 2021b). Both are conducted by the Centre for Longitudinal Studies. Both panels selects individuals born within a very short period of time: the British Cohort Study collects information for individuals born in one week of 1970; the Next Steps Study follows the lives of individuals born in 1989-80[8].

Both of these surveys started following the individuals very early on in life. Accordingly, for the earlier questionnaires, parents of targeted individuals answered the questions on family. Information on mothers' fertility was gathered. This includes detailed information on all children, such as gender and birth years. This feature allows us to identify the relevant individuals. Further, knowing about their next-born gender makes estimating causal effects first-born. We study the effect of sibling gender on preferences, attitudes, education, and labor market outcomes. We first outline the procedure to obtain the sample of interest in the Next Steps data set, and run some balance tests. We then do the same for the British Cohort Study data set.

1.4.1 Next Steps Data

The Next Steps data (NSD) – previously known as the Longitudinal Study of Young People in England (LSYPE) – has been following individuals born in England in 1989-1990. It began following 16.000 young people who were aged 14 in 2004. The last survey for which data is available took place in 2015, when individuals were 25, and gathers a total of 7,707 respondents. There are 8 waves of data available. Next Steps specifically focusses on information about education and employment. Still, it also

8. In the future we plan to also use data from the Millennium Cohort Study, a further study conducted by the Centre for Longitudinal Studies which follows individuals born in 2000-2002.

Table 1.1: Number of Siblings at Age 14 in the initial wave of NSD

Number of siblings	Frequency
1	2598
2	1216
3	470
4	148
5 or more	58
Total	4490

collects information on economic circumstances, family life, physical and emotional health and wellbeing, social participation and attitudes.

In order to obtain the relevant information on siblings we use the data collected in wave 1 of the the survey, when individuals were 14 years old. Specifically, we use the household grid file of wave one to obtain information on all siblings living in the same household. We complement this with further data on non resident siblings to obtain our sample of interest *i.e.* all first-born individuals with at least one younger sibling. By using data from wave 1, we restrict the sample to only include individuals with siblings who have at most 14 years of age difference, by construction. We then impose the following further sample restrictions. First, we only keep individuals who are not part of a multiple birth themselves and whose next-born younger sibling are also not part of a multiple birth. Second, we exclude individuals for whom information on their own or their next-born sibling's gender is missing.

This selection leads to a sample of 4490 individuals from the first wave of the survey. Table 1.1 shows the distribution of the number of siblings for these individuals[9]. Table 1.39 in Appendix 1.A reports a breakdown of first-born individuals by their gender and the gender of their next-born sibling.

Balance tests on Next Steps Data

Since our identification strategy relies on the fact that conditioning on the gender of the first-born child the gender of the second-born child is as good as random, we perform balance tests to examine whether this assumption holds. For these balance tests, we use variables which we expect to be predetermined, *i.e.* which we expect cannot be influenced by the gender of the second-born child. We would ideally use

9. Due to attrition, this sample gets smaller and smaller. A boost sample has been added in wave 4 of the survey, at age 17/18. However, we decide not to use these individuals. We assume that the older sibling of an individual sampled at age 17/18 might no longer be part of the household, making the identification of first-borns imprecise.

outcomes measured at birth of the first-born individual. However, the NSD wave 1 was only conducted at age 14 of the first-born individual. Hence, we use variables which were either determined at birth or are likely to have been fixed already at birth of the first-born child for balance tests. We use almost all of the variables which we use for balance tests also as covariates in all regressions. The only variables, which we include in the balance tests, but do not include as covariates are age difference to the second-born sibling and the number of siblings. We do include the number of siblings as a covariate because it is part of the effect of interest: having a next-born same gender sibling also increases the total number of siblings and hence, it is part of the effect of interest. We do not include the age difference to the next-born sibling as a covariate, because we plan to potentially split the sample further based on this variable.

Table 1.2 shows the results of the balance tests for first-born men. All covariates are balanced, which indicates that our identifying assumption probably holds. Regarding the variables, which we do not include as covariates, which can be seen in the lower half of Table 1.2, the number of total siblings, as was to be expected is not balanced.

For first-born women however, as Table 1.3 shows, not all covariates are balanced. Specifically the educational qualification of the mother, and whether the mother was single at birth seem to differ across the two groups defined by the gender of the next-born sibling. First, one should not that we perform 18 balance tests, it does not come as a surprise that differences are statistically significant. Nonetheless, in Appendix 1.A we shed more light on these results by analyzing each category of the two variables separately by creating a separate dummy variable for each categorical value take by the variable. We find that the t-test for difference in means is only statistically significant for the lowest educational category of the mother, i.e. when the mother's qualification is unknown. This suggests that there is some imbalance across the two groups. This could potentially be due to the case that the variable of mother's education was measured only at age 14 of the individual rather than at birth. Also for the relationship status of the mother at birth only one group, namely when the mother is single at birth, creates imbalance across the two groups.

1.4.2 British Cohort Study

The British Cohort Study (BCS) follows 17.000 children born in a single week of April 1970. Individuals (or their parents) have been interviewed up to 11 times from ages 0 to 50. We use the information collected in wave 2 – at which point individuals were

Table 1.2: **NSD Balance Tests for Men using Covariates**

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Birth weight	1178	316.224	1102	313.820	2.404
		(3.023)		(3.208)	
Uk born	1178	0.896	1102	0.894	0.002
		(0.012)		(0.012)	
Mother's age at birth	1178	2.504	1102	2.438	0.065
		(0.038)		(0.040)	
Mother single at birth	1178	0.111	1102	0.085	0.026
		(0.017)		(0.017)	
Ethnicity	1178	1.940	1102	2.075	-0.135
		(0.074)		(0.092)	
Mother's qualification	1178	3.561	1102	3.410	0.152
		(0.069)		(0.071)	
Father's qualification	1178	2.148	1102	2.332	-0.184
		(0.084)		(0.086)	
Age Difference	1178	3.316	1102	3.371	-0.054
		(0.056)		(0.061)	
Number of Siblings	1178	1.611	1102	1.460	0.151***
		(0.028)		(0.025)	

Notes: The value displayed for t-tests are the differences in the means across the groups. Observations are weighted using variable W1FinWt as pweight weights.***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.3: NSD Balance Tests for Women

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Birth weight	1109	307.627	1101	306.745	0.882
		(2.726)		(2.823)	
UK born	1109	0.902	1101	0.917	-0.015
		(0.012)		(0.011)	
Mother's age at birth	1109	2.481	1101	2.461	0.020
		(0.038)		(0.040)	
Mother single at birth	1109	0.079	1101	0.128	-0.049*
		(0.017)		(0.018)	
Ethnicity	1109	2.040	1101	2.196	-0.157
		(0.091)		(0.084)	
Mother's qualification	1109	3.379	1101	3.648	-0.269***
		(0.073)		(0.070)	
Father's qualification	1109	2.046	1101	2.230	-0.184
		(0.088)		(0.090)	
Age Difference	1109	3.250	1101	3.446	-0.195**
		(0.059)		(0.062)	
Number of Siblings	1109	1.436	1101	1.633	-0.198***
		(0.023)		(0.030)	

Notes: The value displayed for t-tests are the differences in the means across the groups. Observations are weighted using variable W1FinWt as pweight weights.***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.4: Number of Siblings at Age 5 in the BCS

Number of Siblings	Frequency
1	2935
2	478
3	48
4	5
Total	3466

5 years old – to select the individuals relevant to our analysis. We keep first-born individuals who have at least one younger sibling. Since individuals are surveyed at age 5, by construction, the younger sibling can be at maximum 5 years younger. Note that this is a smaller time interval than in the NSD. We impose similar restrictions with respect to multiple births.

The selection procedure results in a final sample of 3.466 individuals. Table 1.4 shows the distribution of the number of siblings for these individuals. While quite different from the distribution of the number of siblings in the NSD, we attribute most of this difference to the fact that the number of siblings is measured at age 5. Table 1.40 in Appendix 1.A then reports the gender structure of the siblings, which is balanced.

Balance Tests for the British Cohort Study

The first wave of the BCS collects data at birth. This data is thus optimal for balance tests: the variables we look at cannot be influenced by the gender of the next-born sibling. The variables we use in the balance tests are birthweight of the individual, mother's age at birth, social class of the mother at birth, country (within the UK), and marital status of the mother at birth.

Tables 1.5 and 1.6 display the results of the balance tests for the BCS sample, analogously to the analysis for the NSD. Again, it ignores the fact that some variables are categorical. Here, all variables seem very balanced – indicating that the assumption that next-born sibling's gender is as good as random very likely holds. This result is possibly in part due to the fact that the variables were measured at birth. Even the total number of siblings is balanced across the gender of the next-born child. However this might only be true because we measure the number of siblings at age 5. Furthermore, these two tables show that the age difference to the next-born sibling is also balanced across the gender of the next-born sibling. Again, all variables which we use for balance tests are used as covariates in the regressions apart from the number of siblings and the age difference to the next-born sibling.

Table 1.5: BCS Balance Test for Men

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Birth weight	919	128.576 (3.358)	858	133.217 (4.127)	-4.641
Mother's age at birth	919	22.830 (0.125)	858	22.627 (0.126)	0.203
Mother's age missing	919	0.000 (0.000)	858	0.000 (0.000)	N/A
Birth weight missing	919	0.013 (0.004)	858	0.019 (0.005)	-0.006
Social class at birth	919	6.422 (1.082)	858	5.322 (0.041)	1.101
Region	919	5.690 (0.086)	858	5.779 (0.088)	-0.089
Mother's marital status at birth	919	1.974 (0.011)	858	1.991 (0.011)	-0.017
Number of Siblings	919	1.173 (0.014)	858	1.162 (0.014)	0.011
Age Difference in Months	919	29.439 (0.385)	858	29.096 (0.394)	0.343

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.6: BCS Balance Test for Women

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Birth weight	856	122.866	833	127.615	-4.749
		(3.310)		(3.994)	
Mother's age at birth	856	22.958	833	22.896	0.062
		(0.131)		(0.129)	
Mother's age missing	856	0.000	833	0.000	N/A
		(0.000)		(0.000)	
Birth weight missing	856	0.012	833	0.017	-0.005
		(0.004)		(0.004)	
Social class	856	5.376	833	6.535	-1.159
		(0.041)		(1.194)	
Region	856	5.567	833	5.571	-0.005
		(0.086)		(0.090)	
Mother's marital status at birth	856	1.951	833	3.136	-1.185
		(0.011)		(1.197)	
Number of Siblings	856	1.157	833	1.188	-0.032
		(0.015)		(0.015)	
Age Difference in Months	856	29.585	833	31.054	-1.469**
		(0.392)		(0.427)	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

1.5 Results

We want to study how individuals' attitudes and preferences are influenced by the gender of their next-born sibling. In particular, we aim to explore potential mechanisms behind these effects. The literature on child psychology has put forward some of these mechanisms, such as those on gender roles (see *e.g.* McHale et al. 2000; Hirnstein, Andrews, and Hausmann 2014). In this section, we first present the results from the Next Steps data and second, those from the British Cohort Study.

First, we consider parental interactions with children, specifically focusing on education. We do this in light of the paper by McHale et al. (2000), who find that in mixed-gender households parenting becomes more gender specific. In the Next Steps data and the British Cohort Study, we focus mostly on educational involvement of parents. However, our survey also asks individuals more generally about their relationship with their parents.

Second, we consider educational preferences of individuals. Educational preferences and attitudes are important outcomes for several reasons. Since the existing literature in economics has found effects of sibling's gender on earnings and occupations, we focus on preferences for subjects such as STEM, which could be mechanisms behind these. This also speaks to the child psychology literature on education. For example, Hirnstein, Andrews, and Hausmann (2014) addresses the fact that in mixed gender families, stereotypes become more salient. Thus, we hypothesize that stereotypes could drive individuals subject preferences. Conley (2000) suggests another channel. He finds that same gender siblings compete more with one another, which could lead to better education outcomes for individuals. This further encourages us to also look at gender attitudes towards schooling and education.

Third, we try to find direct measures of gender roles such as hours spent on domestic chores, agreement with statements on women in employment, or split of household duties. However, the data are very limited in this realm. Our survey aims to fill this gap.

Fourth, we consider preferences regarding career goals, willingness to obtain a degree or a family, and overall life satisfaction. These outcomes are not very conclusive. One reason is that they are measured at a very young age in the Next Steps data.

Fifth, we also explore sexual behavior. Specifically, individuals are asked if they have ever engaged in sexual activities without using contraception – a proxy for risk preferences. In addition, to better understand the relationship individuals have with their parents, we look at whether individuals find it easy to talk to their parents about sexual matters.

Sixth, we also consider earnings and other labor market indicators as outcomes. Most of such outcomes, however, cannot be properly measure well in the Next Steps data, since in that panel, individuals are still too young[10]. The results relative to the British Cohort Study concern other labor market indicators and fertility outcomes and are reported in Appendix 1.A.

1.5.1 Results from the Next Steps Data

In order to understand how sibling gender influences such outcomes, we proceed as outlined above. If variables are measured multiple times we proceed chronologically, starting from wave 1 in which individuals were 14 years of age until wave 8 at which individuals were 25 years old. Table 1.7 shows the age of the individual at each wave.

Table 1.7: NSD Waves and Ages

Wave	1	2	3	4	5	6	7	8
Age	14	15	16	17	18	19	20	25

As mentioned above, we include as covariates in all of the regressions: Birth weight, and indicator for being UK born, mother's age at birth in bands, an indicator for the mother being single at birth, ethnicity, the mother's highest qualification and the father's highest qualification. These are the variables we also used for the balance tests reported in Tables 1.2 and 1.3. All these variables, apart from birth weight, are included as categorical variables, thus a category for missing information is included. However, when displaying regression results we only show the coefficient of interest *i.e.* the indicator for the next-born sibling being a sister. All regressions using the Next Steps Data are weighted with the appropriate cross-sectional survey weights. For each variable used as an outcome we not only report the number of observations used in the regression but also the variable mean and standard deviation for each gender, in order to better understand the magnitude of the effect.

10. At the final wave of the Next Steps data individuals are only 25 years old, and hence the household income which is reported might for many individuals still reflect parental income rather than own income.

Parent-Child Interactions in the NSD

Since the psychology literature puts forward that parenting might change due to the presence of an opposite gender sibling, we look not only at outcomes of the individual but also at the answers provided by the parents of the individual. We are interested specifically in how parental involvement in the upbringing of the child is affected by the gender of the next-born sibling. Since the Next Steps Data focuses on education, we use questions on parental involvement in the individuals education as a measure of parent-child relationship. In waves 1-4 there are questionnaires posed to the parent(s) of the individual, however we only use questions contained in wave 1 since only these capture parental involvement in education [11].

In wave 1, we use the following variables to capture parental involvement in education: parents attending parents evening in school, parents attending meetings with school teachers, parents talking to the young person about choice of year 10 subjects, parents giving advice on what to do in year 10. The variables are recoded such that a higher value means stronger involvement. Using a principle component analysis, we combine these variables into a single outcome and extract the first component. The results are shown in Table 1.8, with column 1 displaying the results for men and column 2 for women. As one can see, only for women does the gender of the next-born sibling have a significant effect on parental involvement. Parents seem more involved in the education of their first-born daughter when also their second-born child is a girl. This outcome potentially speaks to the hypothesis that boys require more involvement of parents in school and hence parents can only give girls the desired attention when the second-born is not a boy.

Furthermore, we continue to look at child-parent interactions regarding education, but this time from the child perspective. To do so, we analyze two variables regarding homework: The first is whether an individual young person receives any help doing homework. We code the variable such that a higher number means more help. The second is a variable which asks for the frequency that someone at home checks whether the young person does their homework. Also here we code a higher number to mean more frequent checks. The results for the two variables regarding homework can be seen in Tables 1.9 and 1.10. As one can see from Table 1.9, the gender of the next-born sibling

11. In wave 2 parents are asked specifically about their involvement in choices of vocational subjects rather than in general education, in waves 3 and 4 the questions on educational involvement do not capture general participation in the education of the individual but only whether specific efforts to schedule extraordinary meetings with teachers have been made. These variables do not capture general parental involvement but rather problems the individual faces in education.

Table 1.8: OLS Regression PCA Parental Involvement

	Men	Women
Next-born Sister	-0.0549	0.105*
	(0.0674)	(0.0631)
Observations	1740	1672
Mean	-0.0425	0.0298
Sd	1.396	1.236

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

has no effect on receiving help doing homework. However, as can be seen from Table 1.10 it does have an effect on the frequency of checking or controlling that homework is done for women. Again, parents are more involved in their daughters education, this time through making sure that their daughter does her homework, when also their second-born child is a girl rather than a boy.

Table 1.9: OLS Regression Help with Homework Wave 1

	Men	Women
Next-born Sister	0.0220	0.00498
	(0.0366)	(0.0353)
Observations	2189	2160
Mean	1.591	1.651
Sd	0.798	0.758

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Educational Preferences and Attitudes in the NSD

Next, we want to analyze how having a next-born sister rather than a next-born brother influences preferences. We do so by analyzing a young person's preferences for school subjects. In wave 1, individuals are asked about their favorite subject. From this question, we construct two indicators; one for the favorite subject being a STEM

Table 1.10: OLS Regression Control Homework Wave 1

	Men	Women
Next-born Sister	0.0224	0.187**
	(0.0695)	(0.0753)
Observations	2189	2160
Mean	3.976	3.860
Sd	1.435	1.549

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

subject, the other for the favorite subject being a humanities subject[12]. Table 1.11 displays the results of an OLS regression with the indicator for a STEM subject as the outcome variable. They show that women with a younger sister are more likely to have a STEM subject as their favorite subject than women with a younger brother. Similarly, Table 1.12 shows that women with a next-born sister are less likely to say their favorite subject is humanities, as compared to women with a next-born brother. In order to analyze preferences further, we combine the following six further variables using principle component analysis into an indicator for STEM: liking maths, being good at maths, liking science, being good at science, liking ICT, and being good at ICT. We use the first principle component as an index for STEM preference. The output of the OLS regression with this index as the outcome can be seen in Table 1.13. Again the effect on preference for STEM, measured by the index, of having a sister rather than a brother as a next-born sibling is only significant for women. From these results from wave 1, we conclude that having a sister rather than a brother influences women's preferences and make them less "gender-stereotypical".

In wave 2, we again consider an indicator for the favorite subject being STEM[13] The results of this OLS regression can be seen in Table 1.44 in Appendix 1.A. Just like in wave 1, having a next-born sister rather than a brother significantly increases women's probability of having a favorite subject which is classified as STEM.

12. We consider the following subjects STEM: Mathematics, Science, Design and Technology, ICT, Home Economics, Business studies or Economics. We consider these subjects humanities: History, Humanities and Social Studies, Art, English, Modern Languages, Music, Drama

13. The subject which we categorize to be STEM are the following: Maths, Science, ICT, Design and Technology and electronics, Design and Technology and food technology, Design and Technology and resistant mater, Design and Technology and systems technology, Manufacturing, Engineering, Applied Science, Applied ICT.

Table 1.11: **OLS Regression Favorite Subject is STEM Wave 1**

	Men	Women
Next-born Sister	0.0114	0.0349*
	(0.0219)	(0.0183)
Observations	2280	2210
Mean	0.334	0.197
Sd	0.472	0.398

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.12: OLS Regression Favorite Subject is Humanities Wave 1

	Men	Women
Next-born Sister	-0.00790	-0.0608***
	(0.0213)	(0.0234)
Observations	2280	2210
Mean	0.300	0.560
Sd	0.458	0.497

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.13: OLS Regression PCA for Liking and Good at STEM Wave 1

	Men	Women
Next-born Sister	0.0267	0.198***
	(0.0623)	(0.0703)
Observations	2225	2172
Mean	0.257	-0.465
Sd	1.309	1.456

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Next, we want to analyze more general attitudes and plans. To do so we, first of all, we consider the attitude towards school – an index of how an individual feels about school in general and about their specific school. A higher value on the attitude towards school suggests that an individual likes school better or thinks that school is more beneficial. Moreover, we consider an individual's plan for education for after year 11 – this is a categorical variable which takes higher values when individuals are more certain that they want to continue full-time education. We analyze both these variables. Table 1.14 shows the results for the attitude towards school and Table 1.15 for the plan for education post year 11. Interestingly, the gender of the next-born sibling has a significant effect on the school attitude for women and no effect on whether a woman plans to stay in school after year 11. However, it needs to be noted that at age 14, almost 90% of all women plan to stay in full-time education and hence there is very little variation in this variable. Concluding, it seems that having a sister not only influences women's subject preferences and parental involvement in schooling, but also their overall attitude towards school at age 14.

Table 1.14: OLS Regression Attitude to School Wave 1

	Men	Women
Next-born Sister	0.126	0.658*
	(0.346)	(0.351)
Observations	2196	2150
Mean	34.00	34.97
Sd	7.310	7.106

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Still in wave 2, we look at attitudes towards and plans for education by considering the attitude towards school and the plan for education after year 11. The corresponding regression outputs can be seen in Table 1.16 for the school attitude and in Table 1.17. The effect for women of having a sister rather than a brother on the school attitude is very similar to the effect in wave 1. However, the effect on the plan for full time education after year 11 is now also positive and significant for women.

In wave 3, we do the same and analyze individuals' attitudes towards school and plans for education. To do so, we use the above described variable for the attitude towards school as well as a variable for the plans for education: it asks young individuals

Table 1.15: OLS Regression Plan for School Wave 1

	Men	Women
Next-born Sister	0.0314	0.0295
	(0.0344)	(0.0271)
Observations	2225	2172
Mean	1.602	1.791
Sd	0.760	0.571

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.16: OLS Regression Attitude to School Wave 2

	Men	Women
Next-born Sister	-0.255	0.949**
	(0.392)	(0.373)
Observations	1938	1907
Mean	32.32	33.44
Sd	7.753	7.499

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.17: OLS Regression Plan for Education Wave 2

	Men	Women
Next-born Sister	-0.0237	0.0681**
	(0.0375)	(0.0264)
Observations	1960	1917
Mean	1.594	1.833
Sd	0.767	0.524

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

for their education intentions post 16. The results can be seen in Tables 1.45 and 1.46 reported in Appendix 1.A. Also at age 16 having a sister rather than a brother increases the school attitude of women only. Similarly to what one could see for wave 2, also in wave 3 the plans for education for women are positively influenced by the next-born sibling being female. These results confirm that the gender of the next-born sibling influences women in their educational plans and attitudes. Specifically, having a younger sister rather than a younger brother improves women's attitude towards school and increases their plans of staying in education.

Measures of Gender Roles in the NSD

Since we are interested in gender roles and in whether the sibling gender composition changes adherence to such gender roles, we are also interested in variables which could measure such kind of behavior. However, these are very hard to find with the existing data. In wave 2, we use a variable which measures the number of self-reported hours spent on domestic chores per week by each young person. The output of the regression using this number of hours as an outcome variable can be seen in Table 1.18. The coefficient on the gender of the next-born sibling is much larger for women than for men, however it is not significant. The sign of the effect is in line with the hypothesis that having an opposite-gender sibling makes women conform to more traditional roles and hence also do more hours of domestic chores.

Table 1.18: OLS Regression Hours spent for Domestic Chores Wave 2

	Men	Women
Next-born Sister	0.0519	-0.152
	(0.118)	(0.110)
Observations	1943	1906
Mean	1.821	2.036
Sd	2.220	2.280

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

A further variable we analyze in order to understand whether the gender of the next-born sibling causally influences preferences is the opinion individuals have regarding mothers working. Specifically, individuals are asked whether they agree or disagree with the statement that women with young children should never work full time. For this variable a higher value indicates stronger agreement with the statement and hence a stronger adherence to gender roles. The results of the OLS regression using this variable as an outcome can be seen in Table 1.19. These results indicate that sibling gender has no effect on the agreement with this statement.

Table 1.19: OLS Regression Women with kids should never work full time Wave 6

	Men	Women
Next-born Sister	-0.0130	-0.0409
	(0.0641)	(0.0606)
Observations	1411	1487
Mean	1.610	1.331
Sd	1.075	1.016

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Finally, in wave 7, we use the question on whether women with small children should work full time as a proxy for gender roles. Table 1.20 shows the results for our measure of gender role adherence. Interestingly, having a sister decreases men's adherence to gender roles, whereas there is no effect for women.

Table 1.20: OLS Regression Women with Kids should never work full time Wave 7

	Men	Women
Next-born Sister	-0.150**	-0.00617
	(0.0666)	(0.0667)
Observations	1256	1336
Mean	1.659	1.360
Sd	1.059	1.082

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

In light of our interest on whether the gender of the next-born sibling influences preferences on the liberal-conservative scale, we analyze individuals' perception of the level of discrimination in Britain. A higher value indicates that an individual perceives that there is more discrimination. In Table 1.21 the results of this regression are displayed. The effect of having a sister is not significant neither for women nor for men.

Table 1.21: OLS Regression Discrimination in Britain Wave 5

	Men	Women
Next-born Sister	0.0487	-0.0357
	(0.0364)	(0.0349)
Observations	1518	1563
Mean	2.307	2.369
Sd	0.641	0.620

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Preferences regarding Goals and Satisfaction in the NSD

Tables 1.22 to 1.24 show the results from wave 7 on attitudes towards what is important for individuals in the future. For all three of these variables a higher value signifies a higher importance. Table 1.22 shows the results for the importance of having a career, Table 1.23 for the importance of having a family and Table 1.24 of the importance of studying to obtain a qualification. Interestingly, sibling gender seems to not have an effect on any of these attitudes. However, one has to say that individuals at age 20 are still very young and hence the importance of these different ideas might not be very strong.

Table 1.22: OLS Regression Importance of Having a Career Wave 7

	Men	Women
Next-born Sister	-0.00140	0.0355
	(0.0377)	(0.0370)
Observations	854	946
Mean	2.713	2.770
Sd	0.490	0.477

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.23: OLS Regression Importance of Having a Family Wave 7

	Men	Women
Next-born Sister	0.0956	0.0320
	(0.0619)	(0.0584)
Observations	1261	1339
Mean	3.215	3.385
Sd	0.957	0.915

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.24: OLS Regression Importance of Having Qualifications Wave 7

	Men	Women
Next-born Sister	0.0113	0.0138
	(0.0673)	(0.0643)
Observations	1261	1339
Mean	2.962	3.176
Sd	1.059	1.030

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

We again consider the individual's measure of importance of having a career in Table 1.25, this time however at age 25, when such goals have probably developed to become clearer. We see that for women, having a younger sister rather than a younger brother significantly increases how much importance they attribute to having a career. For men instead, the gender of the next-born sibling has no effect.

Table 1.25: OLS Regression Importance of Having a Job Wave 8

	Men	Women
Next-born Sister	-0.0402	0.0997**
	(0.0550)	(0.0396)
Observations	1081	1266
Mean	3.760	3.750
Sd	0.625	0.587

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Next, we consider general life satisfaction of the individual. The variable is coded such that a higher value indicates lower life satisfaction. The results of an OLS regression with life satisfaction as the outcome variable can be seen in Table 1.26. Interestingly, having a sister rather than a brother causally increases women's life satisfaction (by causing a lower value of the outcome variable). This is a very novel result, since existing papers on the causal effects of sibling gender have so far focused almost entirely on outcomes such as education, occupation and earnings but have not considered attitudes of individuals. Only Golsteyn and Magnée (2020) have focused on behavior and attitudes and hence this result confirms that sibling gender might matter on other dimensions than income and occupation.

A further attitude we consider is the attitude towards success. This variable is an additive measure of agreement with several statements on success. A higher value indicates that success can be obtained through hard work, whereas a lower value indicates that individuals think success is a matter of luck. This variable can be seen as a measure of how fair individuals perceive the world and how much hard work pays off. As one can see from Table 1.47 in Appendix 1.A, the gender of the next-born sibling does not have any significant effect on an individual's attitude towards success.

Table 1.26: OLS Regression Life Satisfaction Wave 8

	Men	Women
Next-born Sister	0.107	-0.187***
	(0.0754)	(0.0599)
Observations	1046	1241
Mean	2.186	2.047
Sd	0.964	0.886

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Sexual Behavior in the NSD

In wave 6 of the Next Steps Data, individuals are asked about their sexual behavior for the first time. Specifically, we use the following questions to measure preferences, attitudes and relationship with their parents: first, individuals are asked if when growing up they found it easy or difficult to talk about sexual matters with their parents. For this variable a higher value indicates that the young individual found it easy to talk to his or her parents about sexual matters. Second, we use the question on whether respondents ever have had sexual intercourse, and third the questions regarding wether they ever have had sex without using precautions or contraception. The question regarding talking to parents about sexual relations is specifically important to us as an outcome, since it can capture parental involvement and how easy such involvement is for the young person. The other outcome of particular interest is the one on having unprotected sex. This variable can be interpreted as a measure of risk preferences.

Tables 1.27 and 1.28 display the results for the questions on sex. As can be seen from Table 1.27, having a sister rather than a brother makes it easier for men to talk to their parents about sexual matters, whereas the gender of the next-born sibling does not seem to influence women in their ease of discussing sex with their parents. This result is in line with results from the literature on fathers of daughters such as the paper by Washington (2008). Having a sister makes either the young person or the parents more liberal and hence makes it easier to talk about sexual matters. This would also explain why there is no effect for women; if there is already one daughter, *i.e.* the first-born, then parents should already be more liberal and hence the gender of the next-born sibling should not influence the ease of talking about sexual matters. Table 1.28 displays the results of the OLS regression when the outcome is an indicator for ever having had sex without contraception. This variable is only available for individuals who have had sex before, hence the sample is smaller than for the other two sex outcomes. The gender of the next-born sibling does not seem to influence the probability of ever having had unprotected sex, which could proxy for risk preferences, neither for men nor for women.

Table 1.27: OLS Regression Ease of Talking to Parents about Sexual Matters Wave 6

	Men	Women
Next-born Sister	0.146*	-0.0468
	(0.0858)	(0.0819)
Observations	1304	1394
Mean	1.420	1.571
Sd	1.380	1.360

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.28: OLS Regression Indicator Ever had Sex without Contraception Wave 6

	Men	Women
Next-born Sister	0.0500	0.00459
	(0.0347)	(0.0334)
Observations	1011	1048
Mean	0.464	0.425
Sd	0.499	0.495

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Educational Outcomes in the NSD

Next we turn to education variables, in order to see whether the gender of the next-born sibling influences very concrete outcomes such as whether an individual is in higher education, is in higher education at a Russel Group university and which qualification an individual is studying for at the moment. In the literature on causal sibling gender effects, so far no one has found significant effects on degrees obtained, hence a priori we do not expect any significant effect. Table 1.29 shows the results for the highest qualification obtained as of the time of the survey. A higher value indicates a higher qualification. The highest value is obtained when individuals are in higher education and studying for a degree; lower values indicate that an individual is still at school and studying for A level, even lower values for GSCE levels. We find a positive effect of having a younger sister, as opposed to a younger brother, for women. This is in line with our results on school attitudes. However, such result is novel to the literature. Table 1.30 sheds more light on this result by considering an indicator for being in higher education as the outcome. Here, however, the strong positive effect of having a sister for women disappears. This is more in line with the trends in the existing literature suggesting that the gender of the sibling does not seem to matter for degree completion. Table 1.31 (wave 7) shows that women with a younger sister are more likely to be in higher education at age 20 than women with a younger brother. In order to see whether this result also leads to differences in degree attainment, we turn to wave 8.

Table 1.29: OLS Regression Highest Qualification Studied for Wave 6

	Men	Women
Next-born Sister	0.0517	0.335*
	(0.187)	(0.198)
Observations	1418	1489
Mean	3.903	4.302
Sd	3.661	3.760

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

In wave 8 of the Next Steps data, we are able to analyze a variable which we were not able to analyze before thanks to the fact that individuals are finishing degrees. This now makes looking at their highest qualification possible. We consider the highest qualification an individual has at age 25. This variable is measure in National Vocational Qualification Levels. A higher value indicates a higher qualification. In line with the existing literature such as Brenøe (2021) and Cools and Patacchini (2019) we find no effect on the highest qualification as can be seen in Table 1.32.

Table 1.30: OLS Regression Indicator for Being in Higher Education Wave 6

	Men	Women
Next-born Sister	0.0333	0.0398
	(0.0238)	(0.0254)
Observations	1411	1486
Mean	0.309	0.369
Sd	0.462	0.483

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.31: OLS Regression Indicator for in Higher Education Wave 7

	Men	Women
Next-born Sister	0.0281	0.0484*
	(0.0258)	(0.0270)
Observations	1260	1339
Mean	0.428	0.510
Sd	0.495	0.500

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.32: OLS Regression Highest Academic Qualification in NVQ Equivalence Wave 8

	Men	Women
Next-born Sister	-0.155	-0.0610
	(0.103)	(0.0987)
Observations	1083	1268
Mean	3.661	3.861
Sd	1.642	1.539

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

1.5.2 Results British Cohort Study

We also include the British Cohort Study as part of this project because it allows us to consider outcomes which happen later in life, such as highest educational qualification when education is completed, completed fertility, but also earnings once education is completed. These are also the outcomes which have been considered in most of the existing literature on causal sibling gender effects. Thus, expanding our analysis to the British Cohort Study allows us to compare our results to those of Cools and Patacchini (2019), Brenøe (2021) or Peter et al. (2018). Still, using the BCS data, we also try to examine outcomes which are similar to those we examined in the Next Steps Data regarding preferences, attitudes and gender roles.

Parental Interactions in the BCS

Like for the Next Steps data, we also want to analyze parental involvement. To do so, we look at the following four variables: father's interest in the child's education, mother's interest in the child's education, alone time with the father, and alone time with the mother. These categorical variables are all coded such that lower values mean higher interest or more time. We then proceed to estimate ordered logit models. Table 1.33 displays the results for boys while Table 1.34 show results for girls. Let us first focus on the results for boys: the gender of the next-born sibling does not impact parental interest in the education of boys, nor alone time with parents. For girls, instead, having a next-born sister rather than a next-born brother significantly increases father's interest in education, whereas it has no effect on mother's interest on alone time with parents[14]. Hence, also in the British Cohort Study, like in the Next Steps data, the gender of the next-born sibling influences parental involvement in education, specifically from fathers.

Educational Preferences and Attitudes in the BCS

In order to further explore attitudes and preferences, we focus on attitudes towards the subject of mathematics in school at age 10. Individuals are asked if they have difficulties with math (No, Some, Yes), how they rate their ability in math (Well, Not well) and if they are good at the subject math (Good, Don't know, Bad). We combine these three variables into one by summing them [15]. For this index, we run an OLS regressions, the

14. The results are very similar if one runs an OLS regression instead.

15. We also create and index by running a principle component analysis and then using the first component and the results are the same.

Table 1.33: Ordered Logit Regressions for Parental Involvement for Boys

	Father's interest in Education	Mother's interest in Education	Alone time with dad	Alone time with mom
Next-born sister	-0.0640	-0.0220	-0.123	-0.198
	(0.142)	(0.120)	(0.164)	(0.156)
Observations	910	1267	558	577
Mean	1.490	1.474	3.100	2.936
Sd	0.694	0.641	1.294	1.357

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.34: Ordered Logit Regressions for Parental Involvement for Girls

	Father's interest in Education	Mother's interest in Education	Alone time with dad	Alone time with mom
Next-born Sister	-0.326*	-0.222	-0.198	0.0499
	(0.148)	(0.126)	(0.132)	(0.130)
Observations	817	1180	791	824
Mean	1.474	1.438	3.308	2.385
Sd	0.705	0.657	1.334	1.308

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

results of which we display in Table 1.35; all three of the variables are coded in a way such that a lower value means more confidence in one's own ability in math and hence so is also the index. As one can see, there seems to be no effect on math confidence at age 10 for boys in the British Cohort Study. For girls however, having a next-born younger sister rather than a next-born brother significantly decreases the math index, which implies a higher math preference, ability or confidence. This result is very much in line with our results from the Next Steps data and confirms that for women having a sister changes preferences; specifically having a sister seems to make preferences less gender-role conforming.

Table 1.35: OLS Regression Index on Math Attitudes

	Men	Women
Next-born Sister	-0.0104	-0.151*
	(0.0816)	(0.0822)
Observations	1279	1237
Mean	4.539	4.862
Sd	1.467	1.451

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Measures of Gender Roles in the BCS

In a next step, we focus on outcomes regarding preferences and attitudes. We do so by first considering an index on how chores are split within the household at age 30. This index considers variables on chores such as shopping, cleaning, repairs, and tending to children[16]. This index is constructed in a way that lower values of the index mean less gender conservative chore division, i.e. higher values imply more traditional chore splitting. Unfortunately, the sample size is very reduced, due to the fact that not all people answer all the questions regarding chores. Table 1.49 in Appendix 1.A shows that the gender of the next-born sibling does not matter for women, whereas it might have some effect on men, but due to the very small sample size, this is not statistically significant.

16. The exact variables are shopping, cleaning, washing, tending to the kids when they are ill, taking care of the kids, repairs, finances, and teaching the kids good manners. We standardize all these variables, ensuring that they go in the same direction, and then we sum them to create the index.

Second, we create an index on gender stereotypes, constructed from the explicit opinion on certain questions answered at age 30 in wave 6[17]. In general, there is little variation in these variables, which is also why the result that the next-born sibling's gender does not have any impact on this index is not surprising. This is shown in Table 1.36.

Table 1.36: OLS Regression Index on Gender Stereotypes

	Men	Women
Next-born Sister	-0.0459	0.0363
	(0.185)	(0.148)
Observations	912	1104
Mean	0.958	-0.788
Sd	2.739	2.421

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Education and Labor Market Outcomes in the BCS

We now focus on the variables, which are harder to measure in the Next Steps data since individuals are still young. First, we focus on education variables. Table 1.37 shows that also in the BCS the gender of the next-born sibling does not seem to influence final education outcomes. Column 1 and 2 of Table 1.37 show the results of the estimation of a linear regression with the age at which an individual left full time education as the outcome variable, whereas column 3 and 4 have highest qualification obtained as an outcome variable in an ordered logit model. Both these variables seem are not influenced by the gender of the next-born sibling.

Next, we turn to strict labor market outcomes and consider earnings. To do so we exploit, just like in the Next Steps Data, the earnings percentile of each individual, calculated separately for each gender. This allows us to keep individuals who have zero earnings in the sample. The effects of having a next-born sister rather than a next-born

17. The exact questions are the following: There should be more women bosses in important jobs in business and industry; When both partners work full-time the man should take equal share of domestic chores; Men and women should have the chance to do the same kind of work; If a child is ill and both parents work , it should usually be the mother who takes care. Again we proceed as above by first standardizing, then making sure the variables go in the same direction and then summing them in order to create the index.

Table 1.37: Education Outcomes: Age left full time Education (ols), Highest Qualification (ologit)

	Age left full time Education		**Highest Qualification**	
	Men	Women	Men	Women
Next-born Sister	-0.120	0.0285	0.115	0.0719
	(0.196)	(0.171)	(0.126)	(0.116)
Observations	928	1114	888	1082
Mean	18.53	18.60	2.776	2.738
Sd	3.197	2.986	1.576	1.482

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

brother on income percentiles can be seen Table 1.38 for women and Table 1.50 in Appendix 1.A for men.

As far as men are concerned, the gender of the next-born sibling has no effect on earnings, whereas for women there is a small positive effect at age 38. However, this effect is not statistically significant. This is the age at which women typically have young children, and hence this result is somewhat in line with the results of Brenøe (2021), who finds that earnings decrease after the birth of the first child for women with brothers in comparison to women with sisters.

In Appendix 1.A we also look at an indicator for being employed in any particular wave of the survey. Both for men and women, we look at this employment indicator for five different waves as can be seen from Tables 1.51 for men and 1.52 for women respectively. The effects on employment status are very small and hence can be seen as precisely estimated zeros. As these indicators are just measurements of precise moments in time, these null results do not come surprising to us.

Finally, in Appendix 1.A, we also consider fertility outcomes. In Tables 1.53 and 1.54 for men and women respectively, we look at the number of own children at age 42 as well as at whether an individual has ever been married or in a civil partnership at age 42. Having a younger sister rather than a younger brother reduces the number of children a woman has at age 42.

Table 1.38: OLS Regression Earnings Percentile for Women

	age 26	age 30	age 34	age 38	age 42
Next-born Sister	2.018	1.929	3.954*	0.215	-0.199
	(1.959)	(1.885)	(2.037)	(2.047)	(1.919)
Observations	1049	1216	1137	967	1032
Mean	46.99	46.64	45.59	47.86	48.61
Sd	32.86	33.34	34.34	31.90	31.24

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

1.6 Survey

As explained above, from the analyses of the publicly available panel data from Britain, we find that sibling gender matters for parental involvement, attitudes even preferences. Whether one has a next-born younger sister or brother however does not matter for the final educational degree obtained. We do find some effects of sibling gender on earnings, which are also in line with the research by Brenøe (2021). With the readily available panel data, we are however not able to shed much light on the mechanisms behind these effects and hence we designed a survey ourselves. This survey will help us better understand how the gender of the next-born sibling influences in particular gender roles and preferences as well as attitudes. The survey, at the moment of submission of this thesis, is ongoing and hence, we can only include the survey questions and how we plan to analyze the responses.

To better understand how the next-born sibling's gender shapes attitudes and preferences we collect novel survey data the UK. Specifically in order to compare the results from our own survey to those we found in the Next Steps Data we focus on a sample of the UK population aged 18 to 30. We also chose this age restriction keeping in mind, that we plan to expand this project by also analyzing data from the Millennium Cohort Study, a study which follows individuals born from 2000 to 2002 in the UK, which is very similar to the Next Steps. Due to the fact that again the identification relies upon being a first-born and knowing the gender of the next-born sibling, our final sample will consist only of first-borns who have at least one younger sibling. We will sample around 600 individuals from the relevant population.

We conduct this survey together with Dynata using Qualtrics as a platform.

1.6.1 Survey Questions

The questions we want to investigate is whether the sibling's gender affects an individual in his or her attitudes towards gender norms, liberalism and preferences. To answer these questions, we propose the following survey to try and reveal individuals attitudes. The survey is organized in the following subsections:

1. Questions on siblings and family life during childhood
2. Questions on parents
3. Questions on gender roles

4. Questions on liberalism

5. Questions on own (gendered) characteristics

6. Background question on respondent.

The first subsection of the survey allows us to select the relevant individuals within the population of interest, namely first-borns with at least one younger sibling. We also collect data on the interaction with siblings and their family life during childhood. In the second subsection, we ask respondents about their parents, and the relationship they have with them. The following subsections, from 3 to 5 provide us with our main outcomes of interest. Each of these subsections are presented to the respondents, but their order is randomized. These subsections aim to elicit attitudes on gender roles (regarding family life and employment), liberalism, and preferences. Note that we detail the content of all subsections below. Finally, subsection 6 contains a few questions on respondents' background, which we will partially use as control variables.

Survey Questions on Siblings

The first part of this subsection includes questions that help us in selecting the relevant population. Individuals are asked whether they have siblings and are first-borns. If that is not the case, the survey is terminated early. Otherwise, the survey continues and individuals are asked about their interactions with their next-born sibling during childhood:
Here, please think about the relationship between you and your next-born sibling. Do you agree or disagree with the following statements?

- *As children we played with the same toys.*
- *As children we read the same books/had the same books read to us.*
- *As children we had the same friends.*
- *As children we had the same chores in the household.*

(Never - Always, 1-5)

Survey Questions on Parents

Next, we ask respondents some general questions about their mother and father such as year of birth and educational attainment. Then, we ask the following about the relationship with their parents:

Now we want to ask you a bit about your parents when you were in school. How were you spending your time during most of your childhood?

- *On school days, I would spend free time with my mother.*
- *On school days, I would spend free time with my father.*
- *On the weekend, I would spend free time with my mother.*
- *On the weekend, I would spend free time with my father.*
- *My mother helped me study for exams and do my homework.*
- *My father helped me study for exams and do my homework.*

(Never - Always, 1-5)
We now have some questions regarding your relationship with your parents nowadays. Here, if your mother/father is deceased, please choose "Not applicable".

- *Do you talk to your mother about things you do or things you have experienced?*
- *Do you talk to your mother about things that bother or worry you?*
- *Do you and your mother find a solution together when you have a problem with each other?*
- *Does your mother ask for your opinion before making decisions on family matters?*
- *Do you talk to your father about things you do or things you have experienced?*
- *Do you talk to your father about things that bother or worry you?*
- *Do you and your father find a solution together when you have a problem with each other?*
- *Does your father ask for your opinion before making decisions on family matters?*

(Very often - Never, 1-5)

Survey Questions on Gender Roles

The questions on gender roles can be further split into to the following subcategories of family life and employment. The questions on gender roles in family life correspond very strongly to those asked within the Understanding Society study. The questions which relate to employment are partially related to the British Social Attitudes study.

The first question, that relates to family life is as follows:
For each of the following statements please tell us how much you personally agree or disagree:

- *The family suffers when the woman has a full time job.*
- *A pre-school child suffers if his or her mother works full time.*

- *A pre-school child suffers if his or her mother works part time.*
- *A woman and her family would all be happier if she goes out to work*
- *Both the husband and wife should contribute to household income.*
- *Having a full-time job is the best way for a woman to be an independent person.*
- *Children need a father to be as closely involved in their upbringing as the mother.*
- *A single parent can bring up children as well as a couple.*
- *When a child is born and paid leave is available, the parents should have to split this leave almost equally in order to qualify for the the full length of paid leave.*

(Strongly agree - Strongly disagree, 1-5)

The second question on gender roles, that relates to employment is:
We would like to know your opinion about these four statements. Please tell us how much you agree or disagree with each of them:

- *To achieve a more equal gender ratio in management positions, gender quotas for these management positions should be introduced in all companies starting from a certain size.*
- *Schools, universities and the government should do more in order to promote female participation in STEM (Science, Technology, Engineering and Math) education.*
- *The government should mandate all firms to report the difference in wages of their male employees and their female employees, hence their gender pay gap.*
- *Employers should make special arrangements to help parents combine jobs and childcare.*

(Strongly agree - Strongly disagree, 1-5)

Survey Questions on Liberalism

As explained above, literature on parents has shown that the gender of the child influences political preferences of parents, specifically of fathers. The literature has shown that having a daughter makes fathers more liberal (in their party preferences, but also as policy makers). Similarly, with the following questions (adapted from the British Election Study and the British Social Attitudes study), we want to analyze whether the gender of the younger sibling has an effect on liberalism. For this part of the analysis, we consider the following three outcomes separately.

First, we ask respondents which party they consider is closest to their views. *Can you tell us which party do you regard yourself as being closest to?*

- *Conservative*
- *Labour*
- *Liberal Democrats*
- *Scottish National Party*
- *Plaid Cymru*
- *Green Party*
- *Other parties*
- *Don't know, Prefer not to say*

Second, respondents are asked about their view on government spending. *Here are some items of government spending. For each of the following items, please tell us whether, in your opinion, government should increase or decrease spending on this item.*

- *Education*
- *Defense*
- *Health*
- *Housing*
- *Public transport*
- *Roads*
- *Police and prisons*
- *Social security benefits*
- *Help for industry*
- *Overseas aid*

(Decrease greatly / Decrease slightly / Stay the same / Increase slightly / Increase greatly) A third and final question in this subsection is the following:
How much do you agree or disagree with the following four statements:

- *An abortion should be allowed if a woman decides by herself she does not want the child.*
- *Teenagers should be able to have an abortion without obtaining their parents' consent.*
- *Contraception should be available free of charge to young people.*
- *Two adults of the same sex should be allowed to have a sexual relationship.*

(Strongly agree - Strongly disagree, 1-5)

Survey Questions on Preferences and Characteristics

In this subsection on preferences and characteristics we ask two questions. We ask the question on job characteristics, in order to better understand the effects of sibling on earnings the existing literature has found (Brenøe and Zölitz (2020) and Cools and Patacchini (2019)). We focus on job characteristics, since these are more related to individual preferences, but nonetheless can have large effects on earnings. We then also focus on individual traits, in order to better understand individual attitudes and preferences on an even deeper level. *For the following items, please tell us how important you personally think each item is in a job. How important is ...*

- *... job security*
- *... high income*
- *... good opportunity for advancement*
- *... an interesting job*
- *... a job that allows someone to work independently*
- *... a job that allows someone to decide their times or days of work*
- *... a job that allows someone to help other people*
- *... a job that is useful to society*

(Very important - Not important at all, 1-5)

The second question relates to the respondents own characteristics. The following group of traits are presented to them sequentially:
Here are a few groups of traits that can be used to describe people. For each of the following groups of traits, please tell us how much you think they describe you.

- *Assertive, competitive, achievement oriented, leadership ability*
- *Nurturing, warm, sensitive, gentle*
- *Dominant, aggressive, arrogant, intimidating*
- *Weak, insecure, yielding, easily frightened*
- *Emotional, moody, melodramatic*
- *Intelligent, analytical, competent, rational*
- *Independent, self-reliant, ambitious*
- *Shy, reserved, nervous, soft-spoken*
- *Active, energetic, athletic*
- *Likable, cheerful, enthusiastic*
- *Helpful friendly, cooperative, dependable*
- *Wholesome, polite, naive*

- *Rebellious, stubborn, angry, self-centered*
- *Noisy, boisterous, rambunctious*
- *Sexually active, promiscuous*
- *Interested in things like languages, arts, and helping others*
- *Interested in things like science, math, technology and mechanical objects*

(Not much like me - Very much like me, 1-5)

Background Survey Questions

Finally, in the last subsection we ask a few background questions such as current employment or education status, whether respondents are UK nationals, whether they have children. These questions will help us in assessing the representativeness of our survey and in creating control variables.

1.6.2 Survey Regression Specification

In order to analyze the data from our own survey we will use the same identification and estimation approach as outlined in Section 1.3. This strategy leads to the following regression equation, which is the same as Equation (1.2) in Section 1.3:

$$Y_i = \alpha + \beta sister_i + \delta X_i + \epsilon_i \tag{1.3}$$

The outcome Y_i is a function of an indicator whether the next-born younger sibling is female, i.e. the individual has a next-born younger sister, covariates X_i and an error term ϵ_i. The covariates X_i which we use will be the year of birth, mother's and father's age at birth, mother's marital status at birth, and mother's and father's educational attainment. Furthermore, we will include individual education as a further covariate. We will use robust standard errors for the estimation.

Again, this regression will be estimated for men and women separately, because the effect of having a younger sister is potentially different for women than it is for men. The coefficient of interest is β, which is the coefficient on the indicator for having a next-born younger sister rather than a next-born younger brother.

Furthermore, in an extension, we plan split the sample according to whether the mother worked during childhood in order to see if effects are concentrated within more traditional families.

1.6.3 Survey Index Construction and Technical Details

In all of the survey sections, we collect several variables. However, very often we expect that these variables within one section are very interrelated and together measure an underlying latent variable such as family-gender-role adherence, liberalism or other attitudes. Therefore, we aggregate the answers to questions into indices. For doing so we follow several procedures, which we will all explain using the example of the questions on gender roles in the family:

For each of the following statements please tell us how much you personally agree or disagree:

1. *The family suffers when the woman has a full time job.*
2. *A pre-school child suffers if his or her mother works full time.*
3. *A pre-school child suffers if his or her mother works part time.*
4. *A woman and her family would all be happier if she goes out to work*
5. *Both the husband and wife should contribute to household income.*
6. *Having a full-time job is the best way for a woman to be an independent person.*
7. *Children need a father to be as closely involved in their upbringing as the mother.*
8. *A single parent can bring up children as well as a couple.*
9. *When a child is born and paid leave is available, the parents should have to split this leave almost equally in order to qualify for the the full length of paid leave.*

(Strongly agree - Strongly disagree, 1-5) Before creating an index, we need to make sure that the statements are all coded in the same direction. For the first three statements a low value implies a more traditional gender roles, whereas for statements 4 to 9, a lower value implies less traditional gender roles. Therefore, we first recode statements 1, 2 and 3.

We then start with the first procedure to create an index: As has been done previously when analyzing gender norms (see for example Farré and Vella (2013) and Vella (1994)), we build an index by simply summing the answers to the different questions. The index is then simply the sum of all answers and a higher value indicates more traditional gender-roles.

The second procedure to construct an index is by using principal component analysis. To do so we perform a principal component analysis on all the statements jointly. We then extract the first principal component and use this as an index.

These procedures give us two different indices, which we use to analyze the following questions:

- Questions on interactions with the next-born sibling during childhood
- Questions on interactions with the mother during childhood
- Questions on interactions with the father during childhood
- Questions on interactions with the mother today
- Questions on interactions with the father today
- Questions on interactions with the mother during childhood
- Questions on gender roles in the family
- Questions on gender roles in employment

For the remaining questions namely those on liberalism, preferences and characteristics we follow slightly different procedures, since we expect these statements not only to catch one underlying latent factor or variable, but rather we expect there to be several.

For the question on party preferences we create an indicator equal to one if the individual chooses one of the three parties we classify as liberal – namely, labour and liberal democrats and green party – and zero otherwise.

For the question on government spending we proceed in the following way: From the literature we know that women in general are more affine towards in pro-social topics (Chattopadhyay and Duflo 2004) and hence we create one index on pro-social spending (education, health, housing, public transport, social security benefits, overseas aid). We create the second index for less social goods (defense, roads, police and prisons, help for industry). Again we proceed like we did for gender norms first simply summing answers to the likert scale questions for the first type of index. Furthermore, we run a principle component analysis on all the spending choices together to see how many components have an Eigenvalue larger than 1. We then also construct an index using PCA.

For the remaining statements in the liberalism section - i.e. those on abortions, contraception and same-sex relationships - we run a factor analysis in order to see how many different factors these variables load on. Depending on how many factors there are, we will aggregate the variables into indices again by simply summing those groups of statements which load onto a factor or through exploratory factor analysis. We proceed the same way when it comes to preferences on job characteristics as well as to personal characteristics.

In order to limit noise caused by variables with minimal variation, questions for which 90 percent of observations have the same value within the relevant sample will be omitted from the analysis and will not be included in any index. In the event that omission decisions result in the exclusion of all constituent variables for an index, the index will be not be calculated.

In our survey, we split outcomes into primary and secondary outcomes, in order not to test too many hypotheses. Our primary outcomes are:

- Index on interactions with the mother during childhood
- Index on interactions with the father during childhood
- Index on interactions with the mother today
- Index on interactions with the father today
- Index on gender roles in the family
- Index on party affiliation
- Indices on liberalism statements (abortions, contraception, same-sex relationships)

Our secondary outcomes are:

- Index on interactions with the next-born sibling during childhood [18]
- Index on gender roles in employment
- Indices on government spending
- Indices on job preferences [19]
- Indices on characteristics

Since we are testing for multiple hypotheses, we need to adjust the obtained p-values for this. To do so, we will use the procedure put forward by Aker et al. (2012), which is a type of Bonferroni correction but allows to correct for the correlation of outcomes. Depending on the group of outcomes, we will postulate a different within group correlation of outcomes - as we obtain from the data by correlating the indices, which act as main outcomes.

18. We also plan on analyzing each statement on sibling interaction individually.
19. We also plan on analyzing each statement on job preferences individually.

1.7 Conclusion and Outlook

This paper causally investigates the effect of sibling gender on a series of individual outcomes. Specifically, we investigate the channels through which the gender of the second-born sibling influences a first-born individual. We thereby contribute to a small literature on the causal effects of having an opposite gender, as opposed to a same gender sibling. We focus on how an individual's attitudes and preferences are influenced by the gender of their sibling, and are the first to do so, to the best of our knowledge. We conduct our analysis using two UK panel data sets, namely the Next Steps Data and the British Cohort Study. These provide us with the opportunity to assess the effect of sibling gender on novel measures of attitudes and preferences, especially for young individuals.

We find statistically significant effects for a broad range of measures of attitudes and preferences. However, these results hold only for women. We find that having a next-born sister rather than a next-born brother does not seem to have an effect on men. This substantial heterogeneity across gender is very much in line with recent literature (see *e.g.* Cools and Patacchini 2019; Brenøe 2021). Women with a next-born sister rather than a next-born brother experience a differentially higher parental involvement in their education. Further, having a next-born sister significantly influences their educational attitudes. It increases their preferences for STEM subjects in school, improves their attitudes towards mathematics and even improves their overall attitude towards school. Even though we find an effect of sibling's gender on attitudes towards school, we find no effect on the highest qualification obtained. All of these results are reflected both in the Next Steps data as well as in the British Cohort study and hence hold true for different cohorts of women.

Additionally, the Next Steps data allows us to investigate subtler outcomes. We find that sibling gender has no effect on the probability of engaging in sexual activities without using contraception – a proxy for risk preferences. Interestingly, we find that that men with a second-born sister are more prone to talk to their parents about sexual matters than men with a second-born brother. This result, coupled with the effect of sibling's gender on parental involvement in education make clear that the gender composition of siblings changes family dynamics. Incidentally, looking at life goals, we find no effects on outcomes regarding importance of having a career, family or qualifications at age 20. However, already at age 25, women with a sister place a higher importance on having a career. Besides, women with a sister experience a higher general life satisfaction at age 25 than women with a next-born brother.

All in all, we find that having a sister rather than a brother increases parental involvement in young women's education, shifts their educational preferences towards more STEM oriented subjects and generally improves their attitudes towards education. The presence of a next-born sister also causes women to plan to stay in education longer, even though this does not seem to develop into a differential achievement of a qualification. Our results complement those of two recent papers on the subject. Cools and Patacchini (2019) and Brenøe (2021) study the effect of sibling's gender on earnings. They find that for women, having a sister has a positive effect on earnings, as compared to having a brother. At any rate, our results show that the effect of having a sister rather than a brother influences women's underlying preferences. This effect, in turn, might influence many more outcomes than earnings alone.

Gender roles have been put forward by the literature as one of the main mechanisms behind the effects of sibling gender. However, our results are bound by the limitation of the Next Steps' and British Cohort Study's data. The extent of the questionnaire does not allow us to explore in detail certain attitudes towards gender roles, preferences towards job characteristics or political preferences. Therefore, to measure these attitudes more precisely, we have designed our own online survey, which is ongoing in a population of young adults in the UK. With the analysis of our survey, we hope to contribute to the literature in an even more substantial way: to the best of our knowledge, we are the first to explore the effect of sibling gender on explicit gender stereotypes and gender roles. We additionally plan to consider another panel data set from the UK – the Millennium Cohort Study. This will allow us to examine individuals at an even younger age than those of Next Steps Data.

References

Aker, J. C., R. Boumnijel, A. McClelland, and N. Tierney. 2012. "Zap it to me: The impacts of a mobile cash transfer program." *Unpublished manuscript.*

Anelli, M., and G. Peri. 2014. "Gender of siblings and choice of college major." *CESifo Economic Studies* 61 (1): 53–71.

Angrist, J. D., and W. N. Evans. 1996. *Children and their parents' labor supply: Evidence from exogenous variation in family size.* Technical report. National bureau of economic research.

Black, S. E., P. J. Devereux, and K. G. Salvanes. 2005. "The more the merrier? The effect of family size and birth order on children's education." *The Quarterly Journal of Economics* 120 (2): 669–700.

Breining, S., J. Doyle, D. N. Figlio, K. Karbownik, and J. Roth. 2020. "Birth order and delinquency: Evidence from Denmark and Florida." *Journal of Labor Economics* 38 (1): 95–142.

Brenøe, A. A. 2021. "Brothers increase women's gender conformity." *Journal of Population Economics:* 1–38.

Brenøe, A. A., and U. Zölitz. 2020. "Exposure to more female peers widens the gender gap in stem participation." *Journal of Labor Economics* 38 (4): 1009–1054.

Butcher, K. F., and A. Case. 1994. "The effect of sibling sex composition on women's education and earnings." *The Quarterly Journal of Economics* 109 (3): 531–563.

Chattopadhyay, R., and E. Duflo. 2004. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72 (5): 1409–1443.

Collins, M. 2021. *Sibling Gender, Inheritance Norms and Educational Attainment: Evidence from Matrilineal and Patrilineal Societies.* Technical report. Working paper.

Conley, D. 2000. "Sibship sex composition: Effects on educational attainment." *Social Science Research* 29 (3): 441–457.

Cools, A., and E. Patacchini. 2019. "The brother earnings penalty." *Labour Economics* 58:37–51.

Cyron, L., G. Schwerdt, and M. Viarengo. 2017. "The effect of opposite sex siblings on cognitive and noncognitive skills in early childhood." *Applied Economics Letters* 24 (19): 1369–1373.

Dahl, G. B., and E. Moretti. 2008. "The demand for sons." *The Review of Economic Studies* 75 (4): 1085–1120.

Detlefsen, L., A. Friedl, K. Lima de Miranda, U. Schmidt, and M. Sutter. 2018. "Are economic preferences shaped by the family context? The impact of birth order and siblings' sex composition on economic preferences." *The Impact of Birth Order and Siblings' Sex Composition on Economic Preferences (November 2018). MPI Collective Goods Discussion Paper,* nos. 2018/12.

Dossi, G., D. Figlio, P. Giuliano, and P. Sapienza. 2021. "Born in the family: Preferences for boys and the gender gap in math." *Journal of Economic Behavior & Organization* 183:175–188.

Endendijk, J. J., M. G. Groeneveld, and J. Mesman. 2018. "The gendered family process model: An integrative framework of gender in the family." *Archives of sexual behavior* 47 (4): 877–904.

Farré, L., and F. Vella. 2013. "The intergenerational transmission of gender role attitudes and its implications for female labour force participation." *Economica* 80 (318): 219–247.

Golsteyn, B. H., and C. A. Magnée. 2020. "Does sibling gender affect personality traits?" *Economics of Education Review* 77:102016.

Heckman, J., R. Pinto, and P. Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review* 103 (6): 2052–86.

Hiedemann, B., J. M. Joesch, and E. Rose. 2004. "More daughters in child care? Child gender and the use of nonrelative child care arrangements." *Social Science Quarterly* 85 (1): 154–168.

Hirnstein, M., L. C. Andrews, and M. Hausmann. 2014. "Gender-stereotyping and cognitive sex differences in mixed-and same-sex groups." *Archives of sexual behavior* 43 (8): 1663–1673.

Ichino, A., E.-A. Lindström, and E. Viviano. 2014. "Hidden consequences of a first-born boy for mothers." *Economics Letters* 123 (3): 274–278.

Kabátek, J., and D. C. Ribar. 2021. "Daughters and divorce." *The Economic Journal* 131 (637): 2144–2170.

McHale, S. M., K. A. Updegraff, J. Jackson-Newsom, C. J. Tucker, and A. C. Crouter. 2000. "When does parents' differential treatment have negative implications for siblings?" *Social Development* 9 (2): 149–172.

Peter, N., P. Lundborg, S. Mikkelsen, and D. Webbink. 2018. "The effect of a sibling's gender on earnings and family formation." *Labour Economics* 54:61–78.

Ronchi, M. 2021. *Daddy's Girl.* Technical report. Working paper.

Sacerdote, B. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly journal of economics* 116 (2): 681–704.

Schneeweis, N., and M. Zweimüller. 2012. "Girls, girls, girls: Gender composition and female school choice." *Economics of Education review* 31 (4): 482–500.

University College London, UCL Institute of Education, Centre for Longitudinal Studies. 2021a. *1970 British Cohort Study Response Dataset, 1970-2016.*

———. 2021b. *Next Steps: Sweeps 1-8, 2004-2016.*

Vella, F. 1994. "Gender roles and human capital investment: The relationship between traditional attitudes and female labour market performance." *Economica:* 191–211.

Washington, E. L. 2008. "Female socialization: how daughters affect their legislator fathers." *American Economic Review* 98 (1): 311–32.

Appendix 1.A Appendix

1.A.1 Summary Statistics

Table 1.39: NSD Summary Statistics on Gender of Next-born Sibling

	Gender of Next-born sibling		
Gender of the individual	Male	Female	Total
Male	1178	1102	2280
Female	1109	1101	2210
Total	2287	2203	4490

Table 1.40: BCS Summary Statistics on the Gender of the Next-Born Sibling

	Next-Born Sibling Gender		
Gender of the Individual	Male	Female	Total
Male	919	858	1,777
Female	856	833	1,689
Total	1,775	1,691	3,466

1.A.2 Balance Tests

Furthermore, we also analyze other variables for balance, which however we do not include as covariates. These can be seen in Tables 1.42 and 1.43, respectively for men and women. It is not surprising that for both men and women the number of total siblings differs depending on whether the individual has a same or opposite gender sibling. This is a result well known in the literature (see *e.g.* Angrist and Evans 1996). For this reason, we do not include the number of siblings as a covariate, since it is a byproduct of having an opposite versus a same gender sibling. Hence, the causal effect of having an opposite gender sibling also includes the effect of having less siblings on average. Only for women is the age difference significantly different depending on the gender of the next-born sibling, hence we also exclude this variable as a covariate, since it would otherwise potentially pick up some of the effect of interest. The other variables which are tested for balance in Tables 1.42 and 1.43 are variables which we expect to be influenced by the gender of the second-born sibling: the age at which the mother returned to work might very well depend on the gender of the second-born child, since if this is a boy one might expect mothers to return to work later, due to

Table 1.41: **NSD Detailed Balance Tests for Women for Mother's Education and Single at Birth**

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
g_m_hiqual1	1109	0.086 (0.010)	1101	0.051 (0.007)	0.036***
g_m_hiqual2	1109	0.116 (0.011)	1101	0.108 (0.011)	0.007
g_m_hiqual3	1109	0.124 (0.011)	1101	0.122 (0.011)	0.003
g_m_hiqual4	1109	0.123 (0.011)	1101	0.141 (0.012)	-0.017
g_m_hiqual5	1109	0.312 (0.015)	1101	0.313 (0.016)	-0.001
g_m_hiqual6	1109	0.088 (0.009)	1101	0.089 (0.009)	-0.001
g_m_hiqual7	1109	0.011 (0.003)	1101	0.013 (0.004)	-0.001
g_m_hiqual8	1109	0.139 (0.011)	1101	0.165 (0.012)	-0.025
g_single1	1109	0.107 (0.010)	1101	0.095 (0.010)	0.011
g_single2	1109	0.708 (0.015)	1101	0.682 (0.015)	0.026
g_single3	1109	0.186 (0.013)	1101	0.223 (0.014)	-0.037**

Notes: The value displayed for t-tests are the differences in the means across the groups. Observations are weighted using variable W1FinWt as pweight weights.***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.42: **NSD Balance Tests for Men Additional Variables**

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Age Difference	1178	3.316	1102	3.371	-0.054
		(0.056)		(0.061)	
Number of Siblings	1178	1.611	1102	1.460	0.151***
		(0.028)		(0.025)	
Age Mother when returned to Work	1178	2.154	1102	2.044	0.110
		(0.132)		(0.128)	
Whether attended Nursery	1178	0.799	1102	0.788	0.011
		(0.015)		(0.016)	
Family Type	1178	1.725	1102	1.675	0.050
		(0.039)		(0.039)	

Notes: The value displayed for t-tests are the differences in the means across the groups. Observations are weighted using variable W1FinWt as pweight weights.***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

the fact that boys are said to be harder to parent (see *e.g.* Hiedemann, Joesch, and Rose 2004). This might itself again influence whether a child attended nursery and hence we also exclude this variable as a covariate. The family type is measured only at age 14 of the individual and hence the younger sibling is around 11-12 years old on average. As is known from Kabátek and Ribar (2021) parents of teenage daughters are more likely to divorce and hence we also did not expect balance on this variable a priori and therefore we do not use it as a covariate. We conclude that, apart from the age difference towards the next-born sibling, there are no variables which we consider as being critically imbalanced across the gender of of the next-born sibling.

1.A.3 Results Next Steps Data

We also consider income as an outcome variable however, the only continuous measure of income in wave 8 of the Next Steps Data is weekly household income, hence this is what we use even though it potentially measures parental income for many individuals. To do so, we calculate the income percentile of each individual, separately by gender in order not to exclude individuals who report to have zero income. We then run a OLS regression with income as the outcome with the usual covariates and then again

Table 1.43: **NSD Balance Tests for Women Additional Variables**

Variable	N	Next-Born Man Mean/SE	N	Next-Born Woman Mean/SE	T-test Difference (1)-(2)
Age Difference	1109	3.250	1101	3.446	-0.195**
		(0.059)		(0.062)	
Number of Siblings	1109	1.436	1101	1.633	-0.198***
		(0.023)		(0.030)	
Age Mother when returned to Work	1109	2.396	1101	2.120	0.277
		(0.139)		(0.135)	
Whether attended Nursery	1109	0.803	1101	0.825	-0.022
		(0.015)		(0.014)	
Family Type	1109	1.735	1101	1.653	0.082
		(0.041)		(0.038)	

Notes: The value displayed for t-tests are the differences in the means across the groups. Observations are weighted using variable W1FinWt as pweight weights.***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1.44: **OLS Regression Favorite Subject is STEM Wave 2**

	Men	Women
Next-born Sister	0.0138	0.0373**
	(0.0228)	(0.0183)
Observations	1999	1938
Mean	0.301	0.156
Sd	0.459	0.363

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.45: **OLS Regression Attitudes Towards School Wave 3**

	Men	Women
Next-born Sister	-0.0428	0.735*
	(0.444)	(0.418)
Observations	1799	1782
Mean	32.92	33.64
Sd	8.393	8.084

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.46: **OLS Regression Plans for Education Post 16 Wave 3**

	Men	Women
Next-born Sister	-0.0478	0.0502*
	(0.0396)	(0.0265)
Observations	1811	1797
Mean	1.645	1.860
Sd	0.751	0.495

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

Table 1.47: **OLS Regression Attitude Towards Success Wave 7**

	Men	Women
Next-born Sister	-0.124	-0.0554
	(0.160)	(0.153)
Observations	1250	1332
Mean	10.30	9.763
Sd	2.486	2.432

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

including the individual's highest qualification as an additional covariate. The output of both of these regressions can be seen in Table 1.48. Not only are the effects of having a sister rather than a brother on income insignificant, but they are also very small and hence can be understood as true zeros. This could be due to several factors: First, individuals are still very young at age 25 and the literature has typically looked at income after the birth of the first child. Second, the income measure, namely household income is not the correct measure for income of the individual. Third, it could truly be that for individuals in this age group sibling gender does not affect income.

Table 1.48: OLS Regression Household Income in Percentiles Wave 8

	Men	Women	M with Hiqual	W with Hiqual
Next-born Sister	-1.394	-0.227	-0.851	-0.105
	(1.489)	(1.426)	(1.400)	(1.361)
Observations	1083	1268	1083	1268
Mean	47.66	49.95	47.66	49.95
Sd	29.17	28.16	29.17	28.16

Standard errors in parentheses

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

Notes: Covariates used are birth weight, and indicator for being UK born, ethnicity, mother's age at birth in bands, an indicator for the mother being single at birth, the mother's highest qualification and the father's highest qualification. Standard errors are robust.

1.A.4 Results British Cohort Study

Table 1.49: OLS Regression Index on Chore Division

	Men	Women
Next-born Sister	-0.367	0.0307
	(0.385)	(0.320)
Observations	357	496
Mean	0.194	0.884
Sd	3.502	3.422

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.50: **OLS Regression Earnings Percentile for Men**

	age 26	age 30	age 34	age 38	age 42
Next-born Sister	0.0117	-0.0275	-0.0138	-0.0155	0.0380
	(0.0296)	(0.0337)	(0.0339)	(0.0351)	(0.0367)
Observations	721	911	830	704	763
Mean	6.879	7.134	7.370	7.619	7.677
Sd	0.409	0.510	0.492	0.482	0.513

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.51: **OLS Regression Employment Status for Men**

	age 26	age 30	age 34	age 38	age 42
Next-born Sister	0.0207	-0.00852	0.00165	-0.0151	-0.00619
	(0.0230)	(0.0168)	(0.0139)	(0.0168)	(0.0160)
Observations	931	1207	1046	968	1075
Mean	0.858	0.909	0.945	0.927	0.929
Sd	0.349	0.288	0.229	0.261	0.256

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.52: **OLS Regression Employment Status for Women**

	age 26	age 30	age 34	age 38	age 42
Next-born Sister	1.830	-0.375	-0.429	-1.494	0.568
	(2.031)	(1.831)	(1.922)	(2.033)	(1.996)
Observations	853	1068	1011	775	839
Mean	48.89	49.23	48.64	49.69	49.87
Sd	30.39	30.30	30.91	29.41	29.40

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.53: OLS Regression Fertility Outcomes for Men

	Number of children at age 42	Ever Married at age 42
Next-born Sister	-0.0512	0.125
	(0.0738)	(0.156)
Observations	1062	1078
Mean	1.558	3.566
Sd	1.218	2.538

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Table 1.54: OLS Regression Fertility Outcomes for Women

	Number of children at age 42	Ever Married at age 42
Next-born Sister	-0.147**	0.139
	(0.0715)	(0.142)
Observations	1180	1188
Mean	1.775	3.399
Sd	1.219	2.415

Standard errors in parentheses

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

Covariates used are birth weight, mother's age at birth, a dummy for mother's age at birth missing, as well as for birth weight missing, mother's social class at birth, region of residence at birth and mother's marital status at birth.

Chapter 2

Liberalizing the Morning After Pill - Effects on Young Women

2.1 Introduction

Since the seminal works of Arrow (1963), there has been a long history of health economists thinking about moral hazard. In the insurance example originally presented by Arrow the fear is that insured individuals will display more risky behavior such as not wearing seatbelts or smoking more because they know they can fall back on the insurer to pay the costs of their behavior. More generally, apart from insurers also policymakers worry that policies might have unintended consequences which mean that individuals start taking more risky choices. In such contexts, the state or government worries not only about how individual behavior changes, but also how these behavioral changes affect society as a whole: through costs for the individual, the state as an insurer or for other individuals.

In this paper, I study the causal effects of making emergency hormonal contraception more readily available on the reproductive choices of young women. In particular I focus on births, abortions and sexually transmitted infections. I shed light on this question both from a theoretical and an empirical point of view. To do so, I build a model, extending on Levine and Staiger (2002) and Ananat et al. (2009), where women choose risk and EHC intake and finally have an abortion or a baby. Using English data on abortions, births and sexually transmitted infections, I estimate the causal effect of the reform using a difference-in-differences approach. I find that the liberalization only increased births for the age group of 20-24 year old women. However, it had no effects on abortions or sexually transmitted diseases.

The policy change, which took place in England in January 2001, allowed all women aged 16 and over access to emergency hormonal contraception in pharmacies without needing a physician's prescription. Prior to this liberalization all women needed a physician's prescription in order to access EHC in a pharmacy, whereas afterwards this requirement remained in place only for women under the age of 16. This specific setting makes it possible to use the group of under 16 year old women as a control group, since they were not affected by the reform, thereby creating a perfect situation to obtain a causal estimate. I focus on reproductive choices as the specific risky behavior of interest. In this setting, risky behavior, i.e. if individuals choose not to use contraception, can have a direct effect on them via higher chances of falling pregnant or contracting an STI. However, it can also create externalities for society through higher medical costs and increased possibility of STI transition to others. Furthermore, all of these could increase the costs on the state through, for example, increased financial support to families.

The regulation of EHC is a perfect example of how policies can potentially have unintended effects. Emergency hormonal contraception is a way of preventing a possible pregnancy after non-protected intercourse. It is one of the only options to reduce the probability of a pregnancy for women who either used no contraception, or for whom other contraceptive measures failed. EHC, which is also known as the morning after pill, works by preventing ovulation . Thus, avoiding a pregnancy from happening, rather than ending it. This delay of ovulation works worse the closer a woman is to ovulation. Therefore, EHC is only effective when taken before ovulation and as quickly as possible after the relevant intercourse, so as to avoid getting too close to ovulation.[1]

The time pressure associated with the usage of EHC is one of the arguments which caused the liberalization of EHC: rigid opening hours of doctor's offices potentially cause women to delay their intake of emergency hormonal contraception to a point in time when it is less effective or no longer effective. Since pharmacies offer more flexible hours, women have the chance to access EHC faster when it is offered in pharmacies without prescription rather than on doctor's prescription only. In light of a possible liberalization, some worried about possible adverse effects: opponents argued that making EHC more easily available could cause women to engage in more risky

1. It is important to note that EHC is different from an abortion pill since it only delays ovulation, but once an egg cell is fertilized it is not affected by EHC (Black et al. 2004).

sexual behavior since EHC is perceived as an easy insurance against pregnancies. The particular downside to this is the fact that EHC, while preventing pregnancies cannot protect against sexually transmitted infections. In addition, EHC works less well than other methods of regular contraception, thereby potentially causing more unwanted pregnancies when women switch from regular to emergency contraception.

Therefore, I analyze the causal effects of the liberalization of emergency hormonal contraception on the reproductive behavior of young women in England. The focus lies on young women for several reasons: First, the policy's unintended consequences might be specifically strong: teenage pregnancies are are shown to be a driver of poverty, reduce education of teenage mothers, and also have strong, long lasting effects on the children of teenage mothers[2]. Second, young women are also more likely to use EHC; hence, the impact of the reform is more relevant for them[3]. From this analysis, I am able to learn whether the intended effects of reducing abortions and births, or the unintended effects of increasing STDs, abortions and births dominate.

I explore the issue both theoretically and empirically. I start by building a model of the decision process of a woman when deciding whether or not to participate in a one time sexual encounter. First women choose risk; then, they decide on whether or not to take EHC; finally, women have an abortion or a pregnancy. In this model, a decrease in the cost of EHC leads to an ambiguous effect on pregnancies due to two counteracting effects. On the one hand, EHC reduces the number of pregnancies, by providing an additional means of contraception. On the other hand, the lower costs of EHC lead some individuals to take more risk causing an increase in pregnancies. The reason for this is that the availability of EHC allows for leeway in the risk decision in the first step, by providing a further way to prevent a pregnancy.

In addition, I analyze the reform empirically using a difference-in-differences approach. The fact that EHC was only liberalized for women aged 16 and older allows me to use women under the age of 16 as a control group. I then compare how abortions, births, and sexually transmitted infections (all measured in natural logarithms) change differentially over time for women aged 16 and older in comparison to women aged under 16. This comparison, under the assumption of common trends, allows me to uncover the causal effects of the reform. This assumption requires that absent the reform, all outcome variables for the treated groups would have moved similarly to those of the control group (under 16 year olds). I inspect these common trends visually

2. See Hendrick and Maslowsky (2019) and Hoffman and Maynard (2008)
3. See Daniels and Abma (2013) as well as Figure 2.11

and also test for them, in order to convince the reader that this assumption is satisfied. Under this assumption, my analysis estimates the causal intention to treat effect (ITT) of the reform on young women in England, aged 16 to 25. It is an intention to treat estimate, since I cannot condition my estimate on who took EHC. However, in this context, the ITT is the estimator of interest, since as one can see from my model, the EHC liberalization can affect individuals without them actually taking EHC.

From this empirical analysis I find that births to women aged 20-24 increased significantly after the reform. I find no effects on abortions or STIs, specifically when using Randomization Inference (RI) to obtain robust p-values. Randomization Inference p-values allow me to not make any assumptions on the distribution of error terms and obtain correct p-values for estimates even in light of serially correlated standard errors and few clusters, which is often the case in difference-in-differences estimations (Bertrand, Duflo, and Mullainathan 2004). Randomization Inference, in this context, allows me to test the null hypothesis of no treatment effect. The RI p-values, then do not reflect uncertainty stemming from repeated sampling from an unknown distribution, but rather uncertainty from repeated permutations of treatment.

This paper contributes to the literature in two separate ways: First, I contribute to the theoretical literature on EHC and abortions, by building a model, which allows for an endogenous choice of of contraception risk and EHC intake. Using this model, I find that when making EHC more available, risk taken increases. The effect on pregnancies and hence also on births and abortions however is unclear. This is due to the fact that in the model EHC on the one hand induces higher risks, thereby creating more pregnancies, but on the other hand EHC reduces pregnancies. Therefore, overall effects remain unclear, since the two effects on pregnancies, and hence also on births and abortions, go in different directions. My model is able to reconcile my empirical findings: no effects on abortions at the same time as positive effects on births.

My second contribution lies in my empirical strategy. Other papers, which have explored this issue, often exploit regional variation using difference-in-differences specifications. A potential problem of these studies, is that in order for the assumption of common trends to hold, these regions, states or countries would have to behave similarly in the absence of the treatment. However, in the case of regional variation in the timing of liberalization, it is often the case that the regions, states or countries choose themselves when to liberalize and hence self-select into treatment. This is a problem for the identification because if regions which have increasing abortion rates, choose to liberalize EHC one cannot assume that in absence of the reform they would have behaved similarly to other regions. The advantage of using an age rather than a regional discontinuity is that individuals unlike regions cannot self-select into treatment. Furthermore, in the English setting, all age groups were treated at the same time, which also reduces the fear of selection into treatment based on differing trends. This empirical analysis and causal identification of the intention to treat effect is where my main contribution to the existing literature lies. From this analysis I find that only births to women aged 20 to 24 years old increase significantly due to the liberalization. Effects on births to other age groups are not significant, nor are effects on abortions or STDs. These findings can be explained by model, since a priori a decrease in EHC costs has ambiguous effects on births and abortions. These empirical findings are very much in line with recent empirical research on EHC liberalization in Europe by Pfeifer and Reutter (2020), who also find an increase only in births for women in a similar age range due to liberalization of the morning after pill.

The remainder of this paper is structured as follows. Section 2.2 reviews the literature on Emergency Hormonal Contraception and Section 2.3 details EHC in its workings and regulations. For the theoretical model, see Section 2.4. Sections 2.5 and 2.6 shed light on the data and empirical strategy I use. I report the results in Section 2.8 and additional results and robustness checks are reported in Section 2.9. Section 2.7 discusses how the reform effected EHC consumption, and Section 2.10 concludes.

2.2 Literature Review

The objectives of the existing literature on changes in the accessibility of emergency hormonal contraceptives are often twofold. On the one hand medical literature which tries to estimate by how much pregnancy probabilities are reduced by taking emergency contraception and on the other hand economic but also medical literature which tries to assess how differential access policies can effect reproductive behavior. The literature on differential accessibility can, furthermore, be split into two categories: one using randomized control trials in which the treatment group of women is given an in advance supply of emergency contraception, in practice this means that women, when asking for EHC at the doctor's office are not only given one package of EHC, but are provided with several packages in order to use them in the future the next time they have unprotected sex; the other one using differences in legal access policies. The latter is the smaller but potentially more relevant part of the literature in order to assess population effects of liberalization. My paper aims to contribute to this branch of the literature.

Randomized control trials with advance supply of EHC have to be viewed with much caution because the sample selection. Women who make up both treatment and control group are either clients of family planning centers or already seeking a physician's prescription for EHC. Therefore, the effects of easier access in the form of advance provision in this group might be very different from the effects in the entire population. Furthermore, these randomized control trials cannot take into account how easier access to EHC might influence the behavior of individuals in regards to starting to engage in sexual activities.

The literature more closely related to my paper is the one which exploits differences in access due to legal reasons, many of which are concentrated on the US. Cintina and Johansen (2015), like me use, an age discontinuity which arises due to the FDA ruling of 2006 which made EHC available without a prescription for women aged 18 and older. They find that the liberalization led to a moderate decrease in abortion rates. They do not consider birth rates or rates of infections with sexually transmitted diseases which is where I add to this strand of the literature.

Both Gross, Lafortune, and Low (2014) and Durrance (2013) analyze state policies which were in place before the 2006 FDA ruling to see if pharmacy access decreases abortion rates. Durrance (2013) analyzes only the state policies in place in Washington and finds no significant effect on abortion or birth rates. It is one of the first population studies of EHC, and therefore, is very influential. The pharmacy access granted in Washington only made it possible for some pharmacies to grant access without a prescription. This however, poses a potential downside of the setting analyzed by Durrance (2013): since only some pharmacies were able to provide over the counter access it is not clear how well known the policy was. Also Gross, Lafortune, and Low (2014), who analyze multiple state policies as well as the FDA ruling of 2006 by exploiting the regional variation it created, find a zero effect on abortion and birth rates.

Girma and Paton (2011), like me, use English data to analyze differential access to EHC and its effects on reproductive behavior, but they exploit the fact that between 1999 and 2010 a large project and strategy to prevent teen pregnancy was put in place. Within this project, some local authorities provided pharmacies with the option to give out emergency contraception to teenagers for free and without prescription. Using a difference-in-differences approach, they find that such free of charge and prescription free access to EHC does not decrease teenage conception rates.

Pfeifer and Reutter (2020) use data from Europe in a cross country analysis, which spans 28 countries over 18 years. They find no effect on abortions but a significant increase in births for women aged 25 to 34.

My paper, unlike most of the before mentioned existing studies, which try to look at a difference in prescription requirements by exploiting regional variation in prescription requirements, does not rely on the assumption that the regions are similar. The advantage of using an age rather than a regional discontinuity is that individuals unlike regions cannot self select into treatment. This is where my main contribution to the existing literature lies.

There is also theoretical literature on reproductive behavior. Yet, rarely do these models account for emergency contraception. Many of the models see abortion as a form of insurance against pregnancies such as Levine and Staiger (2002) and Ananat et al. (2009) when modeling behavior of individuals who vary in their preferences for having a baby. I build upon this existing literature when it comes to the theoretical part of my paper, however, adding the EHC into the choice set of women. This endogenous choice of whether or not to take EHC is my novel contribution to the theoretical literature on abortions. Like Gross, Lafortune, and Low (2014) I extend the model ofAnanat et al. (2009) to include a choice to take emergency contraception. However, in their model all women who take EHC decide to do so before their decision to have intercourse, which makes it hard to convey that EHC is for emergencies.

I contribute to this literature by allowing for an endogenous decision of risk in my model, as well as an EHC choice after intercourse has taken place and some more information is revealed.

2.3 Emergency Hormonal Contraception

Emergency hormonal contraception (EHC) is a medication taken to prevent a pregnancy after having had unprotected sex. Hence, the two main uses are when either other contraceptive measures failed, such as when a condom breaks, or when no contraception was used, specifically also in cases of rape. All EHC pills prevent pregnancies, after intercourse has already occurred, by delaying or preventing ovulation through hormones and hence they can only work if a woman has not yet ovulated. This delay works worse the closer to ovulation a woman is. This method of functioning makes the timely intake of EHC extremely important, in order to ensure ovulation does not occur such that unwanted pregnancies can be prevented.

As of 2021, in most European countries two different EHC pills are available: one product based on the active ingredient levongestrel (LNG), the other and newer product based on the active agent ulipristal acetate (UPA). As mentioned above, both products work by delaying or inhibiting ovulation, but they differ in their active agents and hence also in how much time one can let pass after the relevant intercourse before taking them. The LNG based product needs to be taken within 72 hours of unprotected intercourse, whereas the UPA based product needs to be taken within 120 hours. However, at the time of interest for this project, only the LNG based product was available in England, hence I focus on this product and from here on I use EHC solely for LNG based EHC.

EHC was originally a product which could be bought in pharmacies only with a doctor's prescription. Yet, over the last 20 years, the product has be reclassified in most countries allowing women to access EHC medication in pharmacies without prescription. The change from on prescription to without prescription made it easier for women to access EHC in a more timely manner, especially on weekends or at night, when doctor's offices are typically closed. It is important to note that EHC is different from an abortion pill since it only delays ovulation, but once an egg cell is fertilized it is not affected by EHC (Black et al. 2004). This is specifically important in the context of liberalization of access also for young adults.

When thinking about EHC liberalization, it is crucial to not only know how EHC works, but also how effectively EHC works. LNG based EHC is said to have an effectiveness of around 85-88% Weismiller (2004). This effectiveness however, does not mean that 15% of women who take EHC fall pregnant. Rather it means that if 1000 women have unprotected sex during the middle two weeks of their menstrual cycles, approximately 80 of them will fall pregnant. Then the use of LNG-EHC would reduce this by 85 percent and hence only 12 of these women would fall pregnant Weismiller (2004). This effectiveness is however only valid for women who are in the middle two weeks of their cycle, women in either the first or the last week of their cycle are much less likely to fall pregnant and hence taking LNG based EHC will not change this by much.

From the functioning of EHC it becomes clear why a timely intake is important: The time pressure involved in the usage of EHC is one of the arguments which caused a liberalization of EHC. Rigid opening hours of doctor's offices potentially cause women to delay their intake of emergency hormonal contraception to a point in time when it is less effective or not at all effective any longer. Since pharmacies offer more flexible hours, women have the chance to access EHC more quickly when EHC is offered without prescription in pharmacies rather than on doctor's prescription only. However, there are also possible downsides involved in this form of liberalization of EHC: opponents worry that making EHC more easily available could cause women to engage in more risky sexual behaviors since EHC is perceived as an easy insurance against pregnancies. One particular downside to this, is the fact that EHC, while preventing pregnancies cannot protect against sexually transmitted infections. Also EHC works less well than other methods of contraception, thereby potentially causing more unwanted pregnancies when women switch from regular to emergency contraception. A further concern expressed especially by doctors is the fact that traditionally, when women came to ask for emergency hormonal contraception, doctors unlike pharmacists had the option of not only prescribing EHC but also a method of regular contraception, thereby potentially preventing future needs for EHC and future unwanted pregnancies. (See DGGG (2015) for an opinion of doctors.) These arguments of opponents and supporters of the reform show clearly why it is necessary on analyze the liberalization of emergency hormonal contraception.

The specific reform and liberalization I analyze is the one which took place in England in 2001, to do so, it is important to know exactly what the situation regarding EHC at the time looked like: Already in 1982 a specific EHC product became available in the UK Glasier et al. (1996). However, this product and some of the ones following were not a single pill to be taken, but rather several pills to be taken over several days. In 2000 EHC was introduced in the form that it is available now, namely one pill of LNG based EHC to be taken as quickly as possible after unprotected intercourse but at the latest within 72 hours. In January 2001 this medication was reclassified from on prescription only to be available without prescription in pharmacies, for all women at least 16 years of age. I use this age cutoff in order to causally analyze if having easier access to emergency hormonal contraception has effects on the reproductive behavior of young women. To do so I employ a difference in difference strategy comparing the different age groups over time.

2.4 The Model

This model shows how a woman makes sequential choices around a one time sexual encounter, when information is revealed gradually.

The general setup of the model is the following: At $t = 1$ a woman, before having sex, chooses how much risk r to take. Higher risk implies higher utility but at the

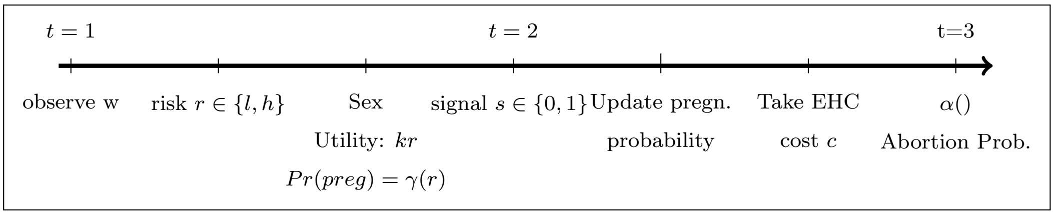


Figure 2.1: Timeline of the Model

same time higher risk also leads to a higher pregnancy probability $\gamma(r)$. If a woman chooses a positive level of risk, sex occurs. During sex a contraception-accident can take place. One can think of this accident as for example the condom breaking. At $t = 2$ a woman receives a signal $s \in \{0, 1\}$. If she receives the signal $s = 1$, she knows that such an accident took place and she knows that she will be pregnant if she does not take further action. If a woman receives $s = 0$ as a signal, she does not know whether or not she is pregnant, but rather updates her probability of being pregnant. After having received the signal and updated her pregnancy probability a woman has the option to take EHC, which has costs c and functions with probability τ. In a final step at $t = 3$ a woman learns if she is pregnant or not. A pregnancy then either ends in a birth or in an abortion, depending on the function $\alpha(w)$.

Figure 2.1, I shows the timing of the game. Women differ in terms of their expected costs of a pregnancy w, on which they base all of their decisions. The expected costs of a pregnancy w follow the distribution $G(w)$ in the population. These expected costs, which differ across women, reflect the fact that women are unsure on how costly a pregnancy is for them, until they learn they are pregnant. The costs of an expected pregnancy can be understood as a combination of costs of a baby and as costs of an abortion. [4] The distribution $G(w)$ has a support ranging from 0 to w_{max}, which reflects the fact that none of these women are women who are actively trying to conceive and hence, none of them have a benefit of having a baby. This can also reflect the fact that abortions are costly and no women draw utility from having abortions.

At $t = 1$, a woman knows her expected costs of a pregnancy w as well as all costs and other parameters of the model. Also at t=1, when choosing risk $r \in \{l, h\}$, a woman in addition to the binary choice always has the option of taking no risk at all ,i.e., having sex. This outside option gives a woman utility of 0. If a woman chooses

4. All women, in my empirical setting and hence also in the model, are very young women for all of them a baby is costly. However, these women differ in preferences towards abortions, and one can simply think of the situation in which these preferences only materialize once a woman knows that she is actually pregnant.

to take high risk, not only does this give her utility $k * r$, but it also increases her probability of becoming pregnant $\gamma(r)$, with $1 \geq \gamma(h) > \gamma(l) \geq 0$.

The risky sexual encounter then takes place. At $t = 2$ a woman then receives the noisy signal $s \in \{0, 1\}$, which tells her whether a "contraception accident" took place. One can think of this accident in different ways: if a woman used a condom as a contraceptive measure, one can think of the condom breaking as the accident. If a woman was using the birth control pill, one could imagine the woman getting severe diarrhea, which inhibits the functioning of the birth control pill. For modeling reasons, I assume that whenever $s = 1$ a woman is sure that if she does not take any further action, she will be pregnant. The signal s is noisy in the sense that it never signals pregnancy if there is no pregnancy, but conditional upon being pregnant it signals pregnancy only with probability q. Based on this signal the woman updates her pregnancy probability: if $s = 1$ then as explained above a woman is sure she will be pregnant if she does not take further action, hence $Pr(pregnancy|s = 1) = \gamma_{s=1}(r) = 1 \; \forall \; r$. If instead, $s = 0$ a woman updates her pregnancy probability according to Bayes rule: $Pr(pregnancy|s = 0) = \frac{(1-q)\gamma(r)}{1-q\gamma(r)} \equiv \gamma_0(r)$. Based on these updated pregnancy probabilities and the cost c and the probability τ of EHC working, a woman then takes the decision on whether or not to take EHC. Finally at $t = 3$ each woman learns if she is pregnant. A pregnancy then either ends in an abortion or a birth: A function $\alpha(w)$ assigns a probability of abortion to each level of the expected pregnancy cost w. This leads to the same outcome as revealing the actual costs of the pregnancy and making each woman take a choice on whether or not to have an abortion[5]. Specifically, since all women in my model are very young women, I assume that for all of them babies are costly. However, they differ in the costs of abortions they face. I assume that women with a high w, are women who face, in expectation, a high cost of abortion. Hence, these are also women who are less likely to have an abortion, which is why the function $\alpha(w)$, which assigns an abortion probability to each w, is decreasing in w.

5. It would be possible to specify that the costs of having an abortion differ for each woman: These costs a for each woman follow a distribution. The distributions differ for each woman however each woman knows her distribution, which is how she can calculate w the expected costs of pregnancy. However, once the pregnancy realizes, also an idiosyncratic shock to the costs of an abortion realizes and hence for a woman of type w, there is a probability, coming from the distribution of the idiosyncratic shock, that she will have an abortion.

2.4.1 Solution of the Model

For reasons of tractability, I present the solution of the model with $l = 0$, [6] hence also the pregnancy probability for individuals who take zero risk $\gamma(l) = \gamma(0) = 0$. Therefore, taking low risk is exactly the same as not participating in sex. Hence, there is only one positive level of risk remaining. The results of this simplified model are very similar to those of the more general model, however, this simplification allows me to characterize the solution by hand. Furthermore, with this simplification, it is also possible to perform the comparative statics by hand. This in return allows me to make more general statements, rather than imposing values for each parameter.

Solving this model is done by backward solution, i.e. in each step one compares the cost of choices and the optimum is the choice which leads to the lowest costs. Since at $t = 3$ there is no choice but rather the function $\alpha(w)$ assigning a probability of abortion, one starts at $t = 2$. In order to solve this step, one first has to update the probabilities of being pregnant. If a woman receives the signal $s = 1$ she knows that she is pregnant with probability 1 if she takes no further action. Therefore she then faces the decision $min\{w, c + (1 - \tau)w\}$, i.e. she has to choose between a sure pregnancy and the cost of EHC plus the cost of being pregnant with the probability of EHC not working, namely $(1 - \tau)$. If she does not receive the signal she knows that she is pregnant with probability $Pr(pregnancy|s = 0) = \frac{(1-q)\gamma(r)}{1-q\gamma(r)} \equiv \gamma_0(r)$, therefore she faces the decision $min\{\gamma_0(r)w, c + \gamma_0(r)(1 - \tau)w\}$.

One can then express the value of the choices at $t = 2$ as well as characterize the set of women who take EHC:

$$v(r) = \begin{cases} min\{w, c + (1 - \tau)w\} & \text{if } s = 1 \\ min\{w\,\gamma_0(r), c + \gamma_0(r)(1 - \tau)w\} & \text{if } s = 0 \end{cases} \tag{2.1}$$

$$\text{take EHC if} = \begin{cases} w > \frac{c}{(1-\tau)} & \text{if } s = 1 \\ w > \frac{c}{\tau\gamma_0(r)} & \text{if } s = 0 \end{cases} \tag{2.2}$$

Next one continues to $t = 1$. The decision faced here is to find the $argmax_r\{k * r + \mathbb{E}[v(r)]\}$, with choice being whether or not to take risk. Combining the consecutive

6. Of course, it is possible to solve the model as it is, however this is only possible using a computing tool and specifying values for the parameters, which is what I do in the appendix.

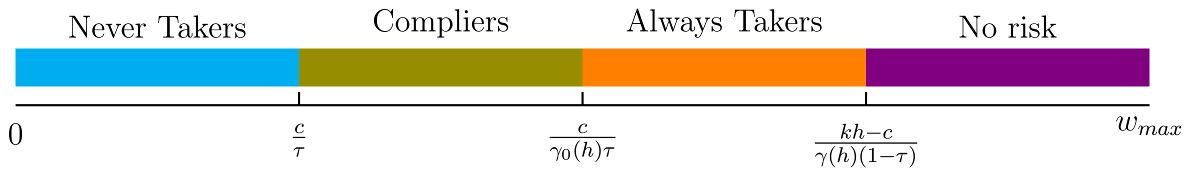


Figure 2.2: Characterization of Women in Terms of w

decisions then leads to the fact that one observes 4 groups of women, who differ in their behavior in terms of EHC intake as well as risk:

- **Never takers** A woman for whom $w < \frac{c}{\tau}$ never takes EHC, irrespective of the signal she receives.
- **Always takers** A woman for whom $w > \frac{c}{\gamma_0(h)\tau}$ always takes EHC, independent of the signal she receives.
- **Compliers** A woman for whom $\frac{c}{\tau} < w < \frac{c}{\gamma_0(h)\tau}$ is a complier in the sense that she only takes EHC, when she receives the signal that there was an accident.
- **Non participators = no risk** A woman for whom $w > \frac{kh-c}{\gamma(h)(1-\tau)}$ decides not to take any risk, and hence not to have sex.

From this characterization one can also write down utility functions for each group of women:

$$U_{NT}(h) = kh - \gamma(h)w \tag{2.3}$$

$$U_{AT}(h) = kh - \gamma(h)(1-\tau)w - c \tag{2.4}$$

$$U_{Co}(h) = kh - \gamma(h)[qc + (1-\tau q)w] \tag{2.5}$$

Comparing the utilities from each group, one can again back out the above specified thresholds, which are also visualized in Figure 2.2:

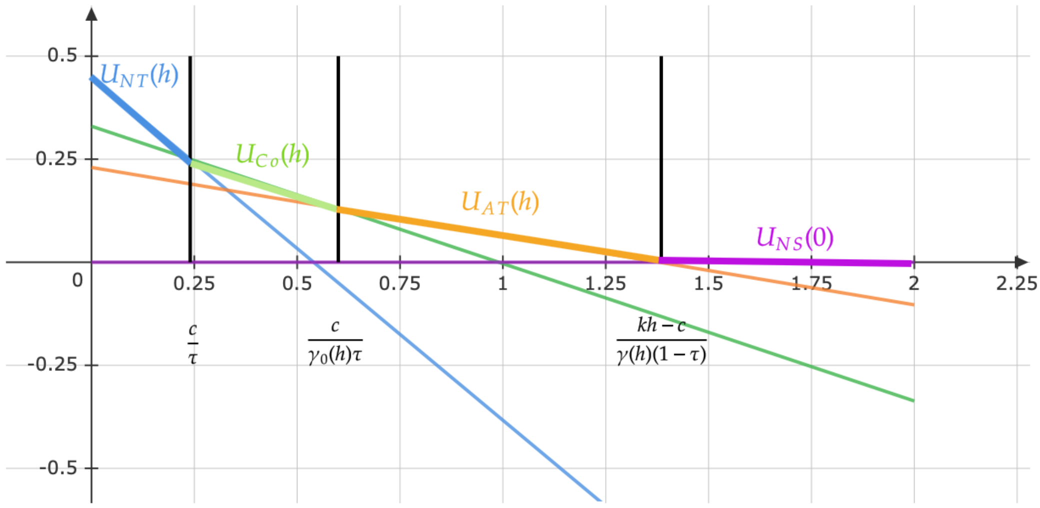


Figure 2.3: Utilities in Simplified Model

I now plot utilities for each group and for each level of w. This makes very clear why women are ordered in the way displayed in Figure 2.2. As one can see, as the expected costs of a pregnancy increase, women are more and more willing to pay the extra cost of EHC more often. At some point the expected costs of a pregnancy become so high that women no longer want to engage in sex, and hence they opt to take no risk at all.

The next step is to analyze how these utilities and choices translate into abortions. As explained above, abortions or births are determined by a probability function which depends on w. Hence, for each pregnant woman of type w, the probability functions assigns a probability that she will have an abortion. This function is decreasing in w, since w, the cost of an expected pregnancy incorporates the costs of both having a baby and of having an abortion. Women are unsure of their true costs until they actually are pregnant.

2.4.2 Comparative Statics: The Reform

I now want to analyze what happens to the amound of EHC taken, as well as to births and abortions, following a crease in c. Since there is only one level of positive risk, there is no intensive margin on which individuals change their behavior in terms of risk, but

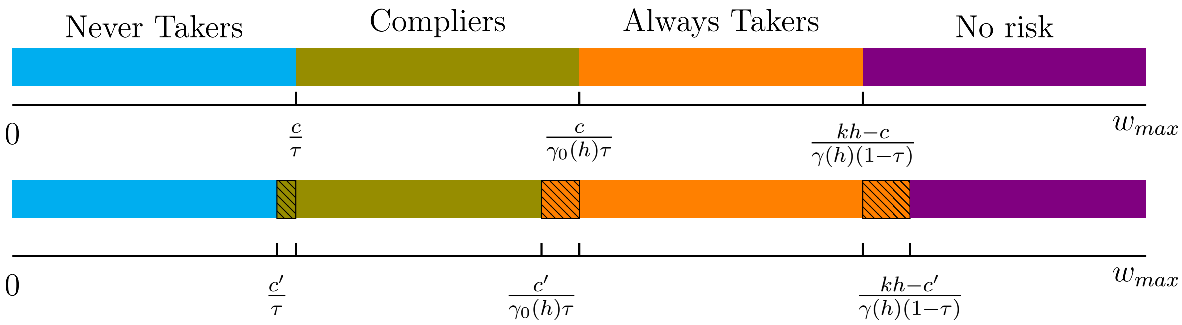


Figure 2.4: Decrease in c: Characterization of women in terms of w

only an extensive one. This makes this analysis simpler. In order to see what happens when c is decreased to c', I use Figure 2.4.

As one can see in Figure 2.4, as c decreases all thresholds between the groups shift. After the decrease in c, there are fewer Never Takers, there are more Always Takers, and less women who don't have sex. The effect on Compliers is ambiguous, since on the one hand, the reduction in Never Takers is compensated by Compliers, however also a part of the increase in Always Takers has to be compensated. Only when making concrete assumptions on the distribution $G(w)$ is it potentially possible to something about whether the absolute number of Compliers increases or decreases. [7]

As a next step, one can evaluate how this decrease in c and therefore the shifting of the thresholds affects EHC intake, risk taken, and pregnancies. Let us first consider the effect on EHC taken:

$$\underbrace{\int_{\frac{c'}{\tau}}^{\frac{c}{\tau}} q\,\gamma(h)dG(w)}_{\text{NT who become Co: } > 0} + \underbrace{\int_{\frac{c'}{\gamma_0(h)\tau}}^{\frac{c}{\gamma_0(h)\tau}} \{1 - q\,\gamma(h)\}dG(w)}_{\text{Co who become AT: } > 0} + \underbrace{\int_{\frac{k-c}{\gamma(h)(1-\tau)}}^{\frac{k-c'}{\gamma(h)(1-\tau)}} 1dG(w)}_{\text{No Sex who become AT: } > 0}$$

Hence, the effect on EHC taken is clearly positive. As a second step, I analyze the effects on risk. This is where the simplification of stetting the low risk level to zero has the largest effect: Since there is only one positive level of risk, changing c does not lead any effects on the intensive margin. The only effect a decrease of c has on risk is

7. Under the assumption of $G(w)$ being a uniform distribution: The magnitude by which the first threshold $\frac{c}{\tau}$ decreases is given by: $\Delta c \frac{1}{\tau}$. This is smaller than the magnitude by which the second threshold shifts to the left $\Delta c \frac{1}{\gamma_0(h)\tau}$. Hence, if one assumes a uniform distribution, one knows that the group of compliers becomes smaller.

through the women who didn't have sex before and become Always takers. Hence, the effect of a decrease of c on risk is:

$$\underbrace{\int_{\frac{k-c}{\gamma(h)(1-\tau)}}^{\frac{k-c'}{\gamma(h)(1-\tau)}} hdG(w)}_{\text{No Sex who become AT: } > 0}$$

As a further step, one can now explore how the change in c affects pregnancies:

$$\underbrace{\int_{\frac{c'}{\tau}}^{\frac{c}{\tau}} \gamma(h)(-q\tau)dG(w)}_{\text{NT who become Co: } < 0} + \underbrace{\int_{\frac{c'}{\gamma_0(h)\tau}}^{\frac{c}{\gamma_0(h)\tau}} \gamma(h)\tau(q-1)dG(w)}_{\text{Co who become AT: } < 0} + \underbrace{\int_{\frac{k-c}{\gamma(h)(1-\tau)}}^{\frac{k-c'}{\gamma(h)(1-\tau)}} \gamma(h)(1-\tau)dG(w)}_{\text{No Sex who become AT: } > 0}$$

The effect on pregnancies cannot be clearly signed. Two effects decrease pregnancies: Never Takers becoming Compliers and Compliers becoming Always Takers. When Never Takers become Compliers they decrease their probability of becoming pregnant, since they start taking EHC whenever the signal $s = 1$. This decreases their probability of being pregnant from $\gamma(h)$ to $\gamma(h)q(1-\tau)$. When Compliers become Always Takers, they also take EHC more often, thereby reducing their probability of becoming pregnant from $\gamma(h)q(1-\tau)$ to $\gamma(h)(1-\tau)$. Only the effect on women who before the reform don't have sex, and then become Always Takers increases the pregnancy probability from 0 to $\gamma(h)(1-\tau)$. This part of the model is where the simplification to only one possible positive level of risk has the biggest effects, since changing the cost of EHC does cannot have any effects on the intensive margin of risk. I explain in more detail how this changes

In a last step one can now analyze what happens to births and abortions. In this step I simply assume that $\alpha(w)$ assigns to each level of w a probability that a woman will have an abortion, and is increasing in w. This can be understood if one assumes that the true costs of an abortion are unclear until one is actually pregnant. Higher costs of a pregnancy stem from higher costs of an abortion. Hence, women with high w are more likely to end up giving birth rather than having an abortion. Let $a = 1$ denote that women have an abortion and $a = 0$ denote that they give birth, then the effect of lower c on abortions can be written as follows:

$$\underbrace{\int_{\frac{c'}{\tau}}^{\frac{c}{\tau}} \gamma(h)(-q\tau)\alpha(w)dG(w)}_{\text{NT who become Co: } <0} + \underbrace{\int_{\frac{c'}{\gamma_0(h)\tau}}^{\frac{c}{\gamma_0(h)\tau}} \gamma(h)\tau(q-1)\alpha(w)dG(w)}_{\text{Co who become AT: } <0} + \underbrace{\int_{\frac{k-c}{\gamma(h)(1-\tau)}}^{\frac{k-c'}{\gamma(h)(1-\tau)}} \gamma(h)(1-\tau)\alpha(w)dG(w)}_{\text{No Sex who become AT: } >0} \tag{2.6}$$

Here it becomes clear that, again in order to sign the overall effect on abortions one needs to make assumptions on parameters as well as on the distribution $G(w)$. The same holds true for the effect on births, which can be seen from the following sum:

$$\underbrace{\int_{\frac{c'}{\tau}}^{\frac{c}{\tau}} \gamma(h)(-q\tau)(1-\alpha(w))dG(w)}_{\text{NT who become Co: } <0} + \underbrace{\int_{\frac{c'}{\gamma_0(h)\tau}}^{\frac{c}{\gamma_0(h)\tau}} \gamma(h)\tau(q-1)(1-\alpha(w))dG(w)}_{\text{Co who become AT: } <0} + \underbrace{\int_{\frac{k-c}{\gamma(h)(1-\tau)}}^{\frac{k-c'}{\gamma(h)(1-\tau)}} \gamma(h)(1-\tau)(1-\alpha(w))dG(w)}_{\text{No Sex who become AT: } >0} \tag{2.7}$$

Now one can try to compare the effects on abortions and births and sign them. For this, I again assume that $G(w)$ is a uniform distribution. Both for abortions and births, there are two parts of the sum which are negative and one which is positive. Hence, it is possible for the total effects to go either way. The sign of the total effects, depends on the parameters, which determine the mass in each of the separate summands, but also on the weight of each summand given by probability of having an abortion. As one can see, the probability of a birth, is 1 minus the probability of having an abortion, i.e. all pregnancies end either in an abortion or a birth.

Since $\alpha(w)$ is an decreasing function in w, the weights on three summands in Equation 2.6 are decreasing: the largest weight is on the term for Never Takers becoming Compliers, the second largest on Compliers becoming Always Takers and the smallest weight is on Non Participators becoming Always Takers. For Equation 2.7 the situation is exactly opposite: the smallest weight is assigned to Never Takers becoming Compliers, the second smallest to Compliers becoming Always Takers and the largest weight is assigned to Non Participators becoming Always Takers. Hence, if the overall effect on births is positive, as I find in the data, then the overall effect on abortions in this model is smaller in magnitude, because the weight on the only positive summand is smaller and the weights on the negative summands are larger.

Hence, even with this very simplified model, it is possible to see how by decreasing the cost of EHC, EHC usage goes up and for some individuals risk taken goes up as well. At the same time, this decrease in costs of EHC also leads to more EHC usage leading to an a priori ambiguous effect on births and abortions. This ambiguous effect is due to the fact that increased EHC availability has two opposing effects: on the one hand EHC leads to more risk taken, and hence higher pregnancy probabilities, on the other hand a lower cost of EHC leads to more EHC intake and hence, lower pregnancy probabilities. Hence, one can say that EHC is a substitute for abortions, but also for regular contraception.

With this model, it is possible to reconcile the effects I find in the data: Positive effects on births and zero effects on abortions even with an increase in EHC usage for some age groups. Furthermore, this model helps to understand how the increased availability of EHC affects different women. Some women take more risk, some women take more EHC, some women more of both. These different effects can be seen even better when solving the model with two levels of positive risk, which I do in the appendix.

2.5 Data

In order to analyze the causal effect of the liberalization of EHC on sexual behavior in terms of abortions, births and sexually transmitted diseases I use publicly available annual data from the English National Health Service and the Office of National Statistics. This is data includes the number of abortions, births and sexually transmitted infections which occurred in each local authority, for each age group in each year from 1998 to 2004. The age groups in my sample are individuals under 16, 16 to 17 years old, 18 to 19 years old and 20 to 24 years old.

I combine this data with data from Girma and Paton (2011) who also used the publicly available data but additionally collected data regarding local authorities which where classified as “Health Action Zones” and thereby had “pharmacy schemes” in place, which allowed all teenagers access to EHC without prescription and free of charge in participating pharmacies. I use the data from Girma and Paton (2011) in order to exclude local authorities whenever they started being classified as such “Health Action Zones”, since from that point onwards in participating pharmacies all teenagers could access EHC without prescription and free of cost. In order for a pharmacy to participate, it had to sign a “Patient Group Directive” for EHC with a local authority and a physician, therefore the participation of pharmacies was not automatic. Hence, in the participating pharmacies the difference between under 16 year olds and women aged 16 and over was no longer present. Therefore, in such pharmacies the effect I identify should be zero. In order not to dilute the effect I might find from non participating pharmacies, I exclude all local authorities once they become health action zones in order not to run into this problem. This of course means, that over time, as more and more local authorities become “Health Action Zones” my sample becomes smaller. As a robustness check, I also perform the regressions using only the smallest possible sample, the local authorities which do not become “Health Action Zones” during the entire period of observation, hence, until at least 2005.

Due to the exclusion of local authorities which become “Health Action Zones”, one has to think about what this means for the estimated effect of EHC liberalization on abortions, births and STIs. Local authorities could decide themselves whether or not to implement “pharmacy schemes” by becoming “Health Action Zones” and hence, it is plausible that local authorities with especially high teenage pregnancy rates were more likely to to implement such schemes. In the data I find marginally significant evidence for this regarding abortions. Local authorities which were among the first to become “Health Action Zones”, had marginally significant higher log numbers of abortions. For births, STIs and conceptions, which are the sum of abortions and births, the difference in means for the two groups is not significant. If one also supposes that the effects of the liberalization might have been stronger in local authorities with higher teenage pregnancy rates, then the effects I find can be considered lower bounds for the causal effects of interest.

In Section 2.7, I explain how I use additional data from KT31 returns on EHC provision in family planning centers in order to show how consumption of EHC was influenced.

2.6 Empirical Strategy

As mentioned above, I exploit the exogenous variation generated by the age requirement needed to obtain EHC in pharmacies without a prescription. This exogenous variation due to age, allows me to use the under 16 year old women as a control group, since they were not affected by the liberalization of EHC from on prescription only to available in pharmacies without prescription. In order to estimate the causal effect I employ a Difference ind Differences strategy. This DiD specification relies on the fact that in absence of the treatment, the control (under 16) and the treated (16 and older) would have moved similarly over time.

I then use the regression specification of Equation 2.8, which I denote as Model 1, where Y_{art} is the outcome for age group a, in the local authority r at time t:

$$Y_{art} = \alpha_a + \lambda_t + \delta_r + \beta_a \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year \geq 2001\} + \gamma X_{rt} + \epsilon_{art} \tag{2.8}$$

The outcome variable Y_{art} is either the log of abortions, log of births or log of sexually transmitted infections in local authority r, at time t for age group a.

α_a are age group fixed effects.

λ_t are year fixed effects.

δ_t are year fixed effects.

β_a are different treatment effects for each age group, which is treated. Using this specification I allow for the different age groups which are treated, i.e. all those above 16, to be effected differently by the reform. The age groups in my sample are under 16 years of age, 16 to 17, 18 to 19 and 20 to 24 years of age. Only for STIs, the age group 16-17 and 18-19 are combined, and hence I only observe three rather than four different age groups for this outcome variable.

X_{rt} is a vector of controls: The rate children aged 15-17 in LA care per 10,000 children; the percentage of individuals who have not obtained any qualification at age 16; and the GP practices per 1000 females under 18 years of age. These controls change across local authorities and over time, but not across age groups. Hence, they do not influence the estimated treatment effect, but only the standard error. Furthermore, for some of these controls, there is a problem of missing data, hence when including the controls the sample gets slightly smaller.

It is important to not that for abortions and STIs the treatment indicator is defined to be 1, whenever the age is larger or equal to 16 and whenever the year is larger or equal to 2001, since that is when the liberalization took place. For births however, this is not the case: Because of the usual duration of nine months of a pregnancy, only one fourth of the births in 2001 are could have potentially been affected by liberalization of EHC in 2001. Hence, for births the treatment variable takes the value 1 only from the year 2002 onwards for the treated age groups.

I cluster the standard errors at the local authority level, because as pointed out by Abadie et al. (2017) clustering is a sampling problem: Since I only sample some of the local authorities, I cluster at the local authority level. As controls, I use the ratio of general practitioners practices to the under 18 female population, the percentage of the population aged 15 to 17 in care and the percentage of teenagers without qualifications at age 16. These controls vary at the local authority level, however they do not vary across age groups, and hence they do not change the estimate of interest but just its standard error.

As explained above, I perform the regression of Equation 2.8 using two different samples: First, using the sample in which I keep all local authorities as long as they are not classified as "Health Action Zones". Once a local authority changes status, I exclude it. Therefore, this sample becomes smaller over time, starting with 134 local authorities in my sample, this number reduces to 68. Second, I use only the sample of 68 local authorities which are never classified as "Health Action Zones" in the time period of observation.

Furthermore, to better understand the dynamics I also estimate the regression of the following form:

$$Y_{art} = \alpha_a + \lambda_t + \delta_r + \sum_{t=2001}^{2004} \beta_{at} \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year = t\} + \gamma X_{rt} + \epsilon_{art} \tag{2.9}$$

Equation 2.9 allows for the treatment effects of liberalization not only to be different across age groups, but also for them to change over time. This specification is especially useful to look at dynamics.

Both Equation 2.8 and 2.9 make clear that the effect I am estimating is an intention to treat effect, because I do not observe which individuals changed their behavior in terms of EHC consumption, due to the change in regulation. But rather I observe all individuals and their final outcomes. In order to be able to say anything about how individuals change their consumption of EHC behavior due to the regulation, on would need sales data on EHC for specific age groups. This data is theoretically feasible before the liberalization, because on a prescription there is usually some patient information such as age. After the liberalization however, this data is not even theoretically feasible, because pharmacies did not need to collect data on the individuals they sold EHC. Hence, I am only able to ever identify an intention to treat effect. Yet, in order to make a case that the reform actually influenced consumption patterns of EHC of young women, I will try to show some evidence in Section 2.7

In Tables 2.1 - 2.3, I display averages of the different outcome variables in raw numbers across local authorities for different age groups in the year 2000, since this is the year before the reform and therefore can be seen as the baseline year. In Tables 2.1 - 2.3 I display the same but rather than using raw numbers I display logs, which are the actual outcome variables I use.

Table 2.1: Number of Abortions by Age Group

Age Group	Mean	Std. Dev.
u16	27.18033	18.22677
16-17	87.19672	59.7728
18-19	126.6066	82.59716
20-24	289.4836	202.6093
Total	132.6168	149.4654
N	488	

This table displays the averages across local authorities for each age group in the year 2000, i.e. the year before treatment.

Table 2.2: Number of Births by Age Group

Age Group	Mean	Std. Dev.
u16	21.47541	15.20614
16-17	115.5492	80.80904
18-19	217.5902	157.4497
20-24	722.2049	527.5792
Total	269.2049	387.7618
N	488	

This table displays the averages across local authorities for each age group in the year 2001, i.e. the year before treatment.

As one can see from Table 2.3 data on sexually transmitted infections is not available for all local authorities, since the number of observations in 2000 shrinks from 488 to 312 local authority age group combinations. In Table 2.6 the sample is even smaller, since when taking logs all observations with zero STIs are dropped.

Table 2.3: **Number of STIs by Age Group**

Age Group	**Mean**	**Std. Dev.**
u16	1.224299	1.11006
16-19	24.11215	16.22968
20-24	38.35514	28.21423
Total	21.23053	24.20776
N	321	

This table displays the averages across local authorities for each age group in the year 2000, i.e. the year before treatment.

Table 2.4: Log of Abortions by Age Group

Age Group	**Mean**	**Std. Dev.**
u16	3.115083	.6157175
16-17	4.289463	.5835865
18-19	4.669858	.5819985
20-24	5.465036	.6303108
Total	4.38486	1.039478
N	488	

This table displays the averages across local authorities for each age group in the year 2000, i.e. the year before treatment.

Table 2.5: Log of Births by Age Group

Age Group	**Mean**	**Std. Dev.**
u16	2.82266	.7971619
16-17	4.540351	.6622366
18-19	5.179345	.6427272
20-24	6.389499	.6099765
Total	4.736886	1.45544
N	488	

This table displays the averages across local authorities for each age group in the year 2001, i.e. the year before treatment.

Table 2.6: **Log of STIs by Age Group**

Age Group	Mean	Std. Dev.
u16	.4159295	.4952776
16-19	2.950327	.7208661
20-24	3.363409	.8120391
Total	2.438553	1.409117
N	290	

This table displays the averages across local authorities for each age group in the year 2000, i.e. the year before treatment.

Since I am interested in identifying the causal effect of the liberalization of EHC on young women's reproductive behavior, I am using a Difference in Differences specification. However, it is not sufficient to use the DiD framework, but rather it also needs to be the case that in the absence of treatment the different age groups would have evolved similarly, which is known as the assumption of common trends. In order to convey this assumption I first plot all outcome variables: Log abortions, log births and log sexually transmitted infections. To do so, I plot the mean over all local authorities for each age group at each point in time. I do this both for all local authorities, which don't have a pharmacy scheme in place in 2001, as well as for all local authorities which never have a pharmacy scheme in place throughout the period 1997 to 2004.

In a next step, I want to convey the assumption of common trends for the different age groups in the years up to 2001. In order to do so, I first plot log abortions, log births and log STIs for each age group, averaged across local authorities over time. From Figures 2.5 to 2.7 one can see how for all the log outcome variables the averages across local authorities differ quite strongly across age groups by what looks like a constant. Hence, to inspect the common trends in more detail, I zoom into each of these figures in Figures 2.8 to 2.10. As one can see, for all the log outcome variables, the averages across Local Authorities seem to follow common trends for the different age groups. From visual inspection only, one should not be concerned about the assumption of common trends for the years 1998-2000.

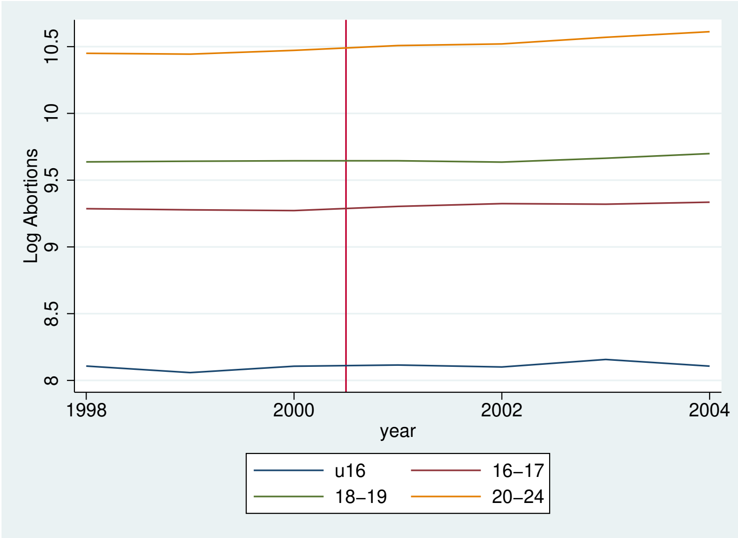


Figure 2.5: Average Log Abortions across Local Authorities per Age Group

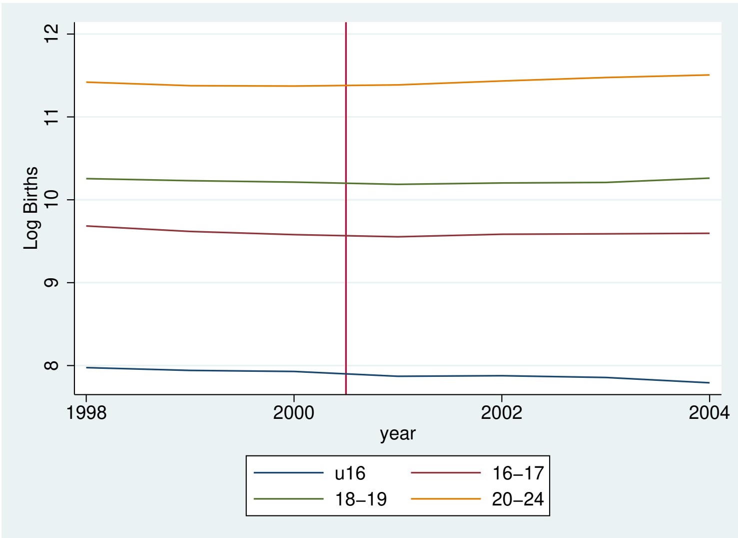


Figure 2.6: Average Log Births across Local Authorities per Age Group

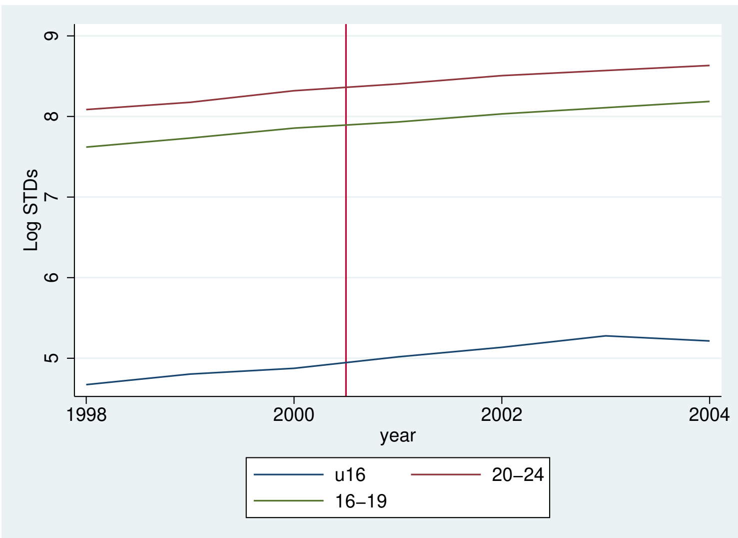


Figure 2.7: Average Log STIs across Local Authorities per Age Group

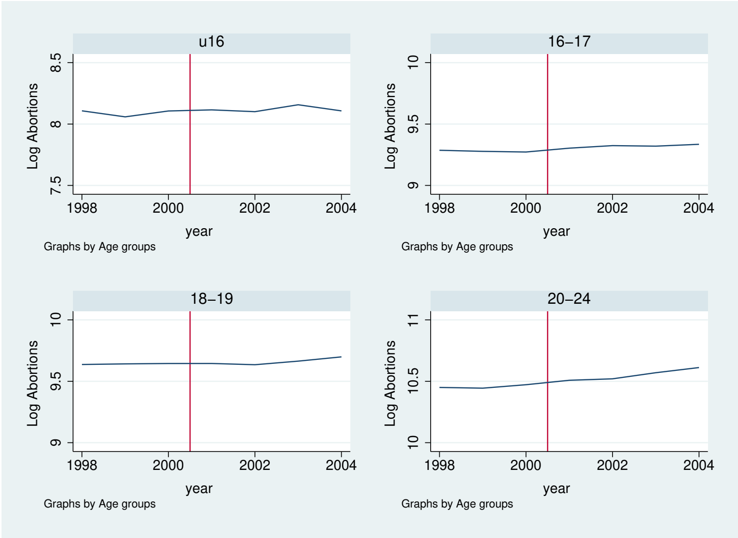


Figure 2.8: Average Log Abortions across Local Authorities per Age Group, detailed

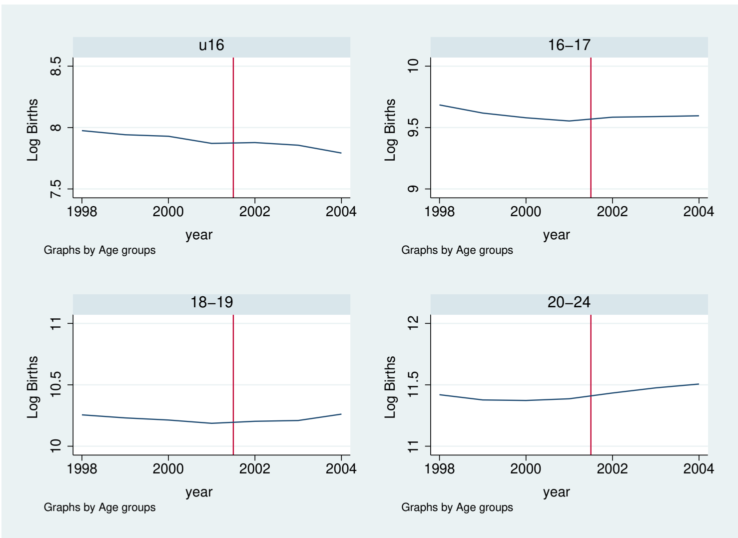


Figure 2.9: Average Log Births across Local Authorities per Age Group, detailed

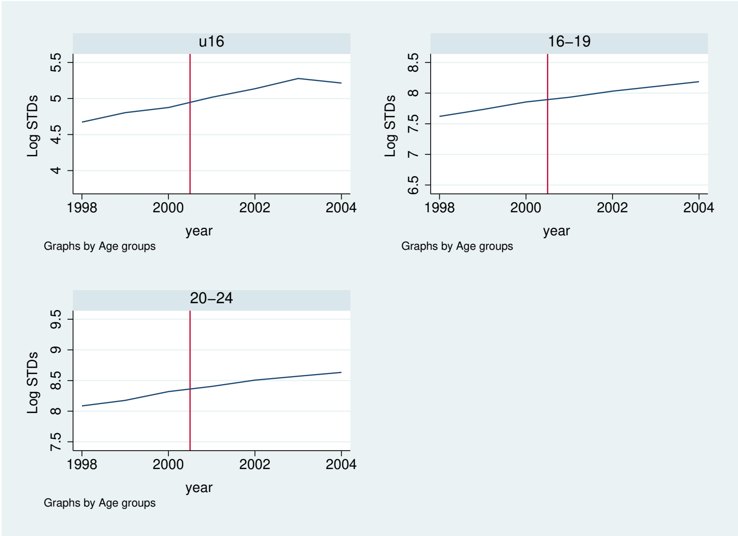


Figure 2.10: Average Log STIs across Local Authorities per Age Group, detailed

However, in order to be more certain about the assumption of common trends for the different age groups in the years up to 2001, I test these. To do so, I run the following regression using only the pre-treatment years:

$$Y_{art} = \alpha_a + \lambda_t + \delta_r + \sum_{t=1997}^{2000} \beta_{at} \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year = t\} + \gamma X_{rt} + \epsilon_{art} \qquad (2.10)$$

In Tables 2.7 to 2.9 I show the results of testing for common trends using only the pre-treatment data. As one can see, for none of the three outcome variables, do there seem to be any year-age-combinations which differ significantly from the under 16 year olds. Hence, also from the inspection of the data using regressions, one should not doubt the assumption of common trends.

A further way of testing for common trends, is to include the interactions of year and age group for the pre-treatment years into the main regression of interest. I show the results of this regression in Section 2.9.

Table 2.7: OLS Regression Common Trends Log Abortions

	(1) Basic	(2) LA-FE cont
Age = 16-17	1.192***	1.192***
	(0.0272)	(0.0285)
Age = 18-19	1.534***	1.534***
	(0.0317)	(0.0332)
Age = 20-24	2.325***	2.325***
	(0.0429)	(0.0448)
Year=1999	-0.0773*	-0.0890*
	(0.0342)	(0.0361)
Year=2000	-0.00808	-0.0429
	(0.0275)	(0.0299)
Year = 1999 × 16-17	0.0648	0.0648
	(0.0387)	(0.0405)
Year = 1999 × 18-19	0.0890*	0.0890*
	(0.0364)	(0.0381)
Year = 1999 × 20-24	0.0724	0.0724
	(0.0368)	(0.0385)
Year = 2000 × 16-17	-0.0177	-0.0177
	(0.0308)	(0.0322)
Year = 2000 × 18-19	0.0208	0.0208
	(0.0304)	(0.0318)
Year = 2000 × 20-24	0.0246	0.0246
	(0.0283)	(0.0296)
Rate kids in LA care per 10,000		-3.006
		(3.365)
Perc. with no quals at 16		-0.0210
		(0.0137)
GP practices per 1000 u18 fem		-0.0376*
		(0.0185)
Constant	3.123***	4.294***
	(0.0541)	(0.247)
Observations	1464	1464

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression to test for common trends using data only for the pretreatment years 1998-2000. Column 1 and 2 use the entire sample of local authorities which did not have a pharmacy scheme in place in until at least 2001. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls.

Table 2.8: OLS Regression Common Trends Log Births

	(1) Basic	(2) LA-FE cont
Age = 16-17	1.730***	1.730***
	(0.0304)	(0.0314)
Age = 18-19	2.325***	2.325***
	(0.0314)	(0.0324)
Age = 20-24	3.498***	3.498***
	(0.0380)	(0.0393)
Year = 1999	-0.00723	-0.0157
	(0.0417)	(0.0395)
Year = 2000	-0.0256	-0.0273
	(0.0376)	(0.0418)
Year = 2001	-0.106*	-0.120*
	(0.0415)	(0.0499)
Year = 1999 × 16-17	-0.0406	-0.0283
	(0.0431)	(0.0414)
Year = 1999 × 18-19	-0.0161	-0.00376
	(0.0443)	(0.0416)
Year = 1999 × 20-24	-0.0344	-0.0221
	(0.0424)	(0.0398)
Year = 2000 × 16-17	-0.0712	-0.0712
	(0.0419)	(0.0433)
Year = 2000 × 18-19	-0.0164	-0.0164
	(0.0399)	(0.0412)
Year = 2000 × 20-24	-0.0222	-0.0222
	(0.0378)	(0.0390)
Year = 2001 × 16-17	-0.0124	-0.00543
	(0.0426)	(0.0436)
Year = 2001 × 18-19	0.0315	0.0385
	(0.0436)	(0.0443)
Year = 2001 × 20-24	0.0693	0.0763
	(0.0408)	(0.0414)
Rate kids in LA care per 10,000		6.973*
		(2.761)
Perc, with no quals at 16		0.00338
		(0.00827)
GP practices per 1000 U18 fem		-0.0194
		(0.0127)
Constant	2.929***	4.094***
	(0.0656)	(0.175)
Observations	1950	1950

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression to test for common trends using data only for the pretreatment years 1998-2001. Column 1 and 2 use the entire sample of local authorities which did not have a pharmacy scheme in place in until at least 2001. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls.

Table 2.9: **OLS Regression Common Trends Log STIs**

	(1) Basic	(2) LA-FE cont
Age = 16-19	2.422***	2.692***
	(0.0655)	(0.0590)
Age = 20-24	2.862***	3.132***
	(0.0696)	(0.0636)
Year=1999	0.0404	0.0515
	(0.0464)	(0.0596)
Year=2000	0.115	0.0957
	(0.0628)	(0.0756)
Year = 1999 × 16-19	0.0832	0.0484
	(0.0492)	(0.0636)
Year = 1999 × 20-24	0.0682	0.0334
	(0.0465)	(0.0597)
Year = 2000 × 16-19	0.112	0.0901
	(0.0663)	(0.0773)
Year = 2000 × 20-24	0.0854	0.0633
	(0.0680)	(0.0756)
Rate kids LA care per 10,000		-1.760
		(6.710)
Perc. with no quals at 16		-0.0577*
		(0.0278)
GP practices per 1000 U18 fem		0.0301
		(0.0400)
Constant	0.301***	0.934
	(0.0522)	(0.528)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression to test for common trends using data only for the pretreatment years 1998-2000. Column 1 and 2 use the entire sample of local authorities which did not have a pharmacy scheme in place in until at least 2001. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls.

2.7 First Stage

Even though it is well known that EHC was liberalized, it is not clear if this liberalization changed the consumption patters of women of EHC. For other countries, in which EHC was liberalized much later, for example Germany, there is sales data available which confirms that after the liberalization sales approximately doubled and stayed at this new high level long term. There is also evidence for Germany that after the deregulation, EHC was consumed much more evenly throughout the week, rather than bunched on Mondays, after the weekend on which even though studies show that roughly 68% of all intercourse takes place on weekends, much less EHC was sold.

Preferably, I would want to show, that the consumption pattern also in EHC were affected by the liberalization, however only for women 16 years and older. This task, however is rather challenging: For the England, it is no longer possible to obtain sales data, due to the fact that the change in regulation was implemented many years ago. However, even with sales data, in order to show that only individuals above 16 years of age were affected, I would need prescription data by age before the change, and sales data by age after the changes. The first of which, is impossible to obtain, due to the long time period which has passed since the reform and the second being even more impossible, due to the fact that EHC was available without prescription after the reform, and hence there was no reason to even register who and at which age bought EHC.

For these reasons of data unavailability, I use data from family planning clinics on EHC use. This data is collected through the annual return form KT31 and aggregated by the department of health. From this data one can only observe emergency contraception prescribed/obtained in family planning centers. Hence, this data can be used in order to learn how the reform affected EHC consumption. However, when using this data it is crucial to know that this data is not annual data, but rather collected across two years. Hence, specifically the datapoint of interest in 2001 contains a mixture of 2000 and 2001 data. Nonetheless, from Figure 2.11 in which I plot EHC given out in family planning centers by age group and year, how the EHC consumption patters in such family planning centers changed. From 1998 to 2000 there is a clear increase in EHC dispensed for all age groups. In 2001 for the age groups, which could now obtain EHC in pharmacies without a prescription namely those aged 16-19 and 20-24, EHC dispensed in family planning centers drops and drops even further in 2002. For women under the age of 16 the upward trend continues. In Figure 2.11 one can see that the drop is

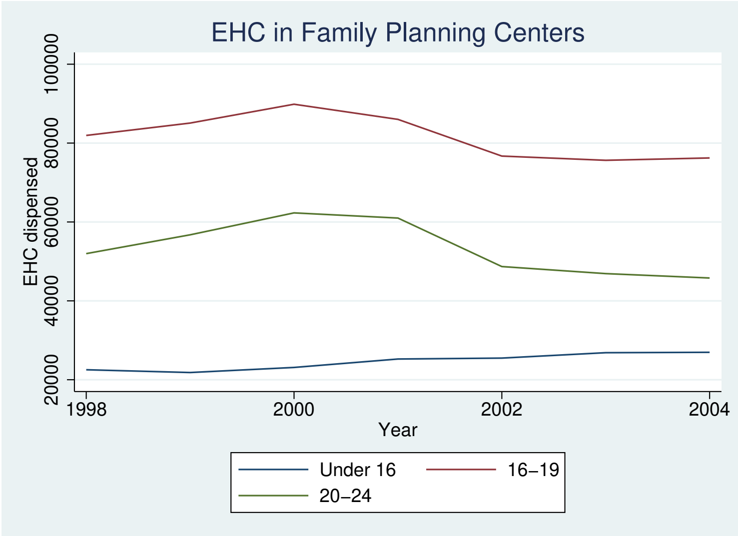


Figure 2.11: EHC dispensed in Family Planning Centers

gradual, but this is probably simply due to the fact that the 2001 datapoint actually also includes numbers from 2000 and hence from when EHC was not yet liberalized.

Therefore, this data can be used as suggestive evidence that the reform had an effect on the intake of emergency contraception. One can clearly see that while there was a decrease in the overall number of emergency hormonal contraception pills prescribed from 2000 to 2001, there was a constant increase in the same time of EHC prescribed to under 16 year olds. This can be seen as suggestive evidence that women of 16 years of age or older, obtained less EHC in family planning centers due to the reform. If one assumes that EHC consumption overall did not change due to the reform, it means that women over 16 must have consumed more EHC from pharmacies, and therefore more likely in a timely manner. Therefore it is reasonable to conclude that the reform had an effect even if the total number of EHC pills consumed did not change. If however the situation in England was similar to the one in Germany, it is possible that EHC consumption actually increased on impact due to the reform. Hence, one would have an even stronger first stage.

2.8 Estimation Results

In this Section, I first show the results obtained when running Regression 2.8 and then from Regression 2.9.

In Tables 2.10 to 2.12 I show the regression results from Equation 2.8 for four slightly different regression specifications. In column 1 and 3, I show the most basic regression without controls or local authority fixed effects. In column 2 and 4 both local authority fixed effects as well as controls are included. Column 1 and 2 show these results using the largest possible sample: in each year the sample includes all local authorities which up until that point have not had a pharmacy scheme to access EHC as a health action zone. Whereas, column 3 and 4 use the smallest possible sample, only including local authorities which until at least 2005 did not have a pharmacy scheme in place.

From Table 2.10, one can see that there seem to be no significant effects on abortions for the age groups 16-17 and 18-19, however there seem to be some positive effects on abortions for the oldest individuals in my sample. These effects on 20-24 year olds, however diminish in magnitude as one switches to the smallest sample as can be seen in column 3 and 4 of Table 2.10.

In Table 2.11, one can see that there seem to be significant effects on births for all individuals: Due to the liberalization births increase and this effect is larger in magnitude for the oldest age group.

In Table 2.12, I display the results of the Model 2.8 regression when using Log STIs as an outcome. Here one can see that there seem to be positive effects on STIs for both the 16-19 year olds as well as the 20-24 year olds. These effects however are not very robust in the sense that they are no longer significant when using the smallest sample and including controls as well as local authority fixed effects. This could be partially due to the small sample size which I have for analyze STIs.

In order to analyze in more detail, what happens after the reform, in Tables 2.13 to 2.15, I report the coefficients of interest of Equation 2.9. Again in column 2 and 4 both local authority fixed effects as well as controls are included. Column 1 and 2 show these results using the largest possible sample: in each year the sample includes all local authorities which up until that point have not had a pharmacy scheme to access EHC as a health action zone. Whereas, column 3 and 4 use the smallest possible sample, only including local authorities which until at least 2005 did not have a pharmacy scheme in place. In Tables 2.13 to 2.15, I only report the β coefficients, which are the coefficients on treatment, for simple reasons of visibility.

Table 2.10: OLS Regression Log Abortions Model 1

	Basic	LA-FE cont	Never Pharm Basic	Never Pharm LA-FE cont
Age = 16-17	1.208***	1.208***	1.229***	1.229***
	(0.0195)	(0.0200)	(0.0288)	(0.0293)
Age = 18-19	1.571***	1.571***	1.585***	1.585***
	(0.0259)	(0.0265)	(0.0394)	(0.0402)
Age = 20-24	2.358***	2.358***	2.405***	2.405***
	(0.0390)	(0.0399)	(0.0595)	(0.0606)
Year = 1999	-0.0207*	-0.0229*	-0.0169	-0.0186
	(0.0105)	(0.0112)	(0.0127)	(0.0136)
Year = 2000	-0.00116	-0.0181	-0.000478	-0.0176
	(0.0108)	(0.0137)	(0.0140)	(0.0172)
Year = 2001	-0.0130	-0.0378	0.00156	-0.0232
	(0.0215)	(0.0245)	(0.0283)	(0.0308)
Year = 2002	-0.0782	-0.0448	0.00280	-0.0282
	(0.0402)	(0.0275)	(0.0280)	(0.0326)
Year = 2003	-0.0964*	-0.0280	0.0256	-0.00883
	(0.0474)	(0.0327)	(0.0314)	(0.0378)
Year = 2004	-0.0138	-0.0177	0.0428	-0.00114
	(0.0524)	(0.0339)	(0.0293)	(0.0387)
Treatment × 16-17	0.0351	0.0351	0.0176	0.0176
	(0.0197)	(0.0202)	(0.0243)	(0.0247)
Treatment × 18-19	-0.00632	-0.00632	-0.0169	-0.0169
	(0.0209)	(0.0214)	(0.0246)	(0.0251)
Treatment × 20-24	0.0860***	0.0860***	0.0573*	0.0573*
	(0.0227)	(0.0232)	(0.0249)	(0.0254)
Rate kids in LA care per 10,000		6.259		6.244
		(3.588)		(4.431)
Perc. with no quals at 16		-0.0114		-0.0110
		(0.00709)		(0.00890)
GP practices per 1000 u18 fem		-0.00994		-0.0101
		(0.0143)		(0.0175)
Constant	3.102***	3.724***	3.039***	3.700***
	(0.0529)	(0.149)	(0.0624)	(0.169)
Observations	2932	2932	1904	1904
Mean Log Abortions	4.372	4.372	4.360	4.360
Sd Log Abortions	1.046	1.046	1.037	1.037

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.8. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls.

Table 2.11: OLS Regression Log Births Model 1

	Basic	LA-FE cont	Never Pharm Basic	Never Pharm LA-FE cont
Age = 16-17	1.699***	1.704***	1.683***	1.688***
	(0.0167)	(0.0176)	(0.0214)	(0.0236)
Age = 18-19	2.325***	2.330***	2.325***	2.330***
	(0.0210)	(0.0220)	(0.0264)	(0.0287)
Age = 20-24	3.501***	3.505***	3.501***	3.506***
	(0.0298)	(0.0309)	(0.0396)	(0.0418)
Year = 1999	-0.0300**	-0.0296**	-0.0242	-0.0242
	(0.0110)	(0.0107)	(0.0173)	(0.0158)
Year = 2000	-0.0530***	-0.0482**	-0.0517***	-0.0411*
	(0.0103)	(0.0144)	(0.0145)	(0.0193)
Year = 2001	-0.0840***	-0.0834***	-0.0808***	-0.0709**
	(0.0123)	(0.0188)	(0.0177)	(0.0261)
Year = 2002	-0.252***	-0.169***	-0.154***	-0.149**
	(0.0491)	(0.0377)	(0.0356)	(0.0455)
Year = 2003	-0.291***	-0.163***	-0.145***	-0.145**
	(0.0544)	(0.0381)	(0.0380)	(0.0453)
Year = 2004	-0.206**	-0.158***	-0.139***	-0.144**
	(0.0617)	(0.0430)	(0.0388)	(0.0496)
Treatment × 16-17	0.0847**	0.0879**	0.0930**	0.0973**
	(0.0287)	(0.0295)	(0.0331)	(0.0345)
Treatment × 18-19	0.115***	0.118***	0.111**	0.116**
	(0.0295)	(0.0307)	(0.0329)	(0.0349)
Treatment × 20-24	0.201***	0.204***	0.191***	0.196***
	(0.0324)	(0.0334)	(0.0337)	(0.0352)
Rate kids in LA care per 10,000		1.952		2.447
		(1.892)		(2.131)
Perc. with no quals at 16		0.00777		0.0120
		(0.00624)		(0.00752)
GP practices per 1000 U18 fem		-0.0169*		-0.0191*
		(0.00848)		(0.00897)
Constant	2.936***	4.111***	2.874***	4.076***
	(0.0621)	(0.120)	(0.0759)	(0.131)
Observations	2928	2928	1901	1901
Mean Log Births	4.741	4.741	4.712	4.712
Sd Log Births	1.445	1.445	1.436	1.436

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.8. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls.

Table 2.12: **OLS Regression Log STIs Model 1**

	Basic	LA-FE cont	Never Pharm Basic	Never Pharm LA-FE cont
Age = 16-19	2.488***	2.729***	2.486***	2.747***
	(0.0570)	(0.0437)	(0.0788)	(0.0592)
Age = 20-24	2.914***	3.155***	2.919***	3.180***
	(0.0633)	(0.0520)	(0.0882)	(0.0715)
Year = 1999	0.0970***	0.0793***	0.130***	0.101***
	(0.0199)	(0.0202)	(0.0277)	(0.0261)
Year = 2000	0.189***	0.151***	0.226***	0.191***
	(0.0304)	(0.0301)	(0.0465)	(0.0385)
Year = 2001	0.144***	0.133**	0.178***	0.174**
	(0.0368)	(0.0426)	(0.0457)	(0.0559)
Year = 2002	0.165**	0.201***	0.269***	0.262***
	(0.0500)	(0.0438)	(0.0466)	(0.0557)
Year = 2003	0.252***	0.309***	0.384***	0.381***
	(0.0565)	(0.0555)	(0.0455)	(0.0652)
Year = 2004	0.278***	0.302***	0.381***	0.361***
	(0.0642)	(0.0511)	(0.0458)	(0.0613)
Treatment × 16-19	0.154**	0.0959*	0.132*	0.0755
	(0.0459)	(0.0370)	(0.0555)	(0.0517)
Treatment × 20-24	0.174***	0.116**	0.143*	0.0871
	(0.0475)	(0.0386)	(0.0572)	(0.0522)
Rate kids in LA care per 10,000		-9.468		-13.83
		(6.882)		(8.106)
Perc. with no quals at 16		-0.0483**		-0.0412*
		(0.0163)		(0.0191)
GP practices per 1000 U18 fem		-0.0206		-0.0238
		(0.0177)		(0.0172)
Constant	0.256***	1.373***	0.172**	1.376***
	(0.0447)	(0.280)	(0.0522)	(0.309)
Observations	1767	1767	1151	1151
Mean Log STIs	2.421	2.421	2.414	2.414
Sd Log STIs	1.434	1.434	1.450	1.450

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.8. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls.

Table 2.13: OLS Regression Log Abortions Model 2

	Basic	LA-FE cont	Never Pharm No Pharm Basic	Never Pharm LA-FE cont
16-17 × Treatment × 2001	0.0111	0.0111	0.00495	0.00495
	(0.0261)	(0.0267)	(0.0380)	(0.0387)
16-17 × Treatment × 2002	0.0562	0.0562	0.0168	0.0168
	(0.0308)	(0.0315)	(0.0367)	(0.0374)
16-17 × Treatment × 2003	0.00486	0.00486	-0.0163	-0.0163
	(0.0362)	(0.0370)	(0.0388)	(0.0396)
16-17 × Treatment × 2004	0.0866*	0.0866*	0.0650	0.0650
	(0.0409)	(0.0418)	(0.0417)	(0.0425)
18-19 × Treatment × 2001	-0.0284	-0.0284	-0.0260	-0.0260
	(0.0260)	(0.0266)	(0.0364)	(0.0371)
18-19 × Treatment × 2002	-0.00756	-0.00756	-0.0408	-0.0408
	(0.0316)	(0.0323)	(0.0347)	(0.0353)
18-19 × Treatment × 2003	-0.0293	-0.0293	-0.0498	-0.0498
	(0.0376)	(0.0384)	(0.0378)	(0.0385)
18-19 × Treatment × 2004	0.0634	0.0634	0.0492	0.0492
	(0.0410)	(0.0419)	(0.0416)	(0.0424)
20-24 × Treatment × 2001	0.0280	0.0280	0.0173	0.0173
	(0.0268)	(0.0274)	(0.0374)	(0.0381)
20-24 × Treatment × 2002	0.0893*	0.0893*	0.0354	0.0354
	(0.0344)	(0.0352)	(0.0346)	(0.0353)
20-24 × Treatment × 2003	0.0923*	0.0923*	0.0461	0.0461
	(0.0399)	(0.0407)	(0.0373)	(0.0380)
20-24 × Treatment × 2004	0.178***	0.178***	0.130**	0.130**
	(0.0451)	(0.0461)	(0.0417)	(0.0425)
Observations	2932	2932	1904	1904
Mean Log Abortions	4.372	4.372	4.360	4.360
Sd Log Abortions	1.046	1.046	1.037	1.037

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.9. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls. All columns only report treatment coefficients $\beta_{a,t}$ even though regressions include other covariates as well.

From Table 2.13 one can see that for the younger age groups, the effects disappear when allowing for a different treatment effect for each age group for each post treatment year. Only for the age group of 20-24 year olds, one could argue that there might be an increase in abortions due to the liberalization.

Table 2.14: OLS Regression Log Births Model 2

	Basic	LA-FE cont	Never Pharm No Pharm Basic	Never Pharm LA-FE cont
16-17 × Treatment × 2002	0.0609	0.0561	0.0470	0.0418
	(0.0353)	(0.0365)	(0.0414)	(0.0433)
16-17 × Treatment × 2003	0.0784	0.0737	0.0907	0.0855
	(0.0502)	(0.0518)	(0.0532)	(0.0554)
16-17 × Treatment × 2004	0.127*	0.151**	0.143*	0.167***
	(0.0550)	(0.0472)	(0.0547)	(0.0468)
18-19 × Treatment × 2002	0.0684*	0.0637	0.0494	0.0443
	(0.0333)	(0.0340)	(0.0390)	(0.0400)
18-19 × Treatment × 2003	0.0966*	0.0918*	0.0840	0.0788
	(0.0433)	(0.0451)	(0.0475)	(0.0504)
18-19 × Treatment × 2004	0.203***	0.227***	0.203***	0.227***
	(0.0558)	(0.0508)	(0.0539)	(0.0483)
20-24 × Treatment × 2002	0.142***	0.138***	0.114**	0.109*
	(0.0370)	(0.0378)	(0.0426)	(0.0437)
20-24 × Treatment × 2003	0.209***	0.204***	0.190***	0.185***
	(0.0453)	(0.0466)	(0.0469)	(0.0490)
20-24 × Treatment × 2004	0.273***	0.298***	0.273***	0.296***
	(0.0586)	(0.0544)	(0.0538)	(0.0485)
Observations	2928	2928	1901	1901
Mean Log Births	4.741	4.741	4.712	4.712
Sd Log Births	1.445	1.445	1.436	1.436

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.9. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls. All columns only report treatment coefficients $\beta_{a,t}$ even though regressions include other covariates as well.

In Table 2.14 one can see how the treatment coefficients of Equation 2.9 evolve when the outcome are log births. On inspection, it is clear that the group of 20-24 year olds experiences an increase in births, which is very much in line with the findings of Pfeifer and Reutter (2020). Also the fact that the effect becomes stronger over time, is in line with what Pfeifer and Reutter (2020) find. Interestingly, also for the age group 18-19 year olds I find an increase in births as a result of the reform.

Table 2.15: **OLS Regression Log STIs Model 2**

	Basic	LA-FE cont	Never Pharm No Pharm Basic	Never Pharm LA-FE cont
16-19 × Treatment × 2001	0.0874	0.0606	0.0737	0.0553
	(0.0482)	(0.0429)	(0.0667)	(0.0627)
16-19 × Treatment × 2002	0.215***	0.132*	0.140	0.0907
	(0.0610)	(0.0534)	(0.0754)	(0.0682)
16-19 × Treatment × 2003	0.107	0.0331	0.0644	-0.0108
	(0.0705)	(0.0642)	(0.0705)	(0.0711)
16-19 × Treatment × 2004	0.237**	0.178**	0.240***	0.161*
	(0.0723)	(0.0607)	(0.0661)	(0.0652)
20-24 × Treatment × 2001	0.0802	0.0535	0.0505	0.0321
	(0.0508)	(0.0456)	(0.0720)	(0.0678)
20-24 × Treatment × 2002	0.247***	0.164**	0.159*	0.109
	(0.0622)	(0.0546)	(0.0753)	(0.0680)
20-24 × Treatment × 2003	0.145*	0.0711	0.0915	0.0163
	(0.0719)	(0.0669)	(0.0712)	(0.0703)
20-24 × Treatment × 2004	0.268***	0.209***	0.263***	0.184**
	(0.0709)	(0.0587)	(0.0652)	(0.0622)
Observations	1767	1767	1151	1151
Mean Log STIs	2.421	2.421	2.414	2.414
SD Log STIs	1.434	1.434	1.450	1.450

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.9. Column 1 and 2 in each year use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes neither local authority fixed effects nor controls. Column 2 includes both local authority fixed effects and controls. Columns 3 and 4 use the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Column 3 includes neither local authority fixed effects nor controls. Column 4 includes both local authority fixed effects and controls. All columns only report treatment coefficients $\beta_{a,t}$ even though regressions include other covariates as well.

In Table 2.15 one can see again that just like in Table 2.15, that there seem to be effects on STIs as well. However, these effects are only present for some years and are less robust to the inclusion of covariates and the reduction of the sample.

From the two different regressions, coming from Equation 2.8 and Equation 2.9 for the three different outcomes, log abortions, log births and log STIs once can conclude multiple facts: The reform seems to not have affected abortions for any of the age groups in the sample. Furthermore, the reform seems to have only affected births for the age group of 20-24 year olds by increasing these. The conclusions one can draw for sexually transmitted infections are rather mixed, since these seem less robust to the inclusion of covariates and the reduction of the sample. Also when allowing for the treatment effects to change over time, the effects on STIs are not present in all years, but rather seem to be switched on and off.

Specifically the results on births are very much in line with one of the most recent papers on the liberalization of EHC by Pfeifer and Reutter (2020). Using a cross country difference in differences specification, they also find an increase in births for a young group of women. They furthermore confirm these results by focussing on Germany, where they find the same result and are able to add to these results by including survey data. Using these additional data on whether pregnancies were planned or unplanned, as well as additional data on regular contraceptive use, they confirm that the increase in births stems from unplanned but not unwanted pregnancies. Pfeifer and Reutter (2020) results are very much in line with my findings on abortions, births, and STIs and hence, confirm that EHC liberalization seems to only increase births for a specific group of women, but not affect any other reproductive outcomes.

However, it is well known in the literature, that specifically with small samples standard errors using Difference in Differences estimations can be much smaller than they should be due to serial correlation (Bertrand, Duflo, and Mullainathan (2004)). Furthermore, in a setting in which I sample all english Local Authorities in which the policy of interest was in place, the uncertainty surrounding the estimates does not stem from sampling, but rather from the design of the experiment (Abadie et al. (2020)). Therefore, not only do I rely on cluster robust standard errors as shown in Tables 2.10 to 2.15, but in addition I also compute Randomization Inference p-values. Furthermore, in Section 2.9, I also try to adjust the estimation in order to take care of serial correlation as suggested by Bertrand, Duflo, and Mullainathan (2004).

2.8.1 Randomization Inference

As a further robustness check, rather than relying on robust standard errors, I perform randomization inference (RI) to obtain p-values for my estimates from my data. The idea behind RI is not to rely on asymptotics for the distribution of errors, which can be very misleading especially in cases of finite samples and complex error structures, but rather to permute treatment in the data and thereby obtain a series of different test-statistics to which one can compare the test-statistic of interest.

In general, randomization inference, follows four steps. First, one runs the regression of interest and stores the test-statistic of interest, which in my case are the three treatment estimates, one for each age groups. Second, one permutes treatment in the data, keeping all other things fixed. Third, one reruns the regression of interest for all possible permutations (or if it is not feasible to try all possible permutation, one randomly permutes treatment for example N= 10,000 times) and again saves the test-statistic of interest. In a fourth and final step one compares the original test statistic of interest, to the ones obtained from permuting treatment, in order to obtain a p-value. The p-value is calculated by comparing the share of test-statistics obtained from permuting are larger in absolute value than the original test-statistic of interest.[8]

In the English case, treatment occurs for three of the four age groups at the same point in time. I want to keep this structure of treatment even when performing randomization inference. I do so since this is the way treatment works in the English setting and I want to test it against the most similar setting so to say. Hence, even in the RI simulations, I do not allow for staggered treatment. Rather, I in a first step I permute the year in which treatment takes place and in a second step I permute the age groups which are treated in that year. By first defining the treatment year, and only then permuting over all combinations of 1,2, or 3 treated age groups, I make sure that in my permutations staggered treatment does not takes place. Furthermore, of course, just like in the data once an age group is treated, it is treated forever.

With this procedure, I obtain all possible $beta_a$ estimates with permuted treatment and then for each age-group I compare them to the $beta_a$ estimates of Model 2.8 without any covariates. I then plot the RI estimates for each outcome variable, the estimates of Model 2.8, as well as the 5th and 95th percentiles of the RI estimates for each outcome variable. Furthermore, I then calculate the RI p-value of each estimate by calculating the share of RI estimates larger in absolute value than the original estimate.

8. For more detailed explanations of RI, see for example Heß (2017)

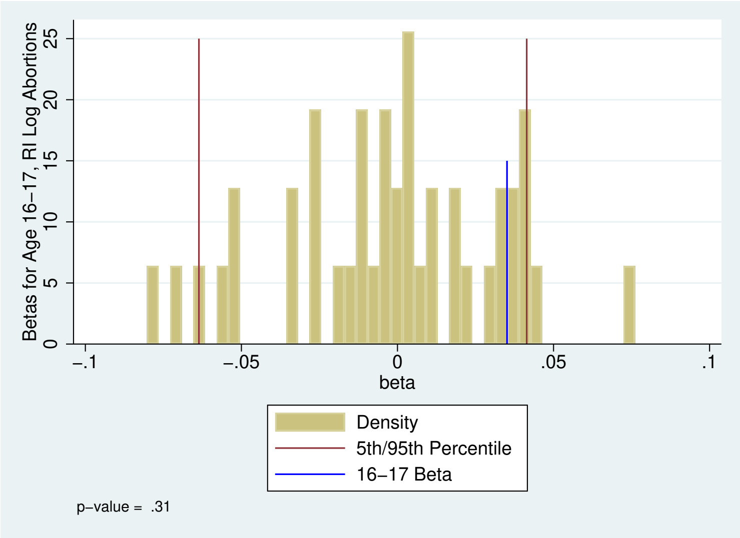


Figure 2.12: Randomization Inference: Abortions, Age 16-17

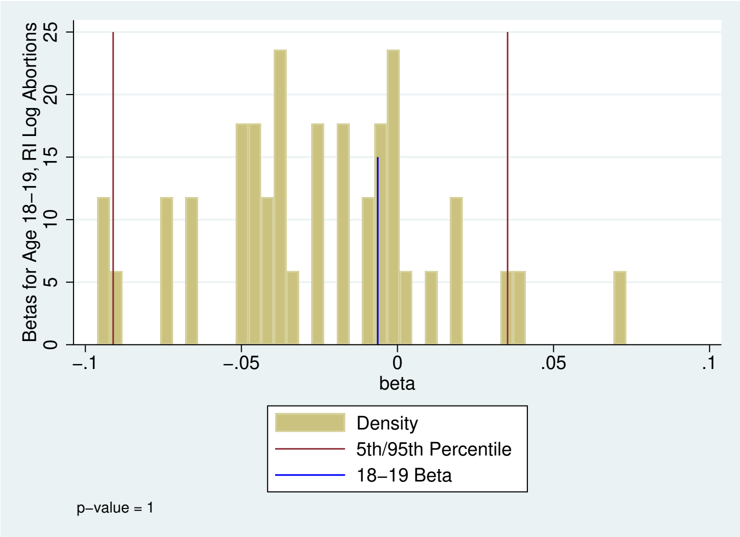


Figure 2.13: Randomization Inference: Abortions, Age 18-19

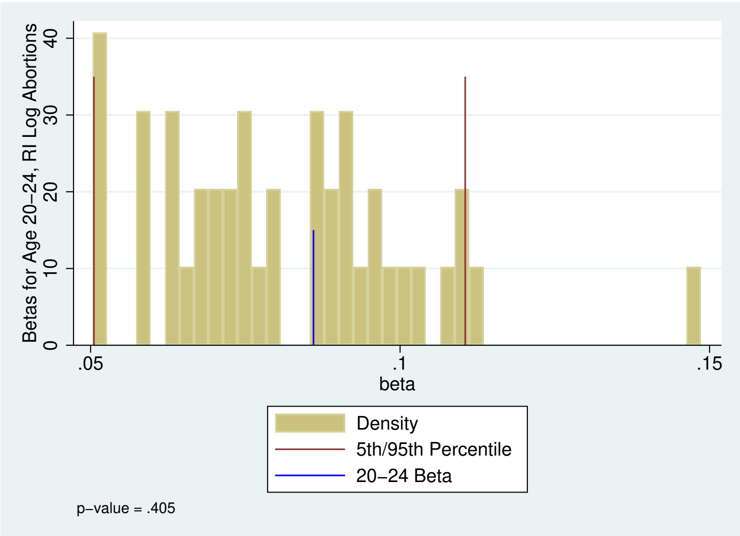


Figure 2.14: Randomization Inference: Abortions, Age 20-24

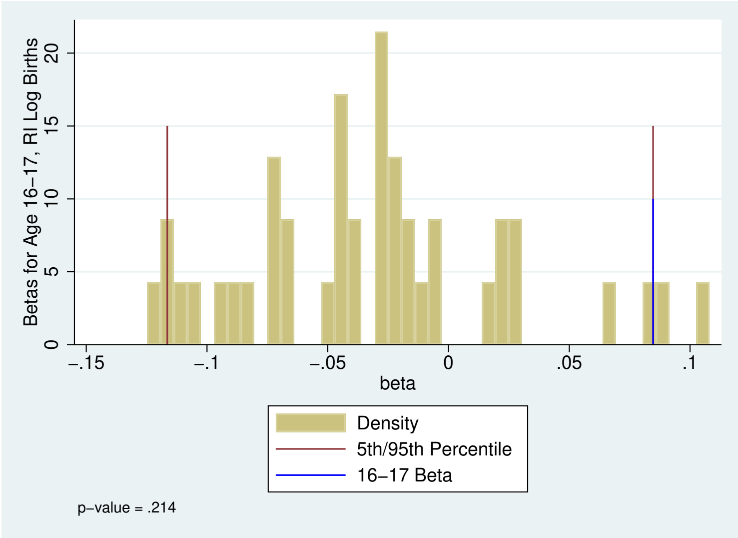


Figure 2.15: Randomization Inference: Births, Age 16-17

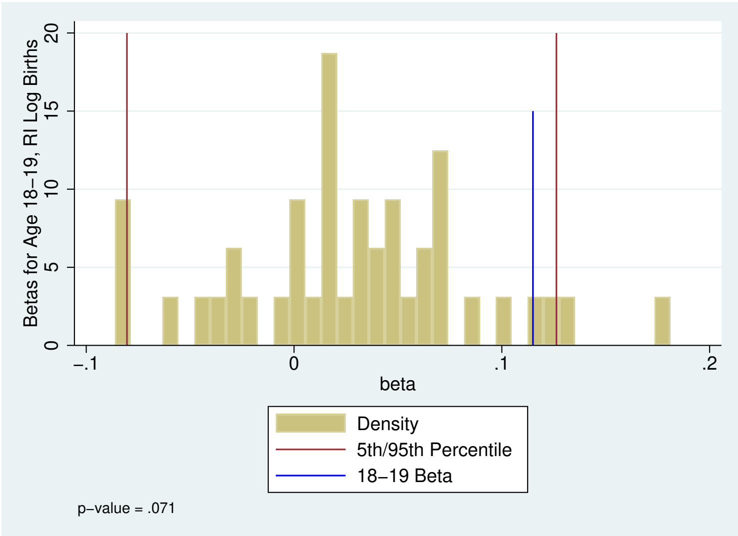


Figure 2.16: Randomization Inference: Births, Age 18-19

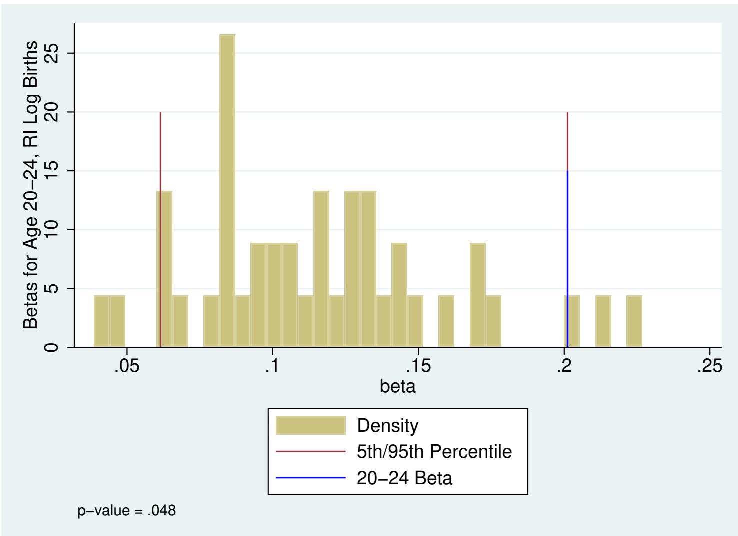


Figure 2.17: Randomization Inference: Births, Age 20-24

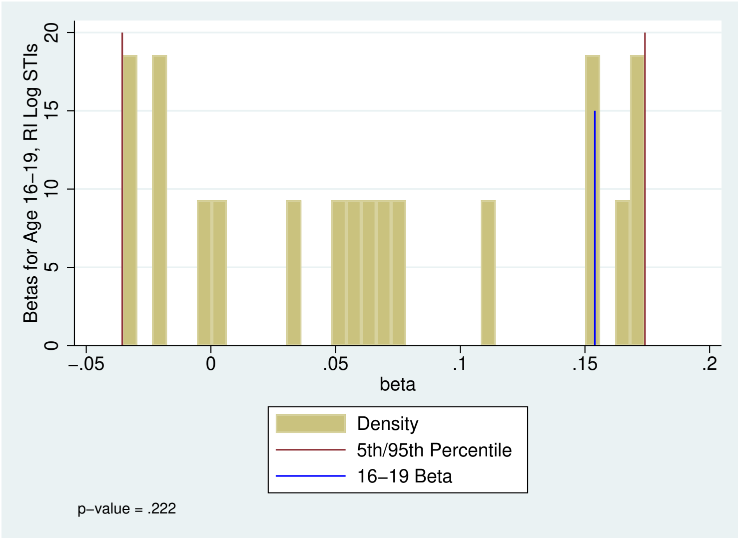


Figure 2.18: Randomization Inference: STIs, Age 16-19

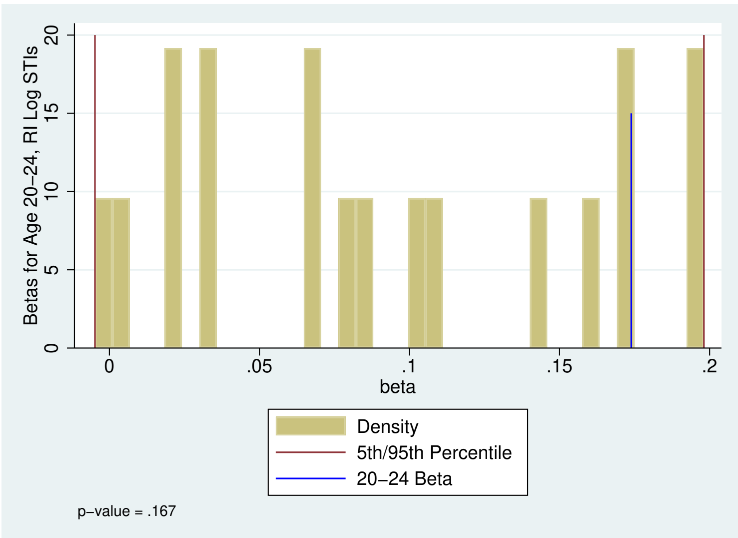


Figure 2.19: Randomization Inference: STIs, Age 20-24

As one can see, from Figures 2.12 to 2.19, for all three outcome variables, log abortions, log births and log STIs one can see the distribution of RI estimates. In red, in each Figure one can see the 5th and 95th percentile of the distribution of RI estimates. In blue, in each of the Figures, I display the actual estimate for the corresponding outcome and age group, and below the Figure, I display the RI p-value of the actual estimate. As one can see, when considering the p-values, only the RI p-value for the effect on logs births for the age group 20-24 remains significant. Again, this is very much in line with the paper by Pfeifer and Reutter (2020) and their findings on EHC only increase births for a young, but not the youngest group of women.

2.9 Robustness Checks

In this section I perform multiple robustness checks, in order to confirm the validity of the results.

2.9.1 Common Trends: Leads in Main Regression

As a first set of robustness checks, I test again for common trends, but this time, rather than performing a separate regression using only the pretreatment, I include the leads in the main regressions of interest. First, as described above, I include the leads in the in Equation 2.8. This inclusion leads to the following regression equation:

$$\begin{aligned} Y_{art} = \alpha_a + \lambda_t + \delta_r + \sum_{t=1998}^{2000} \omega_{at} \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year = t\} \\ + \beta_a \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year >= 2001\} + \gamma X_{rt} + \epsilon_{art} \end{aligned} \tag{2.11}$$

I then also include the leads in Equation 2.9, which gives rise to the following regression equation:

$$\begin{aligned} Y_{art} = \alpha_a + \lambda_t + \delta_r + \sum_{t=1998}^{2000} \omega_{at} \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year = t\} \\ + \sum_{t=2001}^{2004} \beta_{at} \mathbb{1}\{age \geq 16\} * \mathbb{1}\{year = t\} + \gamma X_{rt} + \epsilon_{art} \end{aligned} \tag{2.12}$$

Both in Equation 2.11 and 2.12 β is still the vector of the coefficients of interest, whereas ω simply displays the leads. The difference between the two equations, is the fact that while Equation 2.11 imposes one treatment coefficient for each age group, just as Equation 2.8 does. Equation 2.12 allows for the treatment coefficients of each age group to change over time, just as Equation 2.9 does. For simple reasons of readability, for the regressions using Equation 2.11 and 2.12, I only report the coefficients ω and β.

Table 2.16: OLS Regression Equation 2.11: Log Abortions with leads

β Treatment $\times$ 16-17	0.0508
	(0.0269)
β Treatment $\times$ 18-19	0.0303
	(0.0281)
β Treatment $\times$ 20-24	0.118***
	(0.0288)
Year = 1999 $\times$ 16-17	0.0648
	(0.0395)
Year = 1999 $\times$ 18-19	0.0890*
	(0.0372)
Year = 1999 $\times$ 20-24	0.0724
	(0.0376)
Year = 2000 $\times$ 16-17	-0.0177
	(0.0314)
Year = 2000 $\times$ 18-19	0.0208
	(0.0311)
Year = 2000 $\times$ 20-24	0.0246
	(0.0289)
Observations	2932

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.11. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

One can see from Tables 2.16 to 2.18 that including the leads in Equation 2.8 does not change the results qualitatively. Furthermore, also from these regressions, there seems to be no reason to doubt the assumption of common trends. The same holds true for including leads in Equation 2.9, which can be seen in Tables 2.19 to 2.21.

Table 2.17: OLS Regression Equation 2.11: Log Births with leads

β Treatment $\times$ 16-17	0.0393
	(0.0387)
β Treatment $\times$ 18-19	0.0947*
	(0.0400)
β Treatment $\times$ 20-24	0.167***
	(0.0405)
Year = 1999 $\times$ 16-17	-0.0285
	(0.0409)
Year = 1999 $\times$ 18-19	-0.00405
	(0.0411)
Year = 1999 $\times$ 20-24	-0.0224
	(0.0394)
Year = 2000 $\times$ 16-17	-0.0712
	(0.0428)
Year = 2000 $\times$ 18-19	-0.0164
	(0.0407)
Year = 2000 $\times$ 20-24	-0.0222
	(0.0386)
Observations	2928

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.11. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

Table 2.18: OLS Regression Equation 2.11: Log STIs with leads

β Treatment $\times$ 16-19	0.147**
	(0.0475)
β Treatment $\times$ 20-24	0.153**
	(0.0470)
Year = 1999 $\times$ 16-19	0.0529
	(0.0619)
Year = 1999 $\times$ 20-24	0.0379
	(0.0581)
Year = 2000 $\times$ 16-19	0.0971
	(0.0741)
Year = 2000 $\times$ 20-24	0.0703
	(0.0725)
Observations	1767

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.11. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

Table 2.19: OLS Regression Equation 2.12: Log Abortions with Leads

Year = 1999 × 16-17	0.0648
	(0.0396)
Year = 1999 × 18-19	0.0890*
	(0.0373)
Year = 1999 × 20-24	0.0724
	(0.0377)
Year = 2000 × 16-17	-0.0177
	(0.0315)
Year = 2000 × 18-19	0.0208
	(0.0311)
Year = 2000 × 20-24	0.0246
	(0.0290)
Year = 2001 × 16-17	0.0268
	(0.0363)
Year = 2001 × 18-19	0.00823
	(0.0347)
Year = 2001 × 20-24	0.0603
	(0.0355)
Year = 2002 × 16-17	0.0719*
	(0.0355)
Year = 2002 × 18-19	0.0291
	(0.0368)
Year = 2002 × 20-24	0.122**
	(0.0384)
Year = 2003 × 16-17	0.0206
	(0.0399)
Year = 2003 × 18-19	0.00729
	(0.0429)
Year = 2003 × 20-24	0.125**
	(0.0439)
Year = 2004 × 16-17	0.102*
	(0.0418)
Year = 2004 × 18-19	0.1000*
	(0.0426)
Year = 2004 × 20-24	0.210***
	(0.0465)
Observations	2932

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.12. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

Table 2.20: OLS Regression Equation 2.12: Log Births with Leads

Year = 1999 × 16-17	-0.0285
	(0.0410)
Year = 1999 × 18-19	-0.00405
	(0.0412)
Year = 1999 × 20-24	-0.0224
	(0.0394)
Year = 2000 × 16-17	-0.0712
	(0.0429)
Year = 2000 × 18-19	-0.0164
	(0.0408)
Year = 2000 × 20-24	-0.0222
	(0.0386)
Year = 2001 × 16-17	-0.00546
	(0.0431)
Year = 2001 × 18-19	0.0385
	(0.0438)
Year = 2001 × 20-24	0.0762
	(0.0410)
Year = 2002 × 16-17	0.0298
	(0.0462)
Year = 2002 × 18-19	0.0682
	(0.0454)
Year = 2002 × 20-24	0.145**
	(0.0486)
Year = 2003 × 16-17	0.0474
	(0.0632)
Year = 2003 × 18-19	0.0963
	(0.0572)
Year = 2003 × 20-24	0.212***
	(0.0587)
Year = 2004 × 16-17	0.125*
	(0.0539)
Year = 2004 × 18-19	0.232***
	(0.0589)
Year = 2004 × 20-24	0.305***
	(0.0623)
Observations	2928

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.11. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

Table 2.21: OLS Regression Equation 2.12: Log STIs with Leads

Year = 1999 × 16-19	0.0529
	(0.0620)
Year = 1999 × 20-24	0.0379
	(0.0582)
Year = 2000 × 16-19	0.0971
	(0.0742)
Year = 2000 × 20-24	0.0703
	(0.0726)
Year = 2001 × 16-19	0.111*
	(0.0535)
Year = 2001 × 20-24	0.0904
	(0.0559)
Year = 2002 × 16-19	0.183**
	(0.0588)
Year = 2002 × 20-24	0.201***
	(0.0584)
Year = 2003 × 16-19	0.0839
	(0.0722)
Year = 2003 × 20-24	0.108
	(0.0731)
Year = 2004 × 16-19	0.229**
	(0.0676)
Year = 2004 × 20-24	0.246***
	(0.0623)
Observations	1767

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.11. In each year I use the adaptive sample of local authorities which did not have a pharmacy scheme in place in that year. Column 1 includes local authority fixed effects and controls. I only report treatment coefficients $\beta_{a,t}$ and leads for treatment effects $\omega_{a,t}$ as test for common trends even-though regressions include other covariates as well.

2.9.2 Serial Correlation

A further robustness check I perform is to consider serial correlation. Like pointed out by Bertrand, Duflo, and Mullainathan (2004), when doing difference in difference estimations it is necessary to account for serial correlation in the outcome variable. Since the number of age groups is very small for this study, I follow the simplest approach to account for serial correlation. I aggregate the data for each age group and each local authority into one before and one after treatment observation. I then perform the a similar Difference in Differences estimation as specified in equation 2.8. Since I aggregate into one pre and one post treatment time period, I no longer have year fixed effects, but simply a dummy variable for the post period. Furthermore, I no longer include local authority fixed effects nor include controls. Hence, the regression equation can be formulated as follows:

$$Y_{art} = \alpha_a + \lambda \mathbb{1}\{t = post\} + \beta_a \mathbb{1}\{age \geq 16\} * \mathbb{1}\{t = post\} + \epsilon_{at} \tag{2.13}$$

These estimations of Equation 2.13 give rise to the regression outputs in Table 2.22 for abortions, in Table 2.23 for births and in Table 2.24 for STIs. Similar to above, column 1 uses the largest sample, whereas column 2 uses the smallest possible sample.

One can see that for all of the outcomes all of the treatment estimates are insignificant. However, for births, which seems to be the outcome for which all of the estimates are most robust, the direction of the estimates stays the same: In Table 2.23 the reform still hints at a positive effect on births, even if this is no longer significant. In Table 2.22 also the direction of estimates remains the same, but already in other specifications these results were not as robust as those on births. However, in Tables 2.24 one can see that the sign of the estimates for treatment for STIs switches.

The fact that estimates are no longer significant when aggregating into one pre and one post treatment period is not surprising, due to the large loss of power due to the reduction in sample size. In order to however still learn something from the estimates, without relying on potentially biased standard errors, as a next robustness check, I perform randomization inference.

Table 2.22: OLS Regression Log Abortions: Serial Correlation

	Basic	Never Pharm Basic
Age = 16-17	1.192***	1.215***
	(0.0727)	(0.0865)
Age = 18-19	1.554***	1.569***
	(0.0732)	(0.0881)
Age = 20-24	2.339***	2.387***
	(0.0767)	(0.0968)
Post	0.0122	0.0276
	(0.0755)	(0.0895)
Treatment × 16-17	0.0245	0.0155
	(0.103)	(0.123)
Treatment × 18-19	-0.0137	-0.0196
	(0.105)	(0.127)
Treatment × 20-24	0.0592	0.0551
	(0.110)	(0.138)
Constant	3.116***	3.053***
	(0.0521)	(0.0618)
Observations	976	544
Sd Log Abortions	4.402	4.366
Sd Log Abortions	1.032	1.020

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.13. Column 1 uses the adaptive sample of local authorities which did not have a pharmacy scheme in place in each year. This sample is then collapse into one pre and one post period. Column 2 uses the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Both column 1 and 2 do not include local authority fixed effects or controls.

Table 2.23: **OLS Regression Log Births: Serial Correlation**

	Basic	Never Pharm Basic
Age = 16-17	1.682***	1.666***
	(0.0849)	(0.107)
Age = 18-19	2.305***	2.306***
	(0.0832)	(0.105)
Age = 20-24	3.479***	3.480***
	(0.0818)	(0.103)
Post	-0.211*	-0.123
	(0.0970)	(0.113)
Treatment × 16-17	0.0971	0.108
	(0.131)	(0.155)
Treatment × 18-19	0.120	0.128
	(0.129)	(0.152)
Treatment × 20-24	0.208	0.207
	(0.127)	(0.150)
Constant	2.917***	2.857***
	(0.0609)	(0.0761)
Observations	860	544
Mean Log Births	4.738	4.714
Sd Log Births	1.439	1.431

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.13. Column 1 uses the adaptive sample of local authorities which did not have a pharmacy scheme in place in each year. This sample is then collapse into one pre and one post period. Column 2 uses the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Both column 1 and 2 do not include local authority fixed effects or controls.

Table 2.24: **OLS Regression Log STIs: Serial Correlation**

	Basic	Never Pharm Basic
Age = 16-19	2.799***	2.832***
	(0.0984)	(0.136)
Age = 20-24	3.221***	3.258***
	(0.102)	(0.144)
Post	0.308**	0.381**
	(0.0954)	(0.129)
Treatment × 16-19	-0.0210	-0.0382
	(0.134)	(0.188)
Treatment × 20-24	-0.00746	-0.0196
	(0.140)	(0.197)
Constant	0.0393	-0.0660
	(0.0717)	(0.0962)
Observations	622	347
Mean Log STIs	2.280	2.222
Sd Log STIs	1.569	1.604

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: OLS regression for Equation 2.13. Column 1 uses the adaptive sample of local authorities which did not have a pharmacy scheme in place in each year. This sample is then collapse into one pre and one post period. Column 2 uses the smallest sample of only those local authorities which did not have a pharmacy scheme in place until at least 2005. Both column 1 and 2 do not include local authority fixed effects or controls.

2.10 Conclusion

In this paper I explore the causal effects of a deregulation of emergency hormonal contraception (EHC) on the reproductive behavior of young women in England. Emergency hormonal contraception, also known as the morning after pill, is used to prevent a pregnancy after unprotected intercourse has taken place, thereby acting as a form of insurance against an unwanted pregnancy. The deregulation, which took place in 2001, allowed all women aged 16 and over to no longer require a prescription for obtaining EHC in pharmacies. While however the requirement for under 16 year old women was kept upright.

This liberalization of EHC is a perfect example of how policies, intended to reduce risk, can potentially have unintended effects by increasing risky behavior. These unintended effects can be well understood when considering the theoretical model of individual behavior of a decision process, which I build. First women choose risk; then, they decide on whether or not to take EHC, still being unsure about their pregnancy status; finally, women can have an abortion. In this model a decrease in the cost of EHC, which is equivalent to the liberalization, leads to an ambiguous effect on pregnancies, and hence abortions and births, due to two counteracting effects: on the one hand EHC reduces pregnancies, by being a form of contraception. On the other hand the lower costs of EHC lead some individuals to take more risk causing an increase in pregnancies and thus abortions and pregnancies. The reason for this is that the availability of EHC allows for leeway in the risk decision in the first step, by providing a second chance to prevent a pregnancy. This potential increase in risk is an unintended effect of the liberalization of EHC, which was meant to decrease the risk of unwanted pregnancies.

In addition to the theoretical analysis, I analyze the reform empirically using a difference-in-differences approach: The fact that EHC was only liberalized for women aged 16 and older allows me to use women under the age of 16 as a control group. I then compare how abortions, births, as well as sexually transmitted infections change differentially over time for women aged 16 and older in comparison to women aged under 16. My analysis focusses on young women, those under the age of 25, because these are women who are potentially most effected by not needing a doctor's appointment and prescription any longer due to the following reasons: First, the policy's unintended consequences might be specifically strong when regarding pregnancies: teenage pregnancies are not only a driver of poverty and lead to lower education of teenage mothers, but also have strong long lasting effects on the children of teenage mothers[9]. Second, young women are also more likely to use EHC, and hence the impact of the reform might be stronger for them[10].

From this difference-in-differences analysis I find that births to women aged 20-24 increased significantly after the reform. I find no effects on abortions or STIs, specifically when using Randomization Inference to obtain robust p-values. Randomization Inference p-values allow me to not make any assumptions on the distribution of error terms and obtain correct p-values for estimates even in the light of serially correlated standard errors and few clusters, which is often the case in difference-in-differences estimations Bertrand, Duflo, and Mullainathan (2004).

These findings are very much in line with recent empirical research on Emergency Hormonal Contraception liberalization in Europe by Pfeifer and Reutter (2020), who also find an increase only in births for women in a similar age range due to liberalization of the morning after pill.

This paper contributes to the literature in two separate ways: First, I contribute to the theoretical literature on EHC and more general abortion, by building a model, which allows for an endogenous choice of of contraception effort and EHC intake. Second, unlike existing studies, which try to look at a difference in prescription requirements by exploiting regional variation in prescription requirements, my analysis does not rely on the assumption that the regions are similar. A potentially large problem when using regional variation, even if it is just the comparison of neighboring states in the US, is that if states endogenously choose to offer EHC without prescription because they have higher abortion or birth rates than other states, this will create a biased estimate. The advantage of using an age rather than a regional discontinuity is that individuals unlike regions cannot self select into treatment. This is where my main contribution to the existing literature lies.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge. 2017. *When Should You Adjust Standard Errors for Clustering?* Technical report. National Bureau of Economic Research.

9. See Hendrick and Maslowsky (2019) and Hoffman and Maynard (2008)
10. See Daniels and Abma (2013) as well as Figure 2.11

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge. 2020. "Sampling-Based versus Design-Based Uncertainty in Regression Analysis." *Econometrica* 88 (1): 265–296.

Ananat, E. O., J. Gruber, P. B. Levine, and D. Staiger. 2009. "Abortion and selection." *The Review of Economics and Statistics* 91 (1): 124–136.

Arrow, K. J. 1963. "The american." *The American Economic Review* 53 (5): 941–973.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics* 119 (1): 249–275.

Black, A., D. Francoeur, T. Rowe, J. Collins, D. Miller, T. Brown, M. David, S. Dunn, W. Fisher, N. Fleming, et al. 2004. "Canadian contraception consensus." *Journal of obstetrics and gynaecology Canadaa:JOGC* 26 (4): 347–87.

Cintina, I., and M. S. Johansen. 2015. "The effect of Plan B on teen abortions: evidence from the 2006 FDA ruling." *Contemporary Economic Policy* 33 (3): 418–433.

Daniels, K., and J. C. Abma. 2013. *Use of Emergency Contraception Among Women Aged 15-44, United States, 2006-2010.* 112. US Department of Health / Human Services, Centers for Disease Control and ...

DGGG. 2015. "208. Stellungnahme der Deutschen Gesellschaft fuer Gynaekologie und Geburtshilfe zur Position des Berufsverbandes der Frauenaerzte und der Deutschen Gesellschaft fuer Gynaekologische Endokrinologie und Fortpflanzungsmedizin."

Durrance, C. P. 2013. "The effects of increased access to emergency contraception on sexually transmitted disease and abortion rates." *Economic Inquiry* 51 (3): 1682–1695.

Girma, S., and D. Paton. 2011. "The impact of emergency birth control on teen pregnancy and STIs." *Journal of Health Economics* 30 (2): 373–380.

Glasier, A., E. Ketting, C. Ellertson, and E. Armstrong. 1996. "Emergency Contraception in the United Kingdom and the Netherlands." *Family Planning Perspectives* 28 (2): 49–51.

Gross, T., J. Lafortune, and C. Low. 2014. "What happens the morning after? The costs and benefits of expanding access to emergency contraception." *Journal of Policy Analysis and Management* 33 (1): 70–93.

Hendrick, C. E., and J. Maslowsky. 2019. “Teen mothers’ educational attainment and their children’s risk for teenage childbearing.” *Developmental psychology* 55 (6): 1259.

Heß, S. 2017. “Randomization inference with Stata: A guide and software.” *The Stata Journal* 17 (3): 630–651.

Hoffman, S. D., and R. A. Maynard. 2008. *Kids having kids: Economic costs & social consequences of teen pregnancy.* The Urban Insitute.

Levine, P. B., and D. Staiger. 2002. *Abortion as insurance.* Technical report. National Bureau of Economic Research.

Pfeifer, G., and M. Reutter. 2020. “The Morning After: Prescription-Free Access to Emergency Contraceptive Pills.”

Weismiller, D. G. 2004. “Emergency contraception.” *American Family Physician* 70 (4): 707–714.

Model with two positive risk levels

In order to see better how increased availability of EHC can lead to an increase in risk taken, it is useful to also briefly outline the model when $r = l$ is not restricted to be zero. Hence, when there are two positive levels of risk. If this is the case there are possibly seven different groups of women, Never Takers with low and with high risk, Compliers with low and high risk, Always Takers with low and high risk and Non Participators, i.e. women who choose not to have sex. Depending on the parametrization, not all of these groups exist, in the sense that some of this behavior might not be optimal for any level of w. Furthermore, it is no longer possible to calculate the the thresholds between the groups by hand, but rather one needs a computing tool. From this addition to the model one can learn how risk taken changes when the cost of EHC changes. To show this, I plot utilities for each group once for a high cost of EHC and then for a low cost of EHC.

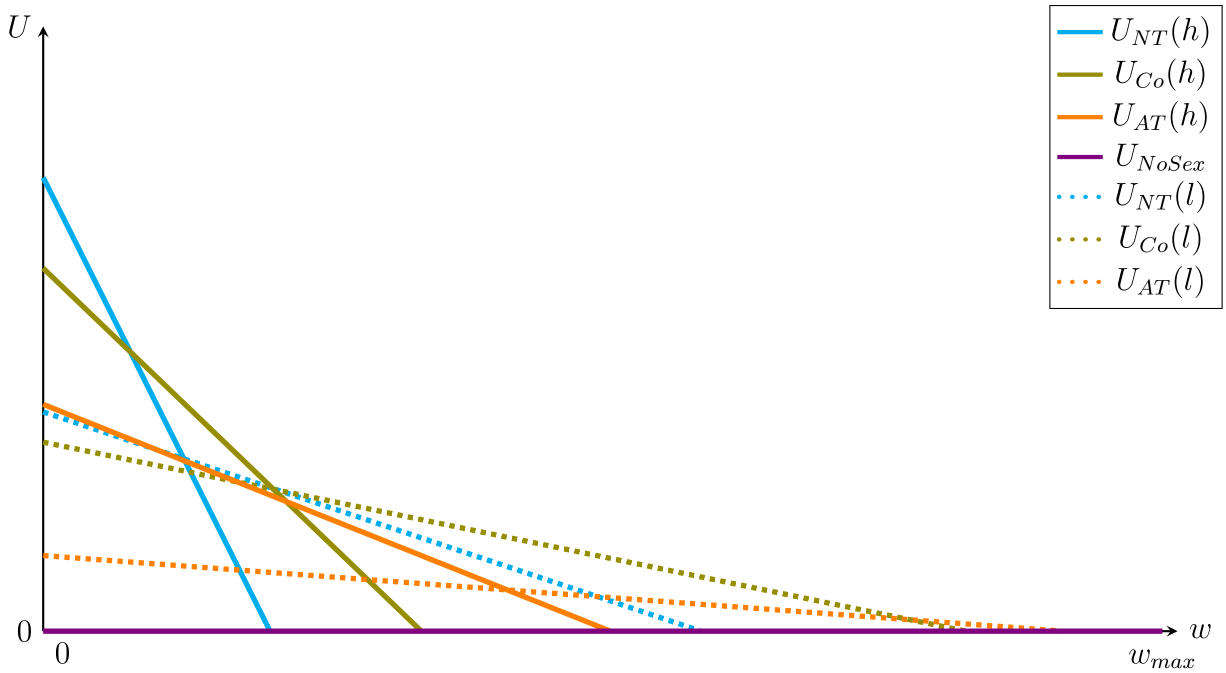


Figure 2.20: Utilities with two levels of risk and high c

Figure 2.20 displays the utilities for the different groups under a high level of c whereas Figure 2.21 displays the same utilities under a low level of c. As one can see, the decrease in c not only changes the thresholds between the groups present under the high level of c, but it also changes whether some groups are ever optimal. In this particular example, the reduction of c leads to the group of Always Takers with high risk becoming optimal for some levels of w. Before the reduction such behavior was

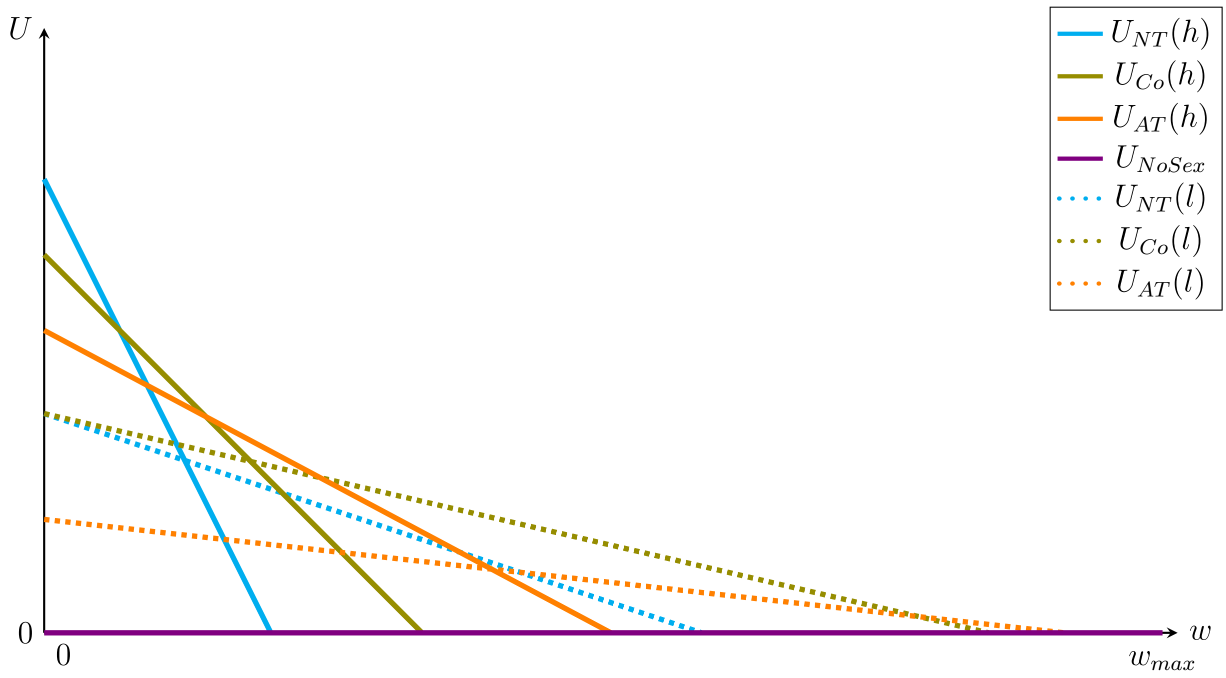


Figure 2.21: Utilities of different types as a function of w

never optimal. This very crude example makes very clear that a reduction in EHC costs can lead to increased risk taking.

Summarizing, already the simplified model with just one positive level of risk makes it possible to theoretically analyze what happens after a reduction in c to EHC consumption, risk taken, pregnancies as well as abortion or birth probabilities. EHC consumption increases, risk taken increases and the effects on the latter three outcomes are ambiguous.

Chapter 3

Multiple Imputation of University Degree Attainment

joint with Nurfatima Jandarova

3.1 Introduction

The second half of the 20th century has seen a massive expansion of university education throughout the world. In the UK, the participation rate in higher education rose from 4.1% in 1960[1] to about 20% in 1990[2]. But not all higher education is made equal. From 1965 to 1992, students in the UK could earn their degrees either from traditional universities or from public sector colleges led by polytechnic institutions. In 1992 the vast majority of the polytechnics were converted to universities. Formally, degrees from polytechnics were of the same standard as university degrees. Nevertheless, the institutions faced different target populations, admission procedures, subjects taught, organization and financing schemes. These differences, together with the elite image of the traditional universities, contributed to a public perception of polytechnics degrees as inferior to that of universities (Willetts 2017; Pratt 1997).

This perceived inferiority hints at something that has been established in the literature: the type of higher education institution can act as a signal of education quality. Thus, types of higher education institutions could serve as signals of education quality. College quality is an important determinant of educational decisions and

1. *The Robbins Report* (1963)
2. *The Dearing Report* (1997)

returns to education. Brewer, Eide, and Ehrenberg (1999), Black and Smith (2004) and Black and Smith (2006) report that attending a higher quality institution is associated with sizeable private wage returns. Dillon and Smith (2017) find that students and their families prefer high-quality universities, even when it is not the best match given the ability of the student. Boliver (2015) finds that a binary divide persists even nowadays based on university characteristics such as research quality, teaching quality and selectivity of admissions. This makes clear that the type of higher education institution can be considered of high importance.

However, common survey data sets often offer limited information about the types of institutions from which individuals earned their degrees. For example, the UK Household Longitudinal Study (UKHLS), the largest panel study in the UK starting from 2009, asked its subjects to identify the higher education institution from which they have received their degree only in the most recent wave 11 (2019-21)[3]. Furthermore, researchers may face additional restrictions before being granted access to such information, as is the case for the UKHLS. This leaves room for imputation-based ways to distinguish between the types of higher education institutions which can be used either as a preliminary or alternative analysis.

In this paper, we try to overcome the issue of missing higher education institution types by using a multiple imputation technique. We rely on the institution type information available in the British Household Panel Study (BHPS), a smaller panel study carried out from 1991 to 2008, as well as the close relationship between the two panel studies. In particular, the BHPS specifically asked its participants to indicate the type of institution last attended, distinguishing between universities and polytechnics. The survey designs are highly comparable between the two studies. Thus, we can transform the lack of institution type into a missing data problem in a combined data set of the BHPS and the UKHLS. In addition, many of the former BHPS respondents are now part of the UKHLS, presenting us with a second strategy of using the BHPS subsample within the UKHLS for imputation.

To properly reflect the uncertainty about imputed values we use a multiple imputation technique (Rubin 1977). By imputing multiple values for each missing observation, we can perform our analysis of interest multiple times and combine the estimated parameters. The combined estimators then reflect both sampling and imputation

3. A first attempt was made in wave 5 (2013-15), but only individuals that earned their degrees from 1995 onwards were eligible to answer the question. In wave 11, all adult participants with a higher degree and interviewed face-to-face were asked for higher education institution details.

uncertainty. For them to deliver valid inference, two crucial assumptions must hold. First, the probability of missing data cannot depend on the missing value. This is the so-called missing at random assumption. We provide evidence from covariate balance tests to support this assumption. Second, the imputation model must be proper. To check this assumption we use the simulation-based evaluation method proposed by Brand et al. (2003).

When constructing the imputation model, another important consideration we take into account is the agreement between imputation and analysis models (Schafer 1997). The agreement means that the imputation model should be consistent with the model that the researchers want to estimate given the research question. In this paper, we adopt the following research question from our companion paper Ichino et al. (n.d.): how did the expansion of the higher education in the UK change the composition of students in terms of their intelligence scores? Differentiating between traditional universities and former polytechnics is crucial for such an analysis. Since the two institution types targeted different types of applicants, it is reasonable to expect that they also faced distinct trends in the composition of the student body. In the context of our companion paper, we are interested in the relationship between the probability of getting a degree from a traditional university and the intelligence score of the student. We also would like to examine how this relationship changed over time. Therefore, the imputation model should ideally control for time trends and intelligence scores. In practice, the intelligence score variable is only available for a subset of the UKHLS panel. Therefore, we test three versions of the imputation model that differ in estimation samples and the inclusion of intelligence score variable.

We find that the imputation models with and without intelligence scores perform similarly across all dimensions. In the simulation-based evaluation, the two models produce combined estimators with similar bias and efficiency statistics for the marginal effect of the intelligence score on the average university degree attainment. We also show that the combined estimators of the average university degree attainment across cohorts is, in general, similar to the benchmark graduation rates computed using the USR and the HESA data set. This similarity could allow us to use a simpler imputation model without the intelligence score in our companion paper.

The rest of the paper is structured as follows. In the next section we discuss the institutional differences between universities and polytechnics. We describe and compare the BHPS and the UKHLS data sets in section 3.3. In section 3.4, we provide a brief introduction to multiple imputation, examine crucial assumptions, construct imputation models and describe the simulation-based evaluation method. Finally, we discuss the results in section 3.5 and conclude in section 3.6.

3.2 Institutional background

The higher education system in the UK was characterised by a binary divide until 1992: tertiary education was provided both by independent universities and public colleges[4]. This division was motivated by the desire of the government to adapt training according to “national economic needs for specific skills” (Willetts 2017, p.52). These public sector colleges were funded and organised by local education authorities and provided vocational and other training necessary to meet local demand for skills.

The role of these public sector colleges was an important one. They included teacher training colleges, nursing colleges and polytechnics, where much of the higher education in technical and scientific subjects took place (Gillard 1998). Already in 1883, the first polytechnics were founded in order to “promote the industrial skill, general knowledge, health and well-being of young men and women belonging to the poorer classes” (Lawson and Silver, cited in Gillard 1998, p.83). However, over time the view on polytechnics and their role in higher education shifted away from being just for the poor, but rather institutions which provided a technical and scientific higher education.

The early 1960s saw two policy changes important for higher education decisions of young individuals: the end of military conscription in 1960[5] and introduction of the centralised applications via the Universities Central Council on Admissions (UCCA) in 1961. As a result, in 1961 the Prime Minister Harold MacMillan announced creation of a committee headed by Lord Robbins “to review the pattern of ... higher education ... and advise ... the government on what principles its long-term development should be based” (*The Robbins Report* 1963, p.1). The report of this committee, also known as the Robbins Report, suggested the unification of the higher education system. Nevertheless, the government followed the idea of the Education Secretary at the time, Anthony Crosland, to adhere to the policy of a binary divide between universities and public sector colleges (Gillard 1998). Within the public sector colleges, polytechnics were the main instrument through which the binary policy was implemented. In 1966 the

4. Most of the public colleges have been present in the UK from at least 1870 onwards alongside the traditional universities, such as Oxford and Cambridge (Gillard 1998). However, many of them were merged in the 1960s to create new polytechnics that were distinctively defined by the government (Pratt 1997).

5. Call-ups for military conscription ended on 31 December 1960.

Government published a White paper detailing creation of 28[6] polytechnic institutions. These new polytechnics were formed by merging over 50 existing colleges (Pratt 1997).

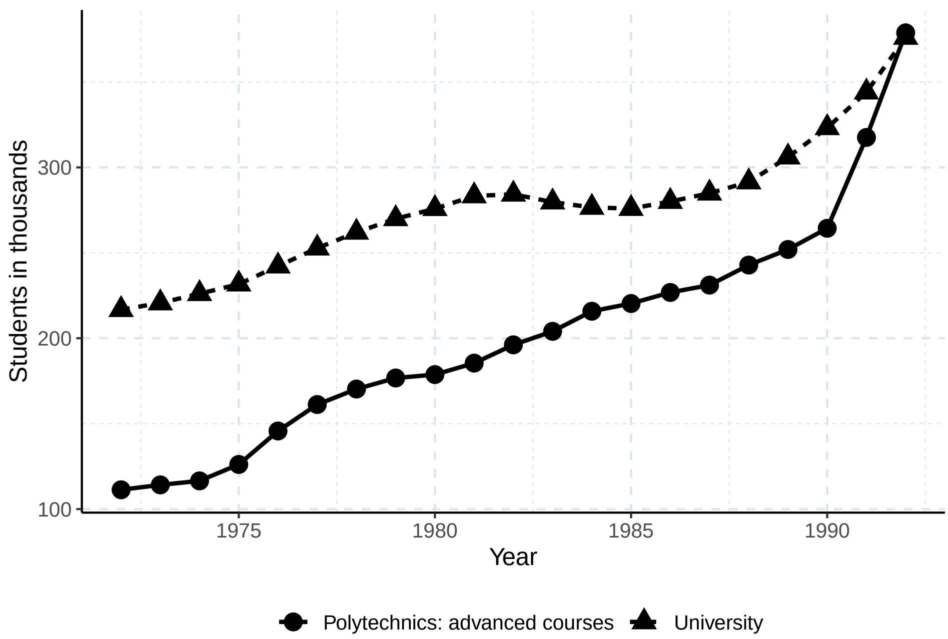
Only in 1992 did the 'binary divide' come to an end with the Further and Higher Education Act, which allowed polytechnics to obtain university status. A majority of the institutions used this option immediately: the number of universities and university students almost doubled over the next two years. The abolishment of the binary system gave rise to what are known as old universities (universities that existed before 1992) and new universities (former polytechnics that became universities after 1992).

In order to better understand the differences and similarities between the different types of higher education, it is important to shed further light on their histories and students. The initial idea of the binary divide was that polytechnics would constitute a parallel form of higher education with a special "commitment to non-degree students and to part-time courses" (Pratt 1997). Thus, polytechnics allowed the government to cater higher education to a wide range of students. The number of students in advanced courses, including degree courses, in polytechnics had been steadily rising over the years, catching up with the number of university students in 1992 (Figure 3.1). Unlike universities, the polytechnics could not award their own degrees. Therefore, most of the degree courses offered by the polytechnics were validated by the Council for National Academic Awards (CNAA). The CNAA was established in 1964 with the aim of granting awards to non-university students who had completed courses of study comparable in standards to university (Pratt 1997).

Despite polytechnics shifting the focus from non-advanced to advanced courses, they were more oriented towards undergraduate and part-time students compared to universities, based on the data reported in Pratt (1997). Although the overall share of part-time students in polytechnics fell from over 70% in 1965 to about 30% in 1988, most of the decline was driven by reduction in non-degree courses. The share of part-time students on advanced courses fell modestly from 40% in 1972 to 30% in 1992. At the same time, part-time students in universities rose from 9% in 1972 to 15% in 1992. In terms of level of study, the share of undergraduate students in degree courses in polytechnics remained at about 87%.

Polytechnics encompassed students from a wider range of backgrounds. For example, non-white students accounted for 14% of students in degree courses in polytechnics in 1991, compared to 8% in universities. Students aged 21 and over constituted about 50% of all full-time and more than 85% of part-time students in polytechnics. The focus on

6. Later the number was increased to 30.



Notes: Figure 3.2 in Pratt (1997).

Figure 3.1: Number of students in polytechnics and universities

a wider population also led polytechnics to have broader admission criteria. While 70% of university students were admitted based on 3 or more GCE A-level passes in 1990, the corresponding share among polytechnics entrants was only 34%.

These figures suggest that the polytechnics did indeed provide access to higher education to a larger group of people. But it came at the expense of quality perception: polytechnics were not viewed as a parallel form of higher education, rather they were seen as secondary to universities (Willetts 2017; Pratt 1997).

All of these factors suggest that distinguishing the degree-holders by types of institutions is important, especially given the context laid out by our companion paper Ichino et al. (n.d.). As explained above, in this companion paper, we are interested in changes in the composition of students during the massive expansion of higher education in the UK from 1960 to 1990. We are especially interested in how individuals of different abilities sort into different types of higher education institutions (HEIs). In this regard, it is important to separate HEI types for the following reasons. First, universities and polytechnics taught different subjects and had different objectives. Second, differences in public perception of the two types of HEIs could also translate to different returns to higher education. These concerns could imply different sorting patterns of individuals in types of higher education institutions based on their abilities. These ability-sorting patterns could have persisted even past 1992.

3.3 Data

We use the UK Household Longitudinal Study[7] (UKHLS), also known as the Understanding Society. This is the largest household panel study in the UK of about 40,000 individuals that started in 2009. The UKHLS is tightly related to a previous smaller longitudinal study, the British Household Panel Survey (BHPS), that was carried out in 18 waves between 1991 and 2009 and covered around 10,000 individuals. The UKHLS questionnaires were built upon those of the BHPS, ensuring continuity of many variables between the two studies. In addition to this, respondents in the last wave of the BHPS were asked if they were willing to join the UKHLS and about 80% of them agreed and were followed within the UKHLS.

The UKHLS covers a wide range of topics and is one of the most popular data sources for research. In the context of our project, however, the UKHLS offers very limited information about the type of institutions respondents received their higher education qualifications from. This limitation comes from the fact that only individuals who were full-time students at the time of the interview were asked about the type of institution attended[8]. We seek to exploit the close relationship between the BHPS and the UKHLS to overcome this limitation. The main advantage of the BHPS for our purposes is that it attempts to distinguish between different types of higher education institutions (HEIs), possibly reflecting the importance of reorganisation in the university sector at the time. In particular, respondents were asked to categorise their further education institution last attended among the following: nursing school, college of further education, other training establishment, polytechnic, university or other. Thus, we can use the information on the institution type in the BHPS to impute the variable in the UKHLS.

Our working sample consists of respondents born in the UK between 1950 and 1984[9] with non-missing information on the highest qualification obtained (from any institution type) and non-zero response weights. We exclude respondents who were still in education. We also exclude respondents from the minority boost samples. This

7. University of Essex, Institute for Social and Economic Research (2020)

8. In wave 5, respondents who completed their studies after 1995 were asked to give the name of the institution they attended. In the currently undergoing wave 11, every respondent is asked about their higher education institution, irrespective of completion status and date of completion. The data from wave 11 is to be released at the end of 2021.

9. That is, we exclude people born before or during the war. We also exclude cohorts born after 1985 as members of these cohorts had not yet completed their education. We verified that the share of respondents in UKHLS wave 3 having a degree as their highest qualification (from any institution type) starts declining past 1985.

results in 6,800 observations in wave 18 of the BHPS and 20,771 observations in wave 3 of the UKHLS, of which 5,117 are former BHPS subjects.

Our main variable of interest is the highest qualification achieved. This is a categorical variable with the following values: no qualification, GCSE or equivalent, A-level or equivalent, other higher degree[10], degree and other qualifications. We construct a binary degree attainment measure equal to one if the respondent has a degree (from any type of HEI). We define a university degree attainment variable as equal to one if a person both has a degree and attended a university as indicated by the type of HEI last attended. The latter variable is only defined for the BHPS subsample and is missing for all the UKHLS respondents.

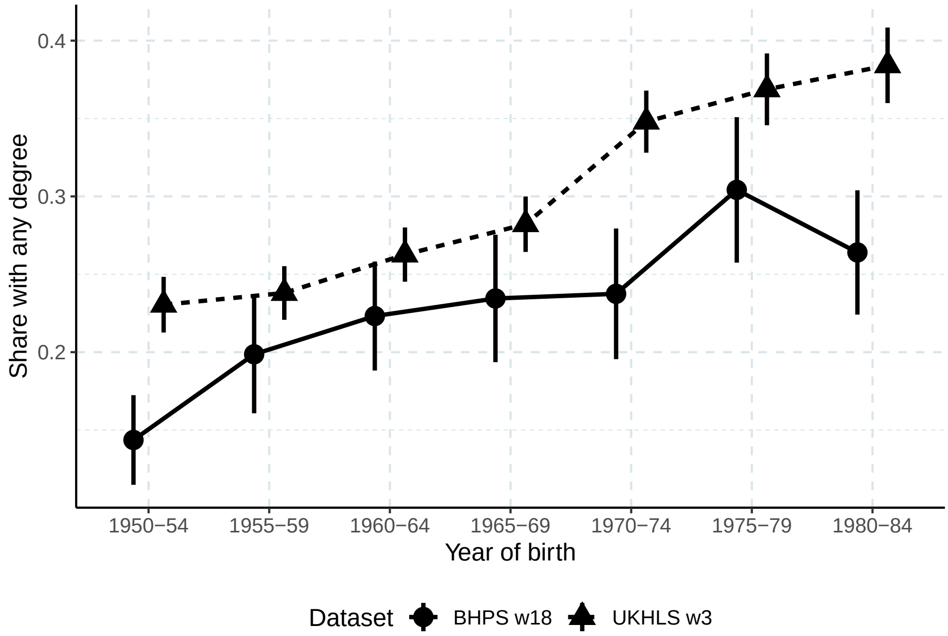
We also need a measurement of cognitive abilitiy of respondents since we adopt the analysis framework of our companion paper Ichino et al. (n.d.). Therefore, we focus on wave 3 of the UKHLS as it is the only wave in which the participants were administered cognitive ability tests. The tests were composed of five parts: word recall, serial 7 subtraction, number series, verbal fluency and numeric ability. We combine the counts of correct answers to each of these tests into a single intelligence score using principal component analysis and extracting the first component[11]. In order to abstract from age-related differences in the test scores, we standardize both the correct answer counts and the resulting intelligence score within each decennial year of birth group.

Furthermore, we use two external data sources on university graduates to establish a benchmark for university degree attainment over time. The Universities Statistical Records (USR) provides detailed annual information about all universities funded by the University Grants Committee during the period from 1972 to 1993. Specifically, we use undergraduate records with student-level microdata. From 1994 onwards this data was released by the Higher Education Statistics Agency (HESA) in the form of aggregated tables. In particular, we use data on the number of first-degree graduates at each university. We combine the two sources to create a time-series of number of graduates from old universities, a measure closely related to university degree attainment we are interested in. Note that the USR only covers the old universities, so we count all first-degree British graduates in the data set. In the HESA tables, we exclude student counts in new universities (i.e., former polytechnics). To compute graduation rates we divide the series by the population of 21-year-olds at the time of graduation. In the rest of the paper, we call this series the graduation rate from old universities.

10. Diploma in HE, teaching qualification and nursing/other medical qualification.
11. It explains 28% of the variation in the data.

3.3.1 Degree attainment in the BHPS and the UKHLS

Before moving onto the discussion of our multiple imputation strategy, we examine degree attainment measures in the UKHLS and the BHPS. In Figure 3.2 we plot the average degree attainment by year of birth in wave 18 of the BHPS and wave 3 of the UKHLS. We can see that for most of the sample the two series are very close to each other with an exception of cohorts born after the mid-1970s. These differences hint that the BHPS might no longer reflect a representative sample of individuals born from 1975 onwards. This could be due to the fact that the BHPS was designed to reflect the population of Great Britain in 1991, whereas the UKHLS reflects the UK population as of 2009[12].



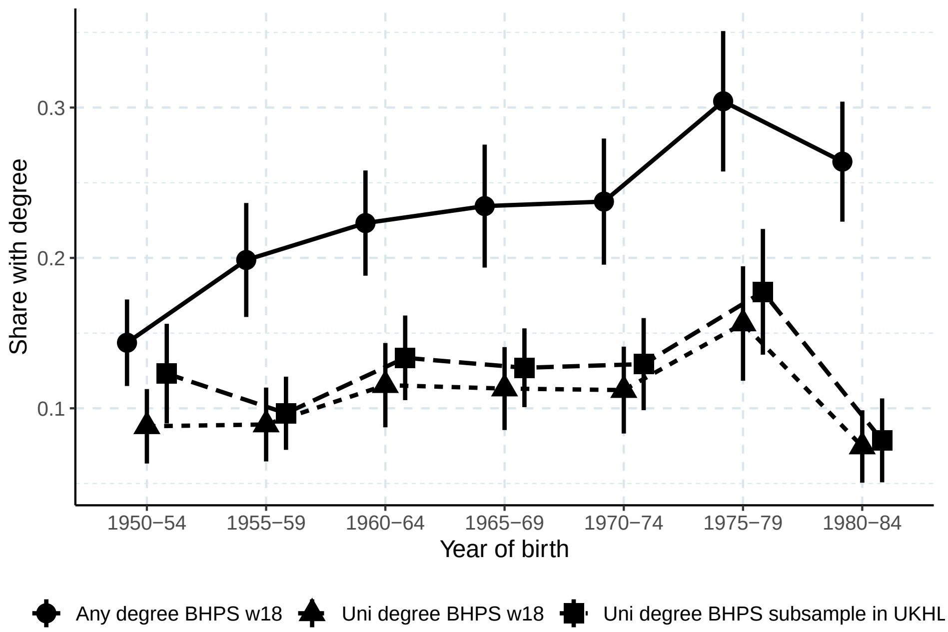
Notes: The figure compares average degree attainment from any type of HEI between the two samples: wave 3 of the UKHLS and wave 18 of the BHPS. The averages were weighted using the respective cross-sectional response weights. The whiskers corresponds to 95% confidence interval due to sampling uncertainty.

Figure 3.2: Degree attainment in UKHLS and BHPS

In Figure 3.3 we compare the degree attainment rates by types of HEI in the two samples. Recall that the UKHLS also follows former BHPS subjects from wave 2

12. The BHPS is a representative sample of the adult population of Great Britain in 1991, Scotland and Wales in 1999 and Northern Ireland in 2001. Ideally, adulthood outcomes of children in the BHPS would be similar to outcomes of adults in the UKHLS born in the respective years. In practice, life events such as migration, institutionalization or death could make the two samples different. For example, it could be that pursuing university education and the subsequent career paths are more likely to involve migration, implying higher chances of these children dropping out of the sample coverage. Therefore, among the BHPS children born after 1975 we see fewer degree-holders than in the corresponding adult population in the UKHLS.

onwards. Thus, we can recover institution types of former BHPS subjects in the UKHLS using their previous responses. Therefore, the figure essentially compares university degree attainment in the full BHPS sample and in the BHPS subsample of the UKHLS. We can see that the two are very close to each other in magnitude and dynamics. This is reassuring as it suggests that the decision to continue from the BHPS to the UKHLS is unlikely to be related to the higher education institution degree type.

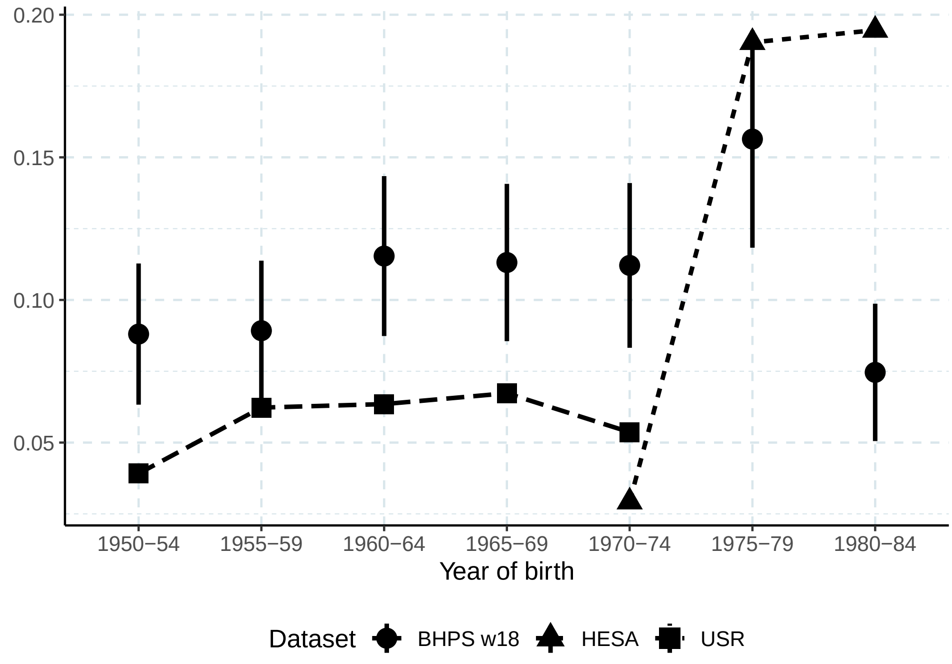


Notes: The figure compares shares of people with degree by type of HEI among BHPS respondents. The degree attainment is calculated using the full sample from wave 18 of the BHPS. The university degree attainment is computed both in the full sample from wave 18 of the BHPS and in the BHPS subsample from wave 3 of the UKHLS. The averages are weighted using cross-sectional response weights from respective waves. The whiskers correspond to 95% confidence intervals due to sampling uncertainty.

Figure 3.3: Degree attainment by HEI type

We also note that among cohorts born in 1960s and 1970s about half of degrees were obtained from universities. But almost all degrees of younger cohorts were obtained from universities, at least in wave 18 of the BHPS. Some of this difference could be related to lower representativeness of BHPS for cohorts born after 1975. But it could also be linked to the questionnaire design. The available options for the HEI type only differentiate universities from polytechnics, but not necessarily from former polytechnics. Therefore, our measure of university degree attainment based on the BHPS data may also contain the graduates who earned their degrees from former polytechnics. By the time people born from the mid-1970s onwards turned 20, 30 polytechnic institutions had already obtained university status.

To have a better understanding of the quality of the university degree attainment measure in the BHPS we further compare it with the benchmark graduation rates from old universities in Figure 3.4. First, we note that the USR ends in 1993 and the HESA starts in 1994. Thus, both data sets do not include the full cohorts of graduates born in 1970-74, which explains the sharp discontinuity in the graduation rates of this birth cohort. Second, we can see that university degree attainment measure in the BHPS is considerably higher than the benchmark graduation rates from old universities, for most of the birth cohorts. One possible explanation is that the benchmark graduation rate from old universities is computed with a larger denominator. We use total population counts of 21-year-olds in the graduation year, which also includes people not born in the UK. Another explanation could be related to the questionnaire design issue, mentioned earlier. Since the survey does not necessarily differentiate universities from *former polytechnics*, respondents may have categorised their institutions as universities based on the status at the time of interview, not at the time of graduation.



Notes: The figure compares university degree attainment in wave 18 of the BHPS with the benchmark graduation rates from old universities computed using data from the USR and HESA. The graduation rate is computed as the ratio of number of British first-degree graduates from old universities to the population of 21-year-olds at the time of graduation. The averages in the BHPS wave 18 are weighted using cross-sectional response weights. The whiskers corresponds to 95% confidence interval due to sampling uncertainty.

Figure 3.4: University degree attainment in the BHPS against the benchmark

3.4 Multiple imputation

We use multiple imputation as a way to deal with imputation uncertainty. Almost any imputation carries a level of uncertainty because the true value is unknown. Imputing a single value for each missing data point and giving it the same importance during the estimation as one gives an observed value fails to account for this uncertainty. In his seminal work, Rubin (1977) presented a multiple imputation technique where imputing multiple values for each missing data allows to incorporate the imputation uncertainty into the main analysis of interest.

The general concept of working with multiply imputed data sets is simple as we explain using Figure 3.5, adapted from van Buuren (2018). First, each missing value in the original data set is assigned multiple imputed values. This could be thought of as creating multiple data sets that only differ in the values assigned to originally missing cells. All these data sets are identical to the original data set in terms of originally observed values. These data sets are referred to as completed data sets. In each of the completed data sets we run the analysis of interest using standard techniques and obtain corresponding estimates and standard errors. That is, in each completed data set we can "forget" that some data points were imputed and use standard regression analysis. Finally, the estimators from each completed data set are combined together into a single estimator using "Rubin's rules". This way the combined estimator accounts for both sampling and imputation uncertainty.

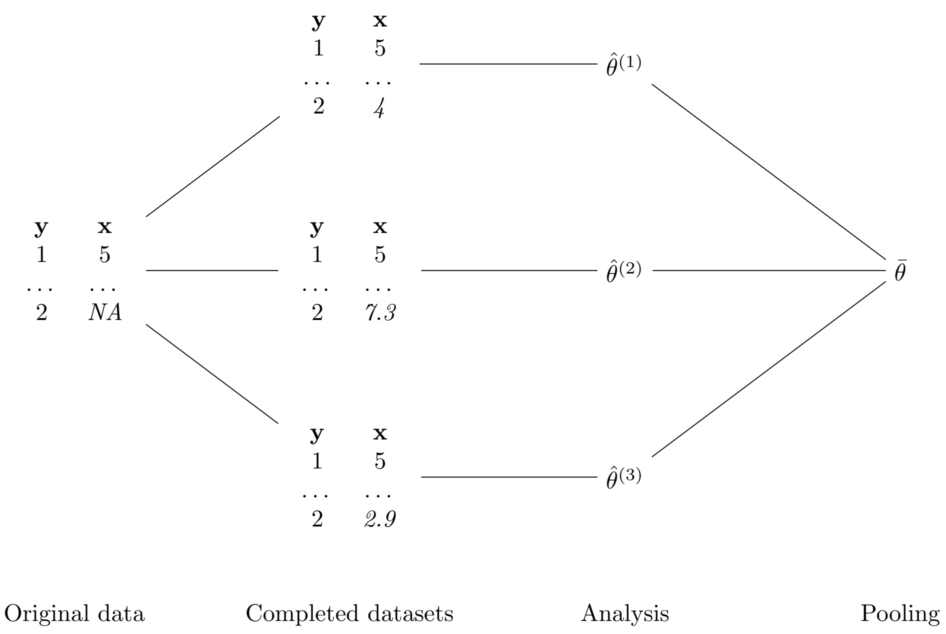


Figure 3.5: General concept of working with multiply imputed data

For a formal definition of the combined estimator, denote the estimator from a completed data set $m \in \{1, \ldots, M\}$ by $\hat{\theta}^{(m)}$ and its variance-covariance matrix by $U^{(m)}$. Then,

$$\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}^{(m)} \tag{3.1}$$

$$T = \bar{U} + \left(1 + \frac{1}{M}\right)B \tag{3.2}$$

$$\bar{U} = \frac{1}{M}\sum_{m=1}^{M}U^{(m)} \tag{3.3}$$

$$B = \frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\theta}^{(m)} - \bar{\theta}\right)^2 \tag{3.4}$$

The combined estimator $\bar{\theta}$ is a simple average of the estimators from M completed data sets. The combined or total variance T of the estimator $\bar{\theta}$ consists of two parts: the within- $\bar{U}$ and between-imputation variance B in equations (3.3) and (3.4), respectively.

Before proceeding to a formal overview of multiple imputation and the necessary assumptions, let us introduce some notation. Suppose we have a sample of size n where a variable Y is missing for some observations. We can construct a corresponding indicator variable $R = \mathbb{1}\{Y \text{ is observed}\}$. It is also convenient to denote the vector with only observed values as Y_{obs} and similarly the vector with missing values Y_{mis}. Note Y_{mis} is a latent variable that contains true values of Y for observations with missing data, but we as researchers do not observe it. Assume the other variables in the sample do not have missing values and denote them by $\mathbf{X}$.

3.4.1 Missing data mechanism

The aim of imputation is to fill in the Y_{mis} values. In most cases, we do not know the true values of Y_{mis} with certainty, but we can characterise their distribution conditional on observed information. In other words, the imputation process could be thought of as drawing observations from the conditional distribution characterised by $\Pr(Y_{mis}|Y_{obs}, X, R)$. In case of multiple imputation, we draw observations from this distribution multiple times. The conditional probability $\Pr(Y_{mis}|Y_{obs}, X, R)$ is the imputation model. How we define the imputation model depends on the processes that generate the missing data.

The missing data mechanisms can be categorized in three ways: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR, as suggested by its name, assumes that the response indicator R is determined completely randomly and depends neither on observed nor missing values. Formally, this can be written as $\Pr(R = 1|Y, \mathbf{X}) = \Pr(R = 1)$. In the context of our paper Y is an indicator for having a university degree. If survey designers only asked a subsample about their HEI type and that subsample was determined randomly, then the missing data would be MCAR.

MAR relaxes the MCAR assumption by allowing the missingness to depend on observed values. That is, $\Pr(R = 1|Y, \mathbf{X}) = \Pr(R = 1|Y_{obs}, \mathbf{X})$. Expanding on the previous example, suppose older people could not correctly remember their institution type and therefore did not provide an answer. In such a case, the missing data would be MAR conditional on age.

Finally, under MNAR whether or not a value is missing may not only depend on observed variables, but also on the value itself. So, $\Pr(R = 1|Y_{mis}, Y_{obs}, X)$ does not simplify. In our example, if specifically people who obtained a degree from former polytechnics did not disclose this information, the missing data would be MNAR.

The missing data mechanism is closely related to the concept of ignorable nonresponse. Nonresponse can be called ignorable if the MAR assumption is satisfied and the parameters of the response model are distinct from parameters of the data-generating model in a sense that knowing one does not provide information about the other (Rubin 1987; Schafer 1997). Ignorable nonresponse allows us to build the imputation model using only observed data without explicitly modelling the missing data mechanism. That is, we can draw imputations from $\Pr(Y|Y_{obs}, X, R = 1) = \Pr(Y_{mis}|Y_{obs}, X, R = 0)$. If nonresponse is non-ignorable, we would have to model the missing data mechanism explicitly because $\Pr(Y|Y_{obs}, X, R = 1) \neq \Pr(Y_{mis}|Y_{obs}, X, R = 0)$.

3.4.2 Imputation model

Notice that the example we give to describe MCAR is very similar to the case of the UKHLS and the BHPS. People in the BHPS were asked about their institution type, people in the UKHLS were not. If the BHPS sample is not systematically different from the UKHLS one, i.e. they are drawn randomly from the same population, then indeed we have a case of MCAR. We examine this claim in Table 3.1 by comparing socio-economic and family characteristics of two samples: wave 18 of the BHPS and wave 3 of the

UKHLS. In wave 18 of the BHPS we consider the respondents with valid information on university degree attainment, i.e., those who have non-missing highest qualification and HEI type information. Panel A of Table 3.1 compares the BHPS sample with the entire wave 3 of the UKHLS including former BHPS respondents; panel B - with wave 3 of the UKHLS excluding former BHPS respondents. From Table 3.1 we conclude that the samples are very different from one another. This suggests that the UK population has changed significantly in the 20 years between the BHPS and the UKHLS. Individuals in the UKHLS are more likely to have continued past compulsory schooling and have some tertiary degree, have slightly lower earnings (although this comparison may be confounded by the financial crisis), and have higher educated parents.

Thus, we need to construct an imputation model that is likely to satisfy the MAR assumption. In order to do so, we need to include covariates that help explain both the university degree attainment and the missing data mechanism. We already know from Table 3.1 that degree attainment, earnings, and parental educational qualifications are highly correlated with the missing data mechanism. Furthermore, the imputation of university degree attainment only makes sense if an individual has a degree. Therefore, we can focus on building an imputation model among degree-holders, setting the university degree attainment variable to zero for everyone else.

Another important consideration in developing our imputation model is its congruence with the analysis model (Schafer 1997). Failure to include the terms of interest means that the imputation model restricts their coefficients to zero. This, in turn, results in attenuated coefficients of interest in the analysis stage. As mentioned earlier, we adopt the analysis context from our companion paper Ichino et al. (n.d.). In particular, we would like to know how the expansion of higher education in the UK changed the composition of the university students in terms of their intelligence scores. In other words, how did the relationship between intelligence scores and university degree attainment change over time. Thus, we are interested in estimating the following equation:

$$\frac{\Pr(U_i = 1)}{1 - \Pr(U_i = 1)} = \exp\left(\alpha + \gamma_y + \boldsymbol{\delta}\mathbf{X}_i + \boldsymbol{\beta}_y\mathbf{X}_i\right) \tag{3.5}$$

where U_i is the university degree attainment variable, $\mathbf{X}_i$ contains gender, intelligence score and their interaction and γ_y are birth cohort fixed effects. Here, $\boldsymbol{\delta}$ describes the relationship between intelligence score and university degree attainment by gender in

Table 3.1: Testing MCAR assumption between BHPS wave 18 and UKHLS wave 3

	UKHLS w3			BHPS w18			Diff			
	Mean	S.D.	N	Mean	S.D.	N	Coef	S.E.	FWER p-val	% mean BHPS
Panel A: full UKHLS sample vs BHPS										
Individual characteristics										
Female	0.525	0.499	20,771	0.523	0.500	6,563	0.002	0.005	1.0	0.4
White british	0.891	0.311	20,771	0.900	0.300	6,563	-0.009	0.007	1.0	-1.0
Born in England	0.747	0.434	20,402	0.776	0.417	6,094	-0.029**	0.008	0.0	-3.7
Age in 2008	41.545	9.817	20,771	42.680	9.330	6,563	-1.134***	0.180	0.0	-2.7
Post-compulsory edu	0.587	0.492	20,771	0.491	0.500	6,563	0.096***	0.010	0.0	19.5
Any degree	0.298	0.458	20,771	0.193	0.395	6,563	0.105***	0.009	0.0	54.4
Ever married	0.793	0.405	20,771	0.809	0.393	6,563	-0.016	0.007	0.3	-2.0
Any children	0.785	0.411	20,771	0.789	0.408	6,563	-0.003	0.007	1.0	-0.4
Working	0.751	0.433	20,771	0.800	0.400	6,563	-0.050***	0.007	0.0	-6.2
Real monthly earnings	17.947	17.732	20,771	20.036	19.481	6,563	-2.089***	0.366	0.0	-10.4
Family characteristics at age 14										
Father has degree	0.117	0.321	16,993	0.093	0.290	5,380	0.024***	0.006	0.0	26.1
Mother has degree	0.075	0.264	17,555	0.055	0.229	5,535	0.020***	0.005	0.0	35.9
Father employed	0.886	0.317	20,403	0.925	0.263	6,158	-0.039***	0.005	0.0	-4.2
Mother employed	0.657	0.475	20,509	0.632	0.482	6,276	0.025*	0.009	0.1	4.0
Father born in Eng	0.689	0.463	20,425	0.717	0.451	6,523	-0.028**	0.009	0.0	-3.9
Mother born in Eng	0.696	0.460	20,491	0.721	0.449	6,545	-0.025*	0.010	0.1	-3.5
Panel B: UKHLS excl. former BHPS respondents vs BHPS										
Individual characteristics										
Female	0.526	0.499	15,654	0.523	0.500	6,563	0.003	0.006	1.0	0.5
White british	0.888	0.315	15,654	0.900	0.300	6,563	-0.012	0.008	1.0	-1.3
Born in England	0.775	0.417	15,651	0.776	0.417	6,094	-0.001	0.010	1.0	-0.1
Age in 2008	41.339	9.840	15,654	42.680	9.330	6,563	-1.341***	0.205	0.0	-3.1
Post-compulsory edu	0.595	0.491	15,654	0.491	0.500	6,563	0.105***	0.011	0.0	21.3
Any degree	0.311	0.463	15,654	0.193	0.395	6,563	0.118***	0.010	0.0	61.2
Ever married	0.790	0.407	15,654	0.809	0.393	6,563	-0.019	0.008	0.2	-2.3
Any children	0.783	0.412	15,654	0.789	0.408	6,563	-0.006	0.008	1.0	-0.8
Working	0.744	0.436	15,654	0.800	0.400	6,563	-0.056***	0.008	0.0	-7.0
Real monthly earnings	17.840	17.921	15,654	20.036	19.481	6,563	-2.196***	0.405	0.0	-11.0
Family characteristics at age 14										
Father has degree	0.121	0.327	12,857	0.093	0.290	5,380	0.029***	0.007	0.0	31.0
Mother has degree	0.078	0.268	13,311	0.055	0.229	5,535	0.023***	0.006	0.0	40.9
Father employed	0.881	0.324	15,597	0.925	0.263	6,158	-0.045***	0.005	0.0	-4.8
Mother employed	0.663	0.473	15,609	0.632	0.482	6,276	0.031**	0.010	0.0	4.9
Father born in Eng	0.712	0.453	15,587	0.717	0.451	6,523	-0.005	0.011	1.0	-0.7
Mother born in Eng	0.720	0.449	15,638	0.721	0.449	6,545	-0.001	0.011	1.0	-0.1

$^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$

Notes: The table compares average characteristics of individuals in wave 18 of the BHPS with those in wave 3 of the UKHLS. The BHPS sample is restricted to individuals with valid university degree attainment variable, i.e., those who have non-missing highest qualification and HEI type information. Panel A uses the entire wave 3 of the UKHLS for comparison and Panel B is restricted to wave 3 of the UKHLS excluding former BHPS respondents. The last four columns report the statistics of the difference in means between the BHPS and the UKHLS samples. The standard errors of the difference are clustered at the sampling strata-wave level. FWER p-values are computed as in Holm (1979) to adjust for multiple inferences and are used to assign significance stars to the coefficient of difference. The last column reports the size of the coefficient of difference relative to the mean in the BHPS sample in %. The estimates are weighted using the cross-sectional response weights in the respective waves.

the base birth cohort group, and $\boldsymbol{\beta}_y$ shows how this relationship has changed across birth cohorts relative to the base group. We acknowledge that the simple specification in equation (3.5) may suffer from an omitted variable bias. For example, it does not control for the educational qualifications of parents, a variable that could explain both the intelligence score of respondents and their university degree attainment status. For the purposes of this paper, we abstract from this issue and treat the intelligence score as if it were randomly distributed in the population. That is, we assume that parameters $\boldsymbol{\delta}$ and $\boldsymbol{\beta}_y$ are true causal parameters describing the effect of intelligence and gender on university degree attainment probabilities. Nonetheless, the implication remains: the analysis and imputation models should be consistent with each other. If parental education enters the analysis model, it should also be part of the imputation model.

Thus, our imputation model should ideally contain all the regressors from equation (3.5), including their interaction terms. Our imputation model can be written as follows

$$\Pr(U_i = 1) = \begin{cases} f\left(\zeta + \eta_y + \boldsymbol{\rho}\ddot{\mathbf{Z}}_i + \boldsymbol{\lambda}_y\tilde{\mathbf{Z}}_i\right) & \text{if } d_i = 1 \\ 0 & \text{if } d_i = 0 \end{cases} \tag{3.6}$$

where d_i is the degree attainment variable. That is, the probability of having a university degree is a function of a constant ζ, birth cohort fixed effects η_y and personal characteristics in $\ddot{\mathbf{Z}}_i$ and $\tilde{\mathbf{Z}}_i$, the latter of which is allowed to have an effect specific to each birth cohort. Given that we specify a non-trivial imputation model only among the degree-holders, the estimation sample consists of 1,540 observations. Therefore, we differentiate between the characteristics that enter the model linearly $\ddot{\mathbf{Z}}_i$ and those that are interacted with birth cohort indicators $\tilde{\mathbf{Z}}_i$. We note that $\mathbf{X}_i \in \tilde{\mathbf{Z}}_i \in \ddot{\mathbf{Z}}_i$. So, in addition to gender, intelligence score and their interaction term, the set of regressors in $\tilde{\mathbf{Z}}_i$ includes country of birth, race, an indicator for whether an individual has ever cohabited, an indicator for whether an individual has ever been married, an indicator for whether an individual has any children, the number of children, the second-order polynomial of real earnings, and the employment status at the time of interview. Besides $\tilde{\mathbf{Z}}_i$, the set of linear regressors $\ddot{\mathbf{Z}}_i$ includes

- *Individual characteristics*: years of education, age when left further education, residence in England at the time of interview, indicator whether current residence is in the country of birth, car ownership, indicator for having a second job, major occupational group of main job, major occupational group of first job

- *Parental characteristics*: countries of birth of father and mother, highest educational qualifications of father and mother, employment statuses of father and mother when the respondent was 14 years old

- *Design variables*: survey design weight, survey response weight.

Unfortunately, cognitive ability tests were only administered in wave 3 of the UKHLS. On the one hand, failing to add intelligence score to the imputation model will attenuate the correlation between university degree attainment and intelligence (Schafer 1997) unless it is not captured by the rest of the predictors. On the other hand, using former BHPS respondents in wave 3 of the UKHLS for estimating the imputation model presents an additional concern. Not only were they sampled from a different UK population, but they have also self-selected to continue into the UKHLS. This could violate the MAR assumption, if the decision to continue from the BHPS is correlated with the HEI type even after controlling for a set of observed characteristics. Due to the potential benefits and drawbacks of both including and excluding the intelligence score terms, we consider different versions of the imputation model depending on inclusion of the intelligence score as one of the predictors, estimation sample and estimator used. These are presented in Table 3.2.

Since we are studying a binary variable describing whether individuals have earned a degree from a traditional university, we fit the corresponding probability using a logit estimator in Models 1-3. In addition, we consider using a machine-learning algorithm in Model 4. Machine learning algorithms are powerful tools for prediction questions. At the same time, the imputation stage in Figure 3.5 can be thought of as a prediction problem. In particular, we use a random forest algorithm to fit the imputation model. A random forest is a tree-based algorithm that can be be applied to both continuous and categorical variables (Schonlau and Zou 2020). A tree is constructed through splitting the sample into various groups based on values of predictors and thresholds. The set of predictors contributing to the tree and their thresholds are determined by splitting criteria[13]. To improve prediction accuracy, the random forest algorithm averages predictions across multiple trees built on bootstrapped samples.

Next, we test if our model specifications violate[14] the MAR assumption in Table 3.3. Determining if missing data is MAR or MNAR is essentially impossible. However,

13. The splitting criterion used for the classification trees is entropy. For more information, see p. 5 in Schonlau and Zou (2020).

14. For this exercise we focus on Models 1-3. We omit Model 4 from this exercise because there is no straightforward way to analyse the importance of a single predictor. Schonlau and Zou (2020) compute a so-called variable importance statistic that captures the average contribution of a given predictor to

Table 3.2: **Imputation model versions**

	Includes IQ	Estimation sample	Estimator
Model 1	No	BHPS wave 18 and UKHLS wave 3	logit
Model 2	No	UKLHS wave 3	logit
Model 3	Yes	UKHLS wave 3	logit
Model 4	No	UKHLS wave 3	random forest

we can test if the missingness indicator is correlated with other observed variables even after conditioning on all the variables included in the model. If so, then the imputation model clearly violates the MAR assumption. According to the results in Table 3.3, the strongest signals of failure of the MAR assumption are observed in Model 1. Interest in politics and neighbourhood characteristics are both strongly correlated with the missing data indicator. Models 2 and 3 display fewer, if any, statistically significant violations. But even models 2 and 3, for some variables the estimated differences between missing and non-missing subsamples are large in magnitude: they constitute more than 10% of the sample mean in the UKHLS. We add these variables to $\ddot{\mathbf{Z}}_i$.

3.4.3 Evaluation

Now that we have defined different specifications of the imputation model, we need some way to choose among them. An obvious selection criteria would seem to be based on prediction accuracy: choose that model which produces imputed values, which are closest to the true value. This could be done using a subset of BHPS subjects for which we know the true value of the university degree indicator. However, as explained by van Buuren (2018) such model selection criteria would choose a model that underestimates imputation uncertainty. Therefore, in the full sample, the parameter estimates following from such an imputation model would generally lead to invalid inferences. Therefore, model selection criteria should be based on the properties of the combined estimator $\bar{\theta}$.

Denote by F_i an indicator which takes the value of 1 if individual i is female and denote by I_i her intelligence score. Then, $\mathbf{X}_i = (F_i, I_i, F_i I_i)$ and the analysis model in equation (3.5) can be rewritten as

the final objective function over the entire forest. The statistic is scaled such that the most important variable is assigned a value of 100%. However, the statistic does not offer a measure of significance of the contribution.

Table 3.3: Testing MAR assumption

	Model 1				Model 2				Model 3			
Dependent variable	$1-R$	N obs	N miss	% mean	$1-R$	N obs	N miss	% mean	$1-R$	N obs	N miss	% mean
Individual characteristics												
Has mobile	0.001 (0.009) [1.000]	1,304	5,049	0.1	0.000 (0.009) [1.000]	1,015	5,049	0.0	0.000 (0.010) [1.000]	967	4,877	0.0
Supports a polit party	-0.006 (0.023) [1.000]	1,304	5,049	-1.6	-0.005 (0.020) [1.000]	1,015	5,049	-1.3	-0.024 (0.021) [1.000]	967	4,877	-6.7
Responsible for child under 16	-0.019 (0.011) [0.595]	1,304	5,049	-8.0	-0.013 (0.011) [1.000]	1,015	5,049	-5.4	-0.009 (0.011) [1.000]	967	4,877	-3.7
Likely to move	-0.045 (0.023) [0.352]	1,304	5,049	-7.3	-0.043 (0.021) [0.431]	1,015	5,049	-6.9	-0.030 (0.021) [1.000]	967	4,877	-4.9
Interested in politics	0.116*** (0.021) [0.000]	1,304	5,049	17.7	0.052 (0.020) [0.126]	1,015	5,049	8.0	0.031 (0.021) [1.000]	967	4,877	4.7
Received interest on savings	0.033 (0.024) [0.819]	1,259	4,848	7.4	0.007 (0.025) [1.000]	967	4,848	1.6	0.004 (0.024) [1.000]	928	4,685	0.9
Interest on savings missing	-0.003 (0.009) [1.000]	1,304	5,049	-7.4	0.006 (0.008) [1.000]	1,015	5,049	14.6	0.015 (0.008) [0.728]	967	4,877	37.5
Good fin situation (subjective)	-0.009 (0.022) [1.000]	1,304	5,049	-1.3	-0.001 (0.021) [1.000]	1,015	5,049	-0.2	-0.011 (0.021) [1.000]	967	4,877	-1.5
Current residence: access to services												
Good shopping services	0.058 (0.026) [0.201]	1,304	5,049	9.4	-0.067* (0.025) [0.093]	1,015	5,049	-10.9	-0.049 (0.026) [0.728]	967	4,877	-7.9
Good public transp services	0.143*** (0.025) [0.000]	1,304	5,049	27.5	-0.015 (0.024) [1.000]	1,015	5,049	-2.9	-0.027 (0.025) [1.000]	967	4,877	-5.2
Good medical services	0.094*** (0.025) [0.001]	1,304	5,049	12.7	0.037 (0.022) [0.827]	1,015	5,049	5.0	0.043 (0.023) [0.728]	967	4,877	5.8
Good leisure services	0.086*** (0.023) [0.001]	1,304	5,049	17.4	0.004 (0.024) [1.000]	1,015	5,049	0.8	0.003 (0.024) [1.000]	967	4,877	0.6
Likes neighbourhood	0.021 (0.010) [0.334]	1,304	5,049	2.2	0.003 (0.008) [1.000]	1,015	5,049	0.3	0.006 (0.008) [1.000]	967	4,877	0.6

$^{*}p < 0.1;^{**}p < 0.05;^{***}p < 0.01$

Notes: The Table reports the results from linear regressions of the dependent variables in the first column on the missingness indicator conditional on all the terms in the corresponding imputation model. In particular, models 1 and 2 do not control for intelligence score terms, while model 3 does. The estimation sample of model 1 includes wave 18 of the BHPS and wave 3 of the UKHLS; that of models 2 and 3 is restricted to wave 3 of the UKHLS only. The estimations are weighted using the respective cross-sectional weights and clustered at the level of sampling strata and wave. Regular standard errors are reported in parentheses and FWER adjusted p-values - in square brackets. The significance stars are assigned based on the FWER adjusted p-values. The Table also reports the size of the estimated coefficients relative to the sample mean in wave 3 of the UKHLS in %.

$$\frac{\Pr(U_i = 1)}{1 - \Pr(U_i = 1)} = \exp\left(\alpha + \gamma_y + (\delta^F + \beta_y^F)F_i + (\delta^I + \beta_y^I)I_i + (\delta^{FI} + \beta_y^{FI})F_i I_i\right)$$

Here, γ_y are cohort fixed effects, $\boldsymbol{\delta} = (\delta^F, \delta^I, \delta^{FI})$ describe the effects of gender, intelligence score and their interaction term on the log odds ratio of the probability of having a university degree in the base birth cohort group. Given our sample restrictions and our definition of birth cohort groups, the base cohort are people born in 1950-54. Then, the parameters $\boldsymbol{\beta}_y = (\beta_y^F, \beta_y^I, \beta_y^{FI})$ capture how the effects of gender, intelligence score and their interaction term change across birth cohorts relative to the base group. We collectively denote these parameters by $\theta = (\gamma_y, \boldsymbol{\delta}, \boldsymbol{\beta}_y)$. The goal in this subsection is to study the properties of the combined estimator $\bar{\theta}$ from the multiply imputed data.

In the absence of missing data and under random sampling we could obtain the consistent estimator $\hat{\theta}$ and rely on its asymptotic distribution to draw inference. The question here is whether we can draw valid inference using the combined estimator $\bar{\theta}$. Rubin (1987, chapter 4) studies the properties of the combined estimator from the random-response randomization-based perspective. In short, he outlines two sufficient conditions for the randomization-validity of the combined estimator $\bar{\theta}$. First, the complete-case estimator $\hat{\theta}$ should be randomization-valid for θ. That is, in the absence of missing data, our estimator $\hat{\theta}$ should be consistent for the parameter of interest θ. Moreover, the 95% confidence interval around $\hat{\theta}$ should contain the true parameter θ in 95 out of 100 samples from the population. Second, the imputation model should be *proper* meaning that $\bar{\theta}$ should be randomization-valid for $\hat{\theta}$.

It is easy to see that the first assumption is unrelated to the problem of missing data. Even if there were no missing data, we would need to satisfy this assumption to be able to draw inference about θ based on the estimator $\hat{\theta}$. When it comes to the second assumption, analytical studies of the properties of the combined estimator $\bar{\theta}$ are usually very difficult (Schafer 1997). Therefore, researchers have developed simulation-based methods. We adopt the algorithm proposed by Brand et al. (2003) to evaluate both bias and efficiency of the combined estimator $\bar{\theta}$ in each imputation model and then compare them across models. The idea is the following:

1. Estimate equation (3.5) in the subsample with both nonmissing university degree attainment information and nonmissing intelligence score. We obtain estimates $\tilde{\theta}$, which we now treat as true parameters. Use the estimated $\tilde{\theta}$ to fill in the missing

values of the university degree attainment variable U. Denote the simulated variable by $\widetilde{U}$, which has no missing values.

2. Specify a missing data mechanism. We assume MAR: $\Pr(R_i = 1|\ddot{\mathbf{Z}}_i, \tilde{\mathbf{Z}}_i)$.

3. Generate L incomplete data sets under the chosen missing data mechanism. This is done by drawing response indicators $R^{(l)}, \forall l \in \{1, \ldots, L\}$ according to the probabilities specified in the previous step and setting $\widetilde{U}_i^{(l)}$ to missing whenever $R_i^{(l)} = 0$. Brand et al. (2003) recommend L in the range between 200 and 1000. We set $L = 500$.

4. Within each of the L simulated incomplete data sets, we multiply impute the university degree attainment variable and estimate $\bar{\theta}^{(l,v)}, \bar{U}^{(l,v)}, B^{(l,v)}, \quad \forall l \in \{1, \ldots, L\}$ and $\forall v \in \{\text{Model 1}, \text{Model 2}, \text{Model 3}, \text{Model 4}\}$.

5. Given the combined estimators, compute the bias and efficiency statistics:

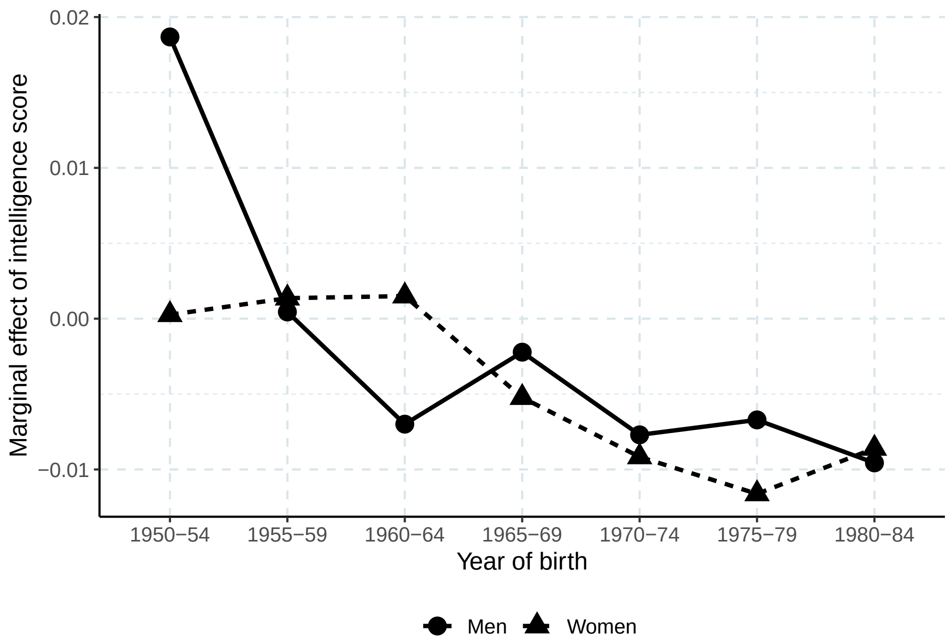
    a. Raw bias: $\frac{1}{L}\sum_{l=1}^{L} \bar{\theta}^{(l,v)} - \tilde{\theta}$

    b. Coverage rate: $\frac{1}{L}\sum_{l=1}^{L} \mathbb{1}\{\tilde{\theta} \in 95\% \text{ CI in data set } l\}$

    c. Average width: $\frac{1}{L}\sum_{l=1}^{L}$ width of 95% CI in data set l

Let us first examine the "true" parameters from step 1. Figure 3.6 plots the marginal effect of a one standard deviation increase in the intelligence score on the probability of having a university degree by gender across birth cohorts. The marginal effects are computed using the estimates of equation (3.5) in the subsample with no missing data. As mentioned earlier, we treat this equation as the true data-generating process for the purposes of this paper. The estimates are likely to change when the analysis model includes possible confounders as controls, such as parental education. Nevertheless, we observe a declining trend. A one standard deviation increase in the intelligence score of individuals born in 1950-54 had a positive effect on university degree attainment probabilities for men. But by the time people born in 1975-79 were of getting their higher education degrees, an increase of one standard deviation in the intelligence score meant a 0.7-1.2pp decrease in the probability of university degree attainment.

We now turn to examining the results of our simulation-based evaluation. We simulated 500 data sets where some observations were randomly set to missing according to the MAR response model. In each of the 500 simulated data sets, the missing values



Notes: The figure plots the marginal effects of intelligence score on the university degree attainment probabilities by gender and birth cohorts. The marginal effects were computed using the estimates of equation (3.5) in the BHPS subsample of the UKHLS. The estimation sample only includes individuals with non-missing university degree variable and non-missing intelligence score. The marginal effects are computed at mean intelligence score.

Figure 3.6: "True" parameters used to initiate the evaluation algorithm

were multiply imputed using the four versions of our imputation model. Thus, we get four combined estimators from each simulated data set.

Table 3.4 reports the bias and efficiency statistics of the combined estimators for the marginal effect of intelligence score on the university degree attainment across imputation models. It is evident from the table that Model 1 performs worse compared to other models. The estimators from Model 1 have the largest in magnitude bias statistic and lowest coverage rate. We can also see that Models 2 and 3 display very similar results, both in terms of bias and in terms of efficiency. This result is somewhat surprising as it would suggest that the imputation model does not need to control for the intelligence score. One possible explanation for such similarity between the results from Models 2 and 3 could be that conditioning on the intelligence score is to a large extent redundant after controlling for the rest of the variables in the imputation model. Finally, combined estimators resulting from Model 4 are to a large extent comparable to those from Model 2. This could mean that the specification in Model 2 controls for most of the possible non-linearities in the data.

Thus, given the results in Table 3.4, we impute the missing observations in the university degree variable in the UKHLS using imputation models 2 and 3. That is, we

Table 3.4: Evaluation results under MAR mechanism

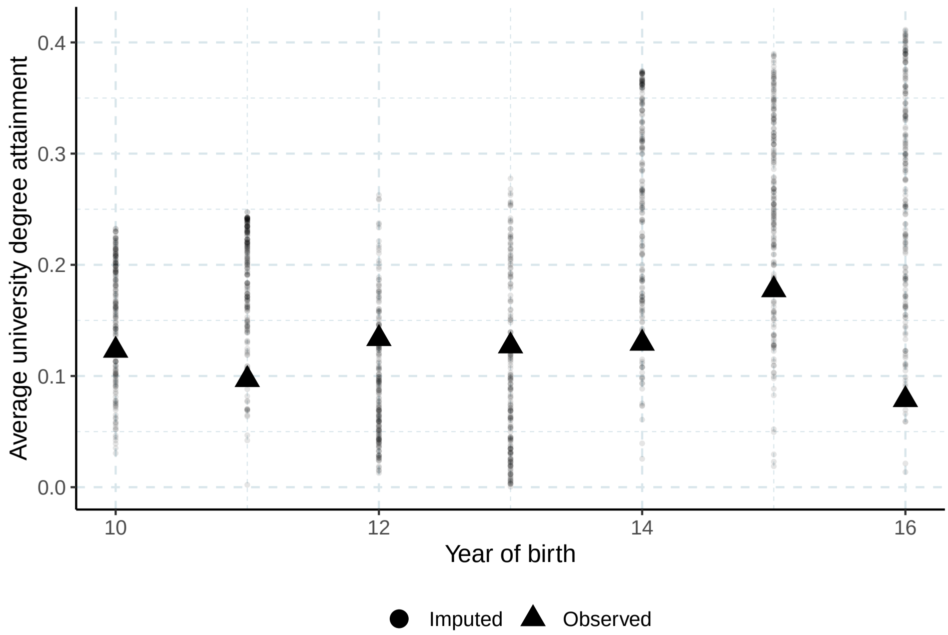
	Model 1		Model 2		Model 3		Model 4	
Birth cohort	Men	Women	Men	Women	Men	Women	Men	Women
Raw bias, pp								
1950-54	7.0	4.2	6.4	3.9	7.6	2.8	6.2	3.6
1955-59	2.7	2.9	2.5	2.7	2.1	1.8	2.4	2.3
1960-64	5.8	6.6	4.8	6.0	5.5	6.3	5.0	5.9
1965-69	6.5	4.9	6.4	4.3	5.4	3.9	5.3	3.8
1970-74	4.4	3.7	4.2	3.4	4.2	1.9	4.4	3.2
1975-79	9.0	5.2	8.5	4.2	10.7	5.5	9.3	4.9
1980-84	4.3	2.6	3.8	2.2	2.2	1.8	4.0	2.4
95% CI coverage rate, %								
1950-54	0.0	14.0	0.0	27.8	0.0	88.4	0.2	47.4
1955-59	81.6	82.6	97.8	97.8	98.6	99.8	98.4	99.0
1960-64	0.0	0.0	0.0	0.0	2.6	0.0	0.0	0.0
1965-69	0.0	0.0	0.0	0.0	9.6	16.8	0.2	2.2
1970-74	23.6	25.8	26.8	52.2	53.4	99.4	23.2	64.0
1975-79	0.0	0.8	0.0	4.2	0.0	0.2	0.0	0.0
1980-84	57.8	99.2	89.8	100.0	99.8	100.0	76.2	100.0
Average width, pp								
1950-54	9.3	7.2	9.2	7.3	9.8	6.9	9.1	7.1
1955-59	6.6	6.6	6.9	6.7	6.9	6.4	6.8	6.4
1960-64	7.7	8.4	7.4	8.0	8.1	8.6	7.5	8.0
1965-69	8.3	7.0	8.3	6.6	8.6	6.8	7.9	6.3
1970-74	8.0	6.8	8.0	6.7	8.4	6.1	8.0	6.5
1975-79	10.8	8.1	10.4	7.4	11.1	8.0	10.5	7.6
1980-84	8.6	7.3	8.4	7.2	7.8	7.5	8.4	7.2

Notes: The table reports the bias and efficiency statistics for the marginal effect of intelligence score on university degree attainment probability computed as described in step 5 of the evaluation algorithm. The table uses the combined estimators from 500 simulated datsets with missing data. The misisng data were generated under the MAR mechanism.

use the UKHLS respondents in wave 3, where only the former BHPS subjects have a non-missing university degree information.

3.5 Results and discussion

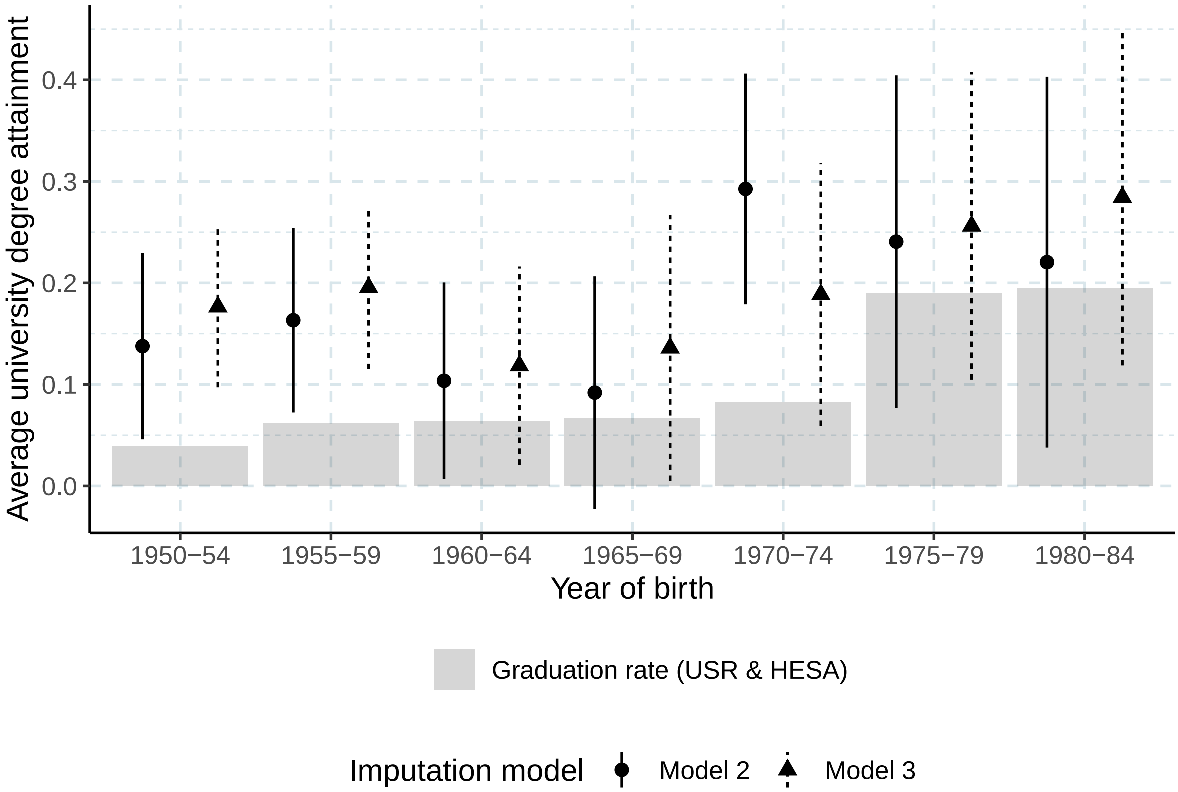
We first examine if the imputed values are plausible. Figure 3.7 plots the average university degree attainment over time among the observed and imputed subsamples. Each dot in the plot corresponds to the average university degree attainment among the imputed observations in each of the completed data sets. The triangles correspond to the average university degree attainment among the BHPS subsample of the UKHLS. Overall, the distribution of the imputed values is within a plausible range of average university degree attainment. Furthermore, the imputed values seem to follow the general trend of rising university education.



Notes: The figure plots the average university degree attainment over time in observed and imputed subsamples. The averages are calculated using wave 3 of the UKHLS and weighted using the corresponding cross-sectional weights. The triangles correspond to the average university degree attainment of the BHPS subsample, i.e. observed. The dots correspond to average university degree attainment among individuals with originally missing data. Their averages are computed within each imputed set out of total M sets.

Figure 3.7: Average university degree attainment over time

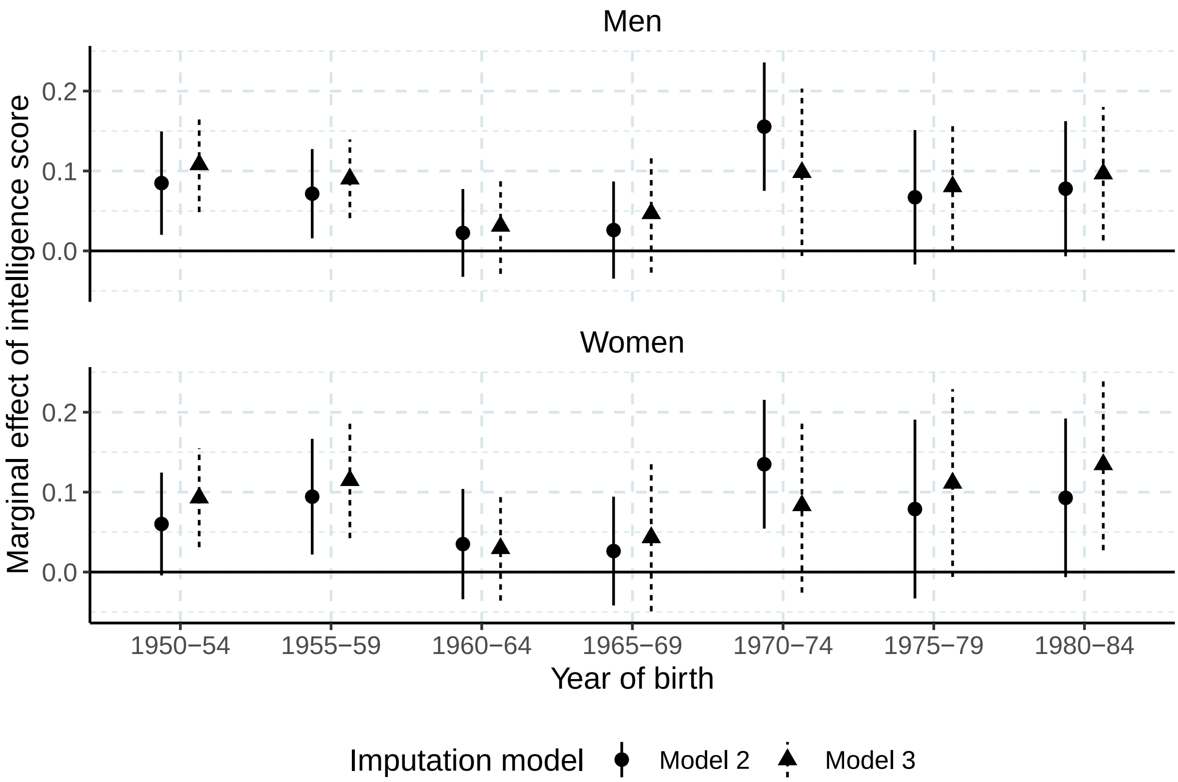
In Figure 3.8 we directly compare average university degree attainment over time with the graduation rates computed using the USR and the HESA data sets. Reassuringly, the estimates of average university degree attainment produced by the two imputation models are, in general, close to the benchmark graduation rates in the USR and the HESA data sets. It is notable that despite the fact that average university degree attainment is considerably underestimated in the observed subsample born in 1980-84 (see Figure 3.4), the multiple imputation yields estimates comparable to the benchmark graduation rates. However, multiple imputation did not perform well in the subsample of older individuals. The combined estimators from both imputation Models 2 and 3 considerably overestimate the university degree attainment rates for this cohort. So, for example, the combined estimators for average university degree attainment among individuals born in 1950-54 are at 14-18%, whereas the graduation rate for that cohort is at 4%, according to the USR data set. For the cohorts born after 1960, imputation Model 3 produces completed data sets that are most consistent with the benchmark graduation rates obtained from the USR and the HESA data sets.



Notes: The figure plots the average university degree attainment by gender over time. The estimations are done in each completed data set separately. The estimation sample is restricted to UKHLS wave 3 subjects. The estimations are weighted using cross-sectional survey weights. The whiskers show the 95% confidence interval based on standard errors clustered at the sampling strata level.

Figure 3.8: Average university degree attainment by gender

Figure 3.9 plots the marginal effect of the intelligence score on the probability of obtaining a university degree estimated from the analysis model in equation (3.5). Similar to the conclusion from the simulation-based evaluation, imputation Models 2 and 3 produce similar estimates of the marginal effect. The results suggest that in the early 1970s, that is when people born in 1950-54 were attending HE institutions, a one standard deviation higher intelligence score raised the probability of having a university degree by 6-11pp. However, within 10 years the marginal effect of the intelligence score went down to 2-4pp. These estimates could suggest that the expansion of higher education in the UK led to a lower threshold in terms of intelligence score for admission and completion of university education. But starting from the 1990s, the intelligence scores again become a significant predictor of university degree attainment. The point estimates of the marginal effect of the intelligence score are at about 7-16pp for cohorts born from 1970 onwards. We also note that the confidence intervals of the estimates for cohorts born after 1970 are wider, compared to older cohorts. This could be due to the fact that the educational attainment of these cohorts were underestimated in the BHPS, leading to a smaller sample size available during the estimation of the imputation model.



Notes: The figure plots the estimates of the marginal effect of the intelligence score evaluated at the mean in logit model in equation (3.5). The estimations for all degrees are done in the original observed data set. The estimations for degrees earned from traditional and new universities are performed in the completed data set related to the imputation model 2. The estimation sample is restricted to respondents in wave 3 of the UKHLS. The estimations are weighted using cross-sectional survey weights. The whiskers show the 95% confidence interval based on standard errors clustered at the sampling strata level.

Figure 3.9: Marginal effect of intelligence score at the mean

3.6 Conclusion

The second half of the 20th century has seen a massive expansion of higher education throughout the world, also in the UK. The share of individuals with a higher education degree has been steadily rising. But not all degrees are made equal. From 1965 to 1992, students in the UK could earn their degrees either from traditional universities or from public sector colleges, led by polytechnics. Despite formal equality of the degrees earned from either type of institutions, these institutions faced different target populations, admission procedures, subjects taught, organization and financing schemes. These differences, together with the elite image of the traditional universities, contributed to a public perception of polytechnics degrees as inferior to that of universities (Willetts 2017; Pratt 1997).

This perceived inferiority hints at something that has been established in the literature: the type of higher education institution can act as a signal of education quality. Therefore, differentiating between types of HE institutions is an important consideration in the analysis of the higher education sector in the UK.

However, common survey data sets often offer limited information about the types of institutions from which individuals earned their degrees.

In this paper, we try to overcome the issue of missing HE institution types by using a multiple imputation technique. We use the two British panel surveys, BHPS (1991-2008) with about 10,000 individuals in each wave and UKHLS (2009-present) with about 40,000 individuals in each wave. The BHPS specifically asked its participants about the type of higher education institution they last attended, distinguishing between universities and polytechnics. Moreover, 80% of the respondents from the last wave of the BHPS continue as part of the UKHLS. We use the close relationship between the BHPS and the UKHLS and transform the lack of institution types in the UKHLS into a missing data problem. To properly reflect the uncertainty about imputed values we use a multiple imputation technique (Rubin 1977).

We build our imputation models taking into account assumptions about the missing data mechanisms and the agreement between the imputation and analysis models. In this paper, we adopt the analysis model from our companion paper Ichino et al. (n.d.), which studies how the expansion of higher education sector in the UK changed the composition of students in terms of their intelligence scores. Thus, the agreement between the imputation and the analysis models requires us to include the intelligence score into the imputation model. However, the intelligence score is only available for a subset of BHPS respondents that continued to the UKHLS. Therefore, we differentiate between three versions of our imputation model differing in the BHPS sample used for the estimation and the inclusion of the intelligence score. For the imputed data sets to produce estimators with properties necessary to draw valid inferences, the imputation model has to be proper. To check whether our imputation models are proper, we use a simulation-based evaluation method.

We find that the imputation models with and without intelligence scores perform similarly across all dimensions. In the simulation-based evaluation, the two models produce combined estimators with similar bias and efficiency statistics for the marginal effect of intelligence score on average university degree attainment. We also show that the combined estimators of the average university degree attainment across cohorts are, in general, similar to the benchmark graduation rates computed using the data sets on the universe of undergraduate students. This similarity could allow us to use a simpler imputation model without the intelligence score in our companion paper.

References

Black, D. A., and J. A. Smith. 2004. "How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics,* Higher Education (Annals Issue), 121 (1): 99–124.

———. 2006. "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics* 24 (3): 701–728.

Boliver, V. 2015. "Are There Distinctive Clusters of Higher and Lower Status Universities in the UK?" *Oxford Review of Education* 41 (5): 608–627.

Brand, J. P., S. Buuren, K. Groothuis-Oudshoorn, and E. S. Gelsema. 2003. "A Toolkit in SAS for the Evaluation of Multiple Imputation Methods." *Statistica Neerlandica* 57 (1): 36–45.

Brewer, D. J., E. R. Eide, and R. G. Ehrenberg. 1999. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *The Journal of Human Resources* 34 (1): 104–123.

Dillon, E. W., and J. A. Smith. 2017. "Determinants of the Match between Student Ability and College Quality." *Journal of Labor Economics* 35 (1): 45–66.

Gillard, D. 1998. *Education in England: A History.*

Higher Education in the Learning Society: Main Report. 1997. Report. London: The National Committee of Inquiry into Higher Education.

Higher Education: Report of the Committee Appointed by the Prime Minister under the Chairmanship of Lord Robbins. 1963. Report. London: Committee on Higher Education.

Ichino, A., N. Jandarova, J. L. Reuter, A. Rustichini, P. M. Visscher, L. Yengo, and G. Zanella. n.d. *Intelligence and Tertiary Education.* Technical report.

Pratt, J. 1997. *The Polytechnic Experiment: 1965-1992.* Buckingham, UK; Bristol, PA, USA: Society for Research into Higher Education & Open University Press.

Rubin, D. B. 1977. "Formalizing Subjective Notions about the Effect of Nonrespondents in Sample Surveys." *Journal of the American Statistical Association* 72 (359): 538–543.

———. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data.* CRC Press.

Schonlau, M., and R. Y. Zou. 2020. "The Random Forest Algorithm for Statistical Learning." *The Stata Journal* 20 (1): 3–29.

University of Essex, Institute for Social and Economic Research. 2020. *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009.*

van Buuren, S. 2018. *Flexible Imputation of Missing Data.* Second. Boca Raton, FL: CRC/Chapman & Hall.

Willetts, D. 2017. *A University Education.* Oxford University Press.