SOLVAY BRUSSELS SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSITÉ LIBRE DE BRUXELLES

Essays on Three Social Insurance Programs

Design & Consequences for Work and Family Decisions

Thesis presented by Sébastien Fontenay

in fulfilment of the requirements of the PhD Degree in Economics
("Docteur en sciences économiques et de gestion")
Academic year 2021-2022

Supervisor: Prof. Ilan Tojerow

Thesis jury :

Prof. Bart Cockx (Ghent University)
Prof. Paula Gobbi (Université Libre de Bruxelles)
Prof. Libertad González (Universitat Pompeu Fabra)
Prof. Olivier Marie (Erasmus Rotterdam & Maastricht University)
Prof. François Rycx (Université Libre de Bruxelles)

DULBEA

Contents

Acknowledgements

"*If I have seen further it is by standing on the shoulders of Giants*"
Isaac Newton

It is of course not for me to say whether I have been able to see further, but as I am nearing the end of my PhD journey I can say that I was lucky to stand on the shoulders of Giants. I have had the chance along the way to meet inspiring scholars, starting with my supervisor Ilan Tojerow, who gave me the opportunity to embark on this journey and always made sure it was a smooth ride. I was also fortunate to have Paula Gobbi onboard, who always took the time to meet me and provide me with great advice. I greatly benefited from the classes of Bart Cockx, who taught me many of the empirical methods that I used in this thesis. This journey also took me abroad, thanks to Libertad González, who always made sure she would find the time to discuss research around a coffee at El Cafe de la Pompeu. François Rycx and Philip Verwimp, who kindly accepted to join my advisory committee from the very beginning. Olivier Marie who was the first ever discussant of one of my papers back in Barcelona in 2019, and will hopefully be able to bear witness to how far I have come. Finally, I remember my conversations with William Parienté, who was the very first person who helped me believe that I could sail the PhD boat.

There are also my personal Giants. Ana, the first reader of all my work, the best listener of my doubts, the partner of my life. My parents, Didier & Sandrine, who always made sure that I could learn, travel and pursue my passions carefree. My sister, Romane, who overtook me on the finish line, becoming a doctor years before I intend to, making me a proud brother. Because the PhD journey is a long one, one needs good friends to sail with, and I was lucky to always find Max and Thomas to share a lunch at Le Grenier d'Elvire. Federico, who, despite the distance, always takes my calls and does wonders for my confidence. Finally, all my colleagues who contributed to make the Dulbea an enjoyable place for research, Benoît, Elisabeth, Emilie, Guillaume, Octave, Laura LF, Laura N, Sébastien A.

General Introduction

Albanesi (2016) defines social insurance schemes as government transfer programs whereby individuals who claim a condition or state that reduces their labor income obtain a transfer from the government. Which "conditions" and "states" should be covered by social insurance, as well as the optimal level of coverage, has been the source of debates among policy-makers and economists for more than a century.

Early on, working women fought for protection against income loss resulting from childbirth. Back in October 1919, delegates of the first International Congress of Working Women gathered in Washington D.C. and successfully convinced the International Labour Organization to adopt a maternity protection convention, which mandated that maternity leave must come with benefits "sufficient for the full and healthy maintenance" of mother and child. Several decades later, in the 1970's, Scandinavian countries adopted gender-neutral parental leave policies, which opened government paid leaves to fathers. They have since been followed by more than 90 countries which offer paid leave mandates to both parents. Probably the largest outlier among non-adopters of parental leave policies is the United States, whose 46$^{\text{th}}$ president Joe Biden failed during the first year of his mandate to fulfill his campaign promise to put in place a nation-wide parental leave program. The century-long reforms across the world, as well as the ongoing debate in the USA, one of the largest labor markets, motivates a deeper understanding of the consequences of parental leave policies for work and family decisions.

The first two chapters of this thesis find their origin in previous work, which studies the labor market trajectory of Belgian women around childbirth. This earlier study demonstrates that Belgian mothers lose about 30 percent of their labor earnings relative to fathers, up to eight years after the birth of their first child (Fontenay, Tojerow, & Murphy, 2021). Figure I at the end of the introduction illustrates this dynamic and has been the common thread of the first two chapters. Several questions can be traced back to this figure: Why do women already experience an income loss when they are pregnant with their first child? How do women react to the income loss resulting from the incomplete wage replacement offered by the maternity allowance after childbirth?

What are the drivers of the persistent drop in wage income up to eight years after the birth of their first child? This thesis tries to offer an answer to all of these questions using empirical methods in the context of Belgium.

From a conceptual point of view, this thesis also makes the case that social insurance programs should not be considered in isolation because they can have important spillover effects. After reading this thesis, the reader should find that social insurance programs shape work and family decisions, but also that changes to one program might have unintended consequences for other branches of social security. From a methodological point of view, the three chapters rely on modern econometric methods for policy evaluation. This work uses both natural experiments and a randomized controlled trial to evaluate the effects of policies and their consequences for workers.

The first chapter originates from the observation that the level of compensation during maternity leave varies significantly across countries, but previous research on the topic provides only limited insights on the consequences of this important design choice. More particularly, the first chapter assesses how the generosity of maternity leave allowance affects first-time mothers' career trajectory and subsequent fertility decisions. It exploits the fact that the allowance is capped in Belgium, so that women with pre-leave earnings above the maximum threshold face drastically lower replacement rates. Using a regression kink design, as well as a rich set of administrative data on mothers from 2002 to 2015, this chapter highlights the consequences for their career and show that mothers who receive higher benefits are more likely to leave salaried employment for self-employment. The study also reveals that mothers who receive a more generous allowance have more children and it provides suggestive evidence connecting this fertility effect to the transition to self-employment.

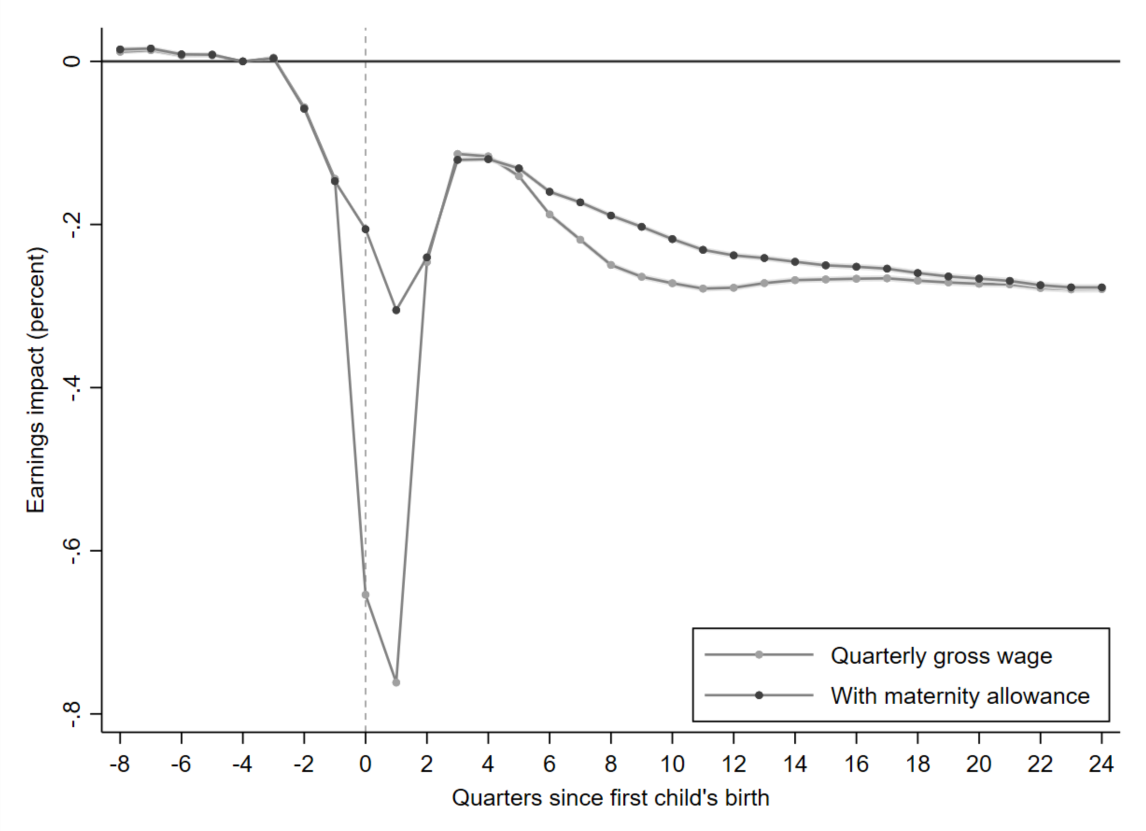
The second chapter, co-authored with Ilan Tojerow, explores the short-run and long-run determinants of the income loss suffered by young mothers. The study shows that a woman's likelihood of claiming disability insurance increases after the birth of her first child, but also reveals that the provision of paternity leave can ease this effect. Using Belgian administrative data, the findings are twofold: the incidence rate of disability is constant across gender up until a woman becomes a mother for the first time, and the provision of paternity leave, even in short intervals, significantly reduces the number of days that mothers spend in disability. Specifically, the regression discontinuity difference-in-differences design shows that mothers with partners eligible for two-week-long paternity leave spend on average 21% fewer days in disability over twelve years. The fiscal consequences of the research suggest that spending on paternity leave would be more than compensated by the savings in mothers' disability benefits.

The second chapter also highlights the increasing number of workers who rely on another social insurance program, called Disability Insurance (DI), as they find themselves unable to work because of a medical condition. A large share of these beneficiaries enter DI because of mental health conditions, which generally start at a young age and too often translate into a lifetime of claiming benefits.

The third chapter, also co-authored with Ilan Tojerow, evaluates the effects of a Supported Employment (SE) program aimed at helping DI recipients with mental conditions to return to work. The program is characterized by a "work first" approach with intensive job counselling and follow-along support. Using a Randomized Control Trial with more than 660 participants over a follow-up period of 18 months, the study compares the benefits of this newly introduced program to regular vocational rehabilitation services traditionally used in Belgium. The results show that SE increases the probability of DI recipients with mental conditions to work while on claim and reduces their reliance on DI benefits. Specifically, the estimates reveal that, 18 months after the start of their return-to-work program, participants in the SE group are 9.6 percentage points more likely to be working and receive 6.5% less in DI benefits than those in the control group. The effects of SE remain substantial even for those participants who were impacted by the COVID-19 pandemic. Cost-benefit analysis suggests that spending on SE could be compensated within less than two years by the savings in DI benefits.

Some important policy implications can be drawn from these studies regarding the career of women and the effects of public policies. First, providing young mothers with more generous maternity leave allowance would help reducing the gender gap in entrepreneurship, while having positive spillover effects on fertility. Second, paternity leave policies, by increasing the involvement of fathers and co-parents, may improve the long-term health of mothers and prevents career breaks due to sickness or disability. Third, for those who are already on disability insurance, supported employment policies may help them return to work and avoid a lifetime of claiming benefits. All these findings are also of particular importance for the economy as a whole since previous research has revealed that a better allocation of women's talent would boost productivity and growth (Andrew, Bandiera, Costa-Dias, & Landais, 2021; Hsieh, Hurst, Jones, & Klenow, 2019).

Figure I: Impact of Children on Belgian Mothers' Earnings



Notes: The data for this figure comes from Fontenay, Murphy and Tojerow (2021). The figure shows event time coefficients (relative to the 4^{th} quarter before the first child's birth) estimated on a sample of mothers who had their first child between 2003-2010 and were eligible for maternity leave (i.e. had sufficient work history). The coefficients are displayed as a percentage of the mean of the outcome measured at t-4. The earnings are measured conditional on labor force participation.

Chapter 1

How Can Paid Maternity Leave Boost Female Entrepreneurship and Fertility?

This article previously circulated under the title "*How Does Maternity Leave Allowance Affect Fertility and Career Decisions?*" - I am grateful to Bart Cockx, Pierre-Edouard Collignon, Bram De Rock, Delia Furtado, Paula Gobbi, Libertad González, Magali Jaoul-Grammare, Jonas Jessen, Scott Kostyshak, Anna Minasyan, Giacomo Pasini, Hanifa Pilvar, Giacomo Ponzetto, Daniel Reck, Maya Rossin-Slater, Ilan Tojerow, Theodor Vladasel and Peter Wang for very helpful comments and suggestions. I also thank seminar participants at Université Libre de Bruxelles, Webinar in Gender and Family Economics (THEMA - CY Cergy Paris Université), Université catholique de Louvain, 10$^{\text{th}}$ ifo Dresden Workshop on Labor Economics and Social Policy, 26$^{\text{th}}$ Society of Labor Economists Annual Meeting, 18$^{\text{th}}$ Augustin Cournot Doctoral Days, 3$^{\text{rd}}$ QMUL Economics and Finance Workshop, 28$^{\text{th}}$ Spanish Meeting on Public Economics, 55$^{\text{th}}$ Annual Conference of the Canadian Economics Association, 23$^{\text{rd}}$ Spanish Applied Economics Meeting, 69$^{\text{th}}$ Annual Meeting of the French Economic Association, 34$^{\text{th}}$ Annual Conference of the European Society for Population Economics, 20$^{\text{th}}$ Journées Louis-André Gérard-Varet, International Association for Applied Econometrics Annual Conference, 96$^{\text{th}}$ Western Economic Association Annual Conference, European Public Choice Society, Arne Ryde Workshop on Gender and Family Wellbeing - Lund University, Universitat Pompeu Fabra, London School of Economics.

Introduction

One in ten women are self-employed in the European Union, and one in twenty in the United States. But according to the International Social Survey on Work Orientations, if women could choose, a large proportion would leave salaried employment to start their own business.[1] These findings prompt questions about what creates the gap between women's aspirations and reality.

In a seminal paper, Evans and Jovanovic (1989) explain that liquidity constraints tend to exclude individuals with insufficient funds at their disposal from self-employment. Meanwhile, we know that young women face a large drop in earnings as they enter motherhood (see e.g., Angelov, Johansson, & Lindahl, 2016; Lundborg, Plug, & Rasmussen, 2017). The resulting earnings gap with men ranges from around 20% in Scandinavian countries to more than 50% in German-speaking countries (Kleven, Landais, Posch, Steinhauer, & Zweimüller, 2019). These two empirical facts give rise to a specific question: Do potential female entrepreneurs remain in salaried employment because they are financially constrained following their entry into motherhood?

In this paper, I study how providing mothers with a higher allowance during maternity leave impacts their decision to become self-employed. My argument is that receiving a higher allowance might help lift women's liquidity constraints, which are particularly salient upon entering motherhood. Indeed, Figure A1 in Appendix shows that Belgian mothers, like mothers in most countries, face a large drop in earnings after giving birth for the first time. For some women, earnings loss even begins in pregnancy, when those facing health-related challenges must temporarily step away from work (Fontenay & Tojerow, 2020). It is therefore possible to think of maternity leave allowance as a consumption smoothing device that allows mothers to transition from salaried to self-employment with limited consequences.[2] This might be important as there is considerable uncertainty about initial earnings for entrepreneurs (Dillon & Stanton, 2017).

My empirical strategy relies on a Regression Kink Design (RKD) to identify the impact of the allowance received by first-time mothers on their subsequent decision to become

[1]In 2019, 11.3% of employed women in the European Union and 5.1% in the United States were self-employed, the OECD reports. The 2005 International Social Survey Programme on Work Orientations found that the share of women who would choose to be self-employed was 55% in the United States, 46% in Canada, 44% in Australia, 43% in Germany, 37% in Great Britain, and 32% in France (ISSP, 2005).

[2]Not all women across the globe, however, benefit from complete protection against earnings loss during maternity leave. In the majority of OECD countries, wage replacement is only partial, so that mothers suffer substantial earnings loss during maternity leave. Figure A2 in Appendix shows that maternity leave benefits completely replace lost earnings in only 13 out of 41 countries. In the vast majority of the remaining countries, the mean replacement rate for a woman with earnings at the level of the national average is around 66%.

self-employed. I exploit the fact that from the 31^{st} day of maternity leave, the allowance is capped in Belgium,[3] so that women with pre-leave earnings above the maximum threshold face drastically lower replacement rates. As shown in Figure 1, this translates into a discontinuous change in the marginal replacement at the threshold set by the Belgian Social Security Administration. I leverage this discontinuity to identify the causal effects of the benefit amount by testing for a change in the slope of the relationship between my outcomes of interest and the assignment variable (i.e., pre-leave earnings) at the kink. My empirical strategy thus allows me to explore the effects of maternity leave allowance generosity while holding constant other policy parameters, including the duration of leave.

I use a rich administrative dataset, which aggregates several registers on social security status and employment of Belgian mothers between 2002 and 2015, as well as detailed household information that includes their co-parent. As such, I am able to paint a comprehensive picture of the consequences of maternity leave allowance on women's labor supply. My data also includes information about children born after the mother's first child, up to five years. I can therefore also study fertility decisions, which as previous research has shown, are intrinsically linked to career decisions. I believe it is a positive feature of my paper, since career and fertility decisions cannot be fully understood in isolation.[4]

This paper's first contribution is to demonstrate that the level of compensation during maternity leave positively affects the probability of young women becoming self-employed. In particular, I estimate that a 900-euro increase in maternity leave allowance raises the probability of becoming self-employed by six percentage points within five years. Heterogeneity analysis shows that the transition to self-employment is larger among women who are not the breadwinner or primary earner at home, which could result from lower liquidity constraints when living with a partner with higher earnings. This study's second contribution is to reveal that mothers who receive higher benefits also have more children in total.

I consider two mechanisms that could generate both results on self-employment and fertility decisions. First, as mentioned above, a more generous maternity leave allowance might help lift young women's liquidity constraints. I call this the *direct effect* of maternity leave allowance on self-employment and believe it could reflect the fact that women with entrepreneurial ideas can venture into entrepreneurship with limited risk for their finances.

[3]As shown in Figure A2 in Appendix, about half of OECD countries cap the benefit amount, which results in a much lower replacement rate for high-earning women

[4]The structural literature has highlighted the interaction between fertility and women's labor supply, and as such, the need to jointly model these decisions. Some of the most recent papers in this literature build on dynamic life-cycle models that account for the effects of family policies (Adda, Dustmann, & Stevens, 2017), including paid leave benefits (Stichnoth, 2020).

This is especially true for the first months of self-employment, when there is considerable uncertainty about entrepreneurial earnings (Dillon & Stanton, 2017). At the same time, because self-employment offers greater flexibility, it is easier for those women to have more children. This is compatible with Wellington's model, (2006) which demonstrates that self-employment offers a better balance between family and career. Second, mothers who receive a higher allowance during their maternity leave have a lower opportunity cost of childbearing, which could influence their decision to have more children in the future. I call this mechanism the *fertility mediator*, whereby mothers who receive more generous benefits are more likely to desire more children in the future. As a consequence, they switch to self-employment to accommodate their desired fertility with the pursuit of their career. It is therefore an indirect effect on self-employment, which is compatible with the previous literature that shows that higher compensation during maternity leave increases subsequent fertility (Cygan-Rehm, 2016; Kluve & Schmitz, 2018; Malkova, 2018; Raute, 2019).[5]

I conduct a number of complementary analyses to distinguish between the direct and indirect channels and find more evidence to support the former. First, when looking at the dynamics of both effects, I find that mothers who switch to self-employment do so in the quarters following the birth of their first child, that is, before they decide to become mother for a second time. Second, if changes in desired fertility were a mediator, one would expect to observe other types of career switches, as suggested by the child penalty literature (e.g., Kleven, Landais, & Søgaard, 2019). However, I find no effects on the probability to remain in the labor force in the long run. The switch to self-employment is the only career shift reflected in my data. Taken together, these supplementary analyses point toward a direct effect of maternity leave benefits on self-employment, with positive spillovers on fertility decisions.

A substantial part of this study is also dedicated to understanding the reasons why young women become entrepreneurs upon entering motherhood. I start by exploring the financial incentives by looking into the consequences of moving to self-employment for the mothers' earnings. I reveal that a small share of women in the right tail of the earnings distribution are financially better off, but for the majority of mothers the impact on their earnings is in fact negligible or even slightly negative. These findings prompt questions about what motivates mothers becoming self-employed if they do not achieve higher earnings. To answer this question, I investigate the characteristics of their pre-childbirth workplace, and more particularly whether they might be required to work at night or during weekends, using the European Working Conditions Survey (EWCS). I find

[5] Another study by Tudor (2020) shows that a 2003 reform in Romania, which switched the earnings-related system to a fixed benefits system with substantial gains for most employed women, did not influence short-term conception rates but significantly decreased the probability of abortion.

that those working in hotels/restaurants or hospitals/clinics, known for their "atypical work schedules" (Eurofound, 2014), are more likely to make the move to self-employment when they receive a higher maternity leave allowance. I argue, after Hamilton (2000), that female entrepreneurs may trade lower earnings for the non-pecuniary benefits of business ownership, such as having a more flexible work schedule.

I believe it is plausible that my results may generalize beyond the particular context of maternity leave. In fact, earnings loss resulting from the child penalty has been shown to be long-lasting (Kleven, Landais, & Søgaard, 2019), which means that policies that aim at lifting the liquidity constraints facing women could be implemented at any stage in their career path. This should also be accompanied by job-protected leave, however, as suggested by Gottlieb, Townsend, and Xu (2021). In recent study, the authors exploit a reform that changed the duration of parental leave in Canada and show that it affected the propensity of mothers to become entrepreneurs after childbirth. The authors' argument is that mothers use the paid time off work to incubate their businesses while the job-protected leave minimizes the risks to their career.[6] My study complements their findings by investigating the effects of the allowance amount and keeping the duration of maternity leave constant. My results suggest that the amount of benefits also matters since it can relax the financial constraints facing mothers who wish to become self-employed.

The importance of liquidity constraints for entrepreneurship has been highlighted in the past (see e.g. Blanchflower & Oswald, 1998; Evans & Jovanovic, 1989; Flèche, Lepinteur, & Powdthavee, 2021; Holtz-Eakin, Joulfaian, & Rosen, 1994a; Holtz-Eakin, Joulfaian, & Rosen, 1994b; Hurst & Lusardi, 2004). Blanchflower and Oswald (1998), for instance, show that receiving 5,000 pounds in inheritance doubles the probability of setting up a business. The effect is particularly large for young adults who received an inheritance before they turned 23, which suggests that capital constraints are more binding for younger individuals (Blanchflower & Oswald, 1998). Lindh and Ohlsson (1996) in Sweden, and more recently Flèche, Lepinteur and Powdthavee (2021) in the United Kingdom, reveal that lottery winners are also more likely to become entrepreneurs. In a different context, Hombert et al. (2020) examine a French reform to Unemployment Insurance (UI). The reform allows UI beneficiaries to keep their benefits if they start a business, resulting in an increased number of entrepreneurs among UI recipients. But while the previous literature stresses the importance of liquidity constraints for aspiring entrepreneurs, it is largely silent on the gender differences in financial capacity, especially after motherhood.[7] My research contributes to this literature by showing that the gap

[6] Another study by Gerards and Theunissen (2018) exploits a reverse reform in Germany, which reduced parental leave generosity, only to find a lower transition to self-employment for affected mothers.

[7] Flèche, Lepinteur and Powdthavee (2021) find no differences in the likelihood of becoming self-employed for female and male lottery winners. Other strands in the literature show that gender differences in entrepreneurship can stem from differences between men and women in terms of skills, traits, or

in self-employment between men and women could be reduced by lifting the financial constraints that motherhood creates for women.

More broadly, I also contribute to the literature on parental leave policies. Much of this research examines the impact of maternity leave duration on mothers' later outcomes (see Olivetti and Petrongolo (2017) and Rossin-Slater (2018) for comprehensive overviews). But only a handful of papers look at the impact of the generosity of maternity leave allowance (Asai, 2015; Bana, Bedard, & Rossin-Slater, 2020; Cygan-Rehm, 2016; Ginja, Jans, & Karimi, 2020; Kluve & Schmitz, 2018; Raute, 2019; Tudor, 2020). This is likely due to the difficulty of finding a credible causal inference design that isolates the benefit amount since it is not randomly assigned and because reforms often combine changes to both the duration of the allowance and the replacement rate. Three of those papers focus on a 2007 German reform that changes parental leave from a flat means-tested system to an earnings-related benefit. Cygan-Rehm (2016) and Kluve and Schmitz (2018) show that the reform reduced subsequent childbearing for low-income mothers who received higher benefits under the old means-tested system. Conversely, Raute (2019) finds a positive effect on the fertility of women at the middle and upper end of the earnings distribution, who were better off after the reform. Looking at the German context, these studies indicate that compensating women according to the opportunity cost of childbearing increases fertility.[8] Three other studies on maternity leave allowance generosity examine the impact on labor market outcomes (Asai, 2015; Bana et al., 2020; Ginja et al., 2020). While Asai (2015) and Bana et al (2020) do not find any effects, Ginja, Jans, and Karimi (2020) show that eligibility for higher parental leave benefits decreases mothers' labor supply, but only in the short run (up to two years). I extend these studies by showing that even though maternity leave allowance has no effect on the employment rate in the long run, it does affect the career paths of mothers by changing their probability to become self-employed.

My findings should provide policy relevant information since women's entrepreneurial potential is widely considered an under-exploited source of economic growth and jobs

preferences (Barber & Odean, 2001; Croson & Gneezy, 2009; Niederle & Vesterlund, 2007). Yet Parker (2018) explains that the gender gap in entrepreneurship remains largely unexplained, even when accounting for these differences. Authors have recently highlighted the importance of exposure to people with entrepreneurial experience, whether neighbors, family members, colleagues or school peers (Baggesgaard, Mertz, Ronchi, & Salvestrini, 2021; Markussen & Røed, 2017; Rocha & van Praag, 2020).

[8] In his seminal work on the economic analysis of fertility decisions, Becker (1960) highlights both the price and income elasticities of the demand for children. Recent empirical findings by Cohen, Dehejia & Romanov (2013) reveal that changes to Israel's child subsidies system have generated substantial price effects on overall fertility, but a more modest income effect. Additional research from the quasi-experimental literature also suggests a positive response of fertility to financial incentives in general and child benefits in particular (González, 2013; Milligan, 2005; Riphahn & Wiynck, 2017). It is expected that maternity leave allowance will also affect both the price/cost of a child, while temporarily increasing the disposable income of the mother. However, my set-up does not allow to measure the contribution of each channel.

(European Commission, 2017). In fact, previous research has revealed that a better allocation of women's talent would boost productivity (Andrew et al., 2021; Hsieh et al., 2019). Furthermore, the level of compensation during maternity leave varies widely across countries. Belgium offers a particularly interesting case for research given that the features of its maternity leave mandate are only slightly above international standards set by the International Labor Organization and are not among the top or bottom performers. Simulations for Belgium suggest that raising the benefit threshold to the 99^{th} percentile would contribute to reducing the gender gap in entrepreneurship, while having positive spillover effects for the fertility of high-earning women.

The remainder of the paper is organized as follows. Section 1.1 provides more details on the Belgian maternity leave program and benefit schedule. Section 1.2 introduces the empirical framework and shows tests of the identifying assumptions of the RKD. Section 1.3 presents the main results, as well as heterogeneity analyses. Section 1.4 provides various robustness checks. Section 1.5 illustrates the impact of moving the benefit threshold in a simulation exercise. Section 1.6 concludes this study.

1.1 Institutional Background

In Belgium, paid, job-protected maternity leave was introduced in 1971. This legislation provides a maximum of 15 weeks of paid leave for mothers of newborn children. Maternity leave is not universal, and women are entitled to paid leave only if they have worked at least 120 days in the last six months.[9]

To some extent, mothers can decide how to distribute these weeks of paid leave before or after giving birth. They must, however, take at least one week before the planned delivery date and cannot return to work earlier than nine weeks after giving birth. In other words, all mothers must stop working during a compulsory period of at least 10 weeks. Payments of benefits are made for a six-day week, so that for a total leave of 15 weeks, the corresponding number of days is 90.

The replacement rate is 82% of pre-leave gross wages during the first 30 days and 75% thereafter. While benefits paid during the first 30 days are not capped, the amount paid for the remaining 60 days is. Figure 1 (Panel A) illustrates this variation in the benefit schedule. One can observe that the daily allowance is a linear function of pre-leave earnings during the first 30 days (solid line), while it is a kinked function for the

[9] "*Loi sur le travail du 16 mars 1971*" Leave can be extended to 19 weeks for multiple births. Part-time workers are required to have worked at least 400 hours. Unemployment insurance recipients are also entitled to maternity leave if they can demonstrate sufficient active days in their job search.

following 60 days (dashed line).

Formally, the average daily allowance $\bar{A}$ received by a woman who maxes out her maternity leave entitlement of 90 days is a fraction τ_1 of her pre-leave daily earnings W during the first 30 days and a fraction τ_2 of the capped pre-leave daily earnings W_{max} during the remaining 60 days:

$$\bar{A} = \begin{cases} (W \cdot \tau_1) \cdot (\frac{30}{90}) + (W \cdot \tau_2) \cdot (\frac{60}{90}) & \text{if } W < W_{max} \\ (W \cdot \tau_1) \cdot (\frac{30}{90}) + (W_{max} \cdot \tau_2) \cdot (\frac{60}{90}) & \text{if } W \geq W_{max} \end{cases} \tag{1.1}$$

Panel B of Figure 1 simulates the benefit function of Equation (1) and illustrates the effect of a cap placed at the January 2007 threshold (i.e., 110.655 euros), right in the middle of my sample window. One can observe the change in elasticity at this threshold. The elasticity of maternity leave allowance to pre-leave earnings is 0.77 below the kink,[10] while it is 0.27 above the kink.[11] The marginal replacement rate faced by women above the kink is therefore 50 points lower than the rate below the kink.

To provide a sense of earnings loss resulting from the cap, Panel B of Figure 1 features the situation of a mother whose pre-leave daily earnings were 175 euros. Her average daily allowance is about 103 euros, that is, a replacement rate of only 59%. In the absence of the kink (i.e., if the benefit schedule was linear), she would receive a daily allowance of about 135 euros (32 euros more per day). If she maxes out her leave entitlement, this mother loses 2,880 euros, compared to a scenario in which there is no cap on the benefit amount.

Figure A3 in Appendix shows that the daily earnings threshold has evolved over time, from 99 euros in January 2003 to 121 euros at the end of 2010, reflecting government decisions as well as automatic adjustments to inflation. It is important to note that the schedule applies based on the start date of the maternity leave, so that a change in the earnings threshold does not affect ongoing spells.

The features of the Belgian maternity leave system described above are very close to the standards set by the International Labor Organization. Indeed, since 2000, the ILO's Maternity Protection Convention (No. 183) mandates a minimum leave period of 14 weeks for women around childbirth. The ILO also recommends that the cash benefit paid to women during maternity leave should amount to at least two-thirds of their previous earnings. Thus, Belgium offers a particularly interesting case, slightly above international standards but not among the top performers. In that sense, this study potentially offers

[10]It corresponds to the following calculation: ((0.82*30/90)+(0.75*60/90))

[11]It corresponds to the following calculation: (0.82*30/90)

good external validity.

1.2 Empirical Strategy

In this section, I start by describing data collected on Belgian mothers who had a child during the period 2003–2010 and provide descriptive statistics on the sample. Then, I explain in detail the estimation strategy based on a Regression Kink Design (RKD). Finally, I discuss several tests that provide support for the validity of the RKD in my particular context.

1.2.1 Data and Sample

I leverage a rich set of administrative data from the Belgian Crossroads Bank for Social Security. It puts together several administrative registers linked at the individual level (via personal identification numbers) and contains information on household composition, labor market outcomes of each member, and social security status. Most importantly, the data allows me to match children with their parents, as well as workers to their firms.

I obtained a large sample of 60% of all births during the years 2003 to 2010, with stratification at the provincial level to ensure representativity. I am able to follow the careers of mothers of children in the period 2002 to 2015. I can therefore build a balanced panel that spans the period from four quarters before the birth of the child to 20 quarters after.

I restrict my sample to mothers who had a first child between 2003 and 2010. The objective is to focus the analysis on maternity leave for firstborns. I then remove mothers who are not entitled to paid maternity leave because they do not have sufficient work history.[12] Finally, I only take into account women who are salaried employees before entering motherhood, since I am interested in the transition from salaried employment to self-employment. In addition, women who are already self-employed at first childbirth receive a flat maternity leave benefit. They will be used as a placebo group in section 1.4.3, dedicated to robustness checks. To summarize, I have a large sample of women who were salaried employees before becoming mothers for the first time in the period 2003–2010 and who are entitled to paid maternity leave, which comprises 222,610 women.

For each maternity leave, I observe the start and end months, the number of days of

[12] I also do not include civil servants who are entitled to a 15-week maternity leave, but with a replacement rate of 100% paid directly by the public administration.

the claim, and the quarterly allowance received by the mother. For each woman, I assign the relevant pre-leave earnings by looking into the last wage payment received over the four quarters that precede the start of maternity leave. I was able to find the pre-leave earnings for more than 82% of the sample.[13] My final sample is therefore composed of 182,923 women for whom I have a complete work history.

I follow these mothers for 20 quarters (i.e., five years) after the start of maternity leave. I can observe their quarterly earnings from both salaried employment and self-employment.[14] For salaried employees, I also have precise information on their volume of work (recorded as full-time equivalent[15]). Finally, I am also able to track the place of residence and composition of their household, including the number of children and their partner[16] and whether or not they are married.

Table 1 presents the means and standard deviations of key variables for women in my sample. I also report descriptive statistics for my "kink sample," that is women with pre-leave daily earnings within 22 euros of the kink point, a sub-sample of 37,906 individuals. As discussed in the next subsection, this is the preferred bandwidth for my estimations. From Figure A4 in Appendix, one can observe the distribution of those pre-leave earnings relative to the kink point located around the 90$^{\text{th}}$ percentile. The 22-euro bandwidth includes most women in the fourth quartile of the earnings distribution.

One can observe from Table 1 that the age of mothers at first birth is on average 28.4 years, slightly below the OECD average of 29.1.[17] Data on maternity leave claims shows that the average duration is 84.7 days, close to the maximum of 90 days. The average benefit amount is 4,808 euros. One can also see that women in the "kink sample" are slightly older, have higher pre-leave earnings, and consequently receive higher benefits. The descriptive statistics also reveal that mothers within the 22-euro bandwidth receive

[13] The missing 18% are women for whom I cannot infer pre-leave earnings based on the limited work history available in my sample. These are very likely women who were receiving unemployment or sickness benefits before delivery. These women are entitled to maternity leave, but their last wage payment might be older than 12 months.

[14] Since self-employment earnings are reported on a yearly basis, I divide the amount by four and impute the value for each quarter of any given year during which the individual had positive self-employment earnings. Earnings from self-employment are net of expenses and social security contributions. Therefore, they correspond to the benefit or loss generated by the self-employed activity.

[15] It is computed as the number of hours worked by an individual during a given quarter, divided by the average time for a reference worker in the same sector. The outcome therefore measures the amount of part-time work but also accounts for the fact that a person has not been working during the entire quarter.

[16] While I can follow the careers of all parents during the entire 2002–2015 period, I can only match them at the moment of the birth of their common children. My sample includes only 60% of all births in Belgium, which implies that I am not always able to match the parents at the birth of their first child. In that case, I match them based on the birth of a subsequent common child. This concerns about 20% of co-parents, a figure that is "imputed" based on their next observed child.

[17] Source: OECD Family Database, 2017.

a higher hourly wage and work more hours per week.

1.2.2 Regression Kink Design

I am interested in identifying the causal impacts of maternity leave allowance on mothers' labor market participation, including self-employment, as well as their subsequent fertility. The challenge is that women with high benefits cannot be directly compared to those with low benefits, since it is likely that there are unobserved variables, which are correlated with both the benefit amount and my outcomes of interest. One can think, for instance, that women who are more career-oriented worked longer hours before entering motherhood and therefore had higher earnings and will consequently receive higher benefits. At the same time, those women might be more inclined to start their own business or have different perspectives on their ideal number of children.

To circumvent this issue, I leverage quasi-experimental variation stemming from a kink in the maternity leave benefit schedule, created by the earnings' threshold set by the Social Security Administration. As explained in Section 1, women below and above the kink face drastically different marginal replacement rates. Following Card, Lee, Pei, and Weber (2015b), I make use of this change in the slope of the benefit function to estimate the causal effects of the benefit amount using a Regression Kink Design (RKD).

The RKD will test for a change at the kink in the slope of the relationship between my outcomes of interest and the assignment variable (i.e. the pre-leave earnings). If one assumes that in the absence of the kink in the benefit function, there would be a smooth relationship between the outcomes and the assignment variable, evidence of a change in the slope would imply a causal effect of the benefit amount on the outcome. I explain more in detail the "smoothness" assumption in Subsection 3.1 and provide tests that support the validity of the RKD in my context.

Mathematically, I want to estimate the marginal effect of maternity leave allowance (A) on the outcome (Y) at the kink point. The allowance is a function of pre-leave earnings W, which have been normalized relative to the threshold set by social security and therefore take the value W_0 at the kink point:

$$E\left[\frac{\partial Y}{\partial A}|W=W_0\right] = \frac{\lim_{W\to W_0^+} \frac{\partial E[Y|W=W_0]}{\partial W} - \lim_{W\to W_0^-} \frac{\partial E[Y|W=W_0]}{\partial W}}{\lim_{W\to W_0^+} \frac{\partial A(W)}{\partial W} - \lim_{W\to W_0^-} \frac{\partial A(W)]}{\partial W}} \tag{1.2}$$

From Equation (1.2), one can see that the RKD estimator is a ratio of two terms. The numerator is the change at the kink point in the slope of the relationship between the outcome Y and the pre-leave earnings W. The denominator is the change at the kink point in the slope of the benefit function. The resulting estimate can be interpreted as a local effect of the treatment on the treated.

In theory, one could infer the denominator directly from the benefit formula. Because of measurement errors, however, there may be small deviations between theoretical and observed values. This may stem from errors in the observed values of the assignment variable or the benefit amount. In addition, not all women max out their maternity leave entitlement, and the average daily allowance will vary with the total duration since the cap applies only from the 31st day. For all those reasons, I follow Card *et al.* (2015b) in using a "fuzzy" RKD and also estimate the slope change in the denominator of Equation (1.2).

I apply local nonparametric regressions on either side of the kink to estimate the slope changes in both the numerator and denominator.[18] The use of local nonparametric methods has been advocated in order to reduce the bias that may result from using data farther from the kink (Lee & Lemieux, 2010). Among the key parameters, one needs to define when implementing the RKD estimator in practice, are the kernel, the order of the polynomial, and the bandwidth.

I follow the literature, notably Card *et al.* (2015b), by using a uniform kernel (i.e., no weighting).[19] As for the polynomial order, I provide results for both local linear and local quadratic specifications.[20] The tests reported in Subsection 4.1 show that the linear specification outperforms the quadratic one in this particular context. It is therefore my preferred specification to study the impact of maternity leave allowance.

There is an active econometrics literature on optimal bandwidth choice in the regression discontinuity design literature (see e.g. Imbens & Lemieux, 2008; Imbens & Kalyanaraman, 2012), but only a handful of papers have explored the question in the case of the RKD (Calonico et al., 2014; Card et al., 2015b). I use the "data-driven" procedure of Calonico *et al.* (2014), which is the only bandwidth selector explicitly designed for the fuzzy RKD, to explore the optimal bandwidth choice for each of my outcomes. The bandwidths picked by this procedure are reported in Table A5 in Appendix. One can observe that they range from 14 to 30 euros around the kink. For comparison purposes,

[18] I implement the local nonparametric regression using the package "rdrobust" developed by Calonico, Cattaneo and Titiunik (2014) for the software Stata.

[19] Card *et al.* (2015b) explain that while a triangular kernel is boundary optimal, efficiency losses from using a uniform kernel are small both in actual applications and in Monte Carlo simulations.

[20] Card *et al.* (2015b) explain that it is commonly assumed that a local quadratic approach is preferred to a local linear fit to estimate derivatives in the RKD because the former is expected to lead to an asymptotically smaller bias compared to the latter. However, Card *et al.* (2015b) also warn against making the quadratic model a universal choice and argue that one should also take into account the characteristics of the dataset of interest, including the sample size.

it is, however, desirable to have a common bandwidth, and therefore similar sample sizes, for the analysis of the different outcomes considered. Therefore, my baseline estimates will use a common bandwidth of 22 euros, that is the median of the suggested bandwidths of Calonico *et al.* (2014)'s selector. As a robustness check, I will also report results for each individual outcome using its own optimal bandwidth, as well as other bandwidths on a 10 to 35 euros interval.

When it comes to inference, I follow Card *et al.* (2015a) in using heteroskedasticity robust standard errors. I also compute bias-corrected confidence intervals (CI), based on the procedure proposed by Calonico *et al.* (2014). Intuitively, the authors suggest to estimate the bias of the estimator by using higher order polynomials on a larger bandwidth. For my linear specification, I will therefore add a quadratic term in the assignment variable on a bandwidth that is four times the one used for the conventional estimates.[21] This bias will then be used to correct the CIs, which in the vocabulary of the authors are therefore "robust" to large bandwidth choices.

1.2.3 Tests of Identifying Assumptions

The identification of treatment effects in the framework of the RKD relies on two main assumptions: (1) the density of the assignment variable —in my case, pre-leave earnings— should be smooth (i.e., continuously differentiable) at the kink (sometimes referred to as the "no sorting" assumption); (2) in the vicinity of the kink, there should be no change in the slope of the underlying direct relationship between the assignment variable and the outcomes of interest (sometimes referred to as the "smoothness" assumption).

The first assumption implies that individuals have not sorted around the kink by manipulating their earnings. This local random assignment condition seems credible in the context of maternity leave benefits since it is unlikely that first-time mothers have perfect knowledge of the benefit schedule. First of all, because the exact location of the kink changes over time reflecting government decisions as well as automatic adjustments for inflation (see Figure A3 in Appendix). Thus, it is arguably difficult for individuals to predict what would be the location of the kink at the time of their maternity leave. In addition, the threshold for maternity leave benefits is distinct from other social programs, such as unemployment insurance or public pensions. Nevertheless, I follow other authors using RKD and provide standard tests for these identifying assumptions (Landais, 2015; Card et al., 2015b; Gelber, Moore, & Strand, 2017; Bana et al., 2020).

[21] In my case, the preferred bandwidth is symmetric and equals to 22 euros. The bandwidth used for bias correction is therefore 88 euros.

I begin by providing graphical evidence that mothers did not engage in sorting around the kink. Figure 2 (Panel A) shows that the density of the pre-leave earnings around the kink point seems smooth. To confirm this graphical diagnosis, I perform a McCrary test as is standard in the regression discontinuity design literature (McCrary, 2008). The test checks for a "jump" in the probability density function (p.d.f.) of the assignment variable at the kink. I report on Figure 2 (Panel A) the estimate for the log difference in the height of the p.d.f. at the kink, as well as the standard error in parentheses. Following Card *et al.* (2015b), I extend the spirit of the McCrary test to check that the first derivative of the p.d.f. is also continuous at the kink. I therefore regress the number of individuals in each bin[22] on polynomials of the assignment variable interacting with a dummy for being above the kink. On the graph I report the coefficient on the interaction term for the first-order polynomial (i.e., testing for a change in the slope of the p.d.f.) and the corresponding standard error. The estimates for both tests are insignificant, which confirms that one cannot detect a lack of continuity at the kink.

Furthermore, I complement the standards tests of the RKD literature and use a novel estimation technique proposed by Cattaneo, Jansson, and Ma (2020) to check for the absence of manipulation. Their approach, based on local-polynomial density estimators, does not require the data to be averaged into bins, and therefore makes no assumption about their size. In addition, the choice of the bandwidth is entirely data-driven.[23] Figure 2 (Panel B) displays the density of the assignment variable on both sides of the kink point. The graphical evidence, as well as the formal test, suggest that one cannot detect manipulation in the neighborhood of the kink.

The second assumption of the RKD cannot be tested directly. Indeed, I cannot observe what would be the direct relationship between earnings and the probability to become self-employed or the number of children in the absence of paid maternity leave since my sample is composed of mothers who all received benefits. I can, however, check that the "smoothness assumption" holds using pre-determined covariates, before the women in my sample went on maternity leave. Figures 3 and 4 plot the mean values of those pre-determined covariates in each bin of the normalized pre-leave earnings for the mother and her partner, respectively. All the figures seem to suggest that covariates evolve smoothly at the kink. Formal tests can also be performed by running local nonparametric regressions on either side of the kink to estimate the slope changes, mimicking the strategy used to compute the numerator in Equation (1.2). Table 2 shows the estimated change in slope in the relationship between the pre-determined outcomes and the assignment variable at the kink point in the linear case. The coefficients for the covariates of both parents

[22] I use 50-euro-cent bins and a 30-euro bandwidth.

[23] The Stata command rddensity (Cattaneo, Jansson, & Ma, 2018) implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo, Jansson, and Ma (2020).

are all statistically insignificant, except for the mother's full-time work equivalent. The coefficient for the latter variable is, however, very small in magnitude. In addition, it likely reflects the fact that the variable is right-censored at one, since individuals in my sample cannot work more than one full-time job. The small negative coefficient therefore reflects that most women on the right of the kink have reached this maximum value, as suggested by Panel D in Figure 3. In subsection 1.4.3, I will use two placebo groups to confirm the absence of a direct relationship: fathers who did not go on leave as well as mothers who were already self-employed before having a first child and were therefore entitled to a flat benefit.

Taken together, the previous figures and formal tests show that there is no evidence of sorting or underlying non-linearities around the kink, thus providing support for the validity of the RKD in my particular context. I can now turn to examining the impact of the maternity leave allowance on my outcomes of interest.

1.3 Results

In this section, I present the main results of the impact of the maternity leave allowance on mothers' propensity to become self-employed, their future earnings, as well as their fertility decisions. Furthermore, I go beyond average effects and estimate quantile regressions to explore the heterogeneity of the results along the earnings distribution. I also conduct heterogeneity analysis based on the wage differential of parents prior to the birth of their first child, as well as the characteristics of the sector of employment of the mother. Finally, I explore the dynamic of the effects and implications for the distinction between the "direct" mechanism and the "fertility mediator."

1.3.1 Impacts of Maternity Leave Allowance

First-stage estimates. I begin by exploring the effect of the kink on the amount of maternity leave benefits that women in my sample received. Figure 5 (Panel A) plots the empirical relationship between the average daily allowance and normalized pre-leave earnings. I use the daily allowance and not the total amount of benefits received to account for the fact that women below/above the kink could have different maternity leave duration, beyond the compulsory 60 days. One can see that the empirical relationship is very similar to the theoretical prediction in Figure 1 (Panel B), with clear evidence of a kink at the threshold set by the Social Security Administration. In the previous section, I showed that according to the benefit formula, the change in slope at the kink could reach

0.5 if the mother maxes out her leave entitlement of 90 days. But because not all mothers take the maximum days of maternity leave, I estimate the so-called "first stage" for the women in my sample, which corresponds to the denominator of my fuzzy RKD estimator in Equation (1.2). One can see from Table 3 (Panel A) that the estimated change in slope for marginal benefits at the kink is 0.45, quite close to the theoretical prediction.[24] Another way to represent the change at the kink is to plot the replacement rate instead, that is, the share of pre-leave earnings replaced by the allowance. Figure 5 (Panel B) plots this empirical replacement rate against the level of pre-leave earnings. One can see that it remains flat, around 74% below the kink, which implies that for each additional euro of pre-leave earnings, the allowance increases by 0.74 euros.[25] From the kink forward, the replacement rate declines linearly with earnings level. Taken together, the graphs and formal estimates prove that my empirical strategy can adequately capture the kink in the maternity leave scheme, which I can use to infer the effects of the benefit amount on my outcomes of interest.

For all outcomes, I start by showing a graph, which plots their mean values in 50-eurocentbins in the assignment variable. The graphs also display a linear trend on each side of the kink, suggesting possible changes in the slope of the relationship. I then report in Table 3 (Panel B) estimates of the treatment effect computed using local polynomial nonparametric regressions of order 1 (i.e., linear), heteroskedasticity-robust standard errors in parentheses, as well as bias-corrected confidence intervals proposed by Calonico *et al.* (2014).

Impact on career. I now turn to the main outcomes of interest and examine the impact of maternity leave allowance generosity on the career of first-time mothers. I create four outcomes, which track, over a period of five years, the probability of being employed (dummy) and quarterly earnings (euros), in both salaried employment and self-employment. I begin by showing the long-term implications of maternity leave allowance and therefore focus on values of those outcomes after five years. I will explore the dynamics in the next section.

In Figures 6 and 7, I show how the generosity of the maternity leave allowance affects the career of women. I begin by showing that their probability of remaining in the labor force, five years after the birth of their first child, seems to be completely unaffected by their maternity leave allowance. The trends on both sides of Figure 6 (Panel A) are

[24] One possible reason for the estimated coefficients to be slightly smaller than the theoretical prediction is that not all women max out their leave entitlement of 90 days. In fact, the average leave duration reported in Table 1 is 85.8 days. Since the cap on benefits applies only from the 31st day, a shorter duration would imply a smaller change in slope at the kink. This provides another argument for using a fuzzy RKD to precisely estimate the denominator in Equation (1.2).

[25] Again, this is very close to the theoretical prediction of elasticity of maternity leave allowance to the pre-leave earnings of 0.77 below the kink.

almost perfectly flat. However, one can observe changes to the type of employment they hold. Indeed, Figure 7 reveals that their probability to be salaried employee (Panel A) tends to decrease with the amount of benefits received during the maternity leave, up to the kink where the trend reverses. Interestingly, Panel B presents an exact inverse relationship, that is, as the amount of benefits increases the probability to become self-employed raises. This is also supported by Panels C and D, which reveal a change in the streams of earnings.

To confirm this graphical exploration, I estimate treatment effects using these outcomes and report the results in Panel B of Table 3. One may first notice that the generosity of maternity leave allowance does not affect the probability of being in the labor force five years after the first childbirth. The coefficient for the outcome "employed" is not statistically significant, and the robust CI includes zero. However, the results suggest changes in career path. In particular, I estimate that an additional 10 euros per day during maternity leave (i.e., 900 euros for a mother who uses her complete leave period) decreases the probability of being a salaried employee after five years by seven percentage points. At the same time, the probability of being self-employed increases by six percentage points. The similarity in the magnitude of the coefficients suggests that women leave salaried employment for self-employment as a result of receiving higher benefits.

One can observe similar effects for quarterly earnings with a change in the streams of income. Estimates reported in Table 3 (Panel B) reveal that for each additional euro in daily allowance, quarterly salaried earnings decreases by 56 euros five years after the first childbirth, while self-employed earnings increases by 115 euros. The net effect on total earnings is positive and amounts to 58 euros for each additional euro in daily maternity leave allowance.

Taken together, the results suggest that an increase in allowance during the maternity leave period has substantial effects on the probability of young women to become self-employed, with positive spillovers on their earnings. It might be interesting to compare my estimates to those found in the literature, albeit in different contexts. For instance, Flèche, Lepinteur and Powdthavee (2021) reveal that top lottery winners in the United Kingdom with gains of about 800 pounds are two percentage points more likely to become entrepreneur, an effect that is similar for men and women. Lim and Michelmore (2018) show that a 1,000-dollar increase in the Earned Income Tax Credit (EITC) in the United States raises the likelihood that married mothers will become self-employed by 5.7 percentage points.[26] In another context, Blanchflower and Oswald (1998) reveal

[26] Of course, the context of the EITC differs vastly, since it affects low-earnings individuals, while the women in my sample are high-earners. One should also notice that the effect might be driven by increased reporting of self-employment activities since the average tax rate decreases when the EITC increases.

that receiving 5,000 pounds in inheritance increases the probability that young women in their twenties will set up a business from 7 to 21 percentage points, a large effect that leads the authors to conclude that capital constraints might be particularly important for the youngest, precisely like the first-time mothers in my sample. Another example that financial incentives matter for the decision to become self-employed comes from Heim and Lurie (2010), who show that increasing the deductibility of health insurance premiums from 60% to 100% in the United States led to a 1.7 percentage point increase in self-employment.

My context, maternity leave, is also highly specific since the allowance comes with job protection, which might be important for women who would like to venture into entrepreneurship. As shown by Gottlieb et al. (2021), the option to return to a previous job is in itself a driver of entrepreneurship. One would therefore expect that the effect of reducing liquidity constraints is larger when combined with job-protected leave. In addition, as shown on Figure A1 in Appendix, the allowance is paid when mothers face a large drop in earnings. For some, the earnings loss starts even before childbirth, when they must stay away from work while pregnant. As such, one could argue that receiving a higher maternity leave allowance permits mothers to keep their savings intact and use them to start their own business.

Impact on fertility. I turn next to examining the implications of maternity leave allowance for subsequent fertility decisions. Figure 8 (Panel A) plots the total number of children five years after the birth of the first one. One can see that the number of children tends to increase with the amount of benefits received during the first leave, up to the kink where the trend reverses. One should note that the outcome is based on the number of children in the household where the mother lives. As such, it also accounts for new children from subsequent adoptions or family recomposition. To be certain that the observed changes reflect fertility decisions, I also provide a measure of new maternity leaves taken by the same mother (Panel B).[27]

Table 3 (Panel B) displays the effect on subsequent fertility decisions with the outcome "Number of children." One can see that the level of maternity leave allowance positively affects the number of children in the household after five years. The estimated coefficient suggests that for each additional 10 euros in daily allowance (i.e., 900 euros in total) the number of children increases by 0.09. As expected, the effect on subsequent maternity leave claims is also positive (Table A1 in Appendix), but the coefficient is slightly lower (0.08). Furthermore, I can rule out that the effect on the number of children in the house-

[27] This additional outcome will, however, underestimate the number of children from the same mother if she has exited the labor force after her first child, meaning she is no longer eligible to take paid maternity leave.

hold is driven by family recomposition, since I do not find any effect on the probability of remaining married (Table A8 in Appendix).[28]

One might want to compare my results to those from Raute (2019) for Germany, who also focuses on high-earning women, although with a different estimation strategy using a difference-in-differences design around a reform introduced in 2007. She estimates that a 1,000-euro increase in total benefits raises the probability that a woman will give birth in each of the five post-reform years by 0.783 births per 1,000 women, that is a 2.1% increase per year (10.5% after five years) compared to the pre-reform situation (Raute, 2019). In comparison, I find that an increase in maternity leave allowance of 10 euros per day (i.e., 900 euros in total) raises the number of children within five years by 0.09, that is, a 4.7% increase compared to the mean reported in Table 3 (1.92). My estimates in the Belgian context are perfectly in line with those from Raute (2019) for Germany.

Impact on maternity leave duration. I study how the amount of maternity leave benefit might affect the time mothers decide to spend on maternity leave. Figure A5 in Appendix shows a slightly increasing relationship, which seems to flatten after the kink. In Table A1 in Appendix, I report the effect on the duration measured in days, as well as taking its natural logarithm.[29] The estimated elasticity is equal to 0.128, which is quite small compared to other social insurance programs. For example, in the case of unemployment insurance, the elasticities estimated in Western Europe usually range from 0.3 to two (Card et al., 2015a). In addition, the bias-corrected confidence intervals include zero, which might suggest that the results are not robust to the inclusion of higher-order polynomials and larger bandwidth definitions. I believe I can therefore rule out that the effects on the decision to become self-employed or to have more children in the future are driven by the fact that mothers who receive higher benefits spend more time on maternity leave.

Impact on co-parent. I test for potential spillover effects on the career of the co-parent. Similar to what I have previously done, Table A2 in Appendix reports RKD estimates for the impact of maternity leave allowance on the co-parent's outcomes five years after the birth of the couple's first child. I find no evidence that higher maternity leave allowance affects the probability that the co-parent takes paternity leave. I also do not find any effect on the co-parent's career in the long run. Their earnings are largely unaffected by the amount of maternity leave allowance received by their partner.

[28] I have also tested for an effect on subsequent adoptions, since I have data on parental leave taken for adopted children. I do not find any effect on the probability of taking a leave for an adopted child over the following five years (results available upon request).

[29] For the latter, I also use the log of benefits for the first stage.

1.3.2 Quantile Effects on Earnings

In the previous subsection, I revealed that mothers who receive a higher maternity leave allowance are more likely to become self-employed and have higher earnings five years after the birth of their first child. However, the positive effect on average earnings may conceal a variety of situations, with some women having created a very successful new business, what Rosen (1981) calls "superstar" entrepreneurs, while others have lower earnings than where they were salaried employee. For this reason, I follow the seminal work of Hamilton (2000) and use quantile regressions[30] to explore the heterogeneity of the effects along the earnings distribution.

Table 4 reports the effects of the maternity leave allowance generosity on total quarterly earnings at the 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$, 90$^{\text{th}}$ and 95$^{\text{th}}$ percentiles, as well as bootstrapped standard errors (after 1,000 replications). The estimate for the 0.75 quantile is smaller in magnitude (11.23 euros) than the average effect reported in Table 3 (58.42 euros) and is not statistically different from zero, while the effects for the 0.25 and 0.50 quantiles are even negative. At the upper end of the distribution, the effects on earnings are positive and large (119.42 euros for the 0.95 quantile), which suggests that a small fraction of female entrepreneurs have substantial earnings, as predicted by the superstar model of Rosen (1981). One may conclude from this new analysis that the positive effects on earnings that were emphasized in the previous subsection are only found in the right tail of the distribution, but they do not characterize the majority of mothers who, in fact, have a null or negative effect on their earnings.

One may therefore wonder what could push women to remain in self-employment five years after the birth of their first child even though the effect on their earnings is not positive. Hamilton (2000) suggests that entrepreneurs might find non-pecuniary benefits that compensate for the lower earnings. In the next subsection, I explore the transition to self-employment across sectors characterized by different working conditions, which could help understanding the non-pecuniary benefits of self-employment for mothers.

1.3.3 Heterogeneous Effects by Sector of Employment

In this subsection, I want to test whether the workplace's characteristics might also play a role in the decision of mothers to become self-employed. I am particularly interested in the temporal flexibility offered by jobs in different industries. According to Goldin (2014), the persistent gender gap in earnings can be attributed, at least in parts, to the fact that

[30] I use the package "sivqr" developed by Kaplan (2022) for Stata.

firms in some sectors disproportionately reward individuals who work "atypical" work schedules, that is at night or during weekends. For young parents, this might translate into "time pressure" that can be difficult to conjugate with household responsibilities.

I use data from the European Working Conditions Survey (EWCS) to observe how frequent are these atypical work schedules among Belgian industries where the mothers in my sample work. I use the data from the 2010 survey to classify NACE sectors according to the share of respondents who work late in the evening, during the night or on weekends. I find that atypical work schedules are dominant in three particular sectors: "Retail trade, hotels and restaurants," "Transport, storage and communication," "Health and social work." Unsurprisingly, these are also the three sectors that score worst on work-life balance (Eurofound, 2014). I believe that becoming self-employed might be a way for mothers in these sectors to regain control of their schedule. One might think, for instance, of a doctor in a hospital that would open her own private practice upon entering motherhood to stop having to do night shifts.

I take advantage of my matched employer-employee data to observe the sector of the firm where the mother was working one year before giving birth for the first time.[31] Then, I rank sectors according to the share of respondents in the Belgian sample of the EWCS who answer that they work at night (question 32), in the evening after 6pm (question 33) or on weekends (questions 34 and 35). I create a dummy that takes on a value 1 for sectors where more than 50% of survey respondents declare working atypical hours.[32]

Table 5 reports the results for the two sub-samples of mothers who use to work in sectors characterized, or not, by atypical work schedules. The last two columns also present z-tests to check whether the difference in the estimated coefficients appears to be statistically significant. One can see that the elasticity between maternity leave allowance and the probability to become self-employed is only positive for those mothers who use to work in sectors with night or weekend shifts. I believe that this is suggestive evidence that mothers who create their own business might also be seeking non-pecuniary benefits, such as work schedules that better conjugate family demands.

[31] I create 8 sectors from the NACE classification Rev. 1.1 for mothers of children born until 2008 and from the NACE classification Rev.2 for those born after.

[32] Alternatively, I rank sectors based on the share of respondents who answer that their working hours do not fit well with their family commitments (question 41 of EWCS 2010 questionnaire) and obtain similar results, available upon request.

1.3.4 Heterogeneous Effects by Wage Gap between Parents

In this subsection, I explore how the effects on the transition to self-employment vary according to the pre-birth characteristics of the parents. I am particularly interested in the wage differential between parents prior to having a first child. Following Bertrand, Kamenica, and Pan (2015), I compute the relative earnings within the household, that is, the mother's earnings divided by the total household's earnings. I then distinguish between two cases: (1) when the relative earnings is lower than 0.5, which implies that the mother contributes less than her partner to household earnings, and (2) when the relative earnings is equal or higher than 0.5, which means that the mother contributes equally or more compared to her partner. The two groups represent, respectively, 45% and 55% of my sample of mothers with earnings in a 22-euro bandwidth around the kink.[33]

Table 6 reports the treatment effects estimated on both sub-samples. When looking at the effects on the career path, one notices that women who used to earn less than their partner before entering motherhood are two times more likely to leave salaried employment for self-employment. The z-tests confirm that the difference between the two groups is statistically significant.

I believe that the marked differences observed between the two sub-samples could stem from at least two self-reinforcing mechanisms. First, women might find it less risky to leave salaried employment for self-employment when their partner is the main breadwinner in the household. Second, mothers who contribute less to household income are often found to assume a larger share of childcare-related activities than fathers (Bittmann, 2015; Pailhé & Solaz, 2008). Switching to self-employment might provide a way for those mothers to reconcile increased family responsibilities and the pursuit of their career. In fact, the literature in management and sociology demonstrates that the balance between work and family demands are among the main motivational factors cited by women who become self-employed (Budig, 2006; Gangl & Ziefle, 2009; Kirkwood, 2009). Self-employment is indeed reported as providing greater flexibility when it comes to mothers' time (Boden, 1999; Georgellis & Wall, 2005). This is compatible with Time Use Survey data, which shows that Belgian women who are self-employed dedicate more time to both paid work (33 minutes per day) and family care (14 minutes per day) than salaried women (statistics are reported in Table A3 in Appendix).

I conclude from the two previous heterogeneity analyses that women who receive higher maternity leave allowance are more likely to transition to self-employment when they are not the main breadwinner in the household and and when they use to work in sectors

[33] Results reported in Table 2 suggest that there is no discontinuity around the kink in the share of household earnings earned by the mother, measured four quarters prior to having a first child.

offering poor work-family balance. As such, it appears that both the households' and workplaces' characteristics play a role in the decision to become self-employed.

1.3.5 Direct vs. Indirect Mechanisms

In this last subsection, I conduct a number of complementary analyses to understand the mechanisms that could generate the results on both self-employment and fertility. I consider two potential channels. First, I argue that a more generous maternity leave allowance might help lift young women's liquidity constraints. I call this the "*direct effect*" of maternity leave allowance on self-employment and believe it could reflect the fact that women with entrepreneurial ideas can use this additional liquidity to start their business. At the same time, because self-employment offers greater flexibility, it is easier for those women to have more children. This is compatible with Wellington's model, (2006) which demonstrates that self-employment offers a better balance between family and career.

Second, I consider the possibility that because mothers who receive a higher allowance during their maternity leave have a lower opportunity cost of childbearing, they could decide to have more children in the future. I call this mechanism the *"fertility mediator,"* whereby mothers who receive more generous benefits are more likely to desire more children in the future, and as a consequence, they switch to self-employment to accommodate their desired fertility with the pursuit of their career. It is therefore an indirect effect on self-employment. The results discussed in the previous subsection, however, seem to go against this channel. Indeed, if the changes in desired fertility were a mediator, one would expect to observe other types of career switches, such as changes in labor supply at extensive margin suggested by the child penalty literature (e.g., Kleven, Landais, & Søgaard, 2019). And yet as discussed in the previous subsections, I find no effects on the probability of being in the labor force of mothers in the long-run. The switch to self-employment is the only career change reflected in my data.

To further distinguish between the two channels, I also explore the dynamics of the effects. To do so, I plot in Figure 9 the coefficients for the treatment effects from separate regressions in each quarter after childbirth and up to five years. Panel C reveals that the total number of children is only statistically different from zero in quarter eight after the first childbirth and continues to increase up to quarter twelve. Interestingly, the effects are remarkably stable from the third year onwards and never converge back to zero, which suggests that it is not due to increased birth spacing. When looking at career dynamics, Figure 9 (Panel A) reveals that the probability of being a salaried employee already becomes negative around quarter three after childbirth and continues to decline until quarter seven. At the same time, the probability of becoming self-employed (Panel

B) increases sharply. Thus, it appears that most of the women who transition from salaried to self-employment do so during the two years that follow their first childbirth, and for the majority, before they decide to have more children. These results indicate that the effect on self-employment precedes the effect on fertility, which offers additional suggestive evidence that the direct mechanism might be at play.

Taken together, these supplementary analyses point toward a direct effect of maternity leave benefits on self-employment, with positive spillovers on fertility decisions.

1.4 Robustness Checks

In the following section, I provide various tests for the robustness of the RKD estimates presented above. I explore the sensitivity of the results to the choice of the polynomial order and bandwidth. Then, I discuss the functional dependence between the assignment variable and the outcomes, before providing tests that corroborate the strength of my findings.

1.4.1 Sensitivity Analysis

Functional form. I begin by analyzing the sensitivity of the results to the choice of the functional form. I start with a graphical exploration and provide in Appendix plots for the main outcomes of interest using both linear and quadratic functions of the assignment variable. One can observe that for both fertility (Figure A8) and employment outcomes (Figure A7), the trends estimated with different polynomial orders suggest similar discontinuities at the kink, and many times the two lines are almost perfectly aligned (the dashed line is for the linear specification, while the solid line is for the quadratic one). To confirm this, I provide estimates for the treatment effects using both linear and quadratic specifications in Table A4 in Appendix.[34] First, one may notice that both specifications report comparable first-stage estimates (first columns of "linear" and "quadratic" panels), even though the Aikake Information Criterion (AIC) suggests that the linear specification almost always dominates the quadratic one (last panel titled "polynomial minimizing

[34]Contrary to the previous specifications using local nonparametric methods for estimation, here I use parametric regressions to report conventional goodness of fit measures. In particular, I show the Aikake Information Criterion (AIC) in square brackets. The last columns of Table A4 in Appendix show which specification (linear or quadratic) minimizes this information criterion. The columns "first stage" and "second stage" are reduced form estimates for the change in slope of the maternity leave benefit amount and the outcomes, respectively. The column "treatment effect" reports coefficients from two-stage least squares estimations, where the benefit amount is instrumented with the interaction between a dummy for being above the kink and the polynomial in the assignment variable.

AIC"). This is not surprising given that the benefit schedule is indeed a linear function. When it comes to the treatment effects (third column of "linear" and "quadratic" panels), in all cases, the sign is the same in both specifications and the coefficients are qualitatively similar. One may note, however, that when controlling for a quadratic polynomial, the standard errors increase drastically. In fact, the AIC advises on using a linear specification for the second-stage estimates in most cases. Taken together, these results confirm the claim of Gelman and Imbens (2019) that controlling for high-order polynomials in regression discontinuity analysis might lead to poor coverage of confidence intervals. For all these reasons, my preferred specification is the linear case, as in most studies using the RKD and surveyed in Ganong and Jäger (2018).

Choice of bandwidth. I now turn to testing the sensitivity of my results to the choice of bandwidth. As mentioned in Subsection 2.2, only a few papers offer guidance on optimal bandwidth choice in the RKD case (Calonico et al., 2014; Card et al., 2015b). The main similarity with standard regression discontinuity design is the tradeoff between bias and variance. Larger bandwidths will likely be more biased, but at the same time, the RKD has been reported to fare poorly with small samples (Landais, 2015). In subsection 2.2, I explain that I use the data-driven bandwidth selector developed by Calonico *et al.* (2014)–CCT selector from now on—as a primary guide. The CCT selector is the only bandwidth selector explicitly designed for the fuzzy RKD and builds on earlier work by Imbens and Kalyanaraman (2012), who proposed an algorithm to compute the MSE-optimal bandwidth. In particular, their procedure involves a regularization term, which reflects variance in the bias estimation and guards against the selection of large bandwidths.

Table A5 in Appendix reports treatment effects for my outcomes of interest using their optimal bandwidth picked by the CCT selector, as well as four different bandwidths in the 15 to 30-euro range. When compared with my baseline estimates using a 22-euro bandwidth (Table 3), one immediately notices that these new results are highly similar. Regarding labor market outcomes, the effects are also quite similar across bandwidths. The coefficient on salaried employment is always negative and ranges from -0.004 to -0.006, which accords with my baseline estimate of -0.007. The coefficient on self-employment is always positive and ranges from 0.003 to 0.006, again highly similar to my baseline result of 0.006. The estimates on earnings also appear relatively stable across the different bandwidth specifications, with standard errors, as expected, that become larger as the sample size decreases. When it comes to the effects on subsequent fertility, they are always positive across the different bandwidth choices. The coefficient on the number of children ranges from 0.005 to 0.01 and the estimate using the data-driven CCT bandwidth is close in magnitude to my baseline estimate.

In order to visualize the relationship between my estimates and bandwidth choice, I also plot the coefficients for the linear case against all possible bandwidths in one-euro increments of normalized pre-leave daily earnings from 10 to 30 euros. Two vertical lines materialize the 22-euro bandwidth (dashed line) used for the baseline estimates and the data-driven CCT bandwidth (dotted line). One can observe that the effects are relatively stable beyond the 15- euro bandwidth for both fertility and employment outcomes (Figures A10 and A9, respectively). Taken together, these tests confirm that the results are consistent across bandwidth choices, but also that the common 22-euro bandwidth offers highly similar results to the ones based on the CCT selector.

1.4.2 Functional Dependence

In Section 2, I explained that the RKD relies on the assumption that the underlying relationship between the assignment variable and outcomes (in the absence of a kink) should be smooth. A legitimate concern could be that the effects captured by the RKD result from non-linearities in this relationship. The graphical evidence I provided earlier seems to exclude this possibility. Indeed, subsequent fertility increases with the benefit amount up to the kink. It is unlikely that this captures a functional dependence with the assignment variable since most studies instead find an inverse relationship between earnings and fertility (see e.g. D. J. Anderson, Binder, & Krause, 2002; Bertrand, Goldin, & Katz, 2010; Hotchkiss, Pitts, & Walker, 2017; Raute, 2019). Nevertheless, to address this concern, I provide several tests that have become standard in the RKD literature.

I begin by including controls in my regressions to account for possible non-linearities. Ando (2017) suggests that controlling for relevant pre-determined covariates should reinforce the credibility of the RKD. I control for the mother's age and region of residence,[35], as well as her partner's earnings, at the time of the birth of the first child. I have chosen those covariates because they are likely to be correlated with both the assignment variable and the outcomes of interest. Indeed, older women, living in Flanders or those with a partner with high earnings, might be less financially constrained to start a business. These variables should therefore help capture potential non-linearities. Table A6 in Appendix shows that when controlling separately and then adding all covariates, the results remain highly similar to the baseline estimates (Table 3).

Finally, to assess the sensitivity of my results to non-linearities in the relationship between the assignment variable and my outcomes of interest, I perform a series of permutation tests, as proposed by Ganong and Jäger (2018). The idea is to estimate RKD

[35] I create an indicator variable that takes on the value one if the mother was living in Flanders, which has the highest GDP per capita in Belgium (after correcting for commuters).

models using placebo kinks at various points along the distribution of the assignment variable. I estimate 300 placebo RKD models around the true kink point, using a 22-euro bandwidth surrounding each placebo kink point. The placebo kinks are situated at a distance of -100 to 50 euros from the true kink point, which covers about 95% of the earnings distribution. I report results for fertility and employment outcomes, all five years after the first childbirth. One should note that the permutation tests are estimated as reduced form models, which correspond to the slope change measured in the numerator of Equation (1.2). As such, the placebo kink coefficients are of the opposite sign of those reported so far, which were scaled by the negative coefficients of the denominator.

The figures plot RKD estimates along placebo kink values, specified in terms of distance from the true kink point (Figures A12 and A11 in Appendix). One can see that for all outcomes, the coefficients estimated at the true kink point (i.e., zero on the horizontal axis) are much larger than those at placebo kinks. Taken together, the results from these multiple robustness checks strongly support the validity of the RKD in my context.

1.4.3 Placebo Groups

In a final exercise, I use two new samples of individuals facing fertility and career decisions similar to my sample of salaried mothers, but who were not affected by the kink in the maternity leave benefit schedule. The first group is composed of fathers who did not take any leave after the birth of their child[36] and therefore did not receive any benefits. The second group is composed of self-employed mothers who by law receive a flat benefit. Intuitively, because these two groups are unaffected by the kink in the benefit schedule, I can use them to test for the absence of an underlying relationship between pre-parenthood earnings and fertility and career decisions in the vicinity of the threshold set by social security.

Starting with the sample of fathers who did not go on leave, I explore their probability of being self-employed five years after the birth of their child. This group is perfectly suitable for the exercise since they did not receive any benefits from the Social Security Administration. They would therefore help capture any underlying relationship between earnings and the probability of being self-employed. One could fear, for example, that there are tax incentives to become self-employed when earnings are above a certain level, or that there are minimum capital requirements to start a business that coincide with the kink. To rule out that this drives the results for mothers in my main sample, I use fathers with similar earnings and compare their probability of being self-employed in the vicinity

[36]Fathers in Belgium were entitled to take two weeks of paternity leave from July 2002 until the end of the period covered by my sample.

of a placebo kink similar to the one used in my main estimates. Figure A13 (Panel B) in Appendix shows that the probability of being self-employed is almost perfectly flat, around 10 percent for all fathers who did not go on leave and have earnings in the vicinity of the placebo kink.

Moving to the second sample of self-employed mothers at first birth, I calculate their earnings before going on leave for the first time and normalize them to a placebo kink that corresponds to the threshold set by the Social Security Administration for salaried mothers. One can see from Panel A in Figure A14 in Appendix that the earnings distribution for self-employed women is highly similar to the one in Figure A4 for the sample of salaried women. The placebo kink is also located at a similar point in the earnings distribution, around the 90th percentile. As mentioned before, self-employed women receive a flat benefit when they go on maternity leave. This is illustrated by Panel B in Figure A14. Self-employed women are therefore a perfect placebo group to observe the direct relationship between pre-leave earnings and subsequent fertility in the absence of variation in financial incentives. One can see from Panel C that the number of children they end up having is completely unrelated to the pre-leave earnings of self-employed women in the vicinity of the placebo kink.

Altogether, these results reinforce the credibility of my main findings that the career and fertility of women are affected by the kink in the maternity leave allowance schedule and that the effects do not stem from underlying relationships around the kink.

1.5 Simulated Effects of Moving the Benefit Threshold

In this last section, I draw policy implications from the estimated link between maternity leave allowance generosity and career and fertility decisions of high-earning women. As explained in Section 1.2.2, the regression kink design allows us to estimate a local effect that is mostly informative in the vicinity of the kink. The behavioral reaction of women around the kink might, in fact, be of interest to policy makers who wish to reduce the gender gap in entrepreneurship. In this regard, the location of the benefit threshold might have important policy implications.

In what follows, I determine the budgetary consequences of moving the benefit threshold, as well as the consequences for the probability of high-earning women to become self-employed, as well as the spillover effects on their fertility decisions. Figure A2 in Appendix suggests that this might be a policy relevant question, since the location of the threshold is highly heterogeneous across countries. For example, both France and the Netherlands offer a 100-percent replacement rate for women below the benefit threshold, but the threshold is placed much lower in the Netherlands, affecting women who already earn 150 percent of the national average, while most French women are unaffected.

In Belgium, the benefit threshold is located around the 90th percentile of the earnings distribution. I can simulate the budgetary consequences of moving this threshold to the 99th percentile as shown on Figure A15 in Appendix and the impact for mothers who, as a result, would receive a higher allowance. Computing the budgetary consequences is rather straightforward. One must know only the pre-leave earnings of those women in the upper quintile of the distribution and estimate the corresponding benefits. Table A7 in Appendix shows the daily wage of women at each percentile between the 80th and the 99th, as well as the corresponding number of women based on administrative information from the Social Security Administration.[37] The panel "Social security threshold" reports the estimated allowance based on the formula described in Section 1.1 if the threshold is set at the level decided by the Social Security Administration in 2007 (about 110 euros), and the corresponding replacement rate. The panel "Simulated threshold at 99 pc." shows the estimated allowance if the benefit threshold was moved to the 99th percentile. One can observe that the corresponding replacement rate would be 77 percent for all women whose daily pre-leave earnings were equal or below 175 euros. For Belgian social security, moving up the benefit threshold would suppose a higher cost, which corresponds to the difference in daily allowance times 90 days for women above the current kink. From this simulation, one can infer that the total cost for social security would be 7,256,451 euros, that is, two percent of total maternity leave payments (INAMI, 2007).

When using the estimates from Table 3, one can infer that raising the earnings threshold to the 99th percentile would bring an additional 484 mothers into self-employment. As a result, the probability for women in the upper decile of the earnings distribution to become self-employed within five years after childbirth would increase from 5.67% to 13%. This simulated reform would almost bridge the gap with the 16% of fathers in my sample who are self-employed. At the same time, the reform would cause the fertility level of mothers in the upper decile to modestly increase from 1.91 to 2.01, as they would have an additional 726 children.

This simulation exercise highlights the complexity of designing maternity leave policies. My results show that both the replacement rate and the location of the benefit threshold (if any) have sizable budgetary consequences, but most important, that they affect the

[37] In 2007, social security reports having paid 6,953,358 days of maternity leave (INAMI, 2007). Since the leave entitlement is 90 days, one can infer the approximate number of women, which is equal to 77,300.

career and fertility of high-earning women. My estimates suggest that raising the benefit threshold to the 99$^{\text{th}}$ percentile in Belgium would contribute to reducing the gender gap in entrepreneurship among high-earners, while having positive spillover effects on fertility.

1.6 Conclusion

OECD statistics show that only one in three entrepreneurs in the European Union is a woman, and only two in five in the United States.[38] However, international surveys reveal that many women would start a business if they had the opportunity,[39] suggesting that the gender gap in entrepreneurship is not primarily driven by preferences or aspirations. In this paper, I study the liquidity constraints that potential female entrepreneurs may face. In particular, I show that providing a higher maternity leave allowance after the birth of their first child, at a time when their finances are strained, increases the probability of young women becoming self-employed thereafter. Further analysis also contributes to the understanding of the drivers of female entrepreneurship. My results reveal that only a few women who make the move from salaried to self-employment achieve higher earnings, but for the majority the impact is negligible or even slightly negative. I argue that female entrepreneurs may trade lower earnings for the non-pecuniary benefits of business ownership. I find that those working in sectors with atypical work schedule (i.e. night or weekend shifts) are more likely to become self-employment, which I interpret as an attempt to find more balance between work and family.

In this paper, I examine how the generosity of maternity leave allowance affects first-time mothers' career trajectory and subsequent fertility decisions. My empirical strategy relies on a Regression Kink Design. I exploit a discontinuity in the maternity leave benefit schedule to estimate the causal effects of the allowance generosity on women's outcomes up to five years. Because of the kink in the benefit formula, women with pre-leave earnings above the maximum threshold set by the Social Security Administration face a lower replacement rate than those below.

Comparing first-time mothers within a small bandwidth around the kink, I find that those who received higher benefits are more likely to leave salaried employment for self-employment. In particular, my estimates show that receiving an additional 900 euros in maternity leave benefits increases the probability of setting up a business within the next five years by six percentage points. Heterogeneity analyses reveal that the transition to

[38] OCDE (2022), "Labour Force Statistics : Summary tables." OECD Employment and Labour Market Statistics (database).

[39] According to the International Social Survey Programme on Work Orientations III, the share of women who would choose to be self-employed was 55% in the United States (ISSP, 2005).

self-employment is higher among women who earn less than their partner prior to having a first child. Subsequently, I explore the consequences for their fertility decisions and establish that the benefit amount positively affects subsequent fertility.

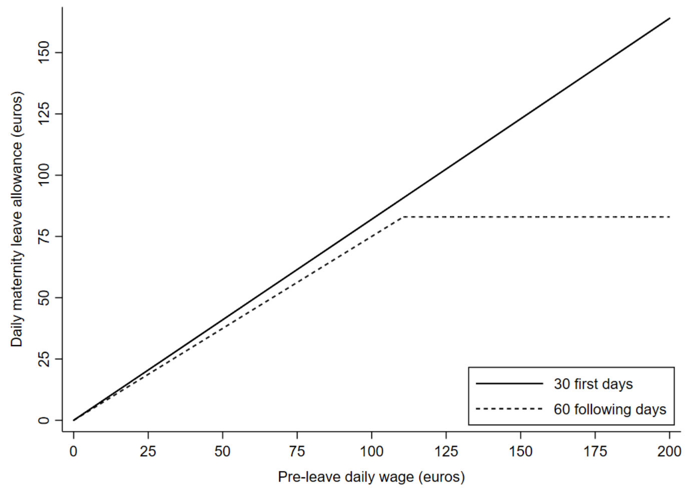
I consider two mechanisms that could generate these results. The "direct mechanism" suggests that receiving higher maternity leave allowance helps financially constrained mothers switch to self-employment. At the same time, because self-employment allows for a better balance between the demands of family and work, those mothers are more likely to have a higher number of children. I call the second mechanism, the "fertility mediator," whereby mothers who receive higher maternity leave benefits are more likely to desire more children because they have a lower opportunity cost of childbearing. To accommodate this higher desired fertility and the pursuit of their career, they move to self-employment, which is considered more family-friendly.

Complementary analyses seem to support the prevalence of the "direct mechanism" for at least two reasons. First, the dynamics of the effects reveal that mothers switch to self-employment in the quarters immediately following the birth of their first child, that is, before they decide to have another child. Second, I do not observe any change in labor supply at the extensive margin, beside the change in career path that leads more women to become self-employed. I posit that if fertility decisions were indeed the mediator, I would have observed other changes in labor force participation as suggested by the child penalty literature.

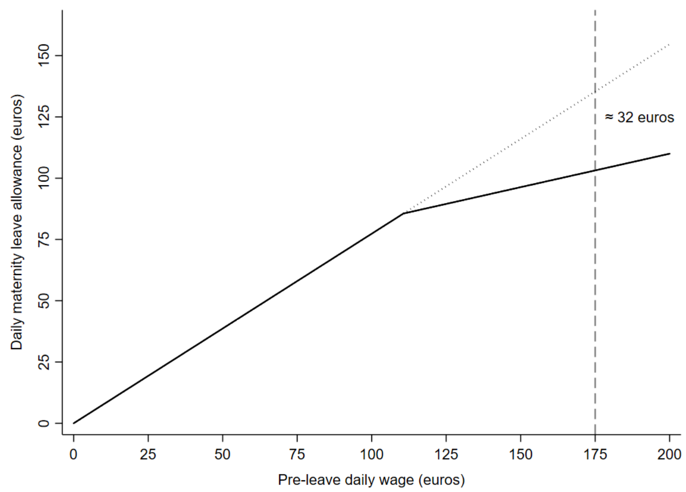
These findings should be of broad interest outside Belgium. First, because the features of the Belgian maternity leave system are similar to the standards set by the International Labor Organization, offering plausibly good external validity. Second, because the majority of OECD countries have placed caps on maternity leave benefits, paving the way for similar studies using a Regression Kink Design in other countries.

Figure 1: Maternity leave allowance as a function of pre-leave earnings (simulation)

Panel A: Daily allowance first 30 days / remaining 60 days



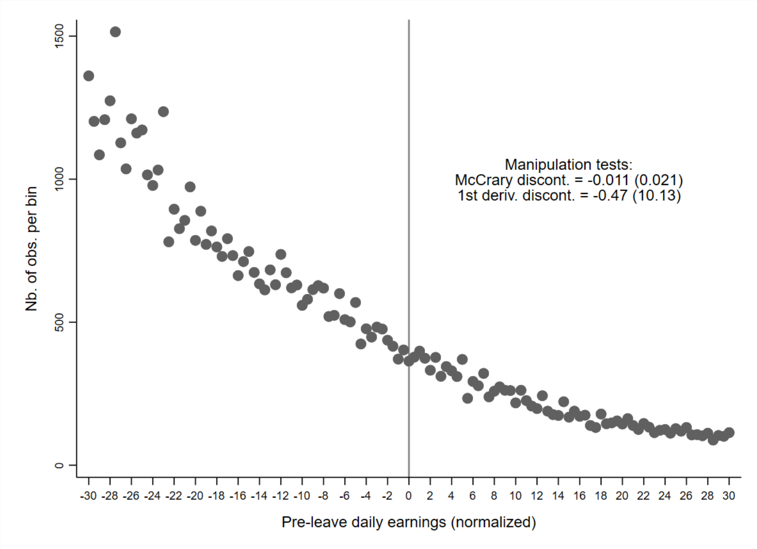
Panel B: Average daily allowance for 90 days



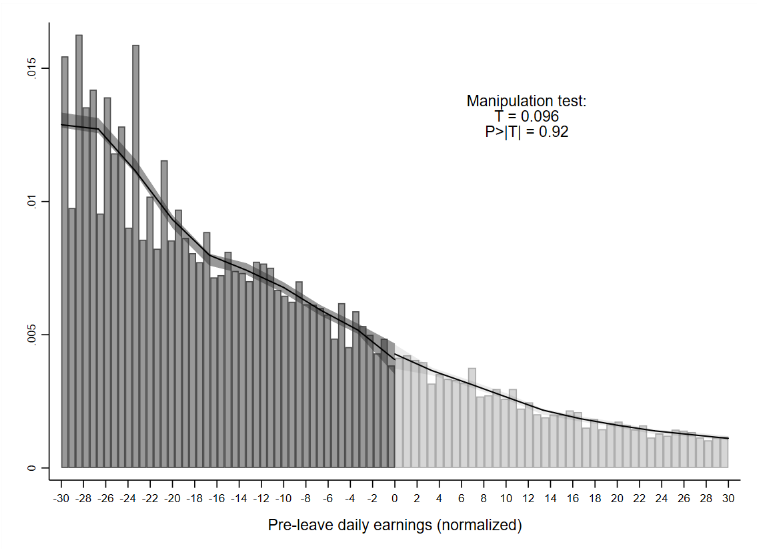
Notes: Both panels simulate the benefit schedule based on the rules set by the social security administration, that is a replacement rate of 82% of pre-leave earnings during the "30 first days" and 75% for the "60 following days" of the maternity leave. Panel A plots the daily allowance paid during the first 30 days (solid line), which is a linear function of pre-leave earning. Panel A also shows the daily allowance paid during the remaining 60 days (dashed line), which is capped. The earnings threshold on both panels is set at 110.655 euros, which corresponds to the one in place on January 1st 2007 in the middle of my sample window. Panel B illustrates the kinked function for a total leave duration of 90 days (i.e. when a mother maxes out her maternity leave entitlement). The average daily allowance is based on Equation (1). The dashed line on Panel B illustrates the situation of a mother with pre-leave earnings of 175 euros, who receives an average daily allowance of about 103 euros. In the absence of the kink (i.e. if the benefit schedule was linear), she would receive a daily allowance of about 135 euros, that is 32 euros more per day.

Figure 2: Manipulation tests

Panel A: Frequency distribution of assignment variable

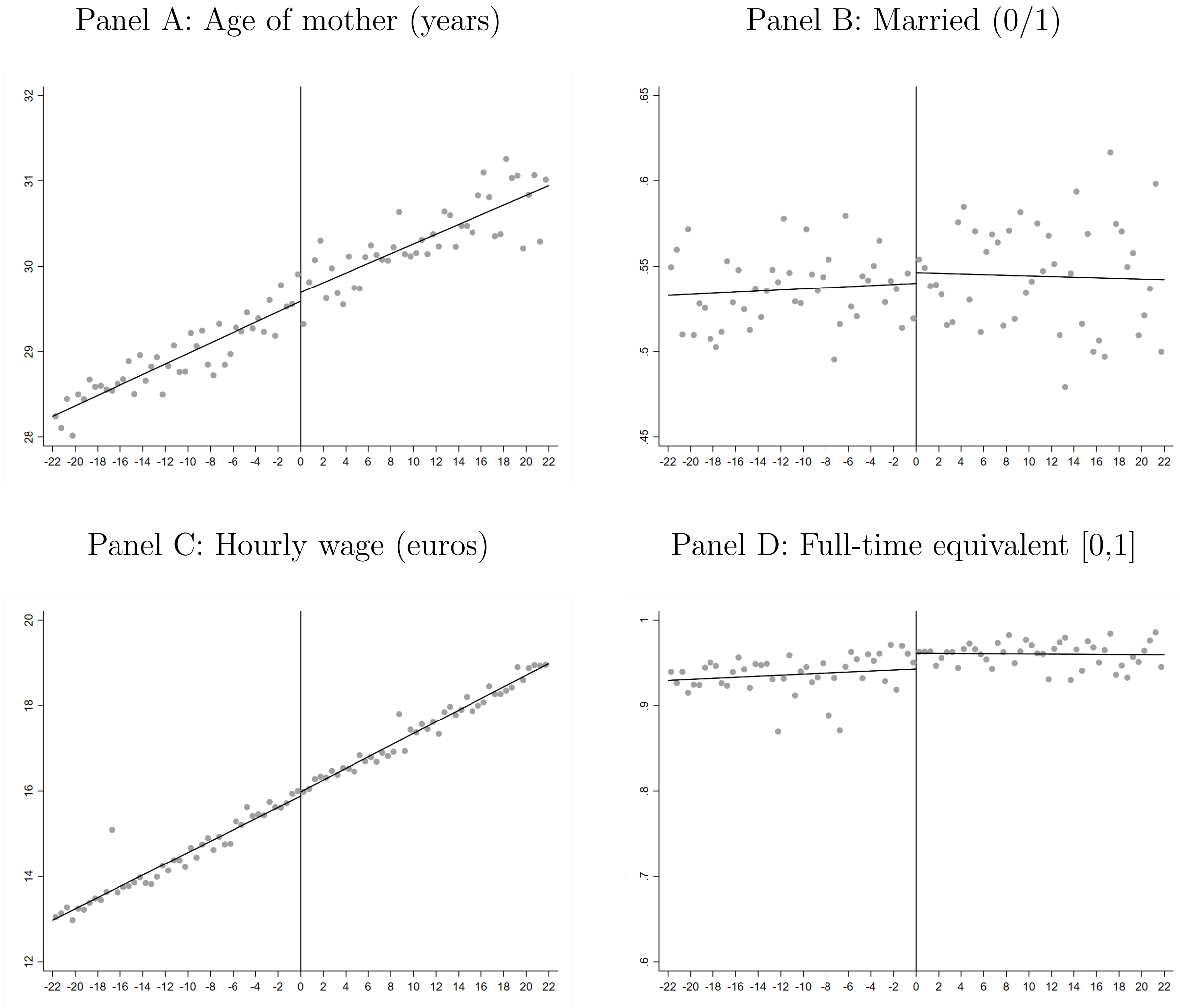


Panel B: Manipulation testing using local-polynomial density estimation



Notes: The graphs assess the validity of the RDK assumption that mothers did not engage in sorting around the kink point. Panel A shows the frequency distribution of pre-leave earnings in 50 euro cents bins, over a 30 euros bandwidth. The graph also displays two manipulation tests: the standard McCrary (McCrary, 2008) test that checks for a "jump" in the p.d.f. of the assignment variable, and the extension proposed by Card et al. (2015b) to test that the first derivative of the p.d.f. is also continuous at the kink. I report the coefficients for both tests, as well as the corresponding standard errors in parentheses. Panel B displays the probability density function of the assignment variable around the kink, but this time estimated using local-polynomials, as proposed by Cattaneo et al., (2020). They also suggest a novel manipulation test, which is reported on the graph with the corresponding p-value. The graphical evidences from both panels, as well as the formal tests, all suggest that the density of the pre-leave earnings around the kink point seems smooth and therefore that one cannot detect manipulation.

Figure 3: Mother's outcomes 4 quarters before the birth of her first child

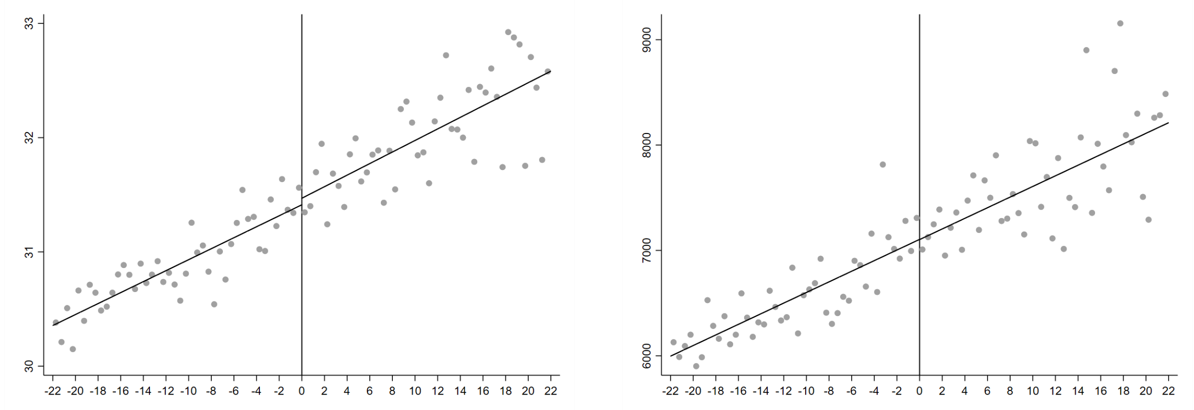


Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure 4: Co-parent's outcomes 4 quarters before the birth of the child

Panel A: Age of co-parent (years)

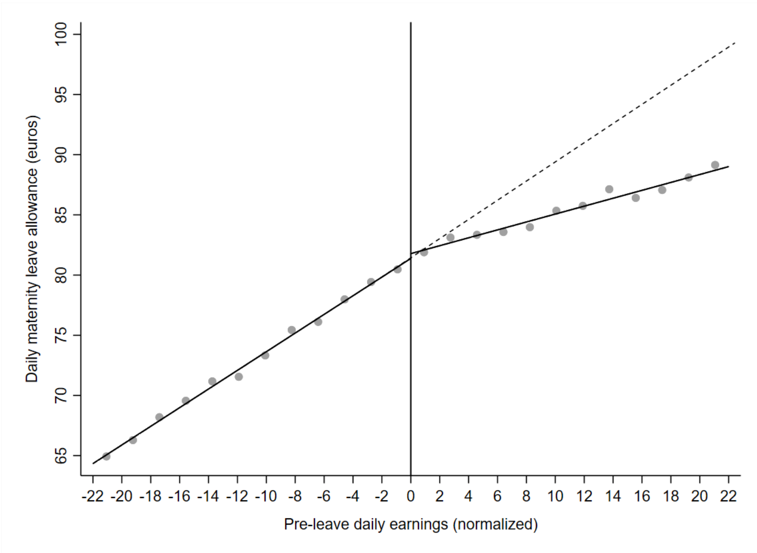
Panel B: Quarterly earnings of co-parent (euros)



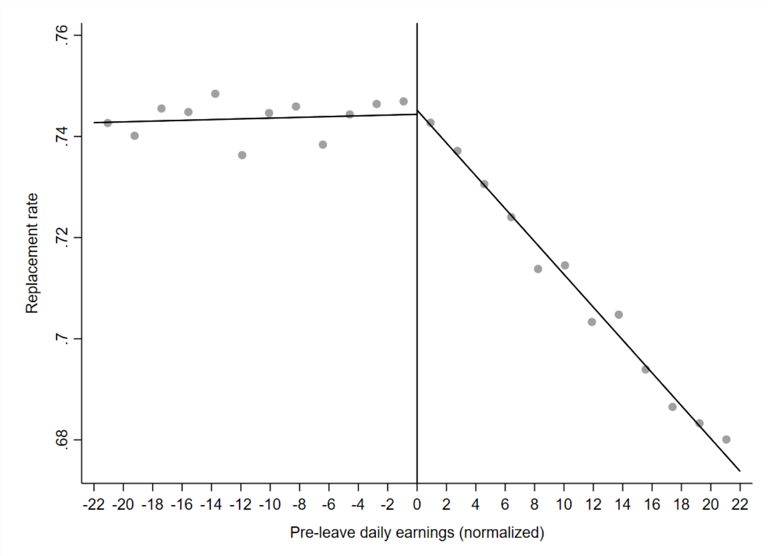
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure 5: Maternity leave allowance as a function of pre-leave earnings

Panel A: Daily allowance (first stage estimates)



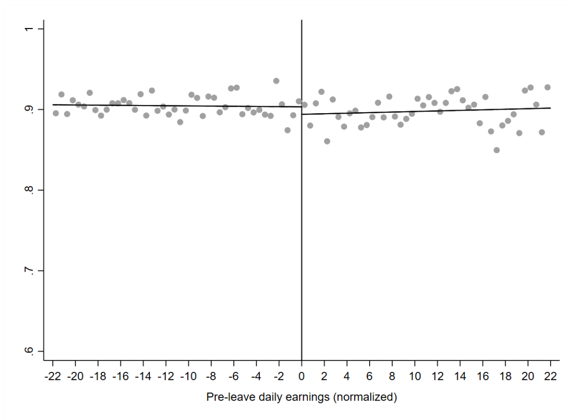
Panel B: Replacement rate



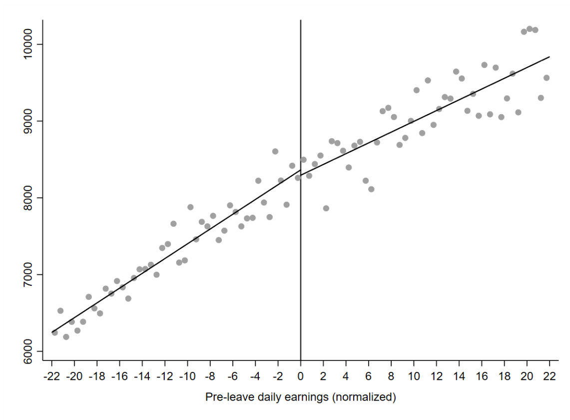
Notes: The first graph (Panel A) shows the empirical relationship between the daily maternity leave allowance and the pre-leave earnings of women within the 22 euros bandwidth around the kink. It corresponds to the "first stage" estimate, that is the change in slopes for the denominator of Equation (1.2). The second graph (Panel B) shows the the empirical relationship with the replacement rate, that is the percentage of pre-leave earnings replaced by the allowance. In both graphs, the horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in 12 bins.

Figure 6: Mother's outcomes 5 years after the birth of her first child

Panel A: Employed (0/1)



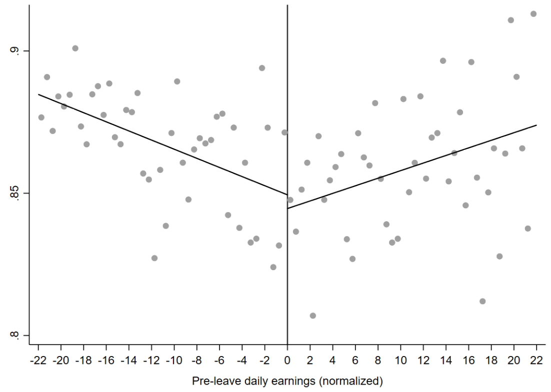
Panel B: Total earnings (euros)



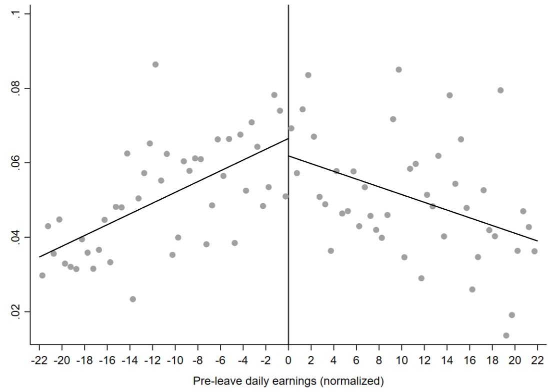
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure 7: Mother's outcomes 5 years after the birth of her first child

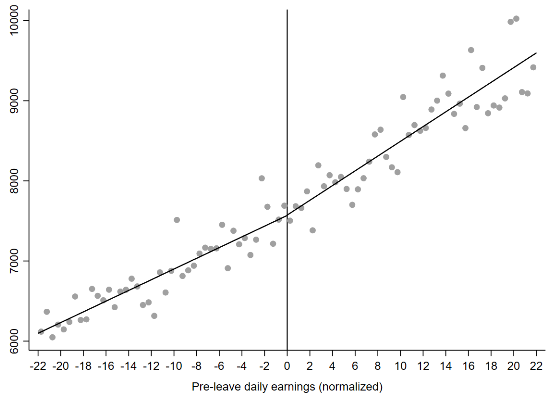
Panel A: Salaried employee (0/1)



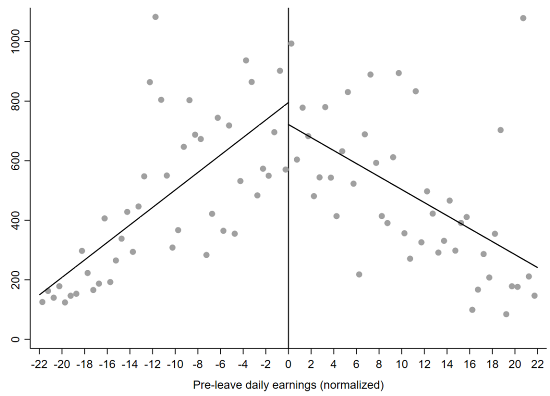
Panel B: Self-employed (0/1)



Panel C: Salaried earnings (euros)



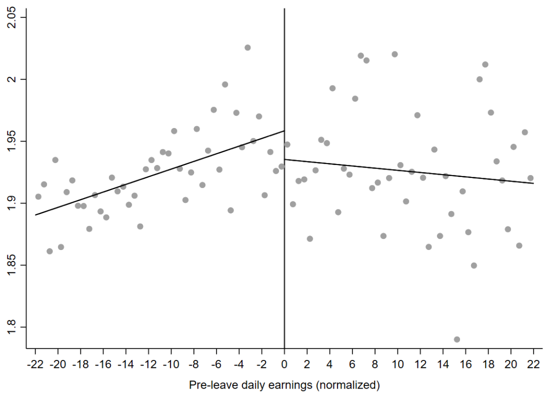
Panel D: Self-employed earnings (euros)



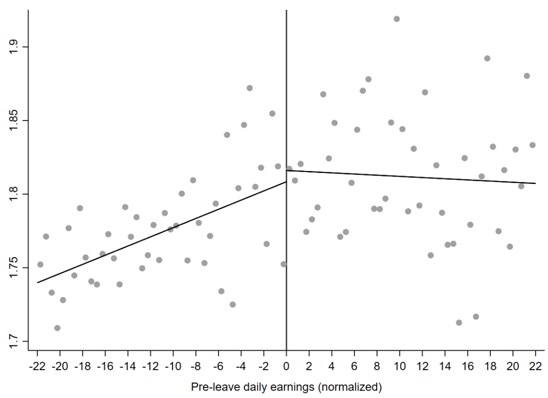
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure 8: Mother's outcomes 5 years after the birth of her first child

Panel A: Number of children



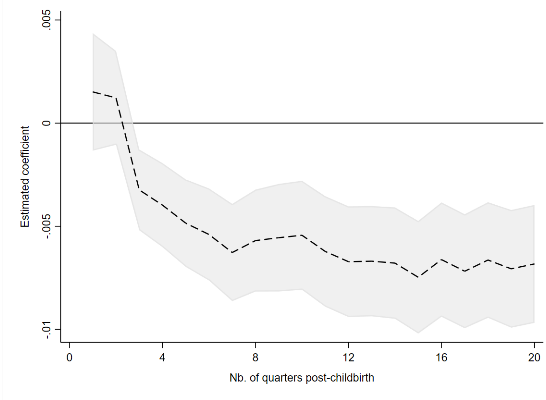
Panel B: Number of maternity leaves



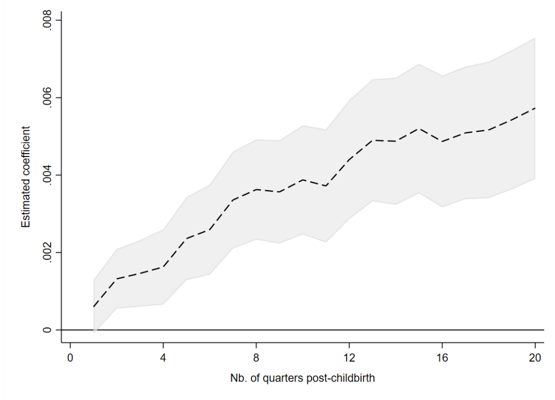
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure 9: Dynamic effects - Mother's outcomes over 5 years

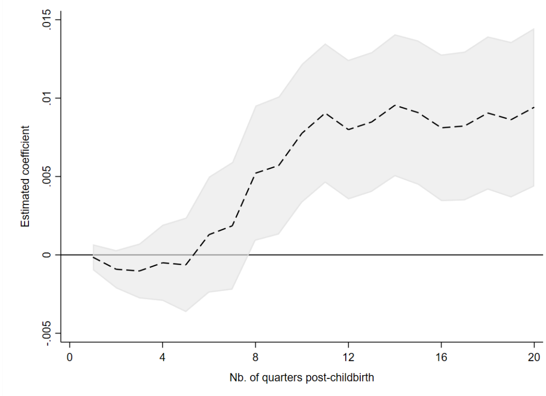
Panel A: Salaried employee (0/1)



Panel B: Self-employed (0/1)



Panel C: Number of children



Notes: These figures show treatment effects (dashed line), based on the RKD estimator of Equation (1.2), as well as 95% confidence intervals (shaded area). The coefficients are from separate regressions for each quarter following the birth of the mother's first child. All specifications use a common bandwidth of 22 euros around the kink.

Table 1: Descriptive statistics

	Full sample Mean (SD)	Kink sample Mean (SD)
Age of mother at first childbirth	28.4	30.2
	(4.0)	(3.3)
Total leave duration (# days)	84.7	85.9
	(13.6)	(11.4)
Total leave benefits (euros)	4806	6491
	(1908)	(1509)
Pre-leave quarterly gross wage (euros)	5237	8000
	(2871)	(1055)
Pre-leave full-time equivalent [0,1]	0.80	0.99
	(0.28)	(0.05)
Pre-leave hourly wage (euros)	12.57	15.92
	(10.18)	(15.60)
Household size (#)	2.0	2.0
	(0.5)	(0.4)
Live in Flanders (0/1)	0.64	0.69
	(0.5)	(0.5)
Married (0/1)	0.50	0.57
	(0.50)	(0.50)
Age of co-parent at childbirth	30.9	32.1
	(5.0)	(4.5)
Co-parent took paternity leave (0/1)	0.57	0.58
	(0.49)	(0.49)
Co-parent employed (0/1)	0.86	0.89
	(0.35)	(0.31)
Co-parent quarterly earnings (euros)	5632	7005
	(4430)	(4662)
Number of observations	182,923	38,255

Notes: This table presents the means and standard deviations (in parentheses) of some of the key variables for women having a first child during 2003-2010. The "kink sample" includes mothers whose daily base earnings are within 22 euros of the kink and corresponds to the sample used in all baseline regressions. All outcomes are measured at the moment of the birth of the first child, except for household size, which is measured before, as well as mothers' outcomes relative to the pre-leave employment.

Table 2: Predetermined outcomes - 4 quarters before the birth of the first child

	Coef. / SE	N
Household size (#)	0.001	38202
	(0.001)	
Live in Flanders (0/1)	0.001	38203
	(0.001)	
Age of mother	-0.004	38255
	(0.006)	
Married (0/1)	-0.001	38241
	(0.001)	
Full-time equivalent [0,1]	-0.001 **	37451
	(0.000)	
Hourly wage (euros)	0.004	37430
	(0.006)	
Age of co-parent	0.003	37405
	(0.008)	
Co-parent quarterly earnings (euros)	0.054	37705
	(10.153)	
Mother's share of household income [0,1]	0.000	37705
	(0.000)	

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The coefficients test for a change in slope at the kink for the predetermined covariates. They provide evidence for the validity of the smoothness assumption of the RKD. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 3: Mother's outcomes 5 years after the birth of her first child

	Coefficient (SE)	Robust CI	Mean
Panel A: First stage	-0.45 *** (0.02)		
Panel B: Treatment effects			
Employed (0/1)	-0.001 (0.001)	[-0.004 , 0.001]	0.90
Salaried employee (0/1)	-0.007 *** (0.001)	[-0.010 , -0.004]	0.87
Self-employed (0/1)	0.006 *** (0.001)	[0.004 , 0.008]	0.05
Quarterly earnings (euros)	58.421 *** (21.527)	[38.984 , 133.183]	7649.82
Salaried income (euros)	-56.342 *** (18.850)	[-74.657 , 8.042]	7193.14
Self-employed income (euros)	114.762 *** (15.948)	[84.393 , 154.388]	456.68
Number of children	0.009 *** (0.003)	[0.004 , 0.015]	1.92
Number of observations	38,255		

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The column "first stage" reports changes in slopes for the denominator of Equation (1.2). It captures the change in marginal benefits at the kink. The column "treatment effect" reports estimates based on the RKD estimator of Equation (1.2). The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the outcomes. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. I also provide bias-corrected confidence intervals ("robust CI") proposed by Calonico et al. (Calonico et al., 2014). The column "Mean" reports the average of the dependent variable within the defined bandwidth. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 4: Quantile Effects on Mother's Total Earnings

Quantile	Quarterly earnings (euros)
0.25	-51.66 *
	(28.50)
0.50	-19.52 *
	(10.40)
0.75	11.23
	(8.50)
0.90	35.75 **
	(15.87)
0.95	119.42 ***
	(31.51)
Number of observations	38,255

*Notes: All coefficients are from separate linear quantile regressions, using a symmetric bandwidth of 22 euros around the kink. The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the mother's quarterly earnings five years after childbirth. The sample includes mothers who had a first child between 2003 and 2010. Bootstrapped standard errors (1000 replications) are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 5: Mother's outcomes 5 years after the birth of her first child
(Heterogeneity analysis by sector)

	Atypical work schedule = No		**Atypical work schedule = Yes**			
	Treatment effect	Mean	Treatment effect	Mean	Diff. (SE)	Z-stat (p-value)
Employed (0/1)	-0.001	0.90	-0.001	0.90	0.000	-0.145
	(0.001)		(0.002)		(0.003)	(0.885)
Salaried employee (0/1)	-0.002	0.88	-0.017 ***	0.83	0.016	4.643
	(0.002)		(0.003)		(0.003)	(0.000)
Self-employed (0/1)	0.000	0.03	0.018 ***	0.09	-0.017	-7.030
	(0.001)		(0.002)		(0.002)	(0.000)
Quarterly earnings (euros)	-14.579	7590.44	215.742 ***	7778.77	-230.321	-(4.135)
	(21.502)		(51.388)		(55.705)	(0.000)
Salaried income (euros)	-28.241	7414.85	-116.113 ***	6675.76	87.872	2.063
	(21.612)		(36.698)		(42.589)	(0.039)
Self-employed income (euros)	13.663	175.59	331.855 ***	1103.01	-318.193	-6.233
	(8.580)		(50.324)		(51.050)	(0.000)
Number of observations	26,592		11,587			

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The column "treatment effect" reports estimates based on the RKD estimator of Equation (1.2). The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the outcomes. The last two columns report results from z-tests to check whether the coefficients estimated on the two sub-samples are statisticaly different. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 6: Mother's outcomes 5 years after the birth of her first child (Heterogeneity analysis by relative income)

	Mother's earnings lower than father's Treatment effect	Mean	**Mother's earnings higher than father's** Treatment effect	Mean	Diff. (SE)	Z-stat (p-value)
Employed (0/1)	-0.003 (0.002)	0.90	0.000 (0.002)	0.90	-0.003 (0.003)	-1.007 (0.314)
Salaried employee (0/1)	-0.010 *** (0.002)	0.87	-0.005 ** (0.002)	0.87	-0.005 (0.003)	-1.621 (0.105)
Self-employed (0/1)	0.009 *** (0.002)	0.05	0.004 *** (0.001)	0.05	0.005 (0.002)	2.517 (0.012)
Quarterly earnings (euros)	65.245 * (39.211)	7562.64	53.262 ** (24.768)	7721.81	11.983 (46.379)	(0.258) (0.796)
Salaried income (euros)	-83.101 *** (31.849)	7042.94	-38.615 (23.528)	7317.18	-44.486 (39.597)	-1.123 (0.261)
Self-employed income (euros)	148.346 *** (33.132)	519.71	91.878 *** (15.820)	404.64	56.469 (36.715)	1.538 (0.124)
Number of observations	17,302		20,953			

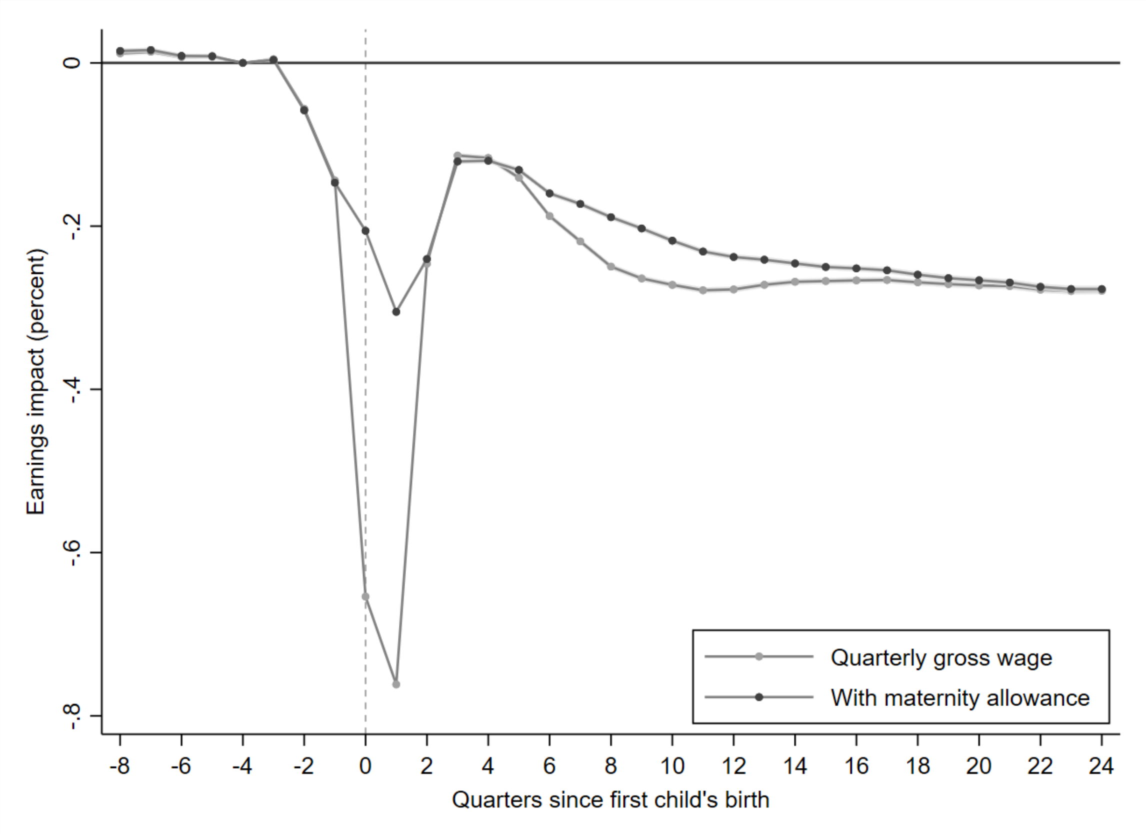
*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The column "treatment effect" reports estimates based on the RKD estimator of Equation (1.2). The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the outcomes. The last two columns report results from z-tests to check whether the coefficients estimated on the two sub-samples are statisticaly different. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Appendix

How Can Paid Maternity Leave Boost Female Entrepreneurship & Fertility?

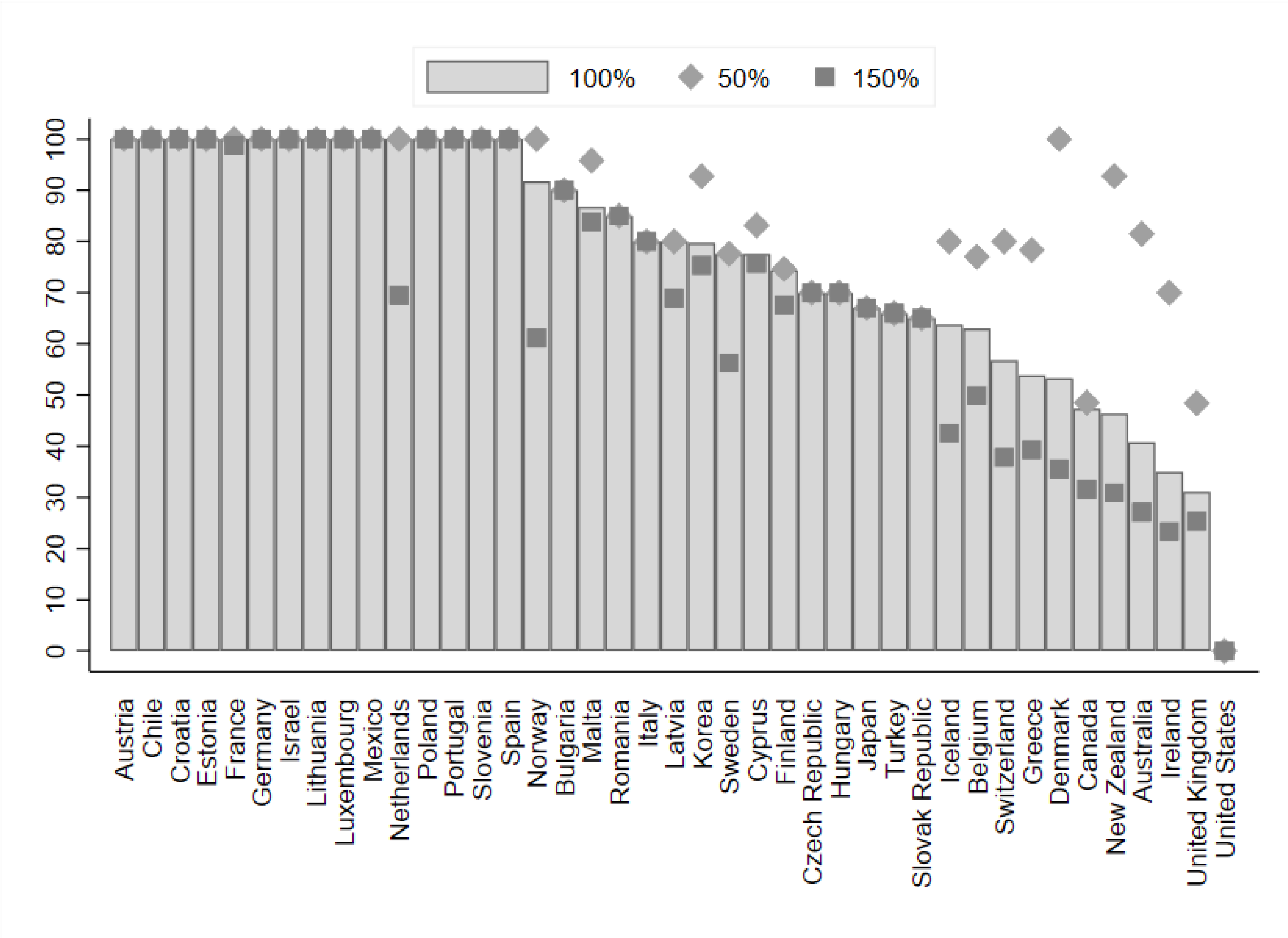
Sébastien Fontenay

Figure A1: Impact of children on mothers' quarterly gross earnings with and without maternity leave allowance



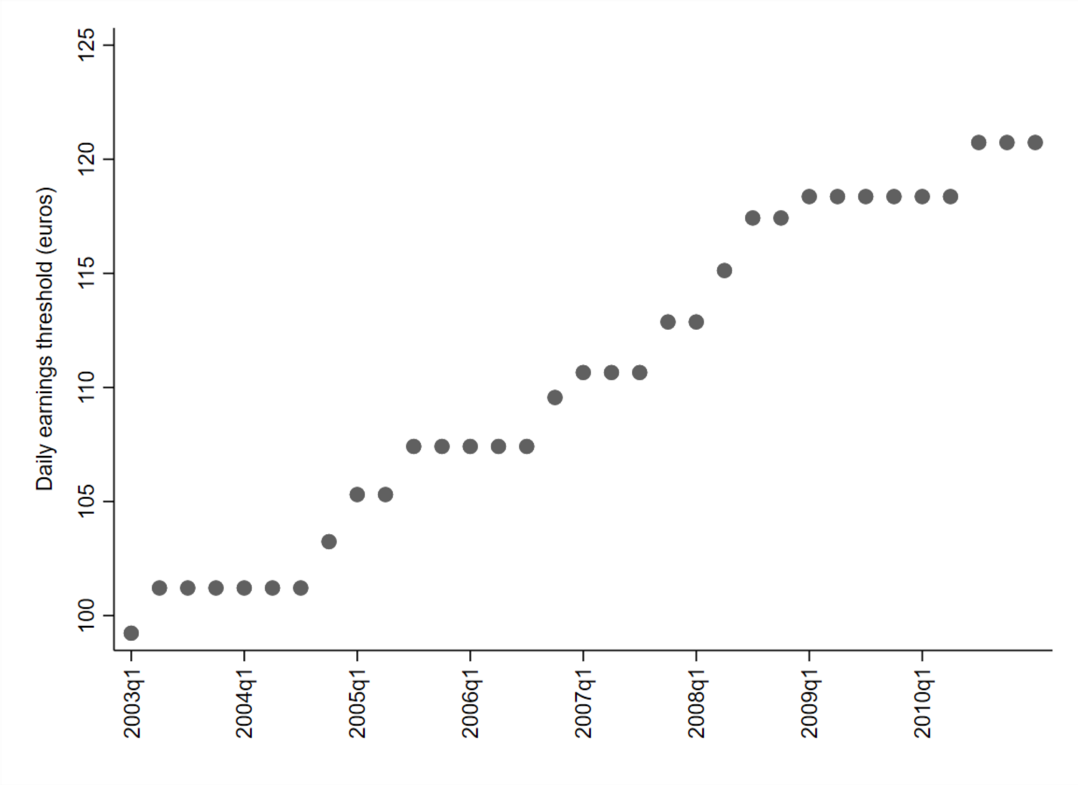
Notes: The figure shows event time coefficients (relative to the 4th quarter before the first child's birth) estimated on a sample of mothers who had their first child between 2003-2010 and were eligible for maternity leave (i.e. had sufficient work history). The coefficients are displayed as a percentage of the mean of the outcome measured at t-4. The earnings are measured conditional on labor force participation. The outcome will therefore not account for women leaving the labor market as a result of having children. The shaded 95% confidence intervals are based on robust standard errors.

Figure A2: Proportion of previous gross earnings replaced by maternity benefits (by level of earnings compared to the national average)



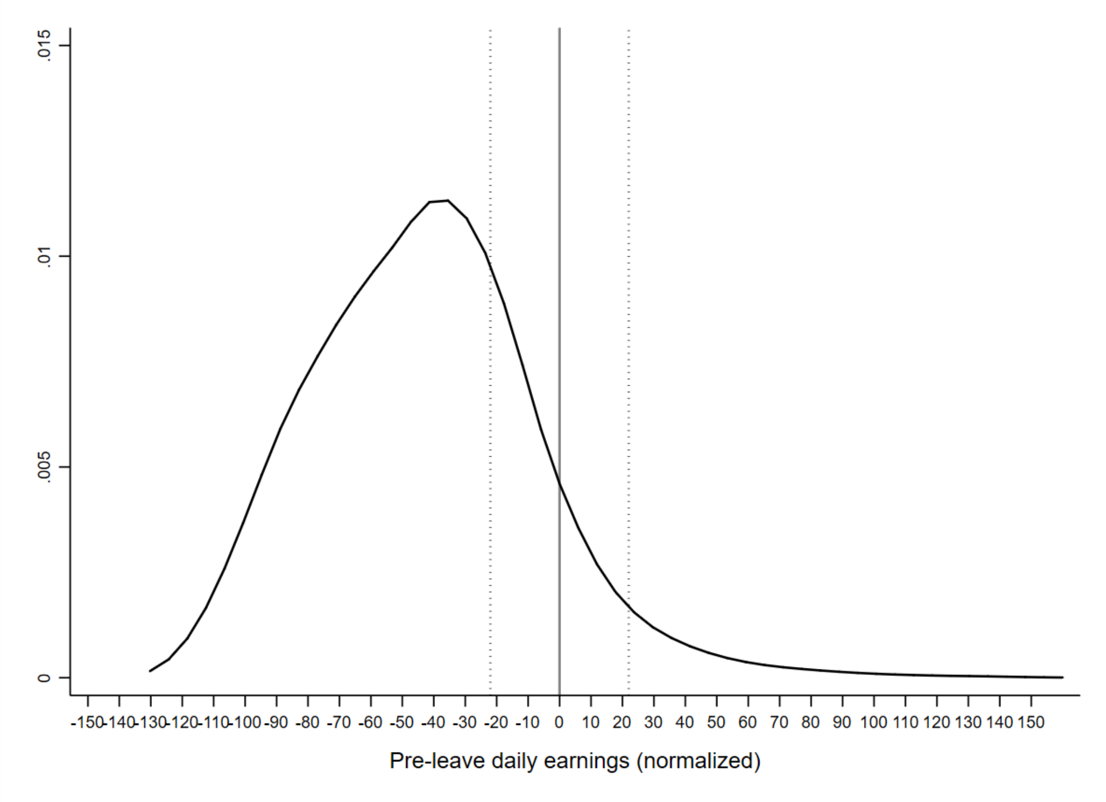
Notes: Data from the OECD Family Database, 2014. In Austria, Chile, and Germany benefits are calculated based on previous net (post income tax and social security contribution) earnings, while in France benefits are calculated based on post-social-security-contribution earnings.

Figure A3: Daily earnings threshold by quarter



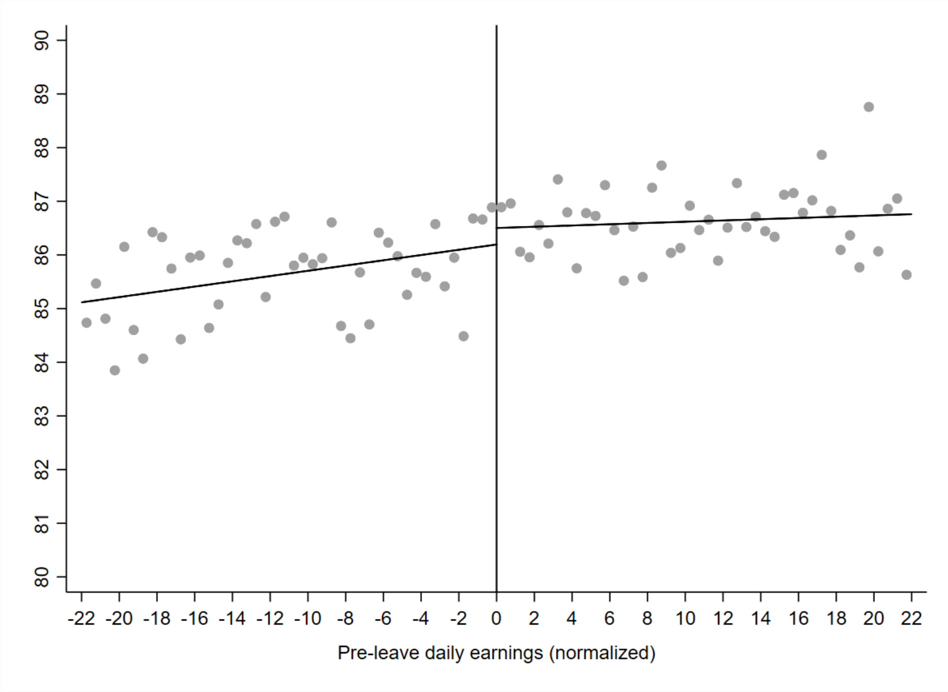
Notes: The figure shows the evolution of the earnings threshold set by the social security administration. The changes reflect government's decisions, as well as automatic adjustment to inflation. Data source: National Institute for Health and Disability Insurance.

Figure A4: Kernel density of pre-leave earnings around the kink



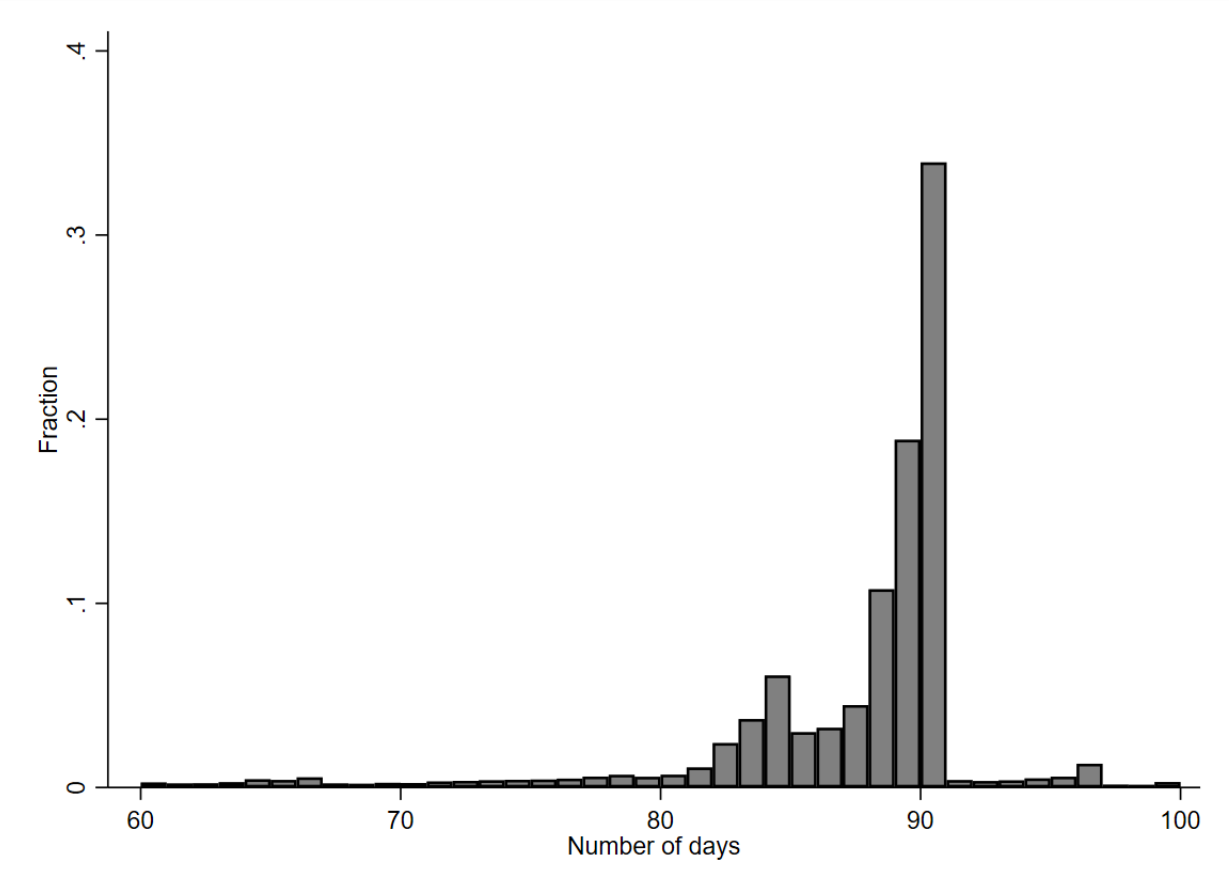
Notes: The graph plots the distribution of the pre-leave earnings using kernel density. The kink is located around the 90^{th} percentile. The dashed lines represent the 22 euros bandwidth used in the baseline specifications.

Figure A5: Duration of maternity leave (# days)



Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

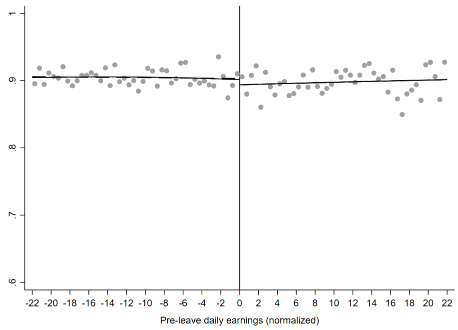
Figure A6: Distribution of total leave duration for women with earnings near the kink point



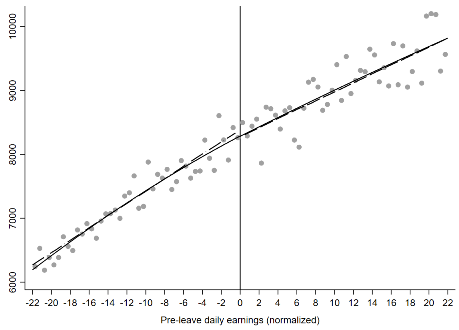
Notes: This figure plots the distribution of maternity leave duration for women with pre-claim earnings within a 22 euros bandwidth surrounding the kink point. The maximum duration of maternity leave in Belgium is 90 days, but it can be extended to 102 days for multiple births. All mothers must stop working during a compulsory period of at least 60 days.

Figure A7: Comparison between linear and quadratic functions of the assignment variable - Mother's employment outcomes

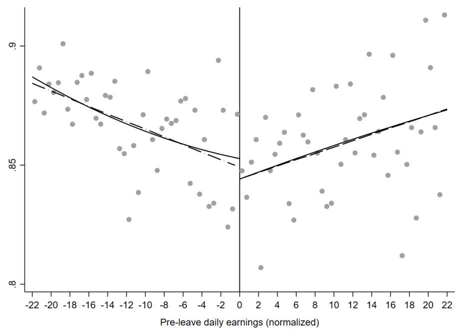
Panel A: Employed (0/1)



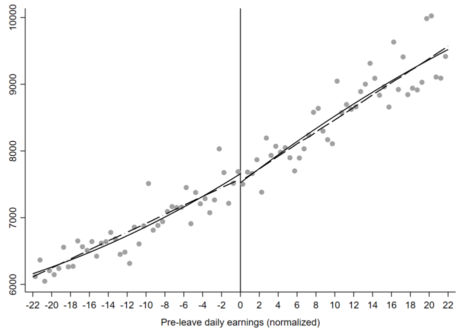
Panel B: Quarterly earnings (euros)



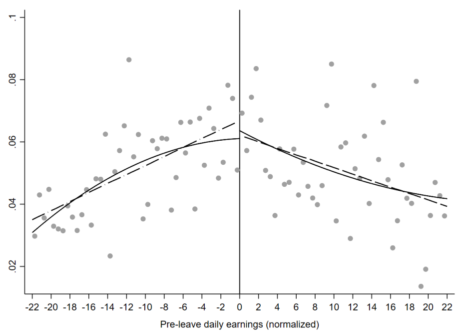
Panel C: Salaried employee (0/1)



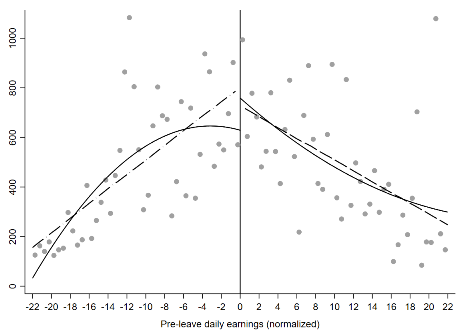
Panel D: Salaried earnings (euros)



Panel E: Self-employed (0/1)



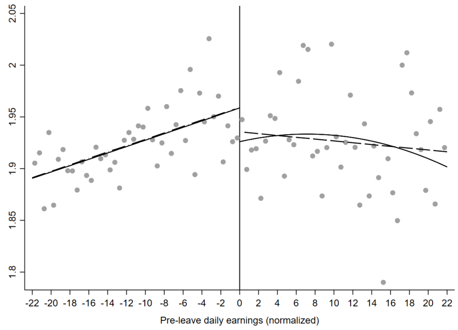
Panel F: Self-employed earnings (euros)



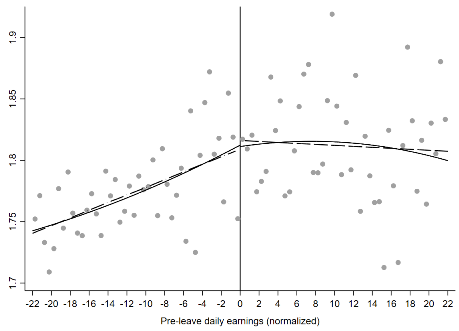
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The dashed lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions of order 1. The solid lines display the underlying quadratic relationship on each side of the kink and are estimated using local nonparametric regressions of order 2.

Figure A8: Comparison between linear and quadratic functions of the assignment variable - Mother's fertility outcomes

Panel B: Number of children



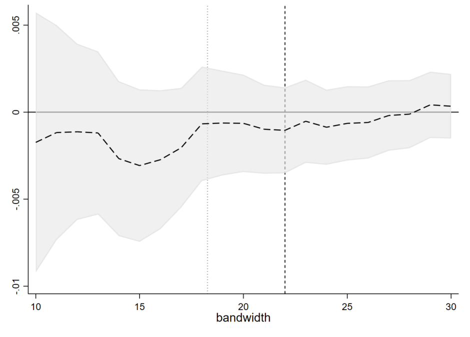
Panel B: Number of maternity leaves



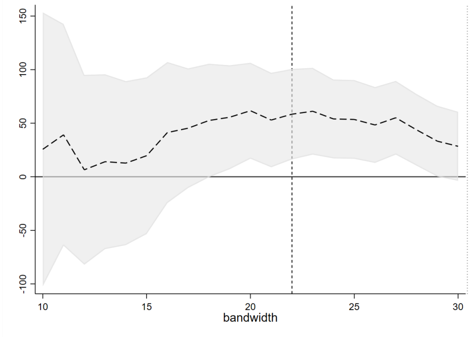
Notes: The horizontal axis plots normalized pre-leave daily earnings (relative to the kink) in bins, using 50 euro cents bins. The vertical axis plots the mean of the outcome in each bin. The dashed lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions of order 1. The solid lines display the underlying quadratic relationship on each side of the kink and are estimated using local nonparametric regressions of order 2.

Figure A9: Varying bandwidth - Mother's employment outcomes

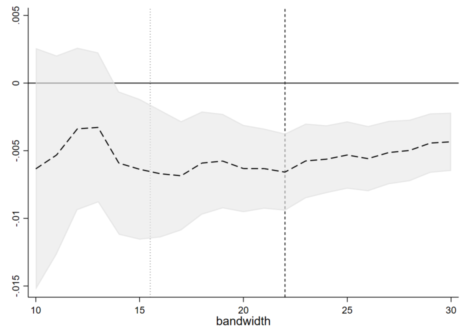
Panel A: Employed (0/1)



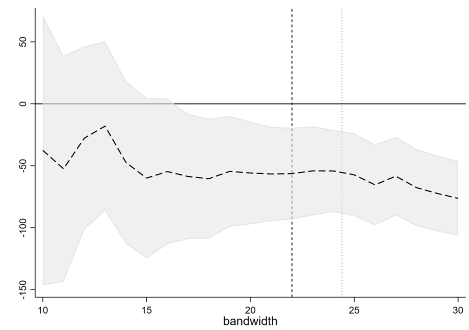
Panel B: Quarterly earnings (euros)



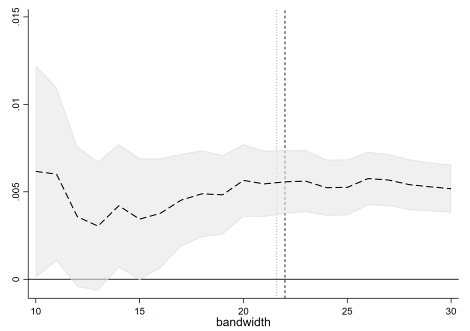
Panel C: Salaried employee (0/1)



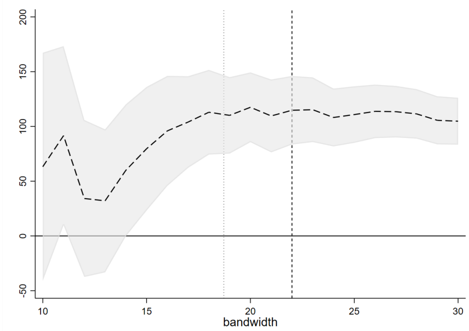
Panel D: Salaried earnings (euros)



Panel E: Self-employed (0/1)



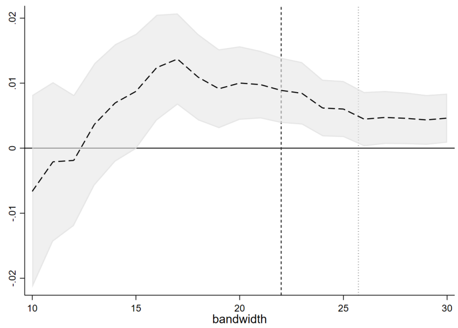
Panel F: Self-employed earnings (euros)



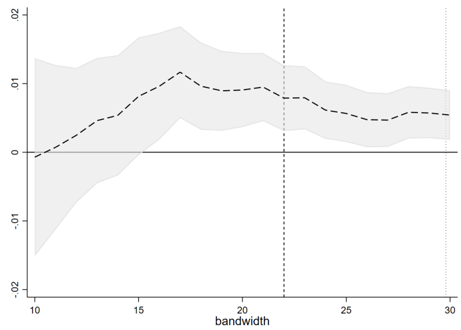
Notes: These figures show treatment effects (dashed line), estimated with local polynomial nonparametric regressions of order 1 (i.e. linear), as well as 95% confidence intervals (shaded area). The coefficients are from separate regressions using all possible bandwidths in 1 euro increments of normalized pre-leave daily earnings from 10 to 35 euros. The dotted vertical line materializes the bandwidth picked by the CCT selector of Calonico et al. (2014). The dashed vertical line materializes the common bandwidth of 22 euros used for the main estimations. All samples include mothers who had a first child between 2003 and 2010. For panels B, D and F, the outcomes are trimmed, replacing the top 1% of the distribution with missing values.

Figure A10: Varying bandwidth - Mother's fertility outcomes

Panel A: Number of children



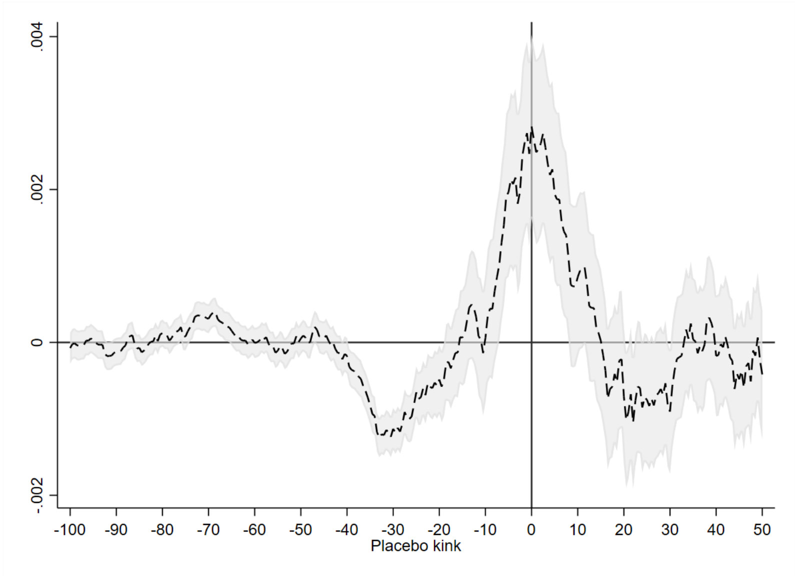
Panel B: Number of maternity leaves



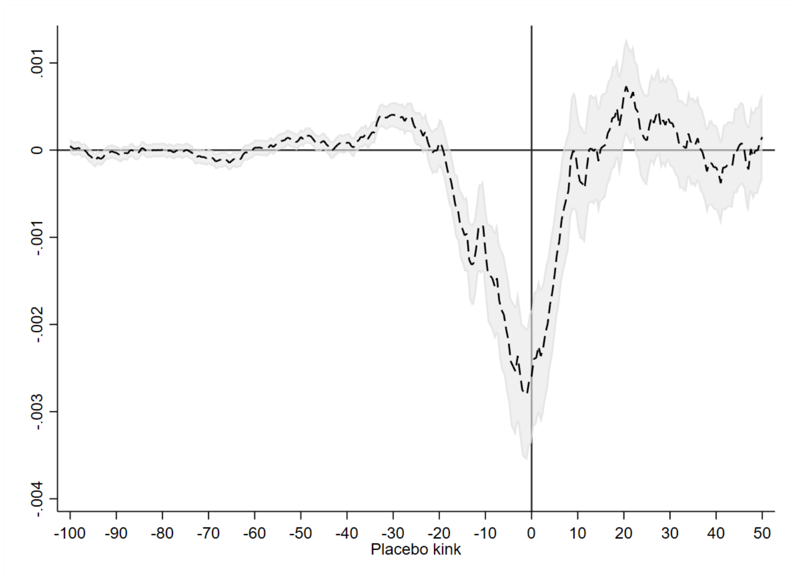
Notes: These figures show treatment effects (dashed line), estimated with local polynomial nonparametric regressions of order 1 (i.e. linear), as well as 95% confidence intervals (shaded area). The coefficients are from separate regressions using all possible bandwidths in 1 euro increments of normalized pre-leave daily earnings from 10 to 35 euros. The dotted vertical line materializes the bandwidth picked by the CCT selector of Calonico et al. (2014). The dashed vertical line materializes the common bandwidth of 22 euros used for the main estimations. All samples include mothers who had a first child between 2003 and 2010.

Figure A11: Permutation tests - Reduced form coefficients and 95% CI

Panel A: Salaried employee after 5 years



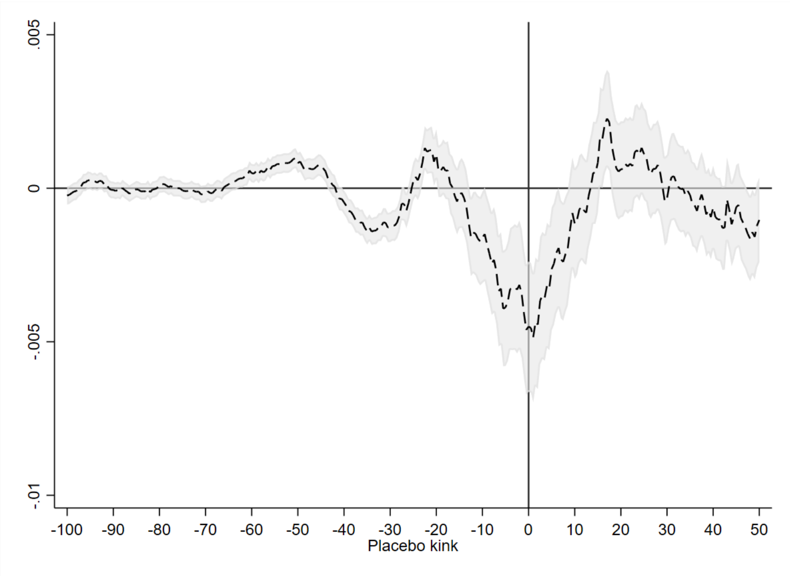
Panel B: Self-employed after 5 years



Notes: The graphs show results from permutation tests, proposed by Ganong and Jäger (2018), to assess the sensitivity of the results to non-linearities in the relationship between the assignment variable and the outcome. The figures plot the coefficients (dashed line) and 95% confidence intervals (shaded area) from 300 RKD models using placebo kinks along the distribution of the assignment variable, with a 22 euros bandwidth. The horizontal axis displays the distance from the true kink point (at 0). Note that those are reduced form estimates that correspond to the numerator of Equation (1.2). As such the placebo kink coefficients are of the opposite sign from those reported in the baseline specifications. One can see that the coefficient estimate at the true kink point is much larger than those at placebo kinks.

Figure A12: Permutation tests - Reduced form coefficients and 95% CI

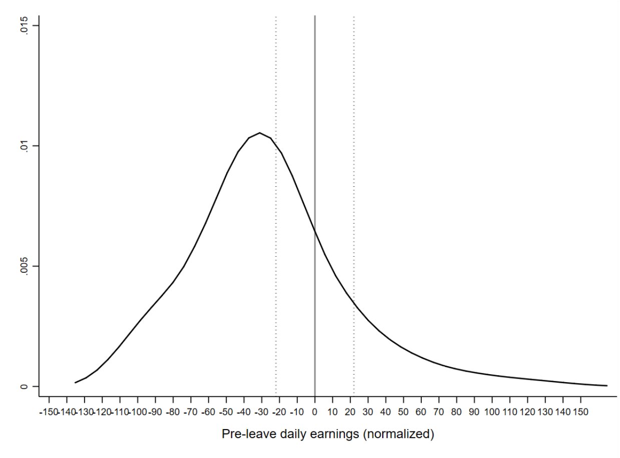
Number of children after 5 years



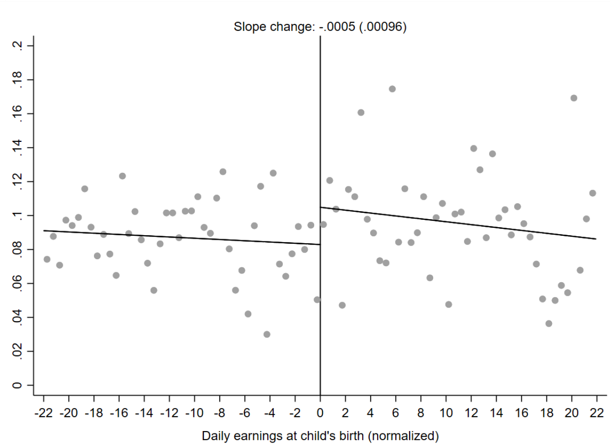
Notes: The graphs show results from permutation tests, proposed by Ganong and Jäger (2018), to assess the sensitivity of the results to non-linearities in the relationship between the assignment variable and the outcome. The figures plot the coefficients (dashed line) and 95% confidence intervals (shaded area) from 300 RKD models using placebo kinks along the distribution of the assignment variable, with a 22 euros bandwidth. The horizontal axis displays the distance from the true kink point (at 0). Note that those are reduced form estimates that correspond to the numerator of Equation (1.2). As such the placebo kink coefficients are of the opposite sign from those reported in the baseline specifications. One can see that the coefficient estimate at the true kink point is much larger than those at placebo kinks.

Figure A13: Placebo group - Fathers who do not go on leave

Panel A: Kernel density of pre-leave earnings around placebo kink



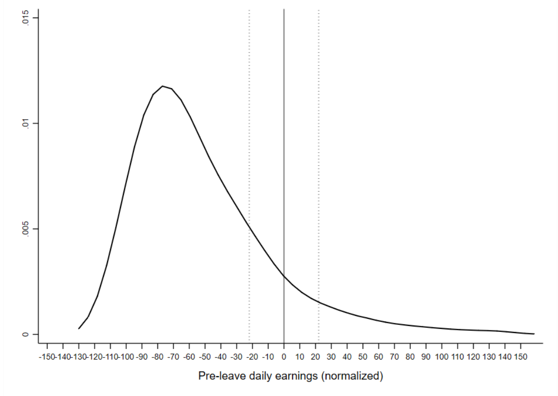
Panel B: Self-employed 5 years after the birth of their child (0/1)



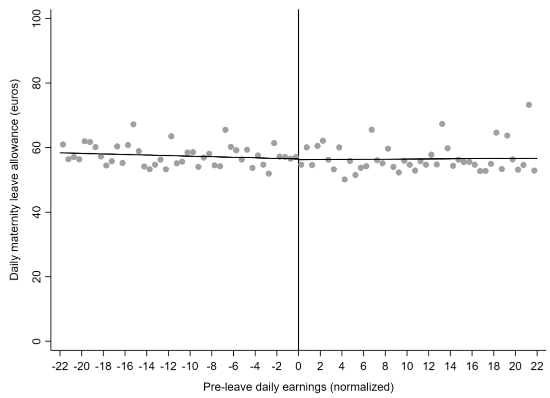
Notes: The sample is composed of fathers who did not go on leave after the birth of their child and therefore did not receive benefits from the social security administration. The horizontal axis plots normalized daily earnings during the quarter of birth of their child (relative to the kink) in 50 euro cents bins. The vertical axis plots the mean in each bin of the outcome variable for the probability to be self-employed after 5 years. The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions.

Figure A14: Placebo group - Mothers already self-employed at first childbirth

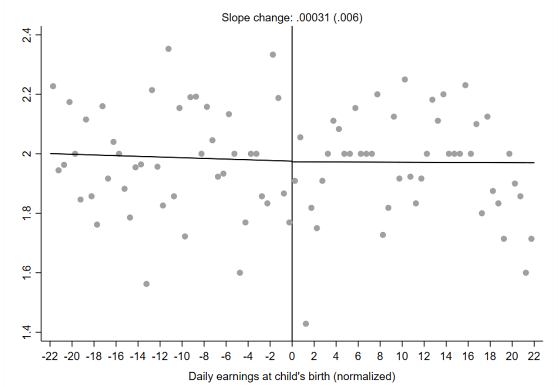
Panel A: Kernel density of pre-leave earnings around placebo kink



Panel B: Maternity leave allowance as a function of pre-leave earnings

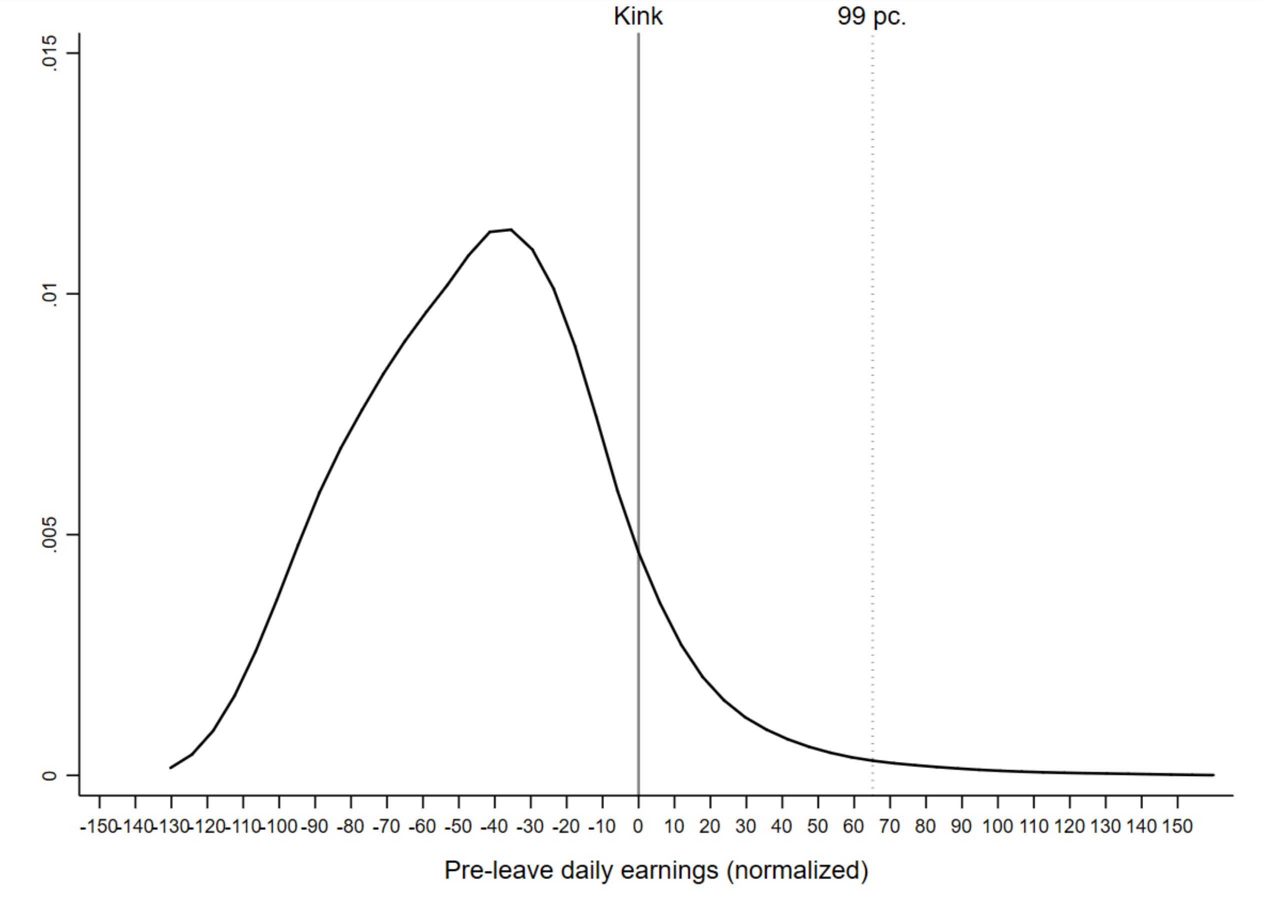


Panel C: Number of children after 5 years



Notes: The sample if composed of first-time mothers who were self-employed before the birth of their child and therefore receive a flat amount of maternity leave benefits. Panel A plots the distribution of the pre-leave earnings for self-employed women using kernel density. The placebo kink is located around the 90th percentile, similar to the main sample. The dashed lines represent the 22 euros bandwidth used in the main specifications. Panel B shows the empirical relationship between the daily maternity leave allowance and the pre-leave earnings of self-employed women within the 22 euros bandwidth around the kink. The lower panel plots normalized pre-leave daily earnings (horizontal axis) in 50 euro cents bins and the mean of the outcome variable for self-employed women (vertical axis): number of children (panel C). The straight lines display the underlying linear relationship on each side of the kink and are estimated using local nonparametric regressions. The change in slope at the kink is reported above the graph with standard errors in parentheses.

Figure A15: Kernel density of pre-leave earnings - Simulated threshold at 99$^{\text{th}}$ percentile



Notes: The graph plots the distribution of the pre-leave earnings using kernel density. The threshold set by the social security administration (solid vertical line) is located around the 90th percentile. The dotted line represents the location of a simulated threshold at the 99th percentile.

Table A1: Mother's outcomes 5 years after the birth of her first child

	Treatment effect	Robust CI	Mean
Duration of maternity leave (# days)	0.084 ** (0.042)	[-0.072 , 0.108]	85.87
Duration of maternity leave (log)	0.128 ** (0.055)	[-0.040 , 0.195]	4.44
Married (0/1)	0.002 (0.002)	[-0.002 , 0.007]	0.62
Number of maternity leaves	0.008 *** (0.002)	[0.001 , 0.012]	1.78
Number of observations	38,255		

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The column "first stage" reports changes in slopes for the denominator of Equation (1.2). It captures the change in marginal benefits at the kink. The column "treatment effect" reports estimates based on the RKD estimator of Equation (1.2). The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the outcomes. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. I also provide bias-corrected confidence intervals ("robust CI") proposed by Calonico et al. (Calonico et al., 2014). The column "Mean" reports the average of the dependent variable within the defined bandwidth. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table A2: Co-parent's outcomes 5 years after the birth of the first child

	Treatment effect	Robust CI	Mean
Paternity leave (0/1)	-0.003 (0.002)	[-0.010 , -0.001]	0.58
Quarterly earnings (euros)	18.295 (32.511)	[-44.865 , 96.604]	9125.72
Number of observations	37,705		

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The column "treatment effect" reports estimates based on the RKD estimator of Equation (1.2). The coefficients show the estimated effect of a 1 euro increase in daily maternity leave benefits on the outcomes. The sample includes co-parents who had a first child with a mother eligible for maternity leave between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table A3: Time Use Survey - Belgian women

	Employees	Self-employed	Difference
Personal care (incl. sleep and eating)	10:56	10:47	- 00:09
Employment	03:51	04:24	+ 00:33
Household and family care	03:28	03:42	+ 00:14
Leisure, social and associative life	04:09	03:31	- 00:38
Other	01:36	01:36	+ 00:00

Data source: Eurostat Time Use Survey, 2010.

Table A4: Mother's outcomes 5 years after the birth of her first child (varying polynomial order)

	Linear			Quadratic			Polynomial minimizing AIC	
	First stage	Second stage	Treatment effect	First stage	Second stage	Treatment effect	First stage	Second stage
Employed (0/1)	-0.437 ***	0.000	-0.001	-0.464 ***	0.001	-0.001		
	(0.020)	(0.001)	(0.001)	(0.080)	(0.002)	(0.004)		
	[298,671]	[15,524]		[298,674]	[15,527]		1	1
Salaried employee (0/1)	-0.437 ***	0.003 ***	-0.006 ***	-0.464 ***	0.002	-0.004		
	(0.020)	(0.001)	(0.001)	(0.080)	(0.002)	(0.005)		
	[298,671]	[26,173]		[298,674]	[26,176]		1	1
Self-employed (0/1)	-0.437 ***	-0.003 ***	0.006 ***	-0.464 ***	-0.002	0.003		
	(0.020)	(0.000)	(0.001)	(0.080)	(0.002)	(0.003)		
	[298,671]	-[8,315]		[298,674]	-[8,314]		1	2
Quarterly earnings (euros)	-0.437 ***	-27.675 ***	63.352 ***	-0.464 ***	-3.788	8.170		
	(0.020)	(9.046)	(20.711)	(0.080)	(36.383)	(78.382)		
	[298,671]	[756,521]		[298,674]	[756,524]		1	1
Salaried income (euros)	-0.437 ***	25.254 ***	-57.809 ***	-0.464 ***	13.043	-28.134		
	(0.020)	(7.791)	(18.173)	(0.080)	(30.125)	(65.456)		
	[298,671]	[742,525]		[298,674]	[742,528]		1	1
Self-employed income (euros)	-0.437 ***	-52.929 ***	121.161 ***	-0.464 ***	-16.830	36.305		
	(0.020)	(6.561)	(15.859)	(0.080)	(28.735)	(62.101)		
	[298,671]	[736,243]		[298,674]	[736,239]		1	2
Number of children	-0.437 ***	-0.005 ***	0.010 ***	-0.464 ***	-0.003	0.006		
	(0.020)	(0.001)	(0.003)	(0.080)	(0.004)	(0.009)		
	[298,671]	[69,961]		[298,674]	[69,964]		1	1

*Notes: All coefficients are from separate regressions using a symmetric bandwidth of 22 euros around the kink. Contrary to the other specifications using local nonparametric methods for estimation, here I use parametric regressions in order to report conventional goodness of fit measures. In particular, I show the Aikake Information Criterion (AIC) in square brackets. The last columns show which specification (linear or quadratic) minimizes this information criterion. The columns "first stage" and "second stage" are reduced form estimates for the change in slope of the maternity leave benefit amount and the outcomes, respectively. The column "treatment effect" reports coefficients from two-stage least squares estimations, where the benefit amount is instrumented with the interaction between a dummy for being above the kink and the polynomial in the assignment variable (i.e. pre-leave daily earnings). The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Table A5: Mother's outcomes 5 years after the birth of her first child (varying bandwidth)

	CCT			BW=15		BW=20		BW=25		BW=30	
	Bandwidth	Treatment effect	N	Treatment effect	N	Treatment effect	N	Treatment effect	N	Treatment effect	N
Employed (0/1)	18	-0.001 (0.002)	30786	-0.003 (0.002)	24942	-0.001 (0.001)	34149	-0.001 (0.001)	44994	0.000 (0.001)	58200
Salaried employee (0/1)	16	-0.006 ** (0.002)	25830	-0.006 ** (0.003)	24942	-0.006 *** (0.002)	34149	-0.005 *** (0.001)	44994	-0.004 *** (0.001)	58200
Self-employed (0/1)	22	0.005 *** (0.001)	37399	0.003 * (0.002)	24942	0.006 *** (0.001)	34149	0.005 *** (0.001)	44994	0.005 *** (0.001)	58200
Quarterly earnings (euros)	30	31.104 * (16.288)	59433	19.665 (37.285)	24942	61.640 *** (22.824)	34149	53.501 *** (18.733)	44994	28.443 * (16.422)	58200
Salaried income (euros)	24	-57.858 *** (17.068)	43773	-59.987 * (33.028)	24942	-55.834 *** (21.123)	34149	-57.320 *** (17.076)	44994	-76.332 *** (15.356)	58200
Self-employed income (euros)	19	116.403 *** (18.698)	31680	79.652 *** (28.776)	24942	117.474 *** (16.226)	34149	110.821 *** (13.143)	44994	104.775 *** (10.899)	58200
Number of children	26	0.005 ** (0.002)	46964	0.009 * (0.005)	24942	0.010 *** (0.003)	34149	0.006 *** (0.002)	44994	0.005 ** (0.002)	58200

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), based on the RKD estimator of Equation (1.2). The first Panel "CCT" reports treatment effects estimated using the data-driven bandwidth proposed by Calonico et al. (Calonico et al., 2014). I use their MSE-optimal bandwidth selector with a regularization parameter that guards against the selection of large bandwidths. The selected common bandwidth (used below and above the kink) is reported on the first column of the "CCT" panel. The other panels report treatments effects estimated from four samples using a bandwidth of 15, 20, 25 or 30 euros. All samples include mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table A6: Mother's outcomes 5 years after the birth of her first child (controlling for pre-determined covariates)

	Cont. = Age Treatment effect	Cont. = Flanders Treatment effect	Cont. = Partner's income Treatment effect	Cont. = All Treatment effect
Employed (0/1)	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Salaried employee (0/1)	-0.007 ***	-0.007 ***	-0.007 ***	-0.007 ***
	(0.001)	(0.001)	(0.001)	(0.001)
Self-employed (0/1)	0.006 ***	0.006 ***	0.006 ***	0.006 ***
	(0.001)	(0.001)	(0.001)	(0.001)
Quarterly earnings (euros)	59.055 ***	58.733 ***	59.082 ***	60.125 ***
	(21.588)	(21.534)	(21.759)	(21.845)
Salaried income (euros)	-56.483 ***	-55.981 ***	-54.776 ***	-54.606 ***
	(18.928)	(18.853)	(19.039)	(19.137)
Self-employed income (euros)	115.538 ***	114.714 ***	113.858 ***	114.731 ***
	(16.021)	(15.956)	(16.146)	(16.240)
Number of children	0.009 ***	0.009 ***	0.008 ***	0.009 ***
	(0.002)	(0.003)	(0.003)	(0.002)
Number of observations	38,255	38,239	37,705	37,692

*Notes: All coefficients are from separate local polynomial nonparametric regressions of order 1 (i.e. linear), using a symmetric bandwidth of 22 euros around the kink. The treatment effects, based on the RKD estimator of Equation (1.2), are estimated controlling for the following pre-determined covariates: mother's age, mother's place of living (indicator for living in Flanders), partner's earnings, all at the moment of the birth of the first child. The last column controls for all the covariates. The sample includes mothers who had a first child between 2003 and 2010. Heteroskedasticity-robust standard errors are in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table A7: Simulated threshold at 99^{th} percentile - Budgetary and fertility consequences

			Social security threshold		**Simulated threshold at 99 pc.**				
Percentile	Nb. mothers	Daily wage	Daily allowance	Replacement rate	Daily allowance	Replacement rate	Cost for social security	New self-employed	Additional children
80	773	93	72	0.77	72	0.77	0	0	0
81	773	95	73	0.77	73	0.77	0	0	0
82	773	96	74	0.77	74	0.77	0	0	0
83	773	97	75	0.77	75	0.77	0	0	0
84	773	99	77	0.77	77	0.77	0	0	0
85	773	101	78	0.77	78	0.77	0	0	0
86	773	102	79	0.77	79	0.77	0	0	0
87	773	104	81	0.77	81	0.77	0	0	0
88	773	106	82	0.77	82	0.77	0	0	0
89	773	108	83	0.77	83	0.77	0	0	0
90	773	110	85	0.77	85	0.77	0	0	0
91	773	113	86	0.76	87	0.77	89,146	6	9
92	773	115	87	0.75	89	0.77	187,205	12	19
93	773	119	87	0.74	92	0.77	303,093	20	30
94	773	122	88	0.72	95	0.77	427,899	29	43
95	773	127	90	0.71	98	0.77	597,276	40	60
96	773	133	91	0.69	103	0.77	802,310	53	80
97	773	141	94	0.66	109	0.77	1,087,578	73	109
98	773	153	97	0.63	118	0.77	1,497,646	100	150
99	773	175	103	0.59	135	0.77	2,264,298	151	226
Total							**7,256,451**	**484**	**726**

Notes: The table reports the daily wage at each percentile of the earnings distribution between the 80^{th} and 99^{th}. The corresponding number of women is based on administrative information from the social security administration (INAMI, 2007). The panel "Social security threshold" reports the estimated allowance based on the formula described in Section 1.1 if the threshold is set at the level decided by the social security administration in 2007 (about 110 euros, that is the 90^{th} percentile). The corresponding replacement rate is equal to the allowance divided by the foregone earnings. The panel "Simulated threshold at 99 pc." shows the estimated allowance if the benefit threshold was moved to the 99^{th} percentile. The cost for the social security corresponds to the difference in daily allowance times 90 days for all the women affected by the change. From this simulation, one can infer that the total cost for the social security would be 7,256,451 euros, that is 2 percent of the total maternity leave payments in 2007 (INAMI, 2007). When using the estimated elasticites in Table 3, this implies an additional 484 women in self-employment and 726 more children born from mothers in the highest decile of earnings.

Chapter 2

Work Disability after Motherhood and How Paternity Leave Can Help

This article is co-authored with Ilan Tojerow. This research project benefited from the financial support of the Thiepolam fund. We thank seminar participants at the Belgian Day for Labour Economists (Maastricht, 2019), IEB workshop Social Expenditure (Barcelona, 2019), ULB Centre Emile Bernheim (2019), DULBEA workshop on the Economics of Disability (2019), Luxembourg Institute of Socio-Economic Research (2019), SOFI Workshop Gender Economics (Stockholm, 2019), Louisiana State University (2019), IWEPS (2019), Université catholique de Louvain (2019), University of Nevada at Reno (2019), IAE Workshop Health & Labour Market (Nantes, 2019), XXVII Meeting Public Economics (Barcelona, 2020), National Bank of Belgium (2020), University of Groningen (2020), Erasmus University Health Economics seminar (Rotterdam, 2020), EALE-SOLE-AASLE World Conference (2020) for comments and suggestions. We also thank Thomas Barnay, Patrick Button, David De la Croix, Delia Furtado, Paula Gobbi, Camille Landais, Erica Lindahl, Olivier Marie, Noemi Peter, François Rycx, Ity Shurtz, Todd Sorensen, Ariane Szafarz, Judit Vall Castello and Philippe Van Kerm for valuable comments.

Introduction

Recent economic research has empirically showed what working mothers have long known: women who have children experience a "child penalty" that negatively impacts the trajectory of their professional career (e.g., Kleven, Landais, & Søgaard, 2019; Lundborg et al., 2017). Despite a burst in research on this topic in the past decade, specific mechanisms behind the so-called child penalty, like the various "states of non-employment," that could more accurately explain how the birth of a woman's first child impacts her broader career remain undocumented. Recent papers that have deepened our knowledge on the dynamics between motherhood and labor market participation have focused on the link between gendered norms and time spent on housework (Boelmann, Raute, & Schönberg, 2020; Cortes & Pan, 2020; Petrongolo & Ronchi, 2020). None of these recent studies nor the ones preceding them, however, have explored a key component of the "child penalty," namely, how claiming disability insurance restricts mothers from full labor force participation and contributes to the stagnation of their career after the birth of their first child. And so, in this paper we find that a woman's first childbirth triggers a significant gender gap in disability claims and demonstrate how paternity leave, a policy that has become increasingly common over the last 20 years, significantly offsets this imbalance.

More broadly, our study offers a more complex understanding of two social programs that impact mothers, their partners and families as well as society overall: Disability Insurance and Paternity Leave. First, we pinpoint large gender inequalities in the context of Disability Insurance (DI) and show that mothers are more likely to experience a work-related disability than fathers after childbirth. This is of particular importance given that the number of persons deemed unable to work for health reasons and thus receiving DI benefits has increased substantially in OECD countries. This increase has particularly impacted women and within this group young working mothers which createes a unique and important challenge for social security funding programs (OECD, 2010). Second, we provide clear evidence that the introduction of a paternity leave policy increases birth spacing within families and mitigates the time mothers spent on disability after childbirth. With this result, our research offers a unique window onto an unobserved and potentially cost-saving dynamic between two distinct benefit programs. It reveals fiscal spillovers between two social security programs and suggests that spending on paternity leave could be more than compensated by savings in mothers' DI benefits.

We now discuss in more detail the child penalty that links motherhood with an increased proclivity towards disability. Our research utilizes an event study approach similar to the one used by Kleven, Landais & Søgaard (2019). This approach is based on individual-level variations in the timing of a woman's first childbirth and the sharp changes that occur around that event. Our analysis using Belgian administrative data reveals a disability-specific child penalty that does not disappear over the long run and, even up to eight years after their first child's birth, mothers are 1.2 percentage points more likely than fathers to enter DI. This represents a child penalty for women of around 40%. We also demonstrate that the impact of children increases with the size of the family, with a gender gap that reaches 2.3 percentage points for parents with three children.

From this analysis, we connect the role of this disability-specific child penalty to the observed reduction in female labor market participation after motherhood. In this regard, we estimate that 17% of women who leave the labor market after having children go on to claim disability insurance. This result provides significant new insights about the career trajectories of young mothers and the specific impact that breaks due to disability have on their labor market attachment.

While postpartum health effects have been highlighted before (Cheng, Fowles, & Walker, 2006; Saurel-Cubizolles, Romito, Lelong, & Ancel, 2000; Spiess & Dunkelberg, 2009), we are among the first to capture the long-term health consequences that ensue many years after childbirth. In line with our findings, Angelov et al. (2020) have relied on within-couple variations to show that mothers more than double their sick leave compared to fathers after the birth of their first child in Sweden. Our study differs because our event study approach allows us to also look at mothers and fathers separately. In another recent study on Norway, Andresen and Nix (2019) consider child penalties after birth for both heterosexual and lesbian couples. While their findings for mothers in heterosexual relationship are similar to ours, the resultss for lesbian couples do not show postnatal differences in sickness absences between mothers who bore the child and the one who did not. This result gives support to our argument that long-term effects may be driven by family arrangements rather than only the biological cost of giving birth. On the downside, all their findings rely on a definition of sickness that also includes absences for caretaking of young children. In contrast, our study relies on disability spells that have been validated by a doctor and concern only the health of the mother. This allows us to capture in our findings any direct health effects without any other interference.

We next argue that the increased probability to enter disability for young mothers might be linked to family arrangements which require employed women to work a "second shift" at home (Hochschild & Machung, 1990) and take on a larger share of domestic work, including child care. This is well-documented in Guryan, Hurst, & Kearney (2008) who have put together time use surveys from 14 countries and have showed that the gender gap in time spent with children varies across countries but is always detrimental to women. Moreover, this across-country imbalance is often combined with another reality whereby working mothers spend less time devoted to childfree leisure and personal care (Craig, 2007; K. Parker & Wang, 2013; Pepin, Sayer, & Casper, 2018).[1] Taken together, this combination of more domestic work and less leisure for working mothers might ultimately affect women's health and career, and explain their increased likelihood to suffer from a work-related disability. Beside this first explanation, it is also possible that after becoming mothers, applying for DI benefits reflects a substitution effect between work and household responsibilities, rather than a strong deterioration in health condition. In other words, some of those mothers might have worked in the absence of a DI scheme, which is a general comment that applies to many DI beneficiaries who are found to retain some capacity to work (French & Song, 2014; Maestas, Mullen, & Strand, 2013).

Building on the initial findings of this study about this specific child penalty, we next look at whether the provision of a paternity leave could be an effective policy to moderate a mother's entry into disability. Our reasoning draws on Becker's (1985) theoretical framework that associates a woman's career with her household responsibilities and relies on numerous studies that have empirically shown how paternity leave policies effectively increase a father's involvement in child care, even when the leaves are short (Farré & González, 2019; Hook, 2010; Kotsadam & Finseraas, 2011; Patnaik, 2019; Tamm, 2019). From there, we argue that if the inequality of family arrangements and a mother's tendency to fall into disability are linked, then even a two-week paternity leave provision could soften mother's child penalty in disability. In the short run, that would translate into more help from the father right after birth. More critically in the long run, it could permanently impact the division of tasks in households.

And so, we exploit a discontinuity in Belgian legislation which established a two-week paternity leave only to fathers of children born after July 1st 2002 to analyze its effect on the probability for women to enter into disability after childbirth. To do so, we use a regression discontinuity difference-in-differences (RD-DiD) framework similar to Avdic & Karimi (2018). This research design relies on the exogeneity of the time of birth to compare households that had a child before and after the July 2002 reform. It also considers using non-reform years to wash out any seasonality in disability through its difference-in-differences dimension (Avdic & Karimi, 2018; Bütikofer, Riise, & Skira, 2021; Cygan-Rehm, Kuehnle, & Riphahn, 2018; Danzer & Lavy, 2017; Farré & González, 2019; Lalive, Schlosser, Steinhauer, & Zweimüller, 2014). We implement this research design to a sample of around 100,000 children born between 2002 and 2004 for which we have administrative data on their parents' labor market history, including disability spells and benefits received.

[1] For Belgium, Table A1 in the Appendix shows that working mothers spent on average more than double the time on childcare than their partners in 2013. At the same time, they had less personal time for leisure.

Overall, we find that the introduction of a two-week paternity leave decreases the number of days on DI for women by 21% up to 12 years after childbirth. The decrease in the number of days in disability is the largest for individuals on DI for more than 12 months. In that case, the decrease is indeed equivalent to a 33% reduction. This result highlights the benefit of paternity leave for a mother's career since individuals on DI for more than 12 months have on average a much lower probability to re-enter the labor market. Results on the impact of the paternity leave reform on DI benefits confirm the different findings on the number of days. By contrast, we find no evidence of any change in days or benefits for fathers that are eligible for a paternity leave. If so, the positive effect on mother's disability rate does not seem to be at the expense of father's, suggesting an overall positive impact of paternity leave at the household level. All these results are robust to alternative specifications including varying the bandwidth selection or the trend definition and to a series of placebo tests.

We next present evidence that our main results are entirely driven by an effect on mothers who had their first child during the reform year. Up to 12 years after childbirth, this specific group experienced 40 percent fewer day on disability. Contrarily, women who gave birth but not to their first child in 2002 did not experience any change after the introduction of the two-week paternity leave. This striking contrast could provide suggestive evidence that, in eligible families, first-time parents are less embedded in fixed roles and are thus more inclined to change their behaviour when the father takes a paternity leave (Patnaik, 2019; Sundström & Duvander, 2002). Focusing on the causes of disability, we show that 50% of the long-term reduction in the number of days on DI happens for mothers with musculoskeletal disorders. These disorders represent up to 40% of all disability cases for mother in the long-run (Saurel-Cubizolles et al., 2000).

Finally, turning our attention to the mechanisms behind our results, we provide evidence linking the role of fertility decisions to the reduction of days on DI for mothers after the reform of the paternity leave system in 2002. In this light, we show that the reform increased birth spacing between the first two children and that the timing of this effect matches the effect related to disability. Indeed, the decline in disability days starts two years after the birth of the first child and correlates with birth spacing between the first and second child. The increase in birth spacing which followed the introduction of paternity leave becomes a key factor in explaining the effect of the reform on disability.

Our results contribute to a fast growing literature concerning the impact of paternity leave policies. An important part of this timely research focuses on how paternity leave impacts labor supply and the wages of both fathers and mothers. While a series of these studies find a positive effect on women's earnings, their probability to participate in the labor force and marital stability after paternity leave, (Andersen, 2018; Druedahl, Ejrnæs, & Jørgensen, 2019; Dunatchik & Özcan, 2019; Farré & González, 2019; Olafsson & Steingrimsdottir, 2020; Rege & Solli, 2013), others do not reach similar conclusive results (Avdic & Karimi, 2018; Cools, Fiva, & Kirkebøen, 2015; Ekberg, Eriksson, & Friebel, 2013). Few studies, however, focus like us on the effect of paternity leave on maternal health and maternal disability spells (Persson & Rossin-Slater, 2019; Ugreninov, 2013).[2] And when they do, they only consider short term effects, while we observe up to 12 years of disability status after childbirth. Ugreninov (2013) focuses on Norway and, contrary to us, does not find any significant effect of paternity leave on mother's health. A possible reason for this difference in the result is that she does not take seasonality in her outcome variables into account like us and the most recent papers on paternity leave.[3] The other study that tackles mother's health focuses more on the impact of a greater degree of flexibility in taking a paternity leave rather than a net effect like we do (Persson & Rossin-Slater, 2019).[4] Using an RD-DiD design like the one we use, Persson & Rossin-Slater (2019) find that increasing a father's leave flexibility reduces a mother's risk of physical postpartum health complications and improves their mental health within the first 6 months after childbirth.

Our paper is also related to the few studies analyzing fertility decisions in conjunction with parental leave policies for fathers (Cools et al., 2015; Duvander, Lappegard, & Johansson, 2020; Farré & González, 2019). In this study, we consider how changes in fertility patterns could explain the effect of paternity leave on maternal disability. On one hand, there is ample evidence from medical literature that short intervals between pregnancies are associated with adverse health conditions (Conde-Agudelo, Rosas-Bermudez, Castaño, & Norton, 2012). Hence, increased birth spacing could directly improve the health of mothers and diminish their likelihood to enter a disability insurance program in the long-term. On the other hand, birth spacing could also have a direct impact on labor market attachment. Karimi (2014) finds, for example, that longer intervals between births have positive long-run effects on income for women.[5] Our paper contributes to these different studies by linking birth spacing, maternal disability status and the labor market attachment of mothers to the provision of paternity leave.

[2]In a recent study, Bütikofer, Riise and Skira (2021) provide evidence that maternity leave has positive effects on maternal health. Chatterji and Markowitz (2012) also show that longer maternity leave is associated with declines in depressive symptoms and an improvement in overall maternal health 9 months after childbirth.

[3]Another explanation could be related to the fact that Ugreninov (2013) does not use a "classic" regression discontinuity design and ends up for that reason with a very small sample of parents who had a child within one month of the reform date.

[4]The paper evaluates a particular Swedish policy called "double days" that allows fathers to take up to 30 days of paid leave on an intermittent basis alongside the mother during the first year of the child, without affecting total leave duration.

[5]At the same time, Troske and Voicu (2013) show that increasing the time between the first and second childbirth reduces mothers' probability to work full-time.

The remainder of the paper proceeds as follows. Section 1 provides information on the institutional settings of the disability insurance system and the parental leave policy in Belgium. Section 2 focuses on the event analysis. Section 3 aims at presenting the regression discontinuity difference-in-differences analysis. Section 4 concludes.

2.1 Institutional Context

2.1.1 Parental Leave

The Belgian parental leave system has gradually developed since the seventies.[6] In 1971, a new law provided for the introduction of a 15-week paid maternity leave around childbirth. This program remains in effect at the time of writing and combines both pre- and post-birth leave with the obligation to be off work at least one week before birth and 10 weeks overall. Prospective mothers qualify for this paid leave if they have worked at least 120 days in the last 6 months.[7] The replacement rate is 82% of their gross salary during the first 30 days (uncapped) and 75% thereafter (capped at a ceiling of 2810 euros per month as of January 1, 2020).

The parental leave system was expanded in July 2002 with the introduction of a comprehensive, job-protected, paid paternity leave for fathers with a salaried contract.[8] Before the introduction of this law, fathers of newly born children were only entitled to 3 days of paid job absence.[9] The new paternity leave program introduced an additional period of 7 working days, which together with the 3 days of job absence, brought the leave period to 2 weeks. Initially, fathers had to take their paternity leave during the first month after childbirth, but the time frame was extended to 4 months in 2009, hence allowing fathers to take their paternity leave after the compulsory maternity leave period of mothers. As for the replacement rate, the first 3 days are fully compensated by the employer, while the remaining 7 working days are compensated like the mothers at 82% of the gross salary. As shown in Figure A1, a substantial number of fathers opted into

[6] Table A2 in the Appendix reports its main features.

[7] The unemployed mothers are also eligible for the same program if they have 120 active days of job search in the last 6 months before birth. Civil servants are entitled to the same program in the same conditions but with a different benefits system. Finally, the self-employed mothers are eligible for a different program paid at a flat rate and offering twelve weeks of maternity leave.

[8] This program includes all private sector workers and contractual employees in the public sector. Similar programs exist for civil servants but are directly managed by the different public administration. A completely separate paid paternity leave program for the self-employed was introduced in 2019.

[9] Since 1978, fathers were allowed to take 3 days off work after the birth of a child, called "congé de circonstance", which is equivalent to specific leaves for attending weddings or funerals offered by the "loi du 3 juillet 1978 relative aux contrats de travail". This category of leave is not fully job-protected and thus not like the paternity leave introduced in 2002.

this policy after it was introduced during the second half of 2002, and kept increasing in the following years.[10]

It might be useful to put the Belgian system in perspective with other countries, notably the Scandinavian countries, which were early adopters of government paid leave policies accessible to the fathers. In Sweden, for instance, the parental leave system was introduced in 1974 and was gender neutral. Both the mother and the father were given an equal amount of paid leave for their children, but with the option of freely transferring paid leave days between each other. The system was reformed in 1995 to encourage fathers to take a bigger share of the parental leave. A so-called "daddy-month" was introduced, reserving 1 month of paid leave to each parent, implying that 1 month of paid leave would be lost if either parent chose not to take any leave.

In Belgium, parental leave has never been transferable between parents. We believe that this feature makes it a particularly interesting case for research, since fathers can take paternity leave without an automatic reduction for mothers. In other words, we can measure the net effect of providing paternity leave. Those studies which use the 1995 "daddy month" reform in Sweden actually measure the combined effect of the paternity leave provision and the reduction of maternal leave (e.g. Avdic & Karimi, 2018; Ekberg et al., 2013). In the context of work disability, the reduction of maternity leave could have detrimental effects on maternal health, which might not be balanced by the provision of paternity leave.

2.1.2 The Belgian Disability Insurance System

In this paper, we focus on the health and labor market attachment of workers by observing how much time they spend on disability benefits. In Belgium, employed workers with a minimum number of working days have access to disability benefits through the National Institute for Health and Disability Insurance (NIHDI).[11] It covers them against health shocks that affect their ability to work for at least one month. The application terms and conditions vary, however, between disability spells that are either less than a year

[10] Finally, on top of the specific maternity and paternity leave programs described above, parents are also individually entitled to 4 months of parental leave that they can take at any time before their children turn 12 years old. The leave can be taken simultaneously by both parents and can also be taken over 8 months for a career interruption of 50% or 20 months for a career interruption of 20%. Workers who decide to use this form of leave receive a fixed amount instead of a percentage of their salary, which could make it less appealing in many cases. In 2017, 70% of the beneficiaries of this program were women (IEFH, 2018).

[11] Full-time workers and unemployed workers must have fulfilled a minimum of 180 working days (or active days of job search for the unemployed) during the last twelve months to be eligible. For part-time workers, the condition is to have worked at least 800 hours over the last 12 months.

and those that are longer.[12] In the remainder of the paper, we will therefore distinguish between these two types by referring respectively to the "short-term disability" spells and the "long-term disability" ones.[13]

In order to qualify for short-term disability coverage, individuals must be recognized as "unable to work" by a doctor designated by their health insurance fund.[14]A worker would be considered eligible when his/her ability to work is reduced by at least 66% with respect to the last occupation.[15] To qualify, the applicant should also have stopped all productive activity as a consequence of a deterioration of his/her health that is not directly related to his/her professional activity.[16] If these two conditions are still applicable after a year, a disabled worker may qualify for long-term disability status. There is, however, no automatic transition from the short-term status to the long-term one. In order to be accepted into the long-term disability program, the applicants' doctor (who oversaw the applicant during the short-term period) has to submit the application to the NIHDI, which can either directly approve the doctor's decision or run its own internal evaluation.

The replacement rate also varies with the duration of the disability spell. During the first year it amounts to 60% of the last wage payment received before becoming disabled. After one year, when one enters the long-term disability program, the replacement rate depends on the last wage payment received, as well as the position of the disabled person in the household. To be precise, this share is 65% for heads of households, 60% for single households and 40% for cohabitants, with defined floor and ceiling amounts.[17]

Figure A2 (Panel A) in the Appendix shows the evolution of the disability rate for the working-age population in Belgium since 1980.[18] Like in many OECD countries, the number of persons receiving DI benefits has increased substantially and particularly among female beneficiaries. It is often argued that this high increase for women reflects, in

[12]Spells shorter than a month are fully paid by the employers and are not covered by this insurance program.

[13]The disability literature sometimes refers to temporary and permanent disability programs.

[14]In Belgium, although the health care system is publicly supported at the national level, the reimbursement of medical expenses and short-term disability benefits are paid through the public health insurance funds called "mutualities", which are funded by the NIHDI and act as intermediaries with the disabled. In short, to benefit from the Belgian medical coverage, individuals must register at a health insurance fund.

[15]Note that an important change occurs after 6 months of disability: the reduction in the ability to work is then evaluated with respect to any occupation that the worker could perform given his/her age, education and experience (instead of his/her previous occupation).

[16]This condition exists to establish a distinction between the DI program and other programs such as the occupational injuries fund and the occupational diseases fund

[17]In 2020, the maximum short-term disability benefits were 2,248 euros per month, while the maximum long-term disability benefits were 2,435 euros per month.

[18]We plot DI beneficiaries from the long-term program, which is more directly comparable to the U.S. Social Security Disability Insurance (SSDI) program, often studied in the DI literature (e.g Autor & Duggan, 2006; Liebman, 2015).

part, their growing labor force participation, which contributed an expansion of the pool of insured workers, as more and more women had sufficient work history to qualify for DI. But according to Autor and Duggan (2006), this would explain only about one-sixth of the increase in the rate of female DI beneficiaries. Consistent with this, Figure A3, which considers eligible workers only, shows that the incidence rate for women is growing faster than for their male counterparts. This is true for Belgium (Panel A), the origin of data used in this study, but also for the United-States (Panel B) and most OECD countries (OECD, 2010).[19]

Another important trend in DI results from reforms that expanded the eligibility criteria and induced major changes in the composition of the beneficiary population, with a notable shift towards younger workers. Autor and Duggan (2006) explain that these new legislations place more weight on "applicants' reported pain and discomfort", making it easier to qualify for certain impairments that used to be "hard to verify", such as back pain or depression (Liebman, 2015). The side effect of these reforms has been an increased incidence rate of disability at younger ages (Congressional Research Service, 2018). Indeed, mental and musculoskeletal disorders tend to have an early onset and low age-specific mortality (Autor & Duggan, 2006). As a result, those beneficiaries are likely to enter early on the DI program and experience a relatively long duration. In 2017, 65.7% of the Belgians on long-term disability benefits were suffering from mental and musculoskeletal disorders.[20]

Hence, while work disability used to concern mostly older men prior to the 1990s, it is now increasingly affecting women, and particularly at younger ages. Our study adds to the existing literature on DI by exploring these gender inequalities among young adults. We show that this gender inequality as related to DI can by partially explained by parenthood and by how couples react to the arrival of children in the household.

2.2 Event Study Analysis of Work Disability after Motherhood

Our first research question evaluates to what extent children can affect the probability of their parents to fall into disability. As explained by Kleven, Landais & Søgaard (2019) the ideal experiment to do so would be to randomize fertility. In the absence of such

[19] Other explanations that have been frequently used emphasize spillover effects from reforms implemented in other social support programs (De Brouwer, Leduc, & Tojerow, 2019; Geyer & Welteke, 2021; Leduc & Tojerow, 2020; Truskinovsky, 2021), as well as from parents to children (Dahl & Gielen, 2020).

[20] Administrative data from NIHDI:
https://www.riziv.fgov.be/fr/statistiques/indemnites/Pages/default.aspx

an experiment,[21] they propose instead an event study approach based on individual-level variations in the timing of the birth of the first child to capture its direct effects on different labor market outcomes. The rationale being that, although fertility choices and the timing of birth are not exogenous, the outcomes of interest should evolve smoothly over time. Thus, any sharp changes around childbirth are likely to be orthogonal to unobserved determinants and seize any causal effects (Kleven, Landais, & Søgaard, 2019). In our case, it might be argued for example that women who invested in their education are more likely to have children later in life and are less likely overall to enter disability. However, the effect of education on those outcomes should not generate any sharp changes and therefore should be disregarded as an explanation linking childbirth to parental disability.

The event study approach has the additional advantage of tracing out the full dynamic trajectory of the effects over time, therefore capturing the impact of the first child, as well as of any subsequent children. Previous studies using instruments for the number of children, such as twin births (Bronars & Grogger, 1994) or the gender breakdown of siblings (Angrist & Evans, 1998), could only succeed in estimating local effects of second or higher order children. Our approach will instead capture the overall impact of having children on the probability to enter DI for mothers relative to fathers.

Event studies have been used in different contexts, those regarding the impacts of hospital admissions (Dobkin, Finkelstein, Kluender, & Notowidigdo, 2018) or family health shocks (Fadlon & Nielsen, 2019). In our specific setting, we foresee one limitation, the fact that this framework will not allow us to measure the impact of choices made before parents had children. For instance, if women invest less in education and career in anticipation of motherhood, then the estimated child penalties represent the lower bounds on the total lifetime impacts of children (Kleven, Landais, & Søgaard, 2019). In other words, our study will be able to identify the post-child effects of children conditional on choices made before parenthood.

2.2.1 Data & Empirical Strategy

We use a rich set of administrative data from the Belgian Crossroads Bank for Social Security (CBSS) to conduct our different empirical analyses. This database puts together several administrative registers linked at the individual level and contains quarterly information on social security status over time, household composition and labor market history. Importantly for our research design, the data allows us to match children with their parents through the National Registry and to observe the exact month of child-

[21] Lundborg, Plug and Würtz Rasmussen (2017) have come up with another very convincing strategy that uses *in vitro* fertilization treatments.

birth. Regarding data on disability, we can observe the disability status during any given quarter, as well as the number of days of each disability spell and the amount of benefits received. As part of this study, we obtained a large sample of 60% of all births during the years 2002 to 2013, with stratification at the provincial level to ensure representativity. From this sample, we were able to identify the parents and build a dataset that tracks their disability status quarterly over the period from 2002 to 2016.[22]

For the event study analysis, we narrow our sample to all individuals who had their first child between 2002 and 2013, without imposing any restrictions on the relationship status of the parents. This leaves us with an estimation sample of 691,922 parents, including 359,657 first-time mothers and 332,265 first-time fathers. We follow those parents over a period of up to 12 years, including up to 4 years prior to the birth of their firstborn and up to 8 years after it.[23] In total, we observe each parent up to 48 quarters.

We now turn to the econometric setting of the event study analysis. For each individual in the data, we first denote by t = 0 the quarter-year in which the father/mother has his/her first child and index all quarters relative to that time period. We then analyze changes in the disability status as a function of event time both in the short- and long-term, estimating the following equation separately for men and women:

$$y^g_{iqt} = \sum_{j \neq -4} \beta^g_j \cdot I[j = t] + \sum_k \gamma^g_k \cdot I[k = age_{iq}] + \sum_y \delta^g_y \cdot I[y = q] + \epsilon^g_{iqt} \tag{2.1}$$

where y^g_{iqt}, our main outcome of interest, is a dummy variable to indicate the receipt of disability benefits during a given quarter q for individual i of gender g and at event time t. On the right side, equation (1) includes a full set of event time dummies (first term on the right-hand side), age dummies (second term) and time period dummies (third term). We omit the event time dummy at t = -4, implying that the event time coefficients β^g_j measure the impact of children relative to four quarters before the first child's birth. We voluntarily chose a date not too close to childbirth, as we suspect that short-term disability would raise for women during their pregnancy. Following Kleven, Landais & Søgaard (2019), we include a full set of age dummies to control non-parametrically for underlying life-cycle trends. Additionally, the age dummies improve the comparison between men and women, as women are on average a few years younger than men when they have their first child. In addition, we include a full set of quarter-year dummies to control non-parametrically for

[22] This corresponds to a sample of 861,344 births and 1,271,079 parents.

[23] Our sample includes parents who had a child between January 2002 and December 2013. We follow those parents until 2016. Our panel is therefore unbalanced because the follow-up period differs according to the birth date of the reference child. For parents who had a child in 2002, we do not have data on the four years before. For parents who had a child in 2013, we have pre-birth outcomes but a reduced follow-up period of 3 years. We ran the estimations on a perfectly balanced panel and found similar results (available on request).

time trends and seasonal effects. Finally, we also control for linear pre-trends to consider potential pre-childbirth differences between men and women that age and quarter-year dummies would not capture and that could bookend the breaks around parenthood. To do so, we follow Kleven, Landais & Søgaard (2019) and estimate a linear trend separately for men and women using only pre-event data (i.e. from quarter -16 to quarter -4 before birth), and then use the result in the main event study specification described in equation (1) to residualize the outcome variable with the estimated pre-trend.

2.2.2 Main Results

Figure 1 (Panel A) plots the gender-specific impacts of children on disability status across event time. The outcome includes both short and long-term disability.[24] As explained above, it corresponds to changes in disability rates at event time t relative to the 4^{th} quarter before the first child's birth (t = -4), having controlled non-parametrically for age and time trends. The figure also includes 95% confidence bands around the event coefficients. Several lessons could be drawn from Figure 1 (Panel A) regarding parenthood, disability and how their interaction could impact men and women differently.

First, we know from our data that the disability rate is equal to 2.8% for both men and women at t = -4, and so there does not seem to be any gender difference in the disability rate before the birth of a first child. From there, however, the situation changes dramatically for women but not that much for men. Indeed, women experience a sharp increase in their probability to enter disability starting 3 quarters before their first child's birth. The timing corresponds to the beginning of the pregnancy and reflects in most of the cases pregnancy-related health issues. This sharp increase peaks in the quarter right before childbirth with an increase of about 6 percentage points in comparison to the 4^{th} quarter before giving birth.[25] From there, the next three quarters show a gradual return to the pre-pregnancy level and to a situation in which both men and women seem to experience the same probability to be on disability benefits. This downward trend around childbirth is to a large part mechanical since all women, sick and eligible for disability benefits or not, slide to compulsory maternity leave for at least 9 weeks after delivery.[26]

[24] The data for short-term disability in 2002 is available only for the four (out of six) biggest health insurance funds. As a robustness check, we also estimate the event study analysis excluding 2002. Results do not change (available upon request).

[25] The effect might be even larger given that women who are sick during the last six weeks of their pregnancy are already covered by their maternity leave and cannot be registered as disabled for that reason.

[26] As we can see in Figure 1, there are still women on disability in the quarter of childbirth and the next one. This situation is because women who are sick during the last six weeks before childbirth are only entitled to 9 weeks of postpartum maternity leave instead of 15. Consequently, a woman who gave birth at the beginning of a given quarter might still enter into disability during the same quarter if she

Moving now to the results between one year and up to eight years after birth, we can learn from Figure 1 (Panel A) that women start to experience another increase of disability during the second and third year following their first child's birth.[27] After the third year, this change stabilizes and eight years after delivery reaches an increase of about 2.1 percentage points compared to the pre-birth disability rate. Bearing in mind that our event study design captures the total effect of all children, this last result suggests the existence of an overall long-term "disability penalty" from having children that only impacts women. In contrast, we also show in Figure 1 (Panel A) that men seem to be largely unaffected by children. We only detect a small increase in their probability to enter disability two quarters after their first child's birth that stabilizes itself at 0.9 percentage points, eight years after it. Most importantly, we observe that the probability of men and women to enter DI never converges back and that eight years after their first child's birth, a 1.2 percentage points gap remains. Since the average disability rate at t-4 was 2.8% for both women and men, this corresponds to a child penalty for women that amounts to 43%. In section 2.5.1 of the Appendix, we show that our results are robust to an alternative specification based on an Instrumental Variable (IV) approach.

This finding suggests that a significant proportion of young women suffer from disability after the birth of their first child. Drawing on this fact, we try to measure the extent to which labor force exits supported by DI contributes to the well-known "child penalty" that impacts the labor force participation of mothers (Kleven, Landais, & Søgaard, 2019). To do so, we first estimate, using equation (1), the child penalty in labor force participation for men and women up to 8 years after the birth of their first child. Our measure of labor force participation includes work in both salaried and self-employment. Figure 1 (Panel B) illustrates these results and shows that mothers suffer in the long-term from a 24% decrease in labor force participation. We do not observe a similar effect for fathers because they do not experience any change in their labor force participation before and after the birth of their first child.[28] Secondly, we estimate equation (1) for an alternative outcome variable that captures both labor market participation and being disabled, thus excluding disability from non-employment.[29] As Figure 1 (Panel B) shows, this "new" measure of the child penalty is equal to 20%, that is 4 percentage points lower than the overall one. This difference implies that 1 out 6 women (17%) who exit the labor market after having children receives DI benefits. This result demonstrates the importance of work disability in explaining the declining labor force participation of women as well as

was only entitled to 9 weeks of maternity leave after birth.

[27]Interestingly, this second increase coincides with the average time of the arrival of subsequent children in an household. We would come back to that dynamic in the next sections.

[28]In another study, we show that the overall child penalty in earnings for Belgian women amounts to 32% (Fontenay et al., 2021).

[29]The new variable takes the value 1 if the individual is either salaried employee, self-employed or disabled.

their career progression more broadly after childbirth.

2.2.3 Heterogeneous Effects by Household Size

In this section, we want to observe how the effects measured in the event study analysis could vary with the total number of children in a given household. Even though our event study is based on parents who had their first child between 2002-2013, the results presented in Figure 1 (Panel A) are based on the full sample, irrespective of the total number of children they end up having. As already explained, this means that the dynamics we observe include the effects of children born after the first one. In other words, the estimated long-run impacts should be interpreted as capturing the total effect of all children. To explore the implications of multiple children, we replicate the event study analysis on 3 subsamples that we split based on the total number of children which parents produce - 1, 2, and 3 children, respectively as of 2016.[30]

Figure 2 presents the results of this analysis by the overall number of children in a household. The figure shows that the sharp increase around the birth of the first child is roughly similar in magnitude for the three subsamples. We also notice that the coefficient for mothers reverts to a level close to zero in the third quarter after childbirth for all types of families. It is only from the fourth quarter after childbirth that trends start differing across households. In families with a single child (Panel A), the trends between parents are only slightly different. The gender gap eight years after the birth of their only child reaches only 0.8 percentage point. In families with two children (Panel B), we observe an increasing gap between mothers and fathers in the second and third year following the birth of the first child. This very likely captures the effect of the second child. The gap between mothers and fathers up to eight years after the birth of their first child reaches 1.4 percentage points. It is expected that the two-child families in Figure 2 (Panel B) look very much like the estimates for the whole sample in Figure 1 (Panel A), since those families make up 50% of our sample. Finally, in families with three children (Panel C), the gap between parents reaches 2.3 percentage points after eight years. Placed end to end, these findings strengthen our conclusions that the probability for women to enter disability depends strongly on having children and increases when they have more than one. These findings also reiterate that dynamic has little to no impact on fathers and their probability to be on disability benefits.

One might ask, however, whether the increased probability for women to enter DI reflects merely the multiple pregnancies and deliveries or corresponds to the larger cost

[30]It gives us three samples of 31%, 50%, and 15%of all the women included in our main analysis. The remaining 4% have more than 3 children.

of having and providing for multiple children. To answer this concern, we replicate our event study analysis around the second child's birth, conditioning our sample on having two children in total, as of 2016. From Figure A4 in the Appendix, we observe a spike in the probability of women to enter DI around the second childbirth that is similar in magnitude to Figure 1 for the first child. We also see a small bump during the four years that precede the second child's birth. This is of course related to the first child's birth. It is a smooth bump rather than a sharp spike because the birth of the first child did not take place during the same quarter for all women. More interestingly, we note, is the increase in DI probability that follows the second child's birth. Since we conditioned our sample on households with two children, this subsequent increase cannot be attributed to other childbirths. We believe that it instead reflects the long-run effects for women of having multiple children. Thus, we conclude that beyond the short-term effects related to giving birth, there are indeed long-term health effects of having children for women, which are reflected in their increased probability to enter DI even eight years after their second child's birth.

2.3 Paternity Leave and Maternal Disability

In the previous section, we provide empirical evidence that children have a large impact on the probability of mothers to enter disability. We now turn to study whether paternity leave could be an effective policy to moderate the entry of women into disability after motherhood. Interested by both the short- and long-term consequences of the policy on women, we focus our analysis on the cumulative effects over a period of up to 12 years after childbirth. Within this framework, we also try to capture any tradeoff the policy could create for fathers and analyze how family planning decisions could play a role in this context.

2.3.1 Empirical Strategy

We use a regression discontinuity design to analyze the impact of paternity leave on maternal disability, exploiting a cutoff in the Belgian legislation, which opened paternity leave only to fathers who had a child after the 1st of July 2002. It relies on the fact that fathers, whose children were born right before that date, did not have access to this newly introduced two-week leave. We implement the method on a sample of parents who had a child in a 6-month window around the reform.[31] Our running variable is the

[31] In Section 2.3.5, we test the sensitivity of our results to different bandwidth selection, i.e. incrementally changing the window's size from 6 months to 1 month around July 2002, the month of the

month of birth. As explained by Imbens (2008), the key assumption of this design is that individuals are unable to manipulate the assignment variable. In our case, this seems like a reasonable assumption since birth dates are arguably difficult to manipulate. If this assumption holds, having a child right before or right after July 1st is as good as random.

All our specifications estimate intent-to-treat (ITT) effects since we observe eligibility (month of birth) but not the actual take-up of paternity leave. Indeed, individual-level data on paternity leave is not available for the second half of 2002, the year that the policy was introduced. We, however, have data on the subsequent years for children born in 2003 and 2004. In our sub-sample of parents who were both working at the time of birth, the take-up was respectively 55.6% in 2003 and 57.5% in 2004. Thus, our estimates suggest effects on the treated that are up to twice as large as our ITT estimates. Finally, to assure that we properly capture a causal effect of the paternity reform at the cutoff, we also need to assume that there are no other important changes of relevance (such as other policy interventions) for parents of children born right after the 1st of July. We are not aware of any such potentially confounding factors.

Taken together, these different elements motivate the estimation of the following regression-discontinuity design model:

$$y_i^T = \alpha + 1[t_i \geq c]\beta + 1[t_i \geq c] \cdot f_r(t - c, \gamma_r) + 1[t_i < c] \cdot f_l(c - t, \gamma_l) + \zeta X_i + \epsilon_i \qquad (2.2)$$

where y_i^T is the outcome of interest, T quarters after birth, for each parent of child i born in month t. c is the reform cutoff month, $1[\cdot]$ is the indicator function, f_l and f_r are unknown functions with parameter vectors γ_l and γ_r, capturing trends in the outcome of interest. We can interpret β as the estimated discontinuity for a given outcome when having children born just before and just after the 1st of July 2002. And if we assume that parents do not have exact control of when their children are born in a period around the 1st of July cutoff, we can interpret the estimated discontinuity as the causal effect of the paternity leave reform. Finally, we include a vector of control variables X_i, for age of parent, number of kids and region of living at the time of the birth of the reference child. We know that those variables might affect the probability of entering disability and should therefore help us get more precise estimates. We test formally at the end of this section that those predetermined outcomes are perfectly balanced between the treatment and control groups.

The only remaining issue concerns the potential seasonality of our main outcome variable that captures maternal time on DI. Panel B of Figure 3 reveals that women who

reform

gave birth during the second part of the years 2003 and 2004 (non-reform years) have, on average, a higher number of disability days, which translates into a positive discontinuity at the July cut-off for all the samples considered.[32] We need to account for this seasonality to properly measure the discontinuity introduced by the paternity leave reform in July 2002. To do so, we combine the regression discontinuity design of equation (2) with a difference-in-differences model in a way similar to other research on the topic of parental leave (e.g. Avdic & Karimi, 2018; Bütikofer et al., 2021; Cygan-Rehm et al., 2018; Danzer & Lavy, 2017; Dustmann & Schönberg, 2012; Farré & González, 2019; Lalive et al., 2014). We then apply this combined approach to the sample of children born during the reform year (2002) as well as during two non-reform years (2003 and 2004).[33] This approach is valid under an additional common trends assumption that the trends of our outcomes are comparable between reform and non-reform years. We cannot think of reasons why the seasonality pattern would change because of the introduction of the paternity leave. In section 2.3.5, we also include two additional non-reform years (2005 & 2006) to our main specification and show that the results do not vary.

Specifically, we extend equation (2), using years 2003 and 2004 on top of 2002, add an indicator R={0, 1}, equal to one for the reform year 2002 and zero otherwise, and interact this new variable with each variable included in the model:

$$\begin{aligned} y_i^T = \alpha + \sum_{s=0}^{1} 1[R_i = s] \cdot \{1[t_i \geq c]\beta_s + 1[t_i \geq c] \cdot f_r(t-c, \gamma_{rs}) \\ +1[t_i < c] \cdot f_l(c-t, \gamma_{ls})\} + \zeta X_i + \lambda_n + \epsilon_i \end{aligned} \tag{2.3}$$

Equation (3) is essentially a fully interacted version of (2) with separate effects for reform and non-reform years, with the exception of fixed effects for each non-reform year, represented by λ_n. Our coefficient of interest is still β_1, which is now the interaction between "having a child after July 1st" and the 2002 indicator (R). By doing so, this new specification controls for systematic differences in outcome across families having a child

[32]This seasonality in disability could be related, for example, to the existence of systematic differences between parents who have children at different times of the year (Buckles & Hungerman, 2013; Currie & Schwandt, 2013) or to the conditions for school entry during the academic year (Danzer & Lavy, 2017). See Appendix 2.5.2 for a more detailed discussion on potential explanation behind the seasonality in birth and long term disability rates.

[33]This setting has the additional advantage of accounting for the fact that we capture the disability status in our data at the quarter level while our running variable is identified monthly. This mismatch could be problematic as it creates mechanically differences in the follow-up period between couples whose children were born at the beginning or at the end of a quarter. For instance, if we observe the outcomes of parents one quarter after birth: those who had a child in June exhibit follow-up periods ranging from 3 to 4 months, while parents of children born in July exhibit follow-up periods ranging from 5 to 6 months. This might be important since the discontinuity will be measured between June and July, which are respectively the end and the beginning of a quarter. Using a regression discontinuity difference-in-differences design will also help solve that problem by washing out any such mechanical correlation between the month of birth and the probability to be on disability benefits (Avdic & Karimi, 2018).

in different (even if close) months of the year.

In terms of data, we restrict the sample used in the event study analysis to households with children born between January 2002 and December 2004 in which both parents are known at the time of birth but do not necessarily form a couple (in the sense of marriage or cohabitation) and might not even live together. Since we are primarily interested in the effects of paternity leave on mothers, we exclude mono-parental families from our analysis. We also restrict the sample to those households in which both parents were working at the time of birth. Since paternity leave is only available for salaried men, we do not want to include in our sample households in which the father was not working at the time of birth. This leaves us with an estimation sample of 101,735 households.[34]

Before moving onto to the presentation of the results, we test the validity of our identification strategy. As explained above, our design relies on when the paternity leave policy was introduced (July 1st 2002) and on the timing of childbirth around that date. Taken together, these two elements imply that being part of the treatment group or the control one is as good as random. First, we show that there is no evidence that parents were able to self-select into the new paternity leave system. If that would have been the case, it would have invalidated our identifying strategy. Manipulating the date for natural births is virtually impossible, but we want to rule out that planned cesarean sections or induced labor were not rescheduled in order for fathers to become eligible for the new paternity leave.[35] To do so, we use data from Statbel, the Belgian statistical office, on the number of daily births in 2002. We start by providing graphical evidence in Figure A6 that the frequency of daily births had not been affected by the reform. We observe that there is no evidence of bunching around the threshold.[36] As a second step, we test for sorting formally by estimating regressions of the form of equation (2), using as an outcome the log number of daily births.[37] Column 1 in Table A3 reports the results using a 7-day window around the threshold. Each subsequent column in the table increases the window by a week up to a specification with a 42-day-long window around the threshold. The coefficients for the different specifications are all small and statistically indistinguishable from zero. These results indicate that there is no discontinuity in the number of births

[34] For the estimations focusing on the fathers in the Appendix, the number of observations is different and stands at 99,502. It is mainly due to the fact that for 2 percent of our sample we do not have information on the fathers for one or more of the three control variables used in the estimations (i.e. number of children, age and region, at the moment of birth of the reference child).

[35] Gans and Leigh (2009) demonstrated the importance of checking for "introduction effects" by showing that Australians parents delayed induction and cesarean section procedures to become eligible for a new child bonus.

[36] We see on Figure A6 that there are always fewer births during the weekend, likely due to fewer scheduled deliveries. The day of the introduction of the policy, July 1st 2002, was a Monday. Mechanically, we observe that there are more births on that day than on the two previous days.

[37] We control for a linear trend in all but the first regressions. We also include dummies for the day of the week (i.e. Monday, Tuesday, . . .).

around the threshold and therefore suggest that families did not manipulate the date of childbirth to become eligible for the new paternity leave policy.

Table 1 further confirms the plausibility of our identifying assumptions by testing for a discontinuity around the reform cutoff for a large array of characteristics of the parents: their region of living, size of household, number of children, whether the reference child is their first child, age, type of employment (i.e. blue/white collar worker, civil servant or self-employed), as well as daily wage. All of these terms are measured in the quarter of birth of the reference child.[38] We test that the parents' characteristics are balanced around the threshold by applying equation (3) to all those observable variables. The right panel of Table 1 reports the results of these regressions and shows that all coefficients are statistically indistinguishable from zero. They confirm that there is no evidence of discontinuity in the characteristics of the parents who had children right before or after the introduction of the paternity leave policy in July 2002.

2.3.2 Main Results

Now that we have presented our RD-DiD framework and established the validity of the design, we can turn to the results of our analysis on the causal impact of paternity leave on the disability status of mothers.

We start with a graphical exploration of the effects of the reform on all mothers, as well as for first-time mothers and experienced ones separately. To do so, Figure 3 present for the reform year (2002) and for two non-reform years (2003 and 2004) the number of days on DI in the 12 years following childbirth. Panel A shows that at the July 2002 cutoff, the number of DI days is lower among women who gave birth after the reform, and more so for the sample of first-time mothers (but standard errors are rather large). On the contrary, we do not observe any discontinuity for experienced mothers. Panel B allows us to compare the estimated discontinuity in 2002 with seasonal patterns during non-reform years. As discussed in the previous sub-section, the discontinuity is positive in all samples during 2003 and 2004. It is only by combining these two effects that we will truly be able to measure the consequences of the paternity leave introduction on maternal disability. Our RD-DiD empirical strategy will therefore compare the 2002 decrease in DI days with the positive discontinuity observed during non-reform years.

Table 2 displays the treatment effects using the RD-DiD strategy for three disability-

[38] The wages are measured the quarter before the quarter of birth of the reference child. We had to limit the sample to a 3-month window (instead of 6 months), because our data starts in 2002. Therefore, we cannot observe outcomes the quarter before birth for those who had a child between Jan. and March 2002.

related outcome variables. The first outcome, reported in Panel A, displays results for the total number of days on DI since childbirth. It captures in this way the effects of the reform both at the extensive and intensive margins. The second outcome for which the results are displayed in Panel B focuses on the total amount of disability benefits received over the 12 years following childbirth and the 2002 reform. This cumulative effect also captures an impact at both the extensive and intensive margins, while accounting for differences in daily allowances. The last outcome, reported in the table as "Ever on DI" (Panel C), displays results for the probability to have entered DI at least once over the 12-year period that follows childbirth. It captures any effect of the reform that would take place at the extensive margin only. For each outcome, we present the overall effect, as well as the effect breakdown between the short-term and the long-term disability programs. Finally, Table 2 also displays the average of the outcome variables in order to give a sense of the size of the impact of the paternity leave reform.

Focusing first on the whole sample ("All mothers" in Table 2), we find statistically significant evidence that mothers who had a child with a father eligible for paternity leave spent on average 22 fewer days on DI in the 12 years following childbirth. Given an average of 105 days on disability, this represents a reduction of 21%. This result suggests an effect on the treated that may be twice as high, since we estimate an ITT effect with a 50% take-up rate. Interestingly, we observe that this effect is most pronounced for the number of days on the long-term disability program. On average, mothers in the treatment group spend 16.1 less days in the long-term disability program and 6.3 less days in the short-term one. This corresponds respectively to a 33% and 11% decrease compared with the baseline average of each group. This particular result more explicitly demonstrates the important role that paternity leave plays on a mother's career since the long-term disability program involves individuals who have been away from the labor market for a long period of time and who do not, in most cases, possess a work contract (De Brouwer & Tojerow, 2018). This suggests that the introduction of paternity leave could be particularly effective at decreasing the number of days in disability that are more consequential for the attachment of women to the labor market.[39]

In line with the results for the number of days, Panel B in Table 2 shows that mothers in the treatment group display an average decrease in disability benefits of the same magnitude (18%, 712 euros) compared with a baseline mean of 4049 euros. As with the number of days, the effect of the paternity leave reform appears much more concentrated on the long-term program with a reduction of around 30% on the benefits received in

[39] Another reason why the time spent in the short-term disability program proves less damaging relates to the fact that employers, at least in Belgium, cannot terminate an open-ended contract during the first six months of a disability leave period. In practice, many workers on short-term disability will therefore go back to the same employer when their health allows. On the other hand, most workers on long-term disability have been laid off and need to find another job when their disability status ends.

the 12 years that follow childbirth. Taken together, our results show that after childbirth mothers whose partners were eligible for a two-week paternity leave saw their health to be significantly less affected over a period of 12 years. This finding shows up through both the number of days spent on disability and the benefits received. Interestingly, this positive effect on the health of mothers does not occur at the expense of the fathers. Table A4 in the Appendix displays the effect of the paternity leave on fathers for our disability-related outcomes. It shows that fathers seem unaffected by the introduction of the new policy, whether it be for the number of days on DI or for benefits received in the 12 years following childbirth. In both cases, we do not observe a statistically significant change one way or another and conclude thus that the reform has not been detrimental to the working health of fathers.

Finally, Table 2 also displays in Panel C results for the probability to have entered DI at least once over the 12-year period that follows childbirth. Reported in the table as "Ever on DI", this outcome allows us to capture any potential effect of the reform that would have taken place at the extensive margin. Unlike the previous results, it indicates no statistically significant change for mothers during the 12 years following childbirth. This seems to indicate that most of the effect of paternity leave concerns rather serious health issues and takes place at the intensive margin rather than the extensive one.

Table 2 displays the aggregate outcome for the cumulative number of days in DI over the 12 year period following childbirth. Our RD-DiD setup also allows us to capture the dynamic effects of the paternity leave reform by estimating equation (3) for each quarter from the birth of the reference child (t=0) to 12 years after childbirth (t=48). Focusing on short- and long-term disability separately, Figure 4 plots the treatment effects of these regressions for the number of days on DI by quarter. In both cases, the dynamic pattern highlights a decreasing trend in the number of days over time, consistent with our previous results that mothers were affected by the paternity leave. As time passes, the beneficial effect of paternity leave is reflected in a commensurate increase in the number of days not on disability. Regarding the timing of the effect in the two programs, we observe that the effect becomes negative around 3 years after childbirth in the short-term program (Panel A) and around 5 years in the long-term one (Panel B). The discrepancy in the timing logically reflects the need to spend a year in the short-term disability program before having access to the long-term one. In addition, it is frequent for sick individuals to have several spells of sick leaves, with various attempts to go back to work in the intervals, before entering the long term disability program (Borg, Hensing, & Alexanderson, 2001; Wallman et al., 2009).

2.3.3 Effect of Paternity Leave by Birth Order

This section investigates how the treatment effects that we have identified could vary as a function of the birth order of children. Table 2 shows estimates separately for first-time mothers, which represent 48% of our sample, and "experienced" mothers who birthed a second or higher order child when the paternity leave reform was introduced in 2002. Regardless of the selected outcome, the effect seems entirely concentrated on those mothers who had their first child during the reform year. Panel A of Table 2 shows that first-time mothers, whose partners had access to paternity leave, spent on average 39 fewer days on disability overall. This result represents a decrease of 40%, that is almost twice as large as the effect observed for the whole sample. In parallel, we do not observe any statistically significant effect of paternity leave on the number of days in disability for mothers who had an additional child during the reform year. Figure 5 illustrates this difference between the two groups in a dynamic way for the short- and long-term disability programs together. While the negative effect on first-time mothers starts after 2 years and slowly builds over time to reach 39 days after 12 years, the effect remains close to zero over the whole period for experienced mothers. Panel B of Table 2 further confirms this result as it relates to disability benefits, indicating a significant decrease in the total amounts for first-time mothers and no effect for the other group of mothers. Interestingly, Table 2 also reports a statistically significant effect at the extensive margin by showing that the probability to enter the long-term disability program is 2 percentage points lower than in the control group. This result is again concentrated on first-time mothers.

Put together with the other results from Table 2, this reinforces our finding that paternity leave seems to have generated important changes in the long run in households lacking childcare experience and in which the respective roles related to child management have not yet coalesced. Those results are consistent with previous findings showing that the division of labor becomes more gender-based only after the birth of the first child (Kleven, Landais, & Søgaard, 2019) and that first-time fathers respond more strongly to policy incentives since their views or habits about child-rearing are not yet anchored (Patnaik, 2019; Sundström & Duvander, 2002). We believe that those results might also reflect subsequent fertility decisions. Indeed, during the follow-up period covered by our dataset, first-time mothers need to decide whether they want additional children, and if this is the case when they would prefer to have them. In the final section of this paper, we deepen the scope of such an argument and look in more detail at how changes in fertility patterns might play a role in explaining the effect of paternity leave on maternal disability.

2.3.4 Effect of Paternity Leave by Medical Condition

Here we explore heterogeneous effects depending on the medical condition for which mothers on DI have obtained their status. We do so with a special focus on mental and musculoskeletal disorders, which account for respectively 37% and 24% of the number of days on disability registered in our sample. Hence, we estimate equation (3) separately to look at the impact of the 2002 reform on the number of days on disability benefits for (1) mental disorders, (2) musculoskeletal disorders and (3) other health issues.[40] Since we only know the medical condition of the beneficiaries once they are categorized as having long-term disability status, we only perform this analysis on this particular program.[41] Table 3 presents results for the three types of health disorders and, as a comparison, for the whole sample now with a follow-up period of 11 years. In sum, these results demonstrate that almost 50% of the reduction in long-term disability days for mothers is related to musculoskeletal disorders. This corresponds to a decrease of 5.7 days in the long-term disability program for the mothers with a child born after July 2002. By contrast, the table indicates no statistically significant change in the number of days on disability for mental health disorders or any other type of health issues. Figure A8 in the Appendix shows dynamic estimates for these different groups and corroborates this result. It shows moreover that long-term disability days for musculoskeletal disorders (Panel B) start decreasing as early as two years after the reference child's birth and slowly accumulate over time to reach a total of about 6 days by the end of the period.

We conclude from this heterogeneity analysis that the long-term reduction in the number of disability days for mothers is largely driven by a decrease in disability related to musculoskeletal disorders. This result is not surprising given that the prevalence of backaches or back pain in general remains high among new mothers, even after the first postpartum year (Saurel-Cubizolles et al., 2000). Thus, the introduction of paternity leave in Belgium in 2002 seems to have unleashed a significant and positive impact on a widespread maternal health problem that historically has led mothers to seek disability benefits. To interpret this result in another way, our study could reveal moral hazard amongst mothers on DI, since musculoskeletal disorders are among the "hard to verify" impairments (Angelov et al., 2020; Liebman, 2015). In that case, paternity leave polices, which encourage fathers to be more involved in childcare, would decrease the occurrence of mothers who use DI to spend time with their children. While we cannot completely exclude this, the fact that we do not observe any effect on other self-reported ailments,

[40] Information on medical condition is based on the International Statistical Classification of Diseases. Using the first 2 digits, we identify 17 categories for ICD-9 until 2015 that we group in 3 categories (1) mental disorders, (2) musculoskeletal disorders and (3) others.

[41] We also restrict the analysis to 11 years after childbirth because there was a change in the ICD classification in 2016 without any possibility to convert the data of that year to the previous classification system.

like mental disorders, makes this explanation highly unlikely. The moral hazard argument is not attuned to this self-diagnostic specificity.

2.3.5 Robustness Checks

In this section, we provide various robustness checks for our empirical design. We first investigate the sensitivity of our results to different trend definitions to obtain an unbiased estimate of the discontinuity at the cutoff. Since we *a priori* do not know the functional forms of f_l and f_r in equation (3), we test for linear trend (our main specification), as well as quadratic and cubic splines. From Table A5 in the Appendix, we see that the reduction in disability days for mothers is very similar whether we use a linear or quadratic trend, respectively -22 and -21. When using cubic splines, the reduction is larger and amounts to -51 days. Regarding disability benefits, the amount varies from -466 (quadratic) to -1541 (cubic). Altogether, our findings appear to be robust to the trend specification.

Subsequently, we test the sensitivity of our results to the bandwidth selection. In Figure 6, we provide estimates for the effect of the paternity leave introduction on maternal disability days using bandwidths that vary from 5 months to 1 month around the reform cutoff. We also run a donut-hole specification, excluding births that took place one month before and after the cutoff, to confirm that parents did not manipulate the birth date. When compared to our "baseline" specification, which uses a bandwidth of 6 months, the other empirical specifications reveal qualitatively similar results with a reduction in disability days that ranges from -19 to -31 days (although with larger standard errors for the donut-hole test, as shown in Table A6 in the Appendix). We conclude from these tests that our findings are robust to the choice of bandwidth.

Furthermore, we provide robustness tests related to the correction of seasonal patterns, that is the difference-in-difference dimension of our identification strategy. As explained in section 2.3.1, we augment the regression discontinuity design with non-reform years (2003 & 2004) and a DiD to consider the seasonality in disability rates in our main specification (see Figure 3). In that framework, it could be argued that our results depend more on the specificities of the two non-reform years than on the discontinuity in the reform year and that we do not properly wash out the seasonality. To address these concerns, we first look at a "classic" regression discontinuity using a tight bandwidth around the cutoff that only covers children born in the same season (June and July 2002). Figure 6 reveals that the simple RDD with a one-month bandwidth ("RDD BW1") provides highly similar effects compared to the RDD-DiD with the same bandwidth definition ("RDD-DiD BW1"). Table A7 in the Appendix displays results for all outcomes, as well as the two sub-samples of first-time and experienced mothers. Again, the effects are highly similar to the main specification, both at the extensive and intensive margins. Second, we include two additional non-reform years (2005 & 2006) to our main specification (RD-DiD) to capture the seasonality over a total of four years. This comes at the price of reducing the follow-up window to 10 years. The right panel of Figure 6 displays effects for this shorter window. One can see that the reduction in maternal disability days is highly similar when using additional non-reform years (coefficient with diamond shape). Table A8 in the Appendix shows that this is true also for the other outcomes. Appendix 2.5.2 discuss in more details these two robustness checks. Altogether, these additional results reinforce our main findings that paternity leave has a long-term positive impact on maternal health in general, and more specifically on the health of first-time mothers.

Finally, following Avdic & Karimi (2018), we use non-reform years in a "randomization inference design" and perform placebo analyses shifting artificially the reform cutoff by one month at a time. We estimate a placebo intervention 43 times between January 2003 and July 2006 using our RD-DiD design defined in equation (3). We estimate effects on our main outcome, that is the cumulative number of days on disability for mothers, but we restrict the period to 10 years after childbirth, which is the maximum follow-up period in our sample for women who had a child in 2006. Figure A9 in the Appendix shows the distribution of point estimates from this procedure (Panel A) and the cumulative distribution of t-values from the series of regressions (Panel B) compared to a standard normal distribution. The point estimates from the placebo interventions are almost always higher than our estimated effect of -13.7 days (indicated by the doted vertical line) and, as expected, centered around zero ($\overline{\beta}_{placebo}$ = -0.1). Furthermore, we perform normality tests on the empirical distribution of the placebo coefficients (Skewness and kurtosis test), as well as the cumulative empirical distribution of the t-values (Kolmogorov-Smirnov test). Both tests cannot be rejected for any conventional significance level. All in all, these placebo tests also reinforce the robustness of our main findings.

2.3.6 Fertility Decisions and the Effect of Paternity Leave on Maternal Disability

In this section, we explore how changes in fertility patterns could explain our central conclusion, which links paternity leave to a reduction in the time that mothers spend on disability benefits after childbirth. We focus particularly on the role of subsequent children in explaining the paternity leave effect for three reasons. First, we know from the event study analysis that the probability to enter disability on the long-run was higher for women with more children, suggesting a link between the number of children and the consequences of motherhood on health. Two, again because of the event study analysis, we associate a significant portion of the overall increase in the likelihood to become disabled to a particular spike that transpires in the second and third years following the birth of the first child. We attributed this second increase to the arrival of more children in the household suggesting again an association between subsequent births and the long-term health of mothers. Finally, our analysis in the previous section highlights that paternity leave impacts maternal disability most significantly when it is taken after the first child. Women in this context are particular because they are in the middle of their fertility window and have the option to decide if and when they want a second child. Together, these findings point to a potential explanation that links the impact of paternity leave on disability to decisions made in the context of family planning. In what follows, we attempt to provide evidence documenting the prevalence of this mechanism.

And so, we analyze how the two-week paternity leave introduced in 2002 may have impacted birth spacing and family planning fertility decisions overall. We estimate our RD-DiD specification using a series of new outcomes indicating if a mother had a second child within 1, 2, 3, 6, and 12 years after the birth of 2002.[42] We complete this series of one-off indicators with an overall indicator measuring the total number of children in the household twelve years after the birth of the reference child. The entire analysis is carried out on a sample of first-time mothers whose reference child was their first child in 2002. We limit our sample to this group of mothers following the results in Section 2.3.3 showing that the effect of paternity leave on maternal disability is only driven by households from this particular type. Since age is a central factor in explaining fertility decision, we also add that dimension in this analysis by estimating our specification by age.

Table 4 displays the estimated effects for all the fertility-related outcomes broken down by the age of the mother at birth. The table also reports averages of outcome variables for comparison purposes. We do not find any statistically significant evidence that paternity leave changes family planning decision for mothers above the age of 30. We do find, on the other hand, that the probability to have a second child for mothers, aged less than 30 and who had a first child with a father eligible for paternity leave, is about 5 percentage points lower in the second and third year after the first child's birth. This represents a decrease of 13.3% and 9.4% in the probability to have a second child in the second and third year, respectively. These results suggest that mother who had a first child with a father eligible for paternity leave took longer to have another child. In other words, birth spacing between the two first children increased for treated women.

[42]While we can observe the first born for each mother in our dataset based on information from the Belgian register of births, we do not observe exactly the subsequent births from the same mother but rather the year her household had one more child. It could thus also result from adoption or family recomposition.

This effect of paternity leave on birth spacing does not seem, however, to engender any overall effect on the likelihood of having subsequent children. Indeed, 12 years after the birth of the reference child, the size of families in our study is not significantly different between the families that were eligible for paternity leave and the ones that were not. Figure 7 (Panel A) illustrates that dynamic with more details by showing those effects on a quarterly basis. We can clearly discern the statistically significant negative effects on fertility around the second and third years and the gradual convergence to zero afterward. Figure 7 (Panel B) also confirms that the subsequent fertility of mothers aged more than 30 years old is unaffected by the introduction of the paternity leave. We believe that since those mothers are closer to the end of their fertile cycle, they cannot easily adjust their birth spacing to the effects of paternity leave on the household. This is compatible with the findings of Karimi (2014) who shows that "late" motherhood induces women to have a second child more closely spaced with respect to the first one, but not fewer children in total. Overall, our results also echo those of Farré & González (2019) for Spain. They found that a similar reform, that is the introduction of a two-week paternity leave, led to delays in subsequent fertility (Farré & González, 2019)[43]

We argue that the increase in birth spacing propelled by the introduction of the paternity leave could be the main mechanism explaining our results related to disability. This assumption rose from the close match in the timing of the effect of paternity leave on birth spacing and on the number of days on disability. If we go back to Figure 5 in Section 2.3.3, we can clearly see that the number of disability days for first-time mothers (Panel A) starts diverging from zero two years after the birth of the reference child, which is also exactly when birth spacing occurs. We thus think that this change in the number of days on disability could be driven by the delay in the birth of a subsequent child, as the dynamic clearly matches the one in Figure 7 (Panel A). This mechanism is all the more plausible given that our main results were driven by first-time mothers in the middle of their fertility window and who could thus consider increasing the time between their first two children.

Interestingly, these new results also put forward the importance of age in understanding the dynamics of paternity leave. For this reason, we went back to our main results concerning the effect on disability and separated first-time mothers below the age of 30 from first-time mothers above the age of 30 (respectively 61% and 39% of the "first-timer" sample). Reported in Table A9 in the Appendix, all the results seem to indicate that the introduction of the paternity leave particularly affected the disability status of younger

[43] Farré & González (2019), however, found that older mothers had fewer children on average, while we find that their total fertility is unaffected. On the same subject but in the different context, two studies focusing on the Nordic countries did not find an effect on fertility following paternity leave reforms (Cools et al., 2015; Duvander et al., 2020).

first-time mothers (below the age of 30 in 2002), while not affecting the older ones at all. At the intensive margin, it corresponds to a negative effect for these younger mothers that amounts to 44.5 fewer days on disability and 1687 euros less benefits, over a period of 12 years (both statistically significant at a 5% level). Those results suggest again that the same population is driving both the results on fertility and disability.

It could be argued, however, that this explanation of our results on the link between paternity leave and disability are driven by the fact that mothers who delay their subsequent fertility are observed during a longer period of time with a single child since the twelve years follow-up period is indexed on the birth of the first child. In that case, if the event of having a second child increases the probability to enter disability and if this event is delayed, this would mechanically reduce the overall number of days on disability benefits and call into question our whole line of reasoning. To study this conjecture, we estimate the effect of the 2002 paternity leave reform on disability-related outcomes from the birth of the second child. In this new setting, the follow-up period is centered around the birth of the second child and all outcome variables capture disability spells that took place after that birth. This new setting is illustrated in Figure A11. By doing so, we can wash out any mechanical effect due to a difference in the period covered after the second birth. In the same vein, we narrow also the period covered to measure the effect of the reform to 8 years after the birth of the second child to assure that both treated and non-treated mothers are followed over the same period. Finally, we estimate this new specification on a sample of mothers with at least two children who had their first one during the reform year. As we do that, we exclude from our analysis all the mothers for whom by construction we cannot observe any variation in birth spacing.

Table A10 in the Appendix reports the estimates of this analysis. It first shows, as previously stated, no statistically significant effect for mothers above 30. On the other hand, the table shows that mothers, aged less than 30 years old and who had their first child with a father eligible for paternity leave in 2002, spent on average 35 fewer days on disability and received 1198 euros less in disability benefits after the birth of their second child. Interestingly, those effects are close in magnitude to the ones measured above for the whole length of the period following the birth of the first child in 2002. In other words, the effect that we observe in Table A9 for first-time mothers in the reform year and over a period of 12 years seems to match the effect we found here after the birth of the second child. This leads us to conclude that most of the overall reduction in disability occurred after the birth of the second child and that the decrease is driven by mothers who delayed the birth of their second child.

The rationale behind this argument is that increased birth spacing could have improved the health of the mothers, as well as their labor market attachment. Both of these improvements would have long-term consequences on disability. Regarding the first aspect, there is ample evidence from the medical literature that short birth spacing is detrimental to a woman's health. A recent review of 58 observational studies has shown that short intervals between pregnancies were indeed associated with several adverse health conditions (Conde-Agudelo et al., 2012). When it comes to labor market attachment, however, evidence is rather limited and inconclusive. An empirical study on Sweden by Karimi (2014), using miscarriages between the first and second births as an instrument, finds that longer birth intervals have positive long-run effects on income and wage rates. On the other hand, Troske and Voicu (2013), using data for the United-States, show that increasing the time between the first and second childbirth worsens labor market outcomes for mothers by reducing their probability of working full-time.

In conclusion, we provide evidence that increased birth spacing, exogenously induced by the introduction of a paternity leave in 2002, might have lowered the time that women spent on DI in the long run. Of course, the association between the timing of the second birth and disability prevalence is correlational. However, the timing of the two effects match perfectly. In addition, the heterogeneity analyses have shown that the subpopulations driving the results are the same, that is young mothers who had a first child during the reform year. We also ruled out the potential mechanical effects by looking only at disability spells following the birth of the second child. Therefore, we conclude that the increase of time between births is the most likely candidate mechanism for the long-term reduction in disability observed after the introduction of paternity leave.

2.4 Conclusion

This paper examines how parenthood and parental gender impacts the probability of experiencing work disability at a young age. While previous work has highlighted the existence of "child penalties" related to women's earnings, we document another child penalty related to work disability that may prove to be equally important for a woman's career in the long run. Notably, the provision of paternity leave softens this child penalty especially for first-time mothers under the age of 30.

Our study proceeds in the following direction. First, we use an event study approach to provide empirical evidence demonstrating that the incidence rate of work disability for women and men only begins to diverge after the birth of their first child. This gender gap in disability culminates over time so much that even eight years after childbirth women are 40% more likely to experience a disability that prevents them from working at their full capacity. We also demonstrate that the impact of children on maternal health increases with the size of the family, with a gender gap in the probability to suffer disability that raises to 2.3 percentage points in families with three children. We believe that these results provide significant new insights into the career trajectories of mothers and the specific role that gaps due to poor health and disability play on their labor market attachment. In this regard, we estimate that 17% of women who leave the labor market after having children go on to claim DI benefits.

Drawing on this result, we next examine how the provision of paternity leave could moderate this child penalty in disability. We argue that if family arrangements related to childcare and a mother's tendency to fall into disability are linked, then paternity leave provisions, which have been found to increase father's involvement in child raising, could have a positive impact on maternal health. We exploit a discontinuity in Belgian legislation, which offered paternity leave only to fathers of children born after July 1st 2002, to evaluate the causal effect of the policy on the prevalence of work disability among mothers. Following a regression discontinuity difference-in-differences (RD-DiD) design, we find that mothers who gave birth to a child immediately after the reform spent on average 21% fewer days on disability over a period of 12 years. This result seems to be largely driven by younger women who had their first child during the reform year. Lastly, and with regards to the specific causes of maternal disability, our results show that mothers with musculoskeletal disorders spent 50% fewer days on DI in the long term.

In conclusion, we provide suggestive evidence that an increase in birth spacing, induced by the paternity leave reform, could have played a large role in the reduction of the time that mothers spent on disability. We demonstrate that results connected to maternal disability and family planning are driven by the same sub-population of younger mothers who decided to delay the birth of their second child. We also provide evidence that both results exhibit similar time dynamics. This leads us to conclude that the timing of births for multiple-children families is key to reducing the problem of work disability of mothers at young ages.

Our study reveals substantial fiscal spillovers between two social security programs that are usually considered in isolation. On the one hand, we show that mothers receive on average 4,049 euros in DI benefits over the decade following the birth of their child; that is, 468 million euros when extrapolated to the 115,618 mothers who gave birth in 2004 in Belgium. Our intent-to-treat estimates reveal that those mothers who had a child with a father eligible for paternity leave receive on average 712 euros less in DI benefits, which represents a saving of 82 million euros to the Social Security budget. At the same time, 52,848 fathers took a paternity leave in 2004, for a total of 28 million euros in allowances paid by Social Security (INAMI, 2007). Hence, a back-of-the-envelope calculation suggests that spending on paternity leave could be more than compensated by the decade-long savings in mothers' DI benefits.

Recent discussions at the European Union level indicate that our findings could provide useful insights in the context of the work-life balance directive, which was adopted by the European Council on June 13, 2019 and should be implemented in all members states within three years. The directive introduces a paternity leave of 10 days for fathers, which corresponds exactly to the laws currently in place in Belgium, making it a particularly interesting case for research. Our findings show that paternity leave policies might favor a convergence in gender inequalities, while reducing public spending on DI programs. Those results are equally important for countries outside Europe, especially the United-States that has not yet adopted a nation-wide paid leave policy.

Table 1: Descriptive Statistics and Balancing Test

		Sample statistics		**RD-DiD**		
		Mean	SD	Coeff.	SE	Obs.
Household†	Live in flanders (0/1)	0.65	(0.48)	0.00	(0.02)	101735
	Size (#)	3.76	(0.96)	-0.01	(0.03)	101735
	Children (#)	1.72	(0.84)	0.01	(0.03)	101735
	First child (0/1)	0.48	(0.50)	-0.02	(0.02)	101735
Mother	Age†	30.22	(4.14)	-0.11	(0.12)	101735
	Salaried employment (0/1)	0.90	(0.30)	0.02	(0.01)	101735
	Blue collar (0/1)	0.16	(0.37)	-0.02	(0.02)	84993
	White collar (0/1)	0.73	(0.44)	0.02	(0.02)	84993
	Civil servant (0/1)	0.11	(0.31)	0.00	(0.01)	84993
	Self-employed (0/1)	0.09	(0.28)	-0.01	(0.01)	101735
	Daily wage (euro)‡	78.90	(51.03)	0.45	(2.32)	51840
Father	Age†	32.50	(4.89)	-0.05	(0.10)	99502
	Salaried employment (0/1)	0.83	(0.38)	-0.01	(0.01)	99502
	Blue collar (0/1)	0.39	(0.49)	-0.01	(0.01)	77946
	White collar (0/1)	0.52	(0.50)	0.01	(0.01)	77946
	Civil servant (0/1)	0.09	(0.29)	0.00	(0.01)	77946
	Self-employed (0/1)	0.16	(0.36)	0.00	(0.00)	99502
	Daily wage (euro)‡	96.20	(56.12)	-0.34	(1.54)	50765

Notes: Columns 1-2 report means and standard deviations. Columns 3-5 report results from RD-DiD regressions based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). All samples include fathers and mothers who were employed at the time of birth. Standard errors are clustered at birth month level. † Outcomes measured on Dec. 31 of each year. ‡ Outcomes measured the quarter before birth; sample limited to 3 months window. Since the data start in 2002, we cannot observe the outcomes for those who had a child between Jan. and March 2002.

Table 2: Effects of Paternity Leave Reform on Maternal Disability
12 years after Reference Child's Birth

	All mothers		**First-time mothers**		**Experienced mothers**	
	Coeff/SE	Mean	Coeff/SE	Mean	Coeff/SE	Mean
Panel A - Cumulative days on DI	-22.3 **	104.6	-38.6 ***	96.8	-7.4	111.7
	(8.9)		(13.4)		(13.3)	
Short-term (less than 12 months)	-6.3 **	55.4	-9.7 **	53.0	-3.1	57.5
	(3.0)		(4.7)		(3.9)	
Long-term (more than 12 months)	-16.1 **	49.2	-28.9 ***	43.8	-4.4	54.2
	(7.0)		(10.5)		(10.4)	
Panel B - Cumulative DI benefits	-712 **	4049	-1322 ***	3806	-159	4270
	(302)		(431)		(453)	
Short-term (less than 12 months)	-157	2194	-298 *	2146	-29	2238
	(119)		(162)		(150)	
Long-term (more than 12 months)	-555 **	1855	-1025 ***	1660	-130	2032
	(227)		(360)		(349)	
Panel C - Ever on DI	0.005	0.399	-0.005	0.407	0.014	0.393
	(0.012)		(0.014)		(0.017)	
Short-term (less than 12 months)	0.007	0.398	-0.004	0.405	0.016	0.391
	(0.012)		(0.014)		(0.017)	
Long-term (more than 12 months)	-0.009	0.061	-0.021 **	0.056	0.001	0.066
	(0.006)		(0.009)		(0.009)	
Number of observations	101,735		48,505		53,230	

Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample.

Table 3: Effects of Paternity Leave Reform on Maternal Disability (long-term only) - 11 years after Reference Child's Birth
(heterogeneous effects by type of disease)

	All mothers	
	Coeff/SE	Mean
All conditions	-12.4 **	39.8
	(5.9)	
Mental disorders	-0.2	14.8
	(3.9)	
Diseases of musculoskeletal system and connective tissue	-5.7 **	9.6
	(2.7)	
Other	-6.4	15.4
	(5.4)	
Number of observations	101,735	

Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes are for the long-term disability program only. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample. Unlike the previous tables, we restrict the analysis to 11 years after childbirth because there was a change in the ICD classification of diseases in 2016 without any possibility to convert the data of that year to the previous classification system.

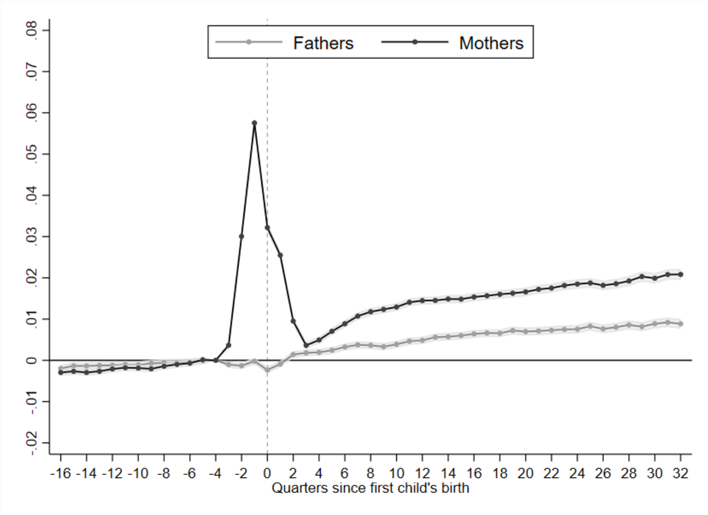
Table 4: Effects of Paternity Leave Reform on Mothers' Subsequent Fertility

	First-time mothers under 30		**First-time mothers over 30**	
	Coeff/SE	Mean	Coeff/SE	Mean
Other child				
After 1 year	-0.016	0.080	-0.021	0.077
	(0.010)		(0.013)	
After 2 years	-0.048 ***	0.360	-0.013	0.309
	(0.016)		(0.021)	
After 3 years	-0.055 **	0.584	-0.011	0.472
	(0.021)		(0.025)	
After 6 years	-0.018	0.775	-0.015	0.607
	(0.017)		(0.029)	
After 12 years	-0.008	0.831	0.003	0.641
	(0.016)		(0.027)	
Nb. children 12 years	-0.056	2.1	0.018	1.8
	(0.037)		(0.056)	
Number of observations	28,449		18,108	

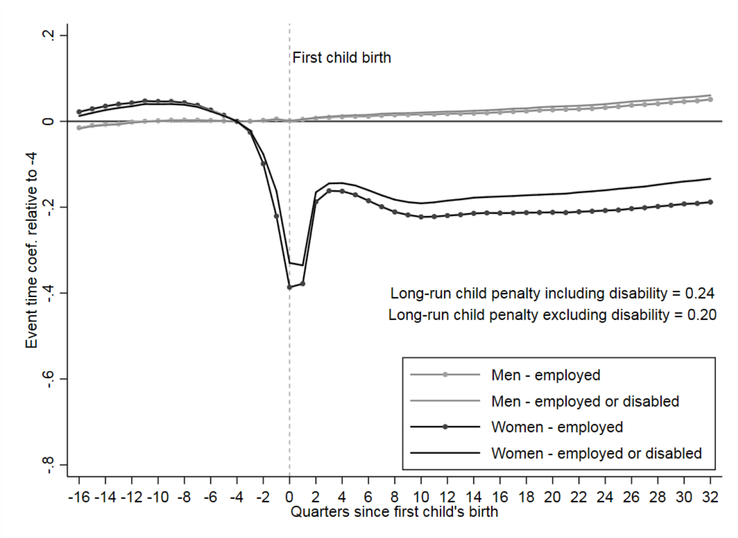
Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Regressions control for mothers' age, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who had a first child between 2002 and 2004 and were employed at the time of the birth of the reference child. The dependent variable "other child" is an indicator for the mother having another child within the following years after the reference child's birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample.

Figure 1: Impact of Children (relative to event time -4)

Panel A: Disability Receipt



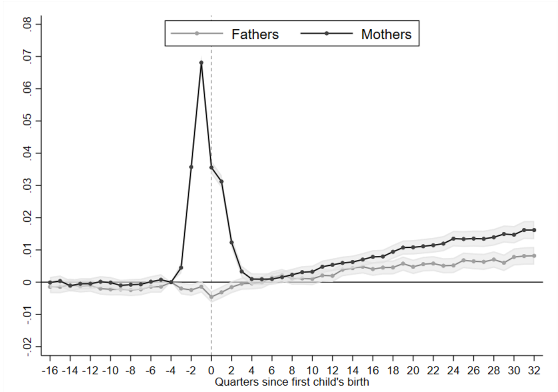
Panel B: Labor Force Participation



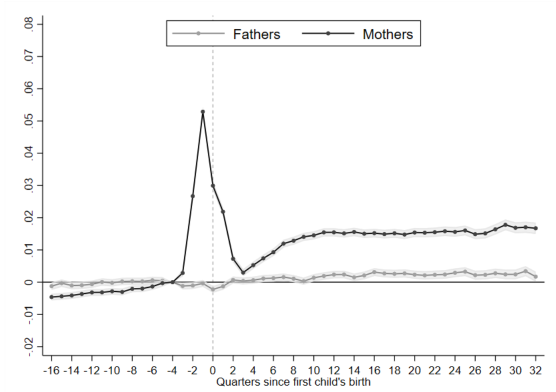
Notes: The figures show event time coefficients estimated from equation (1) relative to the 4^{th} quarter before the first child's birth, for men (gray series) and women (black series) separately. Panel A shows the impact of children on the probability to be on disability insurance (for both the short-term and long-term programs). Panel B displays two different outcomes. The first outcome "employed" takes on the value 1 if the individual receives incomes from salaried employment or self-employment. The second outcome "employed or disabled" takes on the value 1 if the individual receives incomes from salaried employment, self-employment or benefits from the disability insurance system. As such, the second outcome excludes disability from non-employment. The coefficients in panel B are displayed as a percentage of the mean of the outcome measured at t-4. The long-run child penalty - the percentage by which women are falling behind men due to children - is defined as the average penalty across event times 12 to 32. The sample includes all parents who had a first child between 2003 and 2013. The sample includes all parents who had a first child between 2003 and 2013. The shaded 95% confidence intervals are based on robust standard errors.

Figure 2: Impact of Children on Disability Receipt
(relative to event time -4)

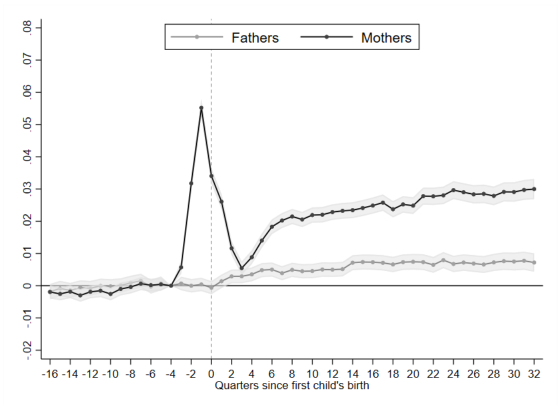
Panel A: One-child Parents



Panel B: Two-child Parents

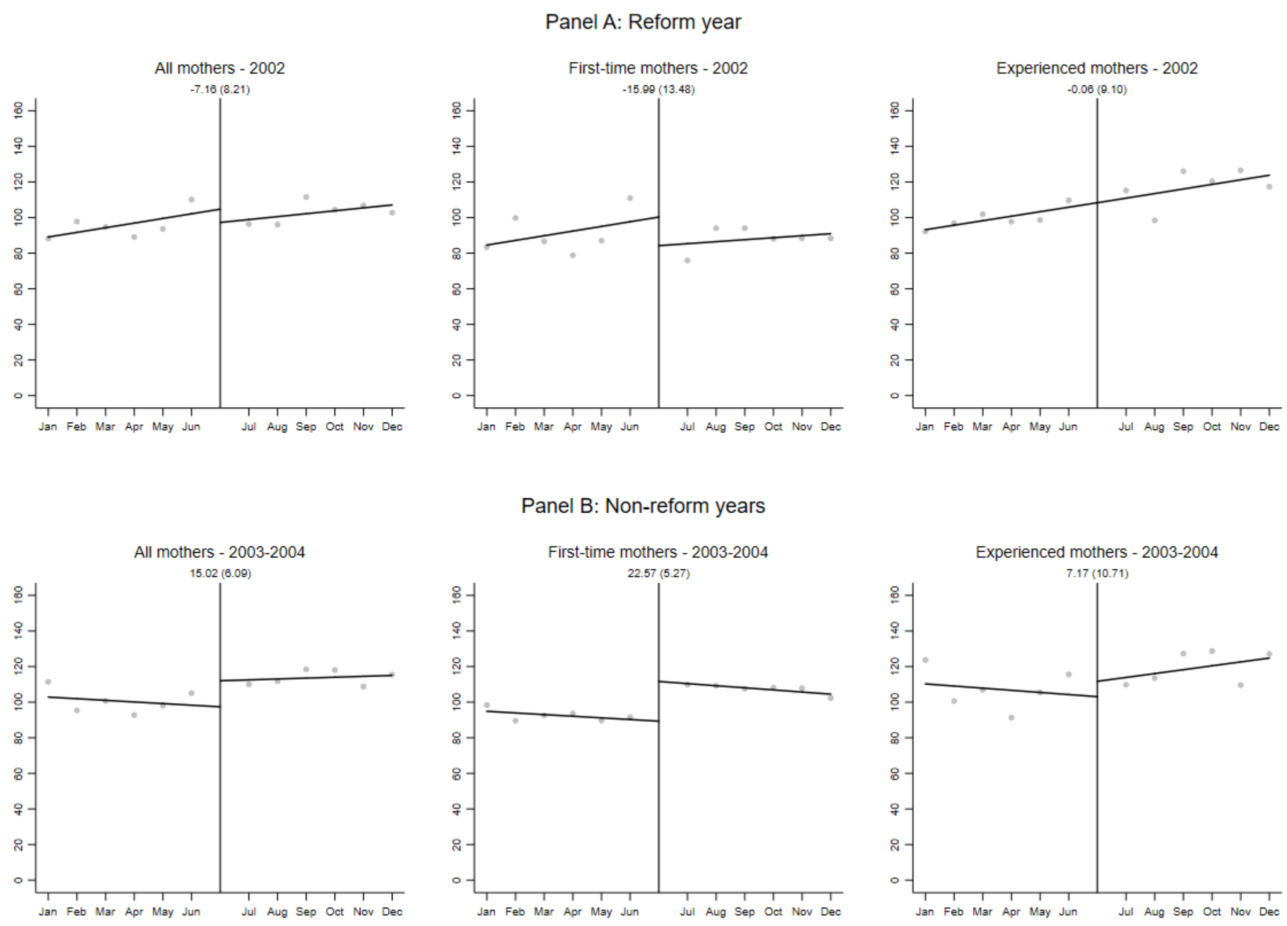


Panel C: Three-child Parents



Notes: The figures show event time coefficients for the probability to be on disability insurance (for both the short-term and long-term programs) relative to the 4^{th} quarter before the first child's birth, estimated from equation (1) for men and women separately. The sample includes all parents who had a first child between 2003 and 2013. For Panels B, C and D, we split the sample by the parents' total number of children as of 2016 (1,2 or 3 children). The shaded 95% confidence intervals are based on robust standard errors.

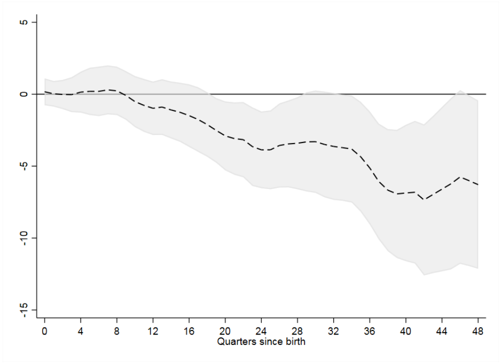
Figure 3: Mothers' Disability Days over the 12-year Follow-up
Comparison Between Reform (2002) and Non-reform Years (2003 and 2004)



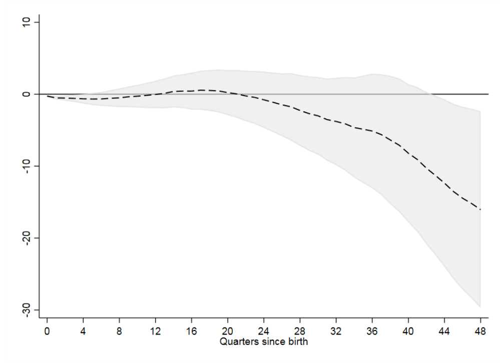
Notes: The figures show the average number of days on DI (for both the short- and long-term programs) for mothers by month of birth of their reference child. The first column includes figures for all mothers who were employed at the time of birth. The second and third columns are sub-samples for first-time and experienced mothers, respectively. Panel A reports figures for the three samples during the reform year (2002), while Panel B reports figures for samples with same characteristics but during non-reform years (2003 - 2004). The vertical bar symbolizes the cut-off in July of each year. On top of each vertical bar are displayed RD estimates based on equation (2) for the linear case and standard errors in parentheses.

Figure 4: Cumulative Effects of Paternity Leave Reform on Mothers' Disability Days

Panel A: Short-term Disability



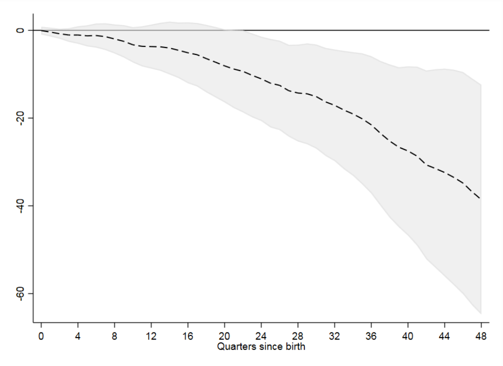
Panel B: Long-term Disability



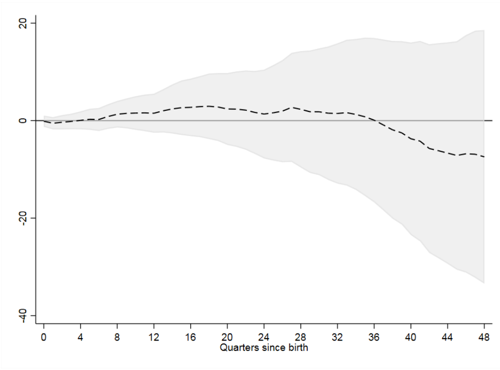
Notes: The figures show RD-DiD estimates from 48 regressions based on equation (3). The sample includes mothers who were employed at the time of birth. The short-term program (Panel A) includes individuals who have spent less than 12 months on DI. The long-term program (Panel B) includes individuals who have spent more than 12 months on DI. The shaded 95% confidence intervals are based on clustered standard errors at birth month level.

Figure 5: Cumulative Effects of Paternity Leave Reform on Mothers' Disability Days (Heterogeneous effects by birth order of reference child)

Panel A: First-time Mothers

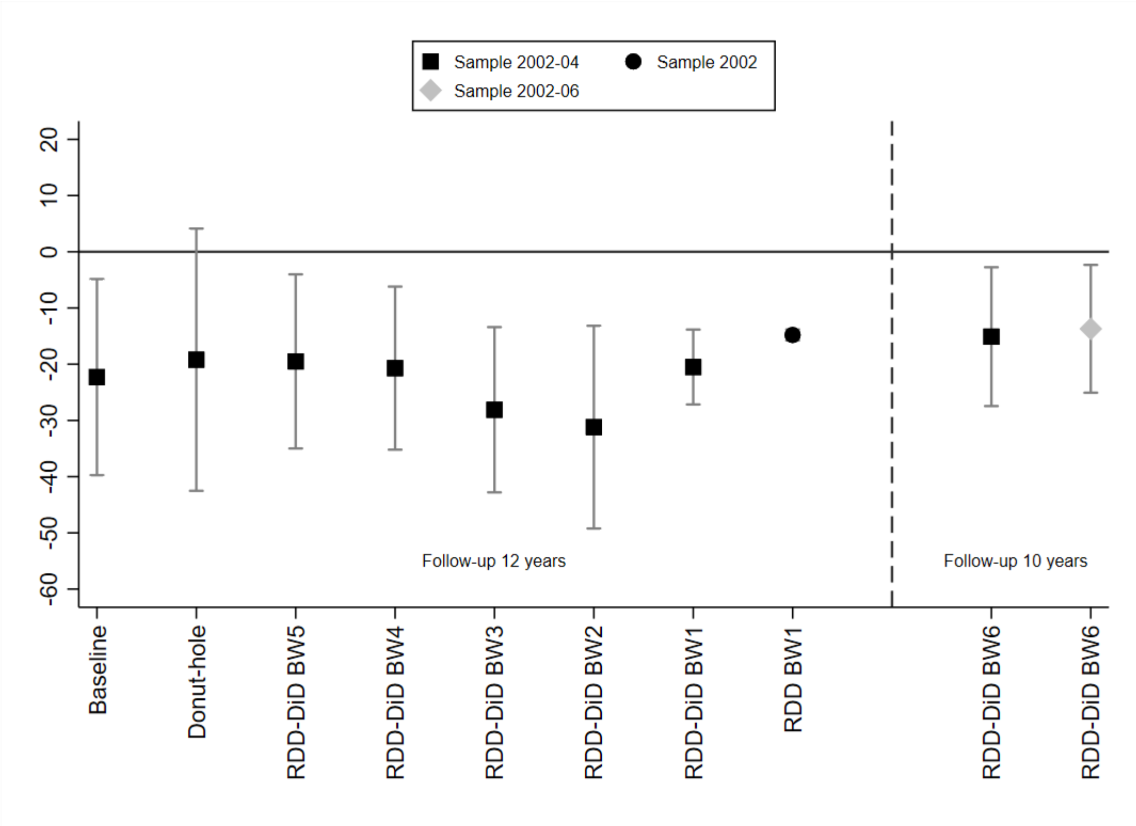


Panel B: Experienced Mothers



Notes: The figures show RD-DiD estimates from 48 regressions based on equation (3). The sample includes mothers who were employed at the time of birth. Results combine effects for both the short- and long-term programs. The shaded 95% confidence intervals are based on clustered standard errors at birth month level.

Figure 6: Cumulative Effects of Paternity Leave Reform on Mothers' Disability Days (Robustness Checks)

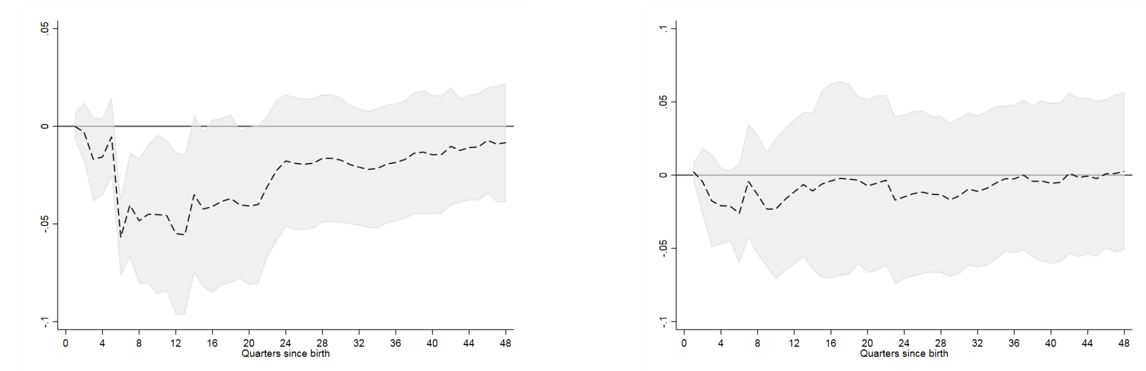


Notes: The figure shows estimates using different empirical specifications for the effect of paternity leave on mothers' cumulative disability days. The "baseline" specification corresponds to the RDD-DiD estimates with a 6 months bandwidth reported in Table 2. We also vary the bandwidth from 5 months ("RDD DiD - BW5") to 1 month ("RDD-DiD BW1") around the threshold, as in Table A6 in the Appendix. The "donut-hole" specification uses a 5-month bandwidth, excluding births that took place one month before and after the cutoff. The "RDD BW1" specification corresponds to a classic regression discontinuity design using a tight bandwidth of 1 month around the cutoff, only including children born in the same season (June and July 2002), similar to Table A7 in the Appendix. The left panel reports cumulative effects over a 12-year follow-up period, while the right panel restricts the follow-up to 10 years. Coefficients reported with a square are computed using the reform year (2002) and non-reform years (2003 -2004), while coefficients reported with a diamond shape are computed using an extended period of non-reform years (2003-2006) as in Table A8 in the Appendix. The sample includes mothers who were employed at the time of birth. The 95% confidence intervals are based on clustered standard errors at birth month level.

Figure 7: Causal Effects of the Paternity Leave on Mothers' Probability to have a Second Child

Panel A: Mothers < 30 years old at the Birth of the Reference Child

Panel B: Mothers ≥ 30 years old at the Birth of the Reference Child



Notes: The figures show RD-DiD estimates from 48 regressions based on equation (3). The sample includes mothers who had a first child during the reform year and were employed at the time of birth. The shaded 95% confidence intervals are based on clustered standard errors at birth month level.

Work Disability after Motherhood and How Paternity Leave Can Help

Sébastien Fontenay, Ilan Tojerow

Appendix

2.5 Appendix

2.5.1 More on the Instrumental Variable (IV) Approach for the Event Study Analysis

Our event study analysis reveals an increased proclivity towards disability for young mothers. We demonstrate the existence of short-term effects, that take place right around childbirth, as well as long-term effects up to 8 years after the arrival of a household's first child. As mentioned in section 2.2, the identification of those effects relies on a smoothness assumption, implying that in the absence of the event the outcomes would evolve smoothly over time. The sharp changes that we observe around childbirth, as well as the absence of pre-trends, provided for a causal interpretation of the short-run effects. However, as Kleven, Landais & Søgaard (2019) rightly point out, the identification of long-run effects requires stronger assumptions. Indeed, as we move further away from the moment of birth, the smoothness assumption becomes less informative. In order to show that our estimates are also valid in the long-run, we provide an additional identification check using an Instrumental Variable (IV) approach. We compare our event study specification with an IV approach using the gender of the first two siblings as an instrument for having a third child, similar to the seminal study of Angrist & Evans (1998). Intuitively, this approach relies on the fact that parents favor mixed sibling-sex composition, so that those who have had two children of the same sex are more likely to decide to have a third one. This strategy should satisfy the exclusion restriction as the preference for mixed sibling-sex should not directly affect the propensity to suffer work disability.

This approach will, however, only capture the marginal effect of having a third child. Thus, we also change our event study analysis, which now measures the effects around the birth of the third child. We follow Kleven, Landais & Søgaard (2019) in adapting the event study design to measure the local impact of a third child for women:

$$\begin{aligned} y_{iqtt't''} = \sum_{j\neq -4} \beta_j \cdot I[j=t] + \sum_k \gamma_k \cdot I[k = age_{iq}] + \sum_y \delta_y \cdot I[y=q] \\ + \sum_{n\neq -4} \zeta_n \cdot I[n=t'] + \sum_{m\neq -4} \eta_m \cdot I[m=t''] + \epsilon_{iqtt't''} \end{aligned} \tag{2.4}$$

The first three terms of equation (4) are exactly the same as those of equation (1), that is a full set of event time dummies with respect to the first childbirth (indexed with t), as well as age and time period dummies. The fourth and fifth terms on the right hand side are new and capture the impact of the second and third child (indexed with t' and t'', respectively). We are primarily interested in the event time coefficients around the birth of the third child, but the event time dummies for the first and second children will help us control for dynamics related to previous children that may matter for the impact of the new child.

As for the main event study analysis in section 2.2, we run this specification on a sample of women who had their first child between 2002 and 2013. This time, however, we select only those mothers who had two or three children when we last observed them in the final quarter of 2016.[44] We follow those parents over a period of 9 years, including up to 4 years prior to the birth of their firstborn and up to 5 years after it.[45]

For the IV specification, we instrument the fifth term of equation (4), that is the event time dummies around the third childbirth, using the gender breakdown of the first two children. Thus, for each event time dummy $I[m = t'']$, we instrument it with the interaction $I[m = t''] \times I[samesexsiblings]$, which takes on the value 1 when the mother is at event time t'' with respect to the third childbirth and her first two children are of the same gender. As such, we follow Kleven, Landais & Søgaard (2019), who have extended the previous IV approaches to account for dynamic effects.

Figure A5 shows the impact of the third child on the probability of their mother to claim DI benefits, compared to the 4$^{\text{th}}$ quarter before that third childbirth. The black line displays OLS coefficients from the updated event study specification, while the gray line show coefficients from the IV approach. The first key insight from this figure is that the estimates from both approaches are almost perfectly aligned. The second key insight is that we observe again a sharp increase in the probability of mothers to suffer from work disability around childbirth, then the effect stabilizes after the first year and remains positive at about 0.0065 (i.e. the average impact across event times 5 to 20). Thus, having a third child has a persistent effect on the probability to be on disability insurance of about 0.65 percentage point. Interestingly, when we estimated the total child penalty for two-child families in section 2.2.3, the long-run effect was 1.4 percentage points. The marginal child penalty for the third child is therefore very close to the combined penalty of the first two.

Taken together, these different elements provide strong support for the validity of our empirical strategy based on individual-level variations in the timing of births. The identification check using an IV approach provides highly similar results and therefore strongly suggests that our event study design causally identified the impacts of children, even in the long-run.

2.5.2 RD-DiD: Additional Identification Checks related to Seasonality

In section 2.3.1, we lay out our main identification strategy to measure the impact of paternity leave on the time mothers spend on DI. There, we explain how we combine a

[44] The sample that we use for this identification check is smaller than the one in section 2.2. Indeed, since we do not have population-level data for all children born in Belgium, but instead a sample of 60% of all births, we can only have detailed information on each child (including birth date and gender) in a two-children family with probability 0.36 (0.6*0.6). Thus, while our sample for the main specification was 359,657 mothers, of which 65% had two or three children by 2016, that is 233,777, our new sample is made only of 87,610 mothers for whom we have detailed information for each child.

[45] The follow-up period is 3 years shorter than the one in the main specification. The reason is that if we were to allow for a follow-up of 8 years, we could only include women who had their third child prior to 2008. Since our sample is made of women who had their first child between 2002 and 2013, we would lose most of our individual observations. We therefore reduce the follow-up period to 5 years to obtain a larger sample and get more precise estimates.

Regression Discontinuity Design, using the birthdate of the child as a running variable, with a difference-in-differences model to account for the seasonality observed in Figure 3. The previous literature has highlighted the importance of considering seasonality for outcomes related to the date of childbirth (Buckles & Hungerman, 2013; Currie & Schwandt, 2013; Cygan-Rehm et al., 2018; Danzer & Lavy, 2017; Dustmann & Schönberg, 2012; Fan, Liu, & Chen, 2017).

In our context, the seasonality could be related to the conditions for school entry and/or to characteristics of new parents that potentially differ over the calendar year. Concerning the school system, Belgian regulations allow children to enter kindergarten when they reach the age of two and a half, but there are only a few dates during the year (after holiday periods) when children can effectively start. This reality also has implications for the childcare system, which can offer more places at some moments of the year (e.g. in September when a large number of children go to kindergarten). For those reasons, some parents may have difficulties finding a childcare spot or may need to wait longer for their child to enter kindergarten, which in turn may affect their ability to work. Regarding potential sample composition effects, previous research has highlighted the existence of systematic differences between parents who have children at different times of the year. Buckles & Hungerman (2013) show that there are substantial differences in maternal characteristics, with younger, less-educated and unmarried women more frequently having winter births.[46] If certain types of parents are more likely to have children in particular months of the year, this might also impact the distribution of disability prevalence across birth months.

For all these reasons, our empirical strategy uses non-reform years to wash out such seasonality in disability through its difference-in-differences dimension. To reinforce the integrity of our research design, we also provide here two additional identification checks. First, we use additional non-reform years, 2005 and 2006, that are further away from the reform cut-off. Second, we provide estimates for a "classic" regression discontinuity design using the smallest feasible bandwidth around the reform date of July 1st, 2002.

Additional non-reform years. Our dataset also includes children born in 2005 and 2006. We can add the parents of those children to our control cohorts and check whether we obtain similar results. This would reinforce the credibility of our assumption that seasonality patterns are stable over time. However, this comes at a cost of reducing the maximum follow-up period, which is only 10 years for children born in 2006. We start with a graphical exploration of the average number of days on disability for all mothers who gave birth in the reform year 2002 and non-reform years 2003-2006, but this time with a 10-year follow-up window. Figure A10 shows again a positive discontinuity that we attribute to seasonality in the non-reform years, while the effect is negative in 2002 when the paternity leave was introduced. In Table A8, we compare the results obtained on our main sample ("cohorts 02/04") and new results with an extended control group ("cohorts 02/06") for a follow-up period of 10 years. We observe that the results on both samples are very much aligned, which suggests that our findings are robust to the inclusion of additional non-reform cohorts and that seasonality patterns seem to be stable across years.

[46] Currie & Schwandt (2013) confirmed this correlation using sibling data for the United-States. Fan, Liu and Chen (2017) find similar evidences from Taiwan and Indonesia.

Simple RDD with tight bandwidth. As suggested by Danzer & Lavy (2018), another way to circumvent the issue of seasonality is to narrow the window of births so as to select only children who are born in the same season. Using only parents of children born very close to the reform date should reduce the issues of school entry conditions and composition effects of parents. As an additional robustness check, we provide a simple regression discontinuity design, based on equation (2), and using the smallest window that is feasible in our dataset that records the month of birth: children born in June and July 2002. For this exercise, we do not include a difference-in-differences component since the issue of seasonality might be less of a problem for the parents of children born very close to each other. Table A7 displays the results for this new empirical strategy using a 1-month window and a simple RDD. We can observe that the results are qualitatively similar to the ones from our main specification in Table 2. This is true for the sample including all mothers, as well as the sub-samples of first-time and experienced mothers.

2.5.3 Additional Tables and Figures

Table A1: Time Use Survey - Belgium, 2013

	Men	Women	Diff.
Paid work	5:01	3:57	- 1:04
Household work	1:54	2:58	+ 1:04
Childcare and raising children	0:33	1:05	+ 0:32
Personal care	2:15	2:24	+ 0:09
Sleep and rest	8:17	8:29	+ 0:12
Education	0:06	0:06	+ 0:00
Social participation	1:14	1:10	- 0:04
Free time	3:13	2:23	- 0:50
Transportation	1:25	1:24	- 0:01
Other	0:03	0:05	+ 0:02

Notes: Household with both parents working and children.

Table A2: Main features of the Belgian parental leave system

	Maternity leave	**Paternity leave**	**Parental leave**
Date introduction	1971	July 2002	1997
Legal texts	”Loi sur le travail”, March 16, 1971	”Loi relative a la conciliation entre l’emploi et la qualité de vie”, August 10, 2001 ”Loi modifiant, en ce qui concerne les coparents, la législation afférente au congé de paternité”, April 13, 2011 ”Loi-programme du 22 décembre 2008”, December 22, 2008	”Arreté royal relatif a l’introduction d’un droit au congé parental dans le cadre d’une interruption de la carrične professionnelle”, October, 29 1997
Duration	Max. 15 weeks (min. 1 before planned delivery + 9 after childbirth)	Max. 2 weeks (initially to be taken within 1 month after childbirth, extended to 4 months in April 2009)	Max. 4 months (per parent per child)
Conditions	Only women who worked min. 120 days in last 6 months	Only fathers (and co-parent in same sex couples since 2011) with salaried contract	Leave may be taken full-time, half-time over 8 months or for one day a week (one-fifth-time) over 20 months Leave may be taken up to the child’s 12th birthday Both parents can take leave at the same time
Replacement rate	82% gross salary (first 30 days) 75% remaining days (capped)	First 3 days fully compensated Remain. 7 days 82% gross salary	Flat rate 802 euros per month

Table A3: Bunching in number of births at the threshold

Window	7 days	14 days	21 days	28 days	35 days	42 days
Log n. of births	0.030	0.029	0.032	0.013	-0.013	0.012
	(0.028)	(0.051)	(0.041)	(0.031)	(0.028)	(0.036)
Linear trend	N	Y	Y	Y	Y	Y
Day of the week	Y	Y	Y	Y	Y	Y
Number of observations	14	28	42	56	70	84

*Notes: This table reports RDD estimates from regressions of the form of equation (2). The outcome variable is the log daily number of births. The reported coefficients are from a binary indicator for birthdates on or after July 1st, 2002. The sample includes all days in the specified window around the date of the introduction of the paternity leave. In all but the first column, we control for a linear trend in date of birth (i.e. the running variable, centered at 0 in July 1st, 2002), interacted with the binary indicator. Robust standard errors are reported in parentheses. Data source: Belgian statistical office - StatBel. *** p<0.01, ** p<0.05, * p<0.1.*

Table A4: Effects of paternity leave reform on paternal disability - 12 years after the reference child's birth

	Coeff/SE	Mean
Panel A - Cumulative days on DI	-2.6	63.8
	(5.968)	
Short-term (less than 12 months)	-4.3	36.9
	(2.749)	
Long-term (more than 12 months)	1.7	26.9
	(4.175)	
Panel B - Cumulative DI benefits	-98.2	2929
	(250.7)	
Short-term (less than 12 months)	-176.4	1844
	(142.7)	
Long-term (more than 12 months)	78.2	1085
	(158.4)	
Panel C - Ever on DI	-0.029 ***	0.312
	(0.0)	
Short-term (less than 12 months)	-0.028 ***	0.311
	(0.0)	
Long-term (more than 12 months)	-0.001	0.034
	(0.0)	
Number of observations	99,502	

*Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for fathers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes fathers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column reports the mean of the outcome for the whole sample. *** p<0.01, ** p<0.05, * p<0.1.*

Table A5: Effects of paternity leave reform on maternal disability - 12 years after reference child's birth
(varying polynomial order)

	Linear Coeff/SE	**Quadratic** Coeff/SE	**Cubic** Coeff/SE
Panel A - Cumulative days on DI	-22.3 **	-20.7 *	-51.5 **
	(8.9)	(11.5)	(19.6)
Short-term (less than 12 months)	-6.3 **	-1.6	-11.1
	(3.0)	(3.2)	(7.2)
Long-term (more than 12 months)	-16.1 **	-19.2 *	-40.4 **
	(7.0)	(10.6)	(16.9)
Panel B - Cumulative DI benefits	-712 **	-466	-1541 *
	(302)	(424)	(782)
Short-term (less than 12 months)	-157	54	-267
	(119)	(147)	(331)
Long-term (more than 12 months)	-555 **	-520	-1275 **
	(227)	(358)	(612)
Panel C - Ever on DI	0.005	0.041 **	-0.057
	(0.012)	(0.019)	(0.035)
Short-term (less than 12 months)	0.007	0.045 **	-0.055
	(0.012)	(0.019)	(0.035)
Long-term (more than 12 months)	-0.009	-0.013 *	0.000
	(0.006)	(0.007)	(0.013)
Number of observations	101,735	101,735	101,735

*Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. *** p<0.01, ** p<0.05, * p<0.1.*

Table A6: Effects of paternity leave reform on maternal disability - 12 years after reference child's birth (varying bandwidth)

	6 Coeff/SE	**5** Coeff/SE	**4** Coeff/SE	**3** Coeff/SE	**2** Coeff/SE	**1** Coeff/SE	**Donut-hole** Coeff/SE
Panel A - Cumulative days on DI	-22.3 **	-19.5 **	-20.7 **	-28.1 ***	-31.2 ***	-20.5 ***	-19.2
	(8.9)	(7.9)	(7.4)	(7.5)	(9.2)	(3.4)	(11.9)
Short-term (less than 12 months)	-6.3 **	-4.5 **	-3.0	-7.3 **	-3.7 *	-3.9 ***	-8.0 *
	(3.0)	(2.1)	(2.1)	(2.7)	(1.9)	(0.2)	(3.9)
Long-term (more than 12 months)	-16.1 **	-15.0 **	-17.7 **	-20.7 ***	-27.4 ***	-16.6 ***	-11.2
	(7.0)	(6.9)	(7.0)	(6.9)	(8.5)	(3.2)	(9.2)
Panel B - Cumulative DI benefits	-712 **	-588 **	-500 *	-833 ***	-925 **	-626 ***	-677
	(302)	(246)	(280)	(282)	(371)	(148)	(425)
Short-term (less than 12 months)	-157	-92	3	-188	-20	-68	-241
	(119)	(86)	(104)	(113)	(118)	(40)	(165)
Long-term (more than 12 months)	-555 **	-496 **	-504 *	-645 **	-905 ***	-559 ***	-436
	(227)	(214)	(248)	(233)	(279)	(109)	(301)
Panel C - Ever on DI	0.005	0.022 **	0.022	0.016	-0.012	0.003	0.003
	(0.012)	(0.010)	(0.013)	(0.014)	(0.013)	(0.006)	(0.017)
Short-term (less than 12 months)	0.007	0.024 **	0.024 *	0.018	-0.008	0.005	0.004
	(0.012)	(0.010)	(0.013)	(0.014)	(0.013)	(0.006)	(0.017)
Long-term (more than 12 months)	-0.009	-0.011 **	-0.011 **	-0.011 *	-0.004	-0.008 ***	-0.011
	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)	(0.000)	(0.010)
Number of observations	101,735	84,987	68,902	51,840	34,784	17,424	84,311

*Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

Table A7: Effects of Paternity Leave Reform on Maternal Disability 12 years after Reference Child's Birth - Simple RDD with 1-month Bandwidth

	All mothers		**First-time mothers**		**Experienced mothers**	
	Coeff/SE	Mean	Coeff/SE	Mean	Coeff/SE	Mean
Panel A - Cumulative days on DI	-14.8 **	103.2	-35.7 ***	93.0	3.3	112.5
	(0.5)		(0.2)		(1.5)	
Short-term (less than 12 months)	-6.5 ***	56.4	-11.0 **	52.0	-2.7 *	60.4
	(0.1)		(0.2)		(0.4)	
Long-term (more than 12 months)	-8.3 **	46.8	-24.7 ***	41.0	6.0	52.1
	(0.4)		(0.0)		(1.1)	
Panel B - Cumulative DI benefits	-349 **	3820	-969 ***	3517	182	4095
	(15)		(7)		(45)	
Short-term (less than 12 months)	-125 ***	2123	-247 **	2028	-27	2210
	(1)		(8)		(7)	
Long-term (more than 12 months)	-224 **	1697	-722 ***	1489	209	1885
	(14)		(1)		(38)	
Panel C - Ever on DI	-0.009 **	0.405	-0.006 **	0.411	-0.012 **	0.399
	(0.000)		(0.000)		(0.001)	
Short-term (less than 12 months)	-0.010 **	0.404	-0.007 **	0.410	-0.013 **	0.398
	(0.000)		(0.000)		(0.000)	
Long-term (more than 12 months)	-0.009 **	0.060	-0.023 ***	0.051	0.002	0.069
	(0.000)		(0.000)		(0.001)	
Number of observations	5,447		2,593		2,854	

Notes: This table reports RD estimates based on equation (2) and using a window of 1 month around the reform cut-off, that is only the mothers of children born in June and July 2002. Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample.

Table A8: Effects of Paternity Leave Reform on Maternal Disability 10 years after Reference Child's Birth - Estimates with Different Cohorts

	All mothers - Cohorts 02/04		**All mothers - Cohorts 02/06**	
	Coeff/SE	Mean	Coeff/SE	Mean
Panel A - Cumulative days on DI	-15.1 **	75.9	-13.7 **	81.2
	(6.3)		(5.8)	
Short-term (less than 12 months)	-6.9 ***	44.0	-6.3 ***	45.8
	(2.4)		(1.9)	
Long-term (more than 12 months)	-8.2	31.9	-7.4 *	35.4
	(4.9)		(4.3)	
Panel B - Cumulative DI benefits	-487 **	2826	-450 ***	3060
	(193)		(167)	
Short-term (less than 12 months)	-202 **	1678	-179 **	1805
	(94)		(69)	
Long-term (more than 12 months)	-285 *	1148	-270 **	1255
	(142)		(125)	
Panel C - Ever on DI	-0.003	0.363	-0.003	0.367
	(0.013)		(0.010)	
Short-term (less than 12 months)	-0.002	0.362	-0.003	0.365
	(0.013)		(0.010)	
Long-term (more than 12 months)	-0.014 ***	0.045	-0.013 ***	0.049
	(0.004)		(0.004)	
Number of observations	101,735		177,194	

Notes: This table reports RD-DiD estimates based on equation (3). The first column "cohorts 02/04" uses the reform (2002) and non-reform years (2003 -2004), while the second column "cohorts 02/06" uses the reform (2002) and non-reform years (2003-2006). Outcomes in Panels A and B capture the cumulative effects over the 10-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample.

Table A9: Effects of paternity leave reform on maternal disability - 12 years after reference child's birth
(heterogeneous effects by age of mother at birth)

	First-time mothers under 30		**First-time mothers over 30**	
	Coeff/SE	Mean	Coeff/SE	Mean
Panel A - Cumulative days on DI	-44.5 **	95.2	-25.6	99.3
	(18.3)		(17.4)	
Short-term (less than 12 months)	-12.0 *	56.6	-4.2	47.4
	(5.9)		(6.2)	
Long-term (more than 12 months)	-32.5 **	38.6	-21.3	51.9
	(13.9)		(14.4)	
Panel B - Cumulative DI benefits	-1687 **	3654	-633	4046
	(644)		(539)	
Short-term (less than 12 months)	-412 *	2205	-65	2053
	(222)		(257)	
Long-term (more than 12 months)	-1274 **	1449	-568	1993
	(480)		(462)	
Panel C - Ever on DI	-0.021	0.437	0.026	0.359
	(0.022)		(0.027)	
Short-term (less than 12 months)	-0.019	0.436	0.027	0.357
	(0.022)		(0.027)	
Long-term (more than 12 months)	-0.023 *	0.054	-0.017	0.058
	(0.012)		(0.012)	
Number of observations	29,648		18,857	

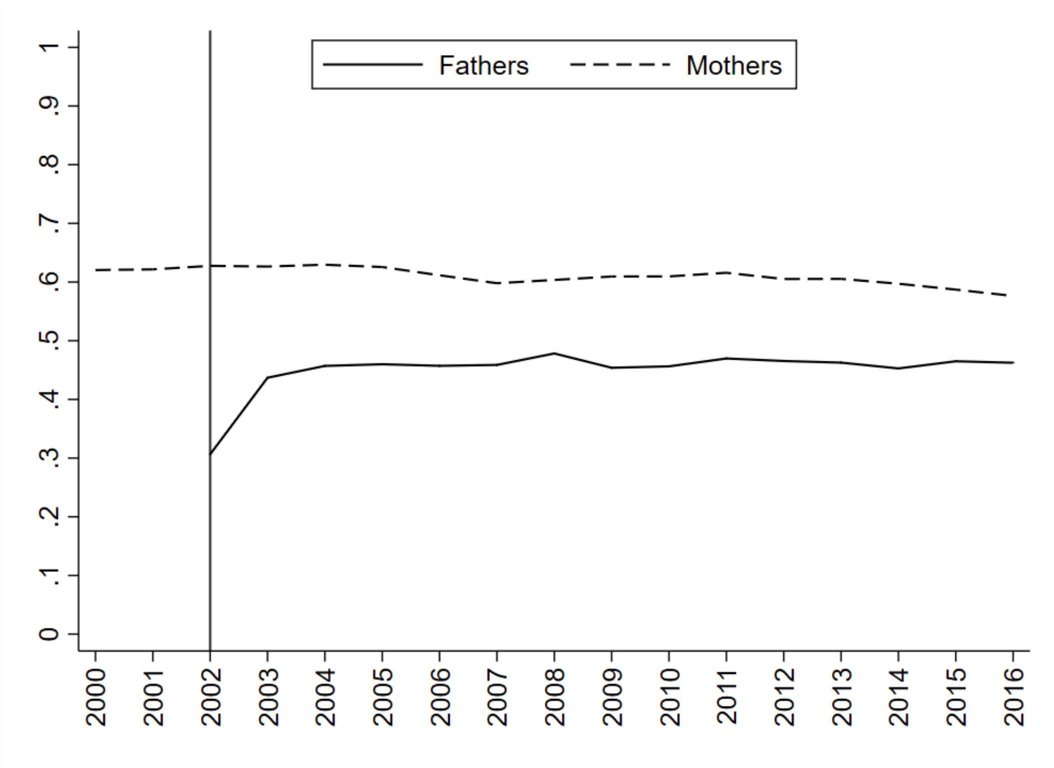
*Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 12-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 12-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who had a first child during the reform year and were employed at the time of birth. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample. *** p<0.01, ** p<0.05, * p<0.1.*

Table A10: Effects of paternity leave reform on maternal disability - 8 years after the second child's birth

	First-time mothers under 30		First-time mothers over 30	
	Coeff/SE	Mean	Coeff/SE	Mean
Panel A - Cumulative days on DI	-34.8 **	55.9	4.5	49.7
	(16.9)		(13.7)	
Short-term (less than 12 months)	-10.7	35.6	-0.4	27.1
	(7.3)		(4.7)	
Long-term (more than 12 months)	-24.2 **	20.4	4.9	22.6
	(10.9)		(11.0)	
Panel B - Cumulative DI benefits	-1198 **	2095	508	2038
	(528)		(579)	
Short-term (less than 12 months)	-367	1368	99	1196
	(251)		(216)	
Long-term (more than 12 months)	-831 **	726	409	842
	(322)		(450)	
Panel C - Ever on DI	-0.010	0.329	0.049	0.254
	(0.021)		(0.038)	
Short-term (less than 12 months)	-0.011	0.328	0.046	0.251
	(0.020)		(0.039)	
Long-term (more than 12 months)	-0.020	0.035	-0.003	0.034
	(0.015)		(0.010)	
Number of observations	21,646		10,909	

*Notes: This table reports RD-DiD estimates based on equation (3) and using the reform (2002) and non-reform years (2003 -2004). Outcomes in Panels A and B capture the cumulative effects over the 8-year period for the number of days and benefits, respectively. The variables labeled "Ever on DI" in Panel C are dummies for the probability to have entered disability insurance at least once over the 8-year period. Regressions control for mothers' age, number of children, as well as region of living, at the moment of the birth of the reference child. The sample includes mothers who had a first child during the reform year, were employed at the time of the first birth, and had at least another child in the years following the reform. Standard errors (reported in parentheses) are clustered at birth month level. The second column of each vertical panel reports the mean of the outcome for the whole sample. *** p<0.01, ** p<0.05, * p<0.1.*

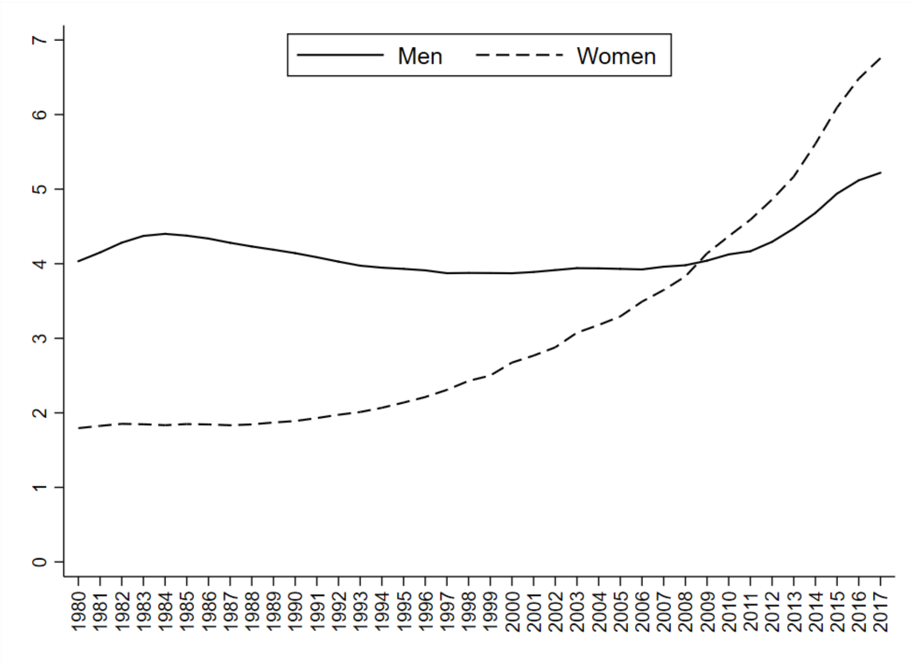
Figure A1: Number of fathers/mothers taking paternity/maternity leave as a fraction of the annual number of births



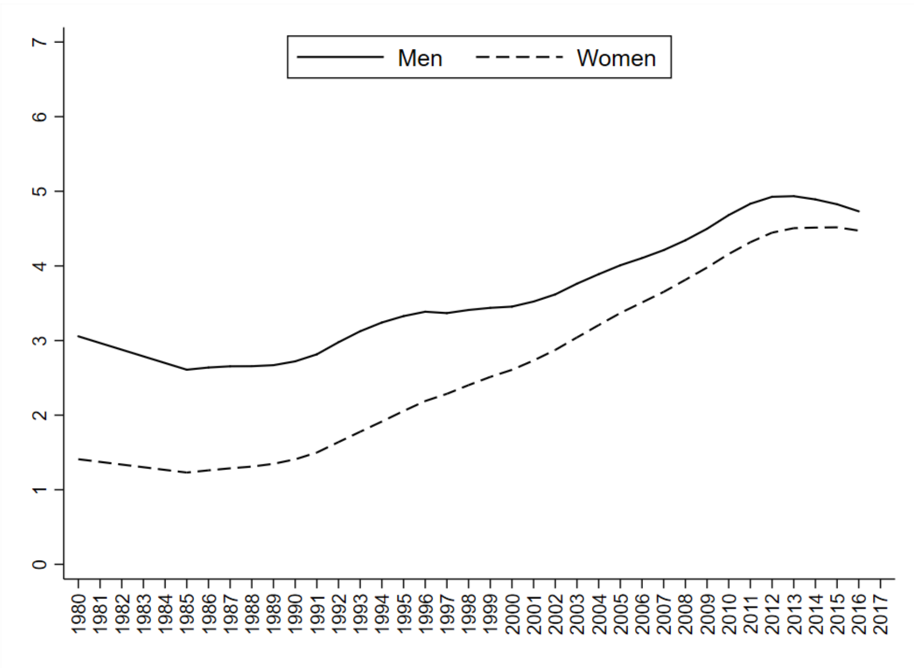
Notes: Statistics for maternity leave do not include civil servants and self-employed workers who benefit from a different system. It should also be noticed that women who have not worked at least 120 days during the last 6 months are not entitled to maternity leave. For all these reasons, only 60% of women on average are reported to have taken a maternity leave over the last two decades. Statistics for paternity leave do not include fathers who stop working for only 3 days or less since they need only to report to their employer. One should also keep in mind that those statistics do not account for civil servants, who benefit from a different system, as well as self-employed workers who were not entitled to paid paternity leave before 2019. For the year of the reform, we only consider births from July to December 2002. Data sources: National Institute for Health and Disability Insurance (leave-takers) and StatBel (births).

Figure A2: Percent of working-age (20–64) population receiving (long-term) DI benefits

Panel A: Belgium



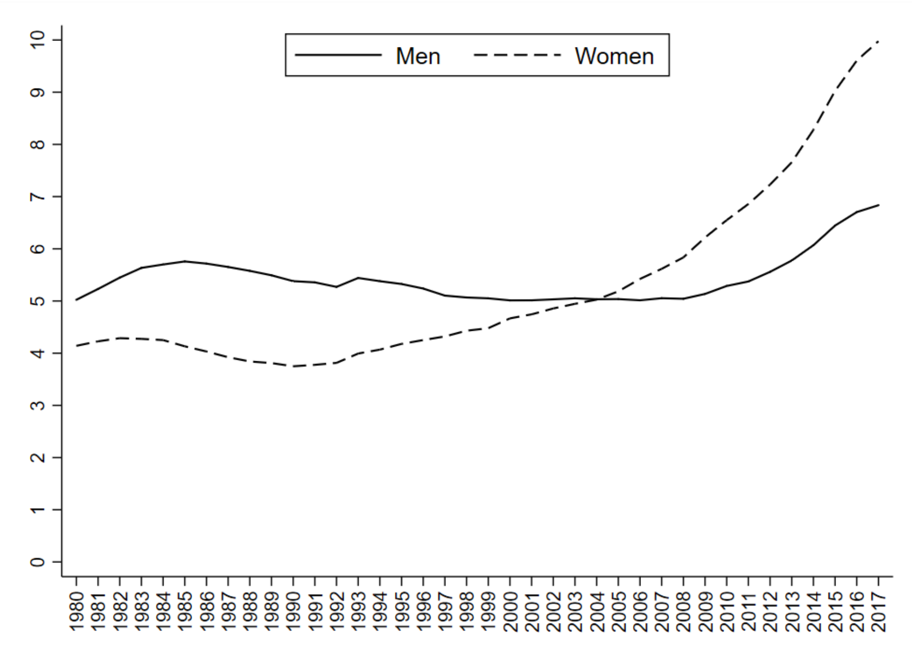
Panel B: United States



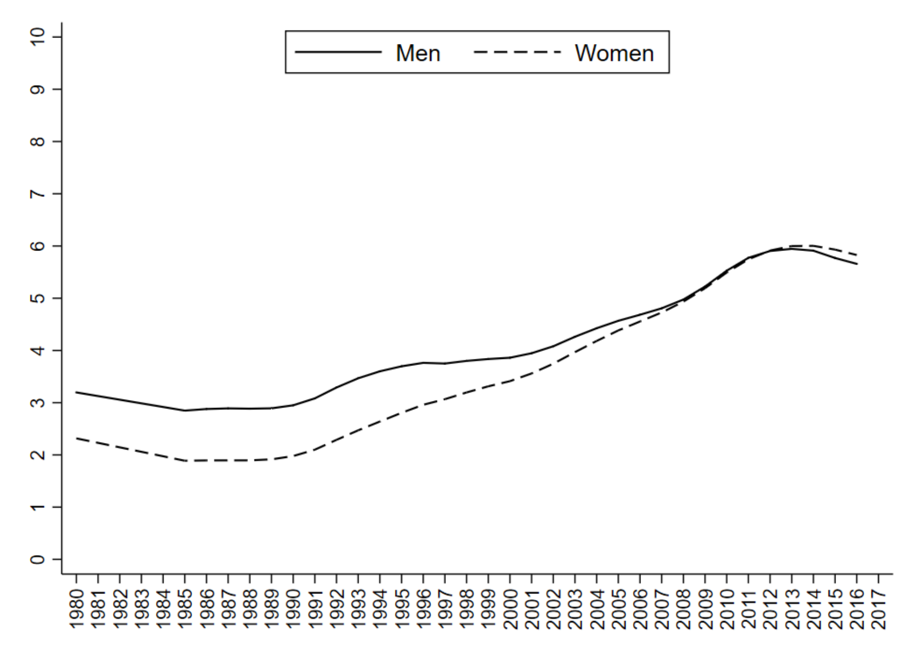
Data source for Belgium: National Institute for Health and Disability Insurance & OECD. Data source for the United States: Social Security Administration, 2017 Annual Statistical Supplement & OECD.

Figure A3: Percent of insured workers receiving (long-term) DI benefits

Panel A: Belgium

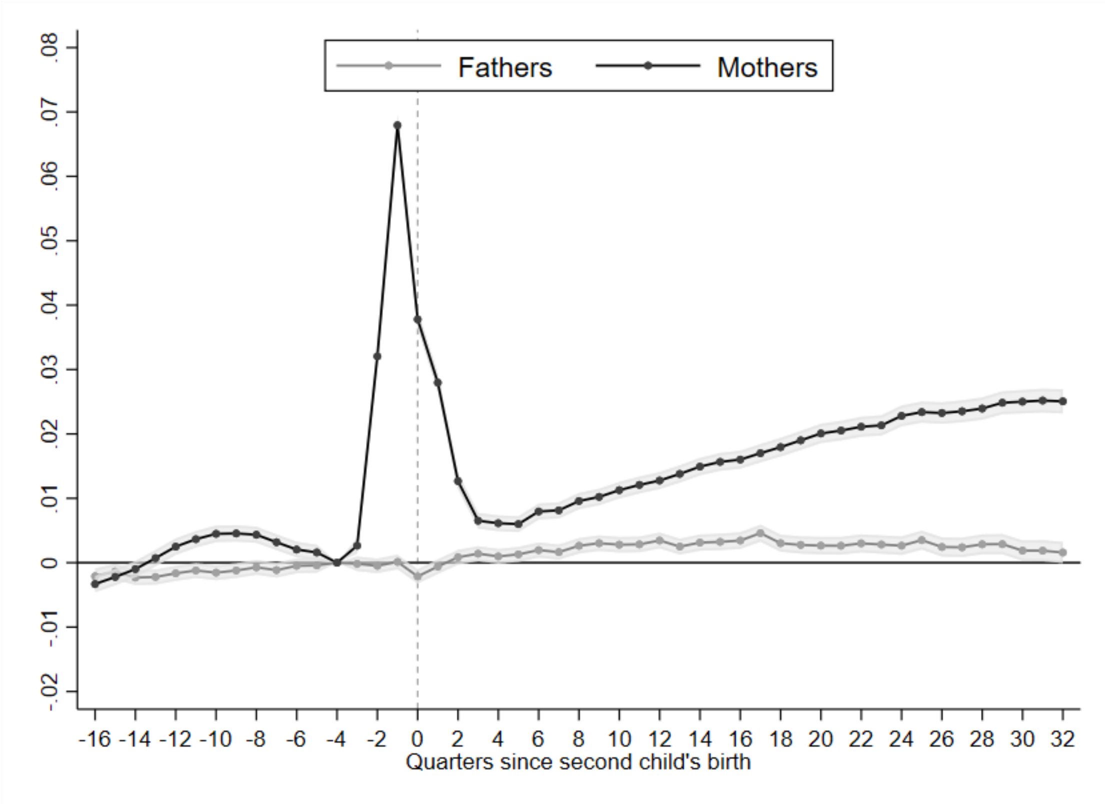


Panel B: United States



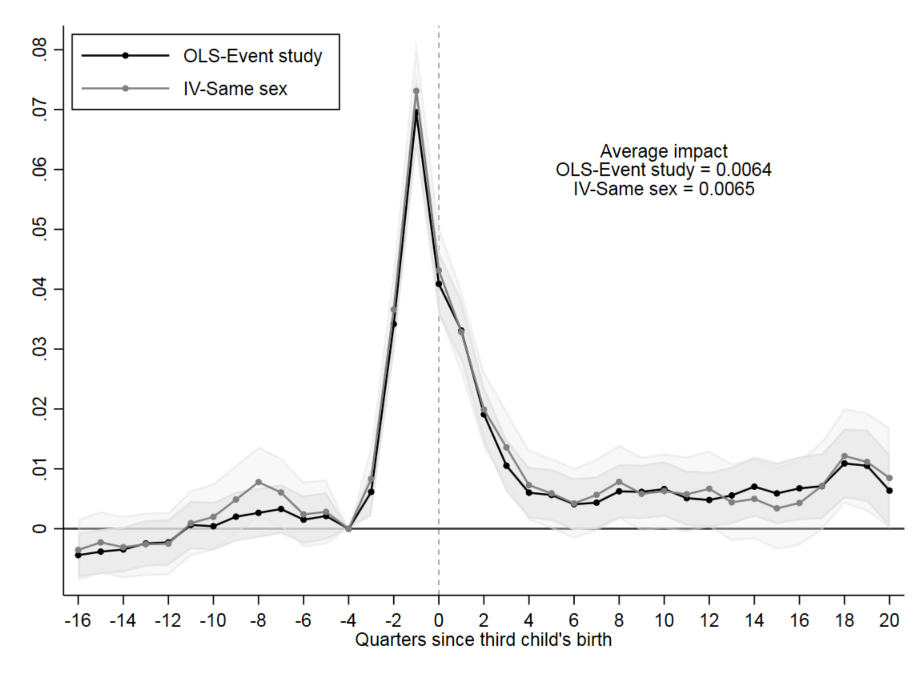
Data source for Belgium: National Institute for Health and Disability Insurance. Data source for the United States: Social Security Administration, 2017 Annual Statistical Supplement.

Figure A4: Event study around second child's birth - Impact on disability receipt (relative to event time -4)



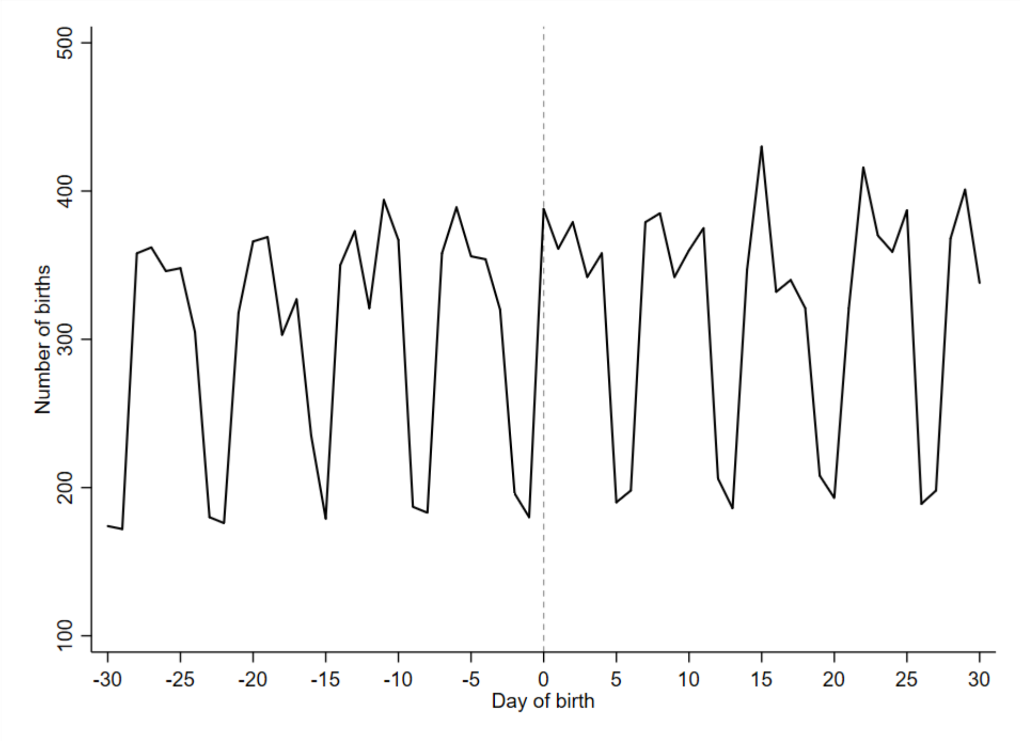
Notes: The figure shows event time coefficients for the probability to be on disability insurance (for both the short-term and long-term programs) relative to the 4^{th} quarter before the second child's birth, estimated from equation (1) for men and women separately. t=0 is now the quarter of birth of the second child. All of these statistics are estimated on a sample of parents who have had two children in total as of 2016. The shaded 95% confidence intervals are based on robust standard errors.

Figure A5: Impact of Third Child on Disability Receipt
Comparison of Event Study and IV Estimates



Notes: The figure shows event time coefficients for the probability to be on disability insurance (for both the short-term and long-term programs) relative to the 4th quarter before the third child's birth, obtained from the OLS event study specification (black series) and the IV samex-sex specification (gray series). The sample includes all mothers who had a first child between 2003 and 2013 and who have completed fertility of two or three when we last observe them in the last quarter of 2016. The shaded 95% confidence intervals are based on robust standard errors. The figure also compares the average impact across event times 5-20 obtained from the event study and the IV approach.

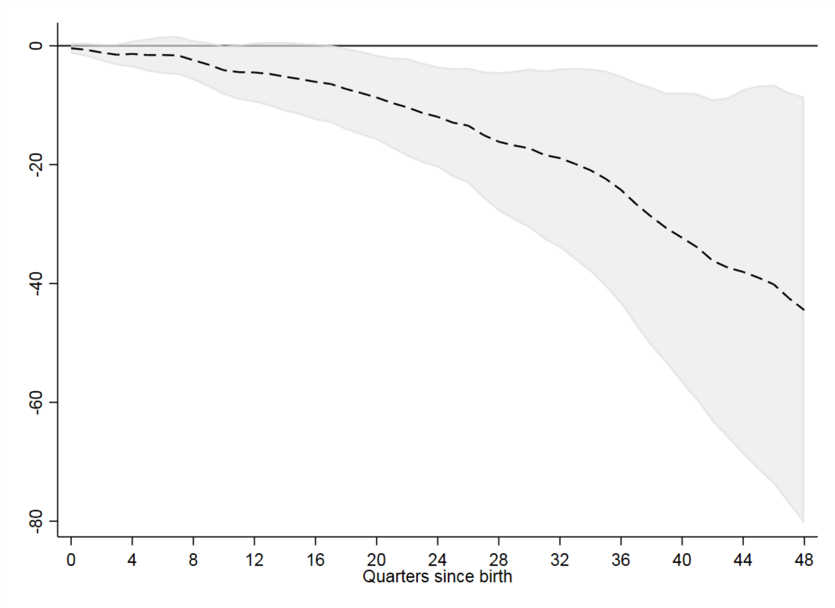
Figure A6: Daily number of births



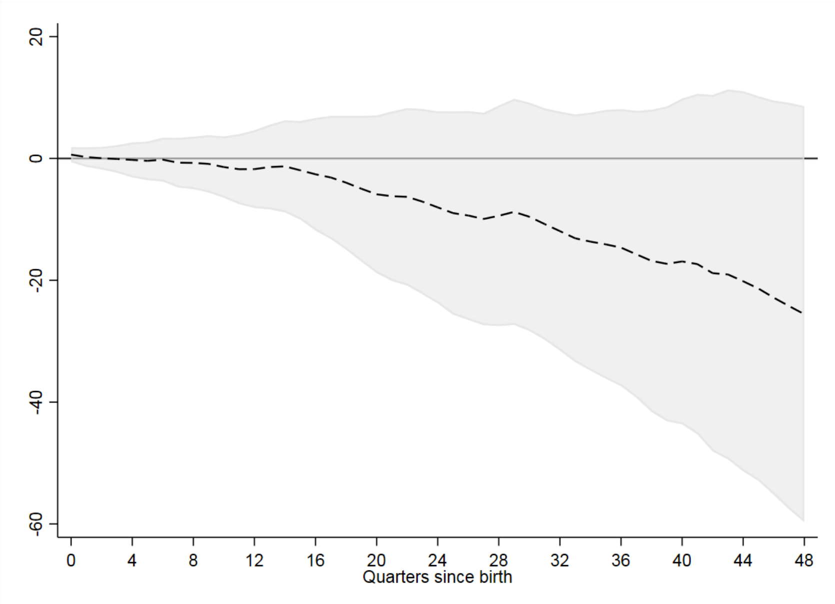
Notes: Daily number of births around the introduction of paternity leave. The day of birth is normalized to 0 for July 1st, 2002. Data source: Belgian statistical office - StatBel.

Figure A7: Cumulative effects of paternity leave reform on mothers' disability days (heterogeneous effects by age of mother at birth of the reference child)

Panel A: First-time mothers < 30 years old



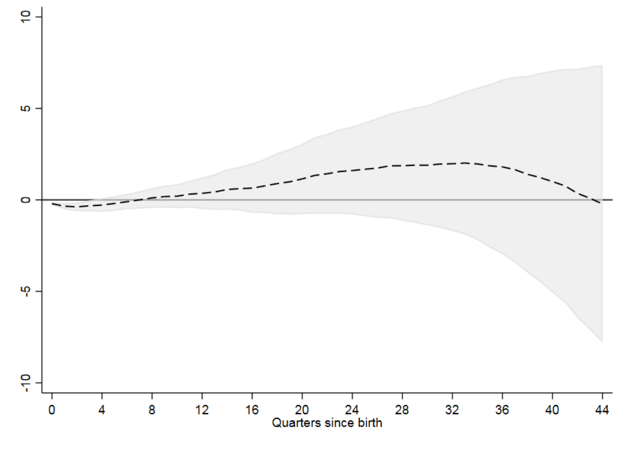
Panel B: First-time mothers ≥ 30 years old



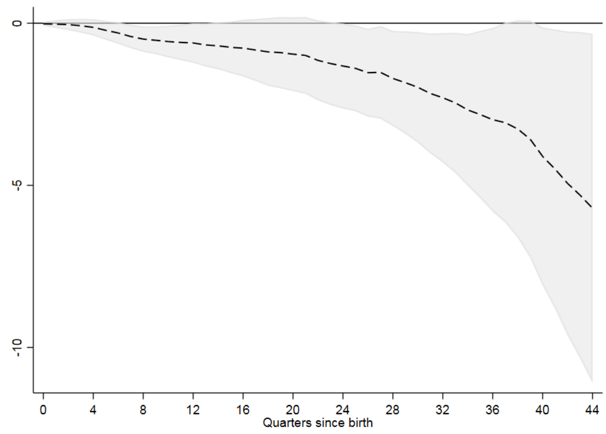
Notes: The figures show RD-DiD estimates from equation (3). All of these statistics are estimated on a sample of mothers who had a first child between 2002 and 2004 and were employed at the time of the birth of the reference child. The shaded 95% confidence intervals are based on clustered standard errors at birth month level.

Figure A8: Cumulative effects of paternity leave reform on mothers' long-term disability days
(heterogeneous effects by medical condition)

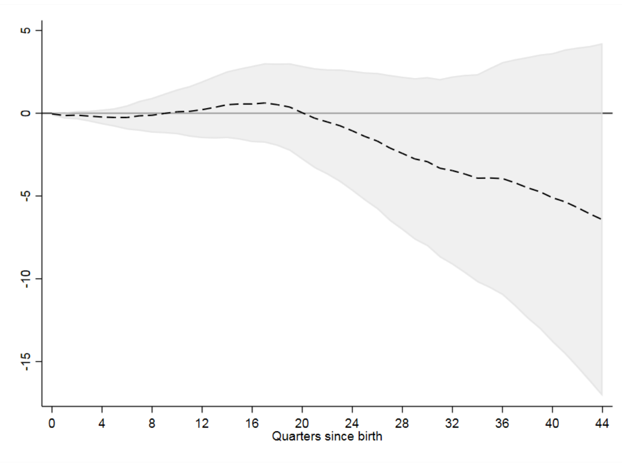
Panel A: Mental and behavioral disorders



Panel B: Musculoskeletal system or connective tissue diseases



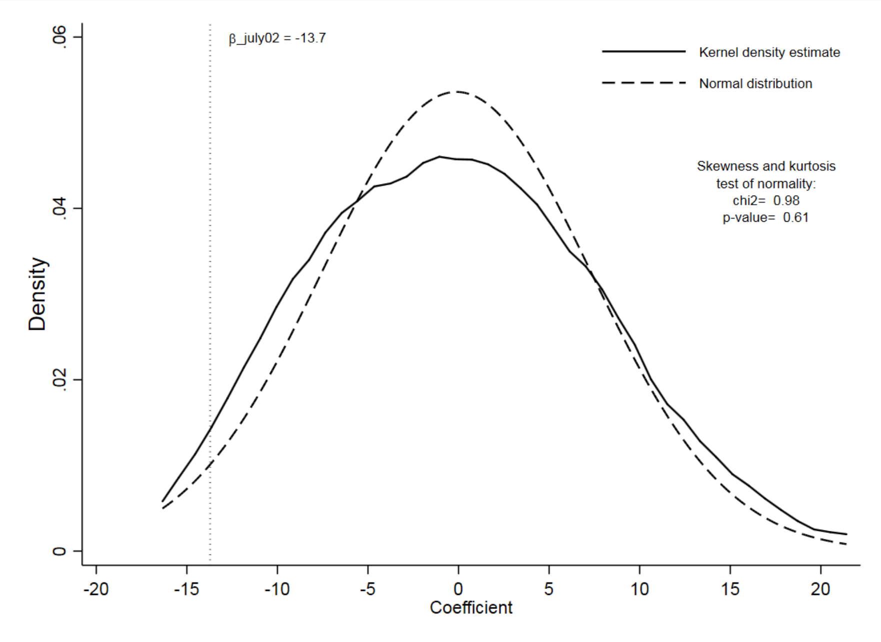
Panel C: Other diseases



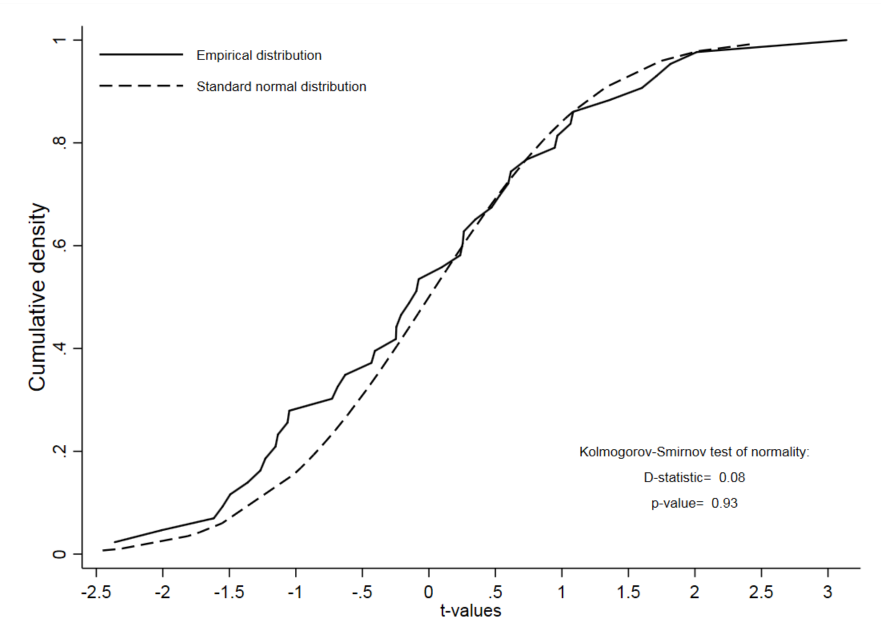
Notes: The figures show RD-DiD estimates from 48 regressions based on equation (3). Outcomes are for the long-term disability program only. All of these statistics are estimated on a sample of mothers who had a child between 2002 and 2004 and were employed at the time of the birth of the reference child. The shaded 95% confidence intervals are based on clustered standard errors at birth month level. Unlike the previous graphs, we restrict the analysis to 11 years after childbirth because there was a change in the ICD classification of diseases in 2016 without any possibility to convert the data of that year to the previous classification system.

Figure A9: Placebo estimates for mothers' cumulative disability days after 10 years

Panel A: Placebo estimates

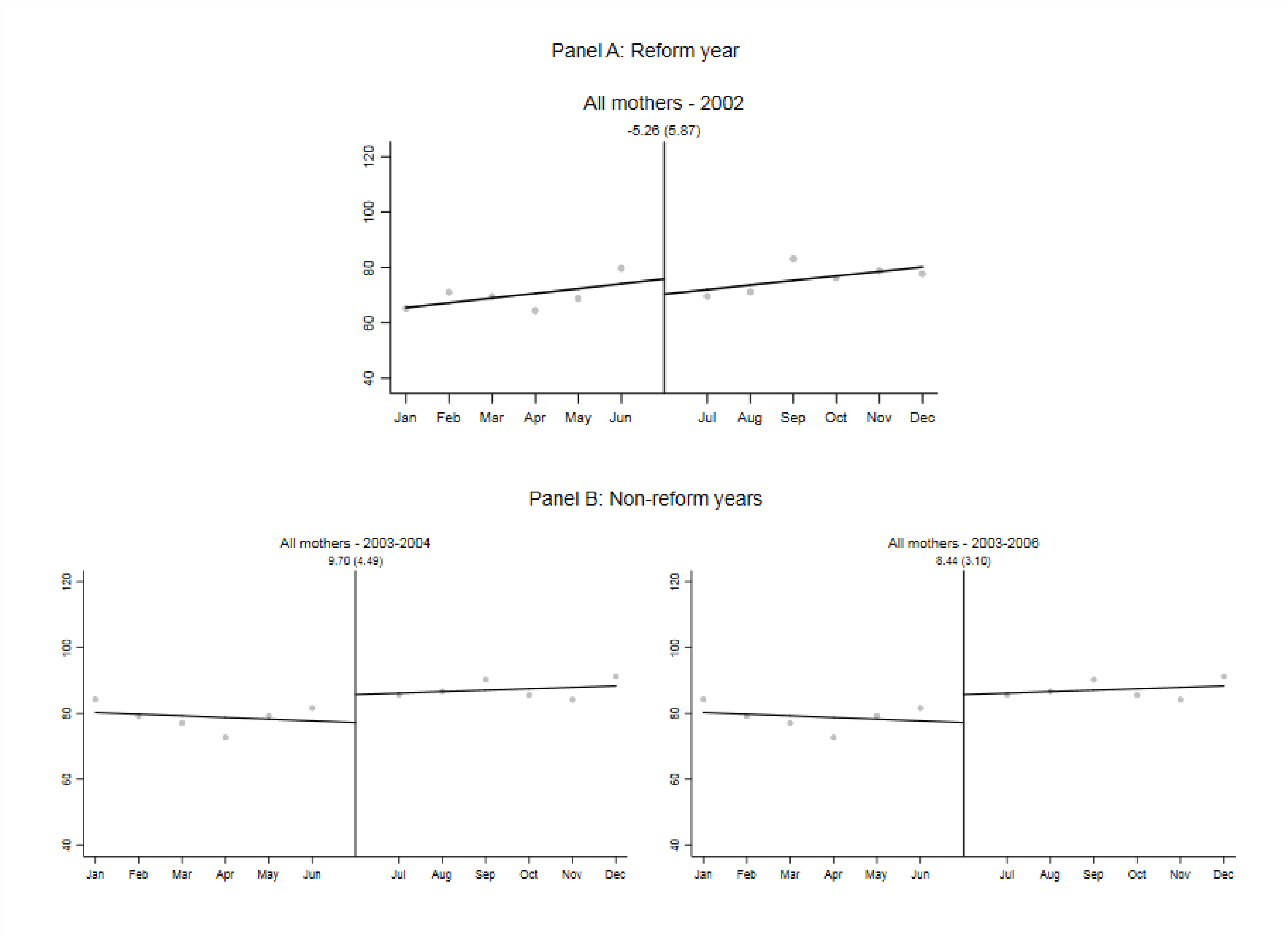


Panel B: t-values from placebo estimates



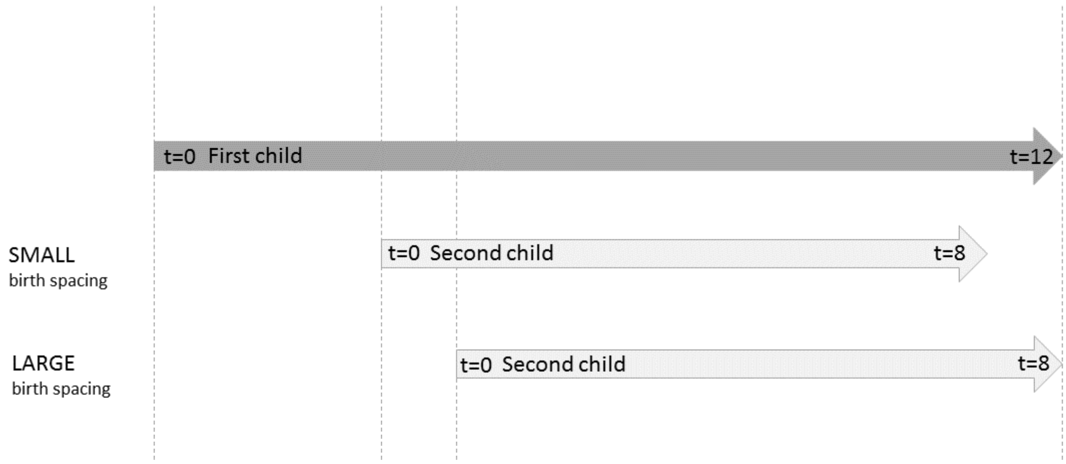
Notes: The figures show RD-DiD estimates from 43 regressions based on equation (3). All of these statistics are estimated on a sample of mothers who had a child between 2002-2006 and were employed at the time of the birth of the reference child.

Figure A10: Mothers' Disability Days over the 10-year Follow-up
Comparison Between Reform (2002) and Non-reform Years (2003-2006)



Notes: The figures show the average number of days on DI (for both the short- and long-term programs) for mothers by month of birth of their reference child. Panel A displays results for the reform year (2002), while Panel B displays results for non-reform years (2003 - 2004) in the left column and non-reform years (2003 - 2006) in the right column. The vertical bar symbolizes the cut-off in July of each year. On top of each vertical bar are displayed RD estimates based on equation (2) for the linear case and standard errors in parentheses.

Figure A11: New follow-up period indexed on second child's birth



Chapter 3

How Does Job Coaching Help Disability Insurance Recipients Work While on Claim?

This article is co-authored with Ilan Tojerow. This study is listed on the AEA registry (AEARCTR-0003378). This research has benefited from the support of INAMI/RIZIV, including Clara-Maria Arbesu Gonzalez, Ellen Wyffels, François Perl, Frank Seynaeve, Isabelle Fransolet, Jean Alves, Kirsten Van Kelst, Marc Casteleyn, Ronny Meert, Saskia Decuman and Thomas Otte. We also thank Actiris, Article 23 asbl, Forem, ISPPC and VDAB who provide support to DI recipients in the field. A special thank also to Natasia Hamarat and Vanessa De Greef. We thank Basile Dabin for excellent research assistance. Finally, we are grateful to Justine Bondoux and William Parienté for very helpful comments and suggestions, as well as seminar participants at Universitat Pompeu Fabra, 43e Journées des Economistes de la Santé Français, CEBRIG Université Libre de Bruxelles, Royal Economic Society 2022, SOLE 2022, 21st Journées Louis-André Gérard-Varet.

Introduction

In recent decades, the number of individuals receiving Disability Insurance (DI) benefits has increased substantially in OECD countries, creating an important challenge for social security funding (OECD, 2010). In Belgium, where this field experiment takes place, the share of working-age population receiving DI benefits from the long-term program reached 6% in 2017.[1] Among the causes associated with this trend, we observe a rise in beneficiaries with mental conditions (OECD, 2010). As argued by Autor and Duggan (2006), mental disorders are characterized by early onset and low mortality, which increase the average duration of disability spells and hence the size of the recipient population. As a result, the fraction of DI beneficiaries diagnosed with mental disorders has increased over time to represent 35% in Belgium (Gerritse, Marcato, Plasman, & Tojerow, 2017).[2]

The rise in mental disorders has raised questions about what the boundaries are between individuals deemed totally and permanently disabled and others who retain some work capacity or can recover it in the future. Indeed, mental illnesses are often characterized by changing productivity levels over time (Kessler et al., 2006, 2008). For this reason, Bound and Burkhauser (1999) believe that since mental health is more difficult to monitor, individuals with mental health conditions would be on average "healthier" and would have worked in the absence of DI schemes. Previous studies have supported this claim, showing that DI beneficiaries have substantial remaining capacity to work (Bound, 1989; French & Song, 2014; Maestas et al., 2013). Maestas, Mullen, and Strand (2013) use examiner assignment as an instrumental variable to show that the employment rate of DI recipients would have been 28 percentage points higher had they not received benefits, with an effect that reaches 37 percentage point for applicants with mental disorders.[3]

The fact that a substantial fraction of DI beneficiaries retains some capacity to carry on economic activities motivates the implementation of active labor market programs to help them return to work. In this study, we evaluate the effects of a new Supported Employment (SE) program aimed at DI beneficiaries with mental conditions. The program is characterized by a "work-first" approach that includes intensive job coaching. To this end, caseworkers have a maximum of 20 beneficiaries at the same time, which allows them to hold frequent meetings (at least every two weeks). The SE program also offers follow-along support after a job has been found, with the aim of helping beneficiaries secure

[1] In comparison, the share of working-age population in the United-States receiving DI benefits from the federal program increased from 2.2% in the late 1970s to 4.6% in 2013 (Liebman, 2015).

[2] In comparison, the share of DI recipients with mental disorders was 25.4% in the United States (Autor & Duggan, 2006).

[3] French and Song (2014) find that DI benefit recipients reduced labor force participation rates by 26 percent in the United States, but that the labor supply of individuals with mental health conditions was less sensitive to benefit recipients (20 percentage points).

long-term employment. Using a Randomized Control Trial, we compare the effects of this newly introduced program with regular Vocational Rehabilitation (VR), used for more than a decade in Belgium (i.e., control condition). Regular rehabilitation services favor "human capital acquisition" through vocational training before attempting a return to the labor market. In addition, regular VR offers support that is limited in both duration and intensity, with caseworkers providing services to about 100 beneficiaries. Table 1 summarizes how the two return-to-work programs differ in focus, intensity, and duration.

Between March 2018 and December 2019, we recruited more than 660 DI beneficiaries from across Belgium who suffered from mental illnesses and were willing to take part in a return-to-work program. Participants were randomly assigned to the new SE program or the regular VR program. We followed them for 18 months, from the start of their program, using data from administrative registers and survey instruments designed for this research.

We find that DI recipients with mental conditions who were randomly allocated to the new SE program are twice as likely to work while on claim compared to those in regular rehabilitation. Specifically, we estimate that 18 months after the start of their return-to-work program, participants in the SE group are 9.5 percentage points more likely to be working part-time than those in the control group. Using survey responses, we observe that they find occupations mostly in the private sector (and not in sheltered workshops). At the same time, increased employment in the SE group translates into a reduction in benefits, paid by the DI, of 6%. In addition, we find that SE was effective in helping DI beneficiaries find and retain a job through the economic turmoil that resulted from the COVID-19 pandemic.

Our rich survey instrument, administered every six months, allows us to explore the channels through which SE achieves a higher re-employment rate than regular rehabilitation, up to 18 months after the start of the program. We find that participants in SE are more likely to be actively looking for a job and dedicate more time to their search (about 2.5 hours per week at follow-up 12) and less time to vocational training (two weeks less, that is, a reduction of about 50% compared to the control group). In addition, DI recipients in SE report that they seek a job with fewer working hours and lower pay (about 10% lower than those in regular rehabilitation). We believe that these differences reflect the fact that SE participants have formed more realistic expectations about their true ability to work, which might only be part-time, while those in regular rehabilitation might seek to fully reintegrate into the labor market. Finally, our survey reveals that SE does not seem to have unintended consequences for the health, perceived self-esteem, or self-efficacy of program participants.

We also take advantage of detailed information provided by the Social Security Administration on their expenditures on both programs in order to perform a cost-benefit analysis. This is particularly important given that the intense supervision provided in the SE program also means higher cost per participant. When taking into account the benefits for the Social Security Administration (in terms of budget saving from the reduction in DI benefits), our most conservative estimates show that the gap between the cost of the two programs would be closed in less than two years. While there is no guarantee that the estimated effects will last beyond the 18-month follow-up period, a dynamic analysis on a monthly basis does not suggest that participants in the control group catch up in terms of their employment rate between follow-ups 12 and 18.

There are only a handful of studies on the effects of programs aimed at helping DI recipients return to work (Broadway & Kassenboehmer, 2019; Fogelgren, Ornstein, Rödin, & Skogman Thoursie, 2021; Markussen & Røed, 2014).[4] A study by Markussen and Røed (2014) for Norway suggests that subsidized employment on the regular labor market improves the long-term perspectives of temporary DI recipients in comparison to sheltered employment or vocational training.[5] They also show that ordinary education (i.e., in schools, colleges, or universities) has a positive impact on re-employment probability, despite a long lock-in period. In contrast, Broadway and Kassenboehmer (2019) find that job coaching requirements for young DI recipients below the age of 35 in Australia had no effect on the probability of working in the long-run.

Likely the closest paper to ours, Fogelgren *et al.* (2021) use a randomized experiment in Sweden to evaluate the effects of an SE program. They find that the program outperforms existing rehabilitation strategies in terms of rate of return to employment. Their control condition is, however, different from ours, since regular vocational rehabilitation in Sweden

[4] A study by Laun and Skogman Thoursie (2014) in Sweden examines whether privatization of vocational rehabilitation can improve labor market opportunities for individuals taking a long-term sickness absence. They find no employment differences between private and public rehabilitation providers. Another study by Dean, Pepper, Schmidt, and Stern (2017) evaluates the effect of vocational rehabilitation for individuals with mental illness, but not necessarily on DI rolls. Using a structural model and data for the State of Virginia, they find that VR services have positive effects on participants' future earnings, but in part due to the increased probability of receiving DI benefits. An additional study worth mentioning is the Mental Health Treatment Study (MHTS) conducted by the Social Security Administration (SSA) in the United States between 2006 and 2010. Using a large-scale randomized experiment with more than 2,000 Social Security Disability Insurance (SSDI) beneficiaries who suffered from severe psychiatric impairment (mostly schizophrenia), the SSA measured the effects of access to supported employment services, coupled with systematic medication management. The scope of the treatment was therefore much larger than vocational rehabilitation and included a significant medical component with fully reimbursed mental health treatments. Frey *et al.* (2011) show that compared to the control group, which received no support, SSDI beneficiaries in the treated group had a higher employment rate over the 24-month follow-up period (although with earnings below the SSA's defined "substantial gainful activity" limit, implying that they would stay on DI rolls), as well as improving mental health and quality of life.

[5] The authors argue that sheltered employment or vocational training courses that target DI recipients carry a stigma and send a negative signal to potential employers (Markussen & Røed, 2014).

is also based on a "work-first" approach but with less intense supervision.[6] In addition, their study focuses on young adults between 19 and 29 years of age, whereas our sample includes DI recipients of all ages. Another unique feature of our study is that in addition to tracking individuals' administrative status and benefits using social security registers, we designed a survey with detailed questions about job search efforts, type of employment, vocational training attendance, as well as health, self-esteem, and self-efficacy indices. As such, we are able to paint a much richer picture of program participants' trajectories and explore potential channels for the observed differences in re-employment success. Finally, because we recruited participants up to December 2019, we can also evaluate the effectiveness of SE during the economic turmoil that resulted from COVID-19.

Our study relates to the broader literature on active labor market policies aimed at the unemployed. In a recent review of over 200 studies, Card, Kluve, and Weber (2018) show that the time profile of impacts varies by type of program. They find that job search assistance programs that emphasize "work first" tend to have relatively stable impacts over time, whereas programs that emphasize "human capital accumulation" through vocational training have larger average effects in the medium and longer runs. They also highlight systematic heterogeneity across groups, with long-term unemployed participants benefiting more from "human capital" programs, while "work-first" programs tend to be more successful for participants with low income and/or low labor market attachment.

The remainder of the paper is organized as follows. Section 1 provides more details on the institutional context and the two return-to-work programs evaluated in this research. Section 2 introduces the experimental framework and data collection process. Section 3 presents the results. Section 4 offers robustness checks. Section 5 compares the costs and benefits. Section 6 concludes this study.

3.1 Institutional Context and Interventions

In this section, we describe the Belgian Disability Insurance (DI) system, focusing on financial conditions. Furthermore, we discuss the differences between the regular Vocational Rehabilitation (VR) program, which has been in place for more than a decade, and the new Supported Employment (SE) program evaluated in this article.

[6] Fogelgren *et al.* (2021) also use a second control group based on "case management," which is a high intensity support program, but whose main goal is to increase the well-being and social integration of individuals with severe mental conditions, without a clear focus on labor market participation.

3.1.1 The Belgian Disability Insurance System

In Belgium, employed workers with a minimum number of working days have access to disability benefits through the National Institute for Health and Disability Insurance (NIHDI).[7] The benefits cover them against health-related events that affect their ability to work for at least one month.[8] Application terms and conditions vary, however, between disability spells that are less than a year and those that are longer. In the rest of this paper, we will therefore distinguish between these two types by referring to "short-term disability" spells and "long-term disability" spells.

To qualify for short-term disability coverage, individuals must be recognized as "unable to work" by a doctor designated by their health insurance fund. Workers would be considered eligible when their ability to work is reduced by at least 66% with respect to their last occupation. To qualify, applicants should also have stopped all productive activity as a consequence of a deterioration in their health that is not directly related to their professional activity.[9] If these two conditions are still applicable after a year, a disabled worker may qualify for long-term disability status. There is, however, no automatic transition from the short-term status to the long-term one. In order to be accepted into the long-term disability program, the applicant's doctor (who oversaw the applicant during the short-term period) must submit the application to the NIHDI, which can directly approve the doctor's decision or run its own internal evaluation.

The replacement rate also varies according to the duration of the disability spell. In the first year, it amounts to 60% of the last wage payment received before becoming disabled. After one year, when one enters the long-term disability program, the replacement rate depends on the last wage payment received as well as the position of the disabled person in the household. To be precise, this share is 65% for heads of households, 60% for single households, and 40% for cohabitants, with defined floor and ceiling amounts.[10] Benefits can be reduced when DI recipients start working while on claim. As shown in Figure 1, the benefits remain unchanged if the ratio between hours worked and full-time employment (38 hours a week in Belgium) is lower than 20%. Above 7.6 hours per week, benefits are reduced by the percentage that exceeds the threshold of 20% (for 19 hours per week, this would be 50%-20%, that is, 30%). If, after attempting to go back to the labor market, the

[7]Full-time workers and unemployed workers must have fulfilled a minimum of 180 working days (or active days of job search for the unemployed) during the last twelve months to be eligible. For part-time workers, the requirement is to have worked at least 800 hours in the last 12 months.

[8]Spells shorter than a month are fully paid by employers and are not covered by this insurance program.

[9]This is to establish a distinction between the disability insurance program and other programs such as the occupational injuries fund and the occupational diseases fund.

[10]In 2020, maximum short-term disability benefits were 2,248 euros per month, while maximum long-term disability benefits were 2,435 euros per month.

beneficiaries become sick again within three months (14 days for short-term program), they automatically requalify for benefits and their previous disability spell is continued.

3.1.2 Regular Vocational Rehabilitation Program

We start by describing the regular Vocational Rehabilitation program, in place since 2009, which constitutes the control condition to which the new program based on Supported Employment is compared. Before this study, it was the only return-to-work program offered to DI recipients who were seeking help to re-enter the labor market.

Following many other countries, Belgium has traditionally favored vocational training to help DI beneficiaries return to work. This approach relies on the assumption that DI recipients should first rebuild working capacity before they re-enter the labor market. In its current form, the VR program includes three phases supervised by the beneficiary's doctor and a caseworker: 1) orientation; 2) training; and 3) job search assistance. This approach favors human capital acquisition before attempting to return to the labor market.

During the *Orientation* phase, meetings are organized between caseworkers and DI beneficiaries at their local public employment service.[11] The goal of those meetings is to assess rehabilitation needs and find adequate vocational training. The next phase begins when the doctor and NIHDI[12] have approved the *Training* program. Upon training completion, participants receive *Job search assistance* for a maximum period of six months.

From administrative records provided by the Social Security Administration, we know that the median duration of the training is 6 months (first quartile is 3 months, third quartile is 11 months). Respondents to our surveys provided similar information, with median training duration equals to 30 weeks (first quartile is 10 weeks, third quartile is 52 weeks). Our questionnaires also provide additional information on the title and content of the training. Text analysis on the information provided by survey respondents reveals that some followed generic training for computer, accounting or administrative skills, as well as language courses (English, Dutch or French). Others followed specific training to access professions that require basic skills such as "medical secretary or assistant," "forklift operator," "esthetician or beauty therapist," while others have started advanced courses on "design of web and mobile applications," "3D modeling or infographic." The variety of training explains why some last a few weeks, while others take up to a year to complete.

[11] Job centers in Belgium are a regional competency, supervised in Flanders by the *VDAB*, in Brussels by *Actiris*, and in Wallonia by the *Forem*.

[12] NIHDI must approve training for it to be financed

3.1.3 New Supported Employment Program

We now turn to describing the new Supported Employment program, which is evaluated in this research. It is, in the words of Card *et al.* (2018), a "work-first" program. Indeed, the emphasis is on a rapid, intensive job search so that participants can have face-to-face contact with potential employers from the very beginning of the process. To this end, caseworkers (called "job coaches") have a maximum of 20 individuals listed at the same time and organize meetings at least every two weeks. Vocational training can be offered within the program framework, ideally in combination with part-time employment, or at least after the beneficiary attempted to find work.

The program builds on the IPS model of Supported Employment that was developed in the United States in the 1990s.[13] In addition to the rapid job search, caseworkers offer follow-along support, even after a job has been found. The goal is to help beneficiaries secure long-term employment, even though mental illnesses are often characterized by ups and downs that can affect their ability to work. The caseworkers also provide financial advice on the consequences of working for disability benefits. They are also involved with employers at all stages: to find vacancies, make appointment for their beneficiaries, request feedback after job interviews, but also to provide support once the beneficiary started working. As such, they serve as a back-up during unexpected health crises and may reduce uncertainty for employers.

Table 1 summarizes the main aspects of the Supported Employment program and how it differs from the regular rehabilitation approach. One may notice that the caseworkers' load is about five times smaller for the new Supported Employment program. In addition, the program favors competitive work on the regular labor market and not sheltered or wage-subsidized employment. Sheltered workshops are organizations who mainly offer positions for persons with disabilities. They pursue a double objective of producing goods and services, while offering integration and rehabilitation for individuals with disabilities (Visier, 1998). In many countries, the legislation authorizes sheltered workshops to offer sub-minimum wages (Visier, 1998), but not in Belgium where it is required since 2003 to pay at least the minimum wage. Caseworkers in the Supported Employment program are specifically asked not to advice participants to join sheltered workshops, but rather try to work in a regular firm.

[13] The Individual Placement and Support (IPS) approach was developed primarily to provide Supported Employment services for people with severe mental health conditions (e.g., schizophrenia, bipolar disorders) at mental health centers. The program was first evaluated by Drake, McHugo, Becker, Anthony, and Clark (1996) at two mental health centers in New Hampshire. Many small-scale randomized controlled trials followed (summarized in Luciano et al., 2014 and Marshall et al., 2014). A recent meta-analysis by Modini *et al.* (2016) reveals that the IPS model was effective in contexts characterized by heterogeneous labor market conditions.

3.2 Experimental Design and Data Collection

In this section, we present the experimental setup used to measure the effectiveness of the SE program in comparison to regular VR. We also describe the outcomes constructed to compare the effects of the two programs using both administrative registers and a survey designed for this research project.

3.2.1 Recruitment of Participants and Random Allocation to Programs

In March 2018, NIHDI notified all healthcare-funded doctors in Belgium that a study was underway to evaluate the effects of a new return-to-work program for DI recipients with mental conditions. The doctors, in turn, informed patients who (1) suffered from mental illness and (2) were willing to re-enter the labor market that they were eligible to participate in the study.[14] Participants were recruited in Belgium's three regions. Wallonia was split to account for two different partners assuming responsibility for the implementation of the Supported Employment program there.[15] In total, the experiment comprises four clusters that cover all of Belgium (i.e., Flanders, Western Wallonia, Eastern Wallonia, and Brussels). Doctors were encouraged to recruit patients with moderate to severe mental disorders. The form sent to doctors explains that moderate disorders may include depression or anxiety, while severe disorders may include bipolar disorders, schizophrenia or obsessive compulsive disorders. The list was not meant to be exhaustive, but doctors were told not to include patients with mild mental health issues who should be able to return to the labor market on their own.

DI recipients who agreed to take part in the study signed an informed consent before their doctor sent their file to NIHDI and the researchers. Their administrative file contained individual characteristics (gender and work experience) used for the stratified randomization, as well as names and contact details for the follow-up surveys. Randomization was performed at the individual level. On a weekly basis, we allocated participants to the treatment or control groups using the randomization list for the correct cluster and stratum.[16] We used stratification on two variables that are important confounding fac-

[14] The doctors were asked not to discuss the specifics of the return-to-work programs so as not to create expectations.

[15] Wallonia was divided between West and East, based on the postal codes of participants. There was thus no overlap between the two sub-regions.

[16] The randomization took place at the individual level and was performed on a weekly basis by the researchers. As participants entered the study over several months between March 2018 and December 2019, a predefined allocation sequence was created in the form of computer-generated randomization lists. The randomization lists were created using the software Stata and the user-written command "ralloc"

tors of return-to-work success according to previous studies: gender (Wewiorski & Fabian, 2004) and work experience in previous years[17] (Drake et al., 1996).

Participants were then invited to a job center in their region to complete the baseline survey. After completing the baseline survey, they met with a caseworker who informed them of the program in which they had been randomly allocated.[18] In total, 667 participants were recruited from March 2018 to December 2019.

3.2.2 Administrative and Survey Data

We rely on two data sources to estimate the effects of the two return-to-work programs: (1) administrative registers and (2) survey instruments designed for this research. The main advantage of using administrative data is that we can observe the complete sample of participants for the duration of the experiment and at a relatively high frequency. At the same time, conducting a dedicated survey allows us to collect detailed information on individual behavior and well-being.

The administrative data comes from the registers of NIHDI. They provided pre-treatment characteristics of participants, including age, gender, place of residence, date of entry into DI, status before entering DI (i.e., a blue-collar or white-collar worker), and whether they have dependents (children or spouse). On a monthly basis, we are able to track: their disability status; benefit amount (in 20-euro bins); participation in vocational training financed by NIHDI; and most important, partial work resumption (with including the exact volume of work). Our administrative dataset allows us to build a balanced panel for our entire sample of participants that spans 12 months before the start of their return-to-work program and 18 months after it.

We complement the data from administrative registers with a survey designed for the purposes of this research. All participants answer a baseline survey before the start of their return-to-work program. The baseline survey instrument allows us to collect a rich set of information on all participants, including their nationality, education, previous work experience, and financial situation. All participants also receive a follow-up questionnaire

from Ryan (1998). The command provides a sequence of treatments randomly permuted in blocks of varying size (treatments are balanced within blocks). In total, 16 randomization lists were created, one for each combination of the four regional clusters and the two stratification variables (i.e. gender and work experience). Each week the research team at the university received a list of new participants that had signed the informed consent form. New participants were allocated to the treatment or control groups using the randomization list of the correct cluster and stratum.

[17] The exact question is: "Have you been working in a paid job in the last two years?" The answer is binary (yes/no), but an additional third category covers cases where the information was not provided by the doctor.

[18] Caseworkers were asked not to give any information about the other program.

every six months, by email or in hard copy. Follow-up surveys provide us with detailed information on meetings with caseworkers, job search efforts, labor force participation (including the type of contract and sector of activity), and earnings sources, as well as subjective evidence on their health and well-being.

3.2.3 Integrity of the Experimental Design

Balancing test: Table 2 presents summary statistics for DI beneficiaries before their program assignment (using both administrative and survey data). The first column displays means and standard deviations in parentheses for the entire sample of 667 participants. The next two columns show the respective statistics for the control and treatment groups. The statistics reveal that participants are perfectly balanced in terms of gender. Their average age is 40 years and they have spent 44 months on disability. Only 20% have dependents, either children or a spouse. 53% were blue-collar workers before entering DI, while 26% had a higher education degree. Finally, the vast majority holds Belgian nationality (87%).

The last column in Table 2 reports results from balancing tests. The latter reveal that at the 10% level, we fail to reject the equality of means of treatment and control groups for any of the 11 outcomes considered. The aggregate test, reported in Panel B, also finds that we are not able to reject equality of means across all 11 variables (p-value = 0.65). Overall, it appears that individuals in both groups showed similar characteristics.

Survey attrition: Table 3 presents an analysis of survey attrition for the follow-up surveys at 6, 12, and 18 months. The follow-up rate is relatively good for a self-administered questionnaire, with about 64% of participants who answered the first survey at 6 months. Panel B presents an analysis of the type of people who were less likely to be surveyed. Those who did not answer were slightly more likely to be men of foreign nationality and without higher education. Panel C presents a test of whether the treatment affected the type of person who completed follow-up surveys, in other words, whether the treatment caused a sample composition bias. The p-values on a full set of baseline characteristics interacted with treatment are 0.62 (follow-up 6 months), 0.31 (follow-up 12 months) and 0.97 (follow-up 18 months). Taken together, these results demonstrate that there is no systematic difference between individuals who answered the follow-up survey in the treatment and control groups.

3.3 Results

In this section, we introduce our empirical strategy and subsequently compare the effects of the two return-to-work programs on the set of outcomes described in the previous section.

3.3.1 Estimation Strategy

We estimate specifications that compare the effects of the new SE program (treatment) with the regular VR (control). We run the following Ordinary Least Squares regression:

$$Y_i = \alpha + \beta Treatment_i + \gamma X_i + R_{region} + S_{month} + T_{year} + \epsilon_i \tag{3.1}$$

where Y_i is the relevant outcome for DI beneficiary i, $Treatment_i$ is an indicator variable equal to one for beneficiaries who were randomly assigned to the new Supported Employment program, X_i denotes a vector of beneficiaries controls that includes the two stratification variables (dummies for gender and work experience in last two years), as well as the individual's baseline value of the outcome variable Y (when available). R_{region} is a vector of dummy variables for the four regions in the study (described in subsection 3.2.1). For estimations using the administrative register data at monthly frequency, we also include month S_{month} and year T_{year} fixed effects to account for seasonality and trends. The coefficient of interest is β and captures the effect of being assigned to treatment. In all estimations, we report heteroskedasticity robust standard errors.

3.3.2 Effects on Administrative Status and Benefits Payment

We start our analysis by comparing the effects of the two return-to-work programs on the DI status and benefit receipt of study participants over time. We use data from NIHDI administrative registers, which allows use to track all participants during a 30-month period, including 12 months before and 18 months after the start of their return-to-work program. We create three dummy variables that track the disability status of individual participants on a monthly basis, and whether they work or follow a training while on claim.[19] We also create a variable to report monthly DI benefits received by participants,

[19]The two variables imply contradictory effects for the social security budget. Indeed, as explained in subsection 3.1.1, individuals who work while on claim will receive reduced benefits if their working hours exceed 20% of a full-time job. In contrast, the cost of training, when approved by the Social Security Administration, is fully covered. In addition, participants receive a bonus for each effective hour

which takes the value 0 if the individual exited DI. As such, the benefits variable will capture effects on DI reliance at both the extensive and intensive margins.

Figure 2 compares the trajectories of DI recipients who randomly joined one of the two return-to-work programs. One can observe in Panel C that their probability to be on disability rolls is highly similar across the 30-month window. Panel A, however, shows that SE participants are more likely to start working part-time while on claim from the 6^{th} month after the start of their program, with a gap that widens up to the 18^{th} month. At the same time, Panel D reveals that SE participants receive lower benefits in the long-run, which is expected when DI recipients work part-time. Finally, Panel B reveals that SE participants are less likely to follow a training while on claim, compared to VR participants whose program focuses on vocational rehabilitation.

Table 4 presents formal estimates for the difference in trajectories between participants in SE and VR. We provide results for the four outcomes described above at months 6, 12, and 18 after the start of a return-to-work program, which match the follow-up periods of the survey and should ease the comparison. Since administrative data are available on a monthly basis, we also produce four graphs that show the evolution of the effects for all outcomes over time (Figures 3, 4, 5 and 6). Figure 3 clearly shows that SE does not seem to increase the probability of exiting entirely from DI compared to the regular rehabilitation program. The estimated intention-to-treat effects are very small and their confidence intervals always include 0. However, Figure 4 suggests that those who took part in the new SE program are more likely to work part-time while on claim. This effect starts around 6 months after program entry and slowly builds, reaching 9.5 percentage points after 18 months (equivalent to two times the control mean). At the same time, participants in the SE program are 2.5 percentage points less likely to follow a training at follow-up 18. This is not surprising given that SE relies on a "work-first" approach, while regular rehabilitation in the control group clearly encourages vocational training. Finally, when looking at the consequences for disability payments, Figure 6 clearly indicates a small reduction in the amount received by individuals in the treatment group. Again, this effect materializes around the 6^{th} month after program entry and slowly increases to reach 70 euros per month at follow-up 18, that is a reduction of 6% compared to the control mean. As explained above, the reduction in DI benefits very likely reflects the fact that individuals working more than 20% of a full-time job automatically incur a reduction in their allowance. Taken together, these results suggest that participation in the SE program does not increase the probability of exiting DI, but rather favors working while on claim. The result is less reliance on DI benefits.

of training, as well as a final bonus if successfully the training is successfully completed.

3.3.3 Effects on Type of Employment and Sector of Activity

In the previous subsection, we showed that SE increases the probability that DI recipients work while on claim. We complement this finding using our survey data. The latter has two additional advantages compared to NIHDI administrative registers: (1) it also tracks the employment status of those who left DI (i.e., individuals who returned to full-time employment); and (2) it contains detailed information on the type of contract and sector of activity.

The first column in Table 5 displays the effect of SE on the probability of working in a paid job at follow-ups 6, 12, and 18 months. When last observed at 18 months, the employment rate of the SE group was more than two times larger, with a treatment effect of 8.7 percentage points compared to a control group mean of 13 percentage points. The effect is relatively aligned with the one measured using administrative registers (9.5 percentage points in Table 4), reinforcing the strength of our findings and the reliability of our survey data.

We now turn to the type of contracts that participants in the study have found in the course of their return-to-work process. We distinguish between three types of contract, permanent (i.e., open-ended contract), temporary contract (i.e., fixed-term contract) or self-employment. We believe that the type of contract reflects the quality of the employment found, as well as the strength of the labor market attachment. Indeed, a potential pitfall of the "work-first" approach of SE is the risk of individuals accepting low quality jobs to quickly re-enter the labor market. Previous research in the context of unemployment insurance reveals that job search assistance does not boost employment in the long-run if the program places participants in lower-quality jobs (Cottier, Flückiger, Kempeneers, & Lalive, 2018). Our survey data allows us to check whether this is the case for our study population of DI beneficiaries.

Table 5 shows that the SE group is 6.4 percentage points more likely to find a permanent contract at follow-up 12 (column (2)), but the effect tends to fade over time, reaching only 2.6 percentage points at follow-up 18 (although with large standard errors). At the same time, we observe a larger probability that SE participants work with a temporary contract at follow-up 18 (column (3)) and no significant difference in terms of self-employment. These results suggest that SE did favor a quicker return to the labor market, but for a significant share of DI recipients, this is only with temporary contracts.

Finally, we take advantage of our survey data to observe the sector of activity for those who found a job. A key feature of the IPS model of SE is to help individuals with mental health conditions find a job in the "regular" labor market, that is avoiding placement in sheltered work. Table 5 reveals that at follow-up 18, the vast majority of those who work in a paid job have a position in the private sector (column (5)), while there is no significant effect on employment in the public or nonprofit sectors. More important, we do not observe any effect on the probability of joining sheltered workplaces. Taken together, our results suggest that participation in SE increases the employment rate of DI recipients with mental conditions, mostly through occupations in the private sector.

3.3.4 Effects on Earnings

In this subsection, we explore the effects of participation in the two return-to-work programs on earnings. In our survey, we ask participants about three earnings sources: work income; DI benefits; and Unemployment Insurance (UI) benefits. If a participant does not receive earnings from a given source, the variable takes on the value "0," as such, capturing effects at both the extensive and intensive margins. We also create an outcome that sums up earnings from all three types and title it "total earnings."

Table 6 reveals that 18 months after the start of their return-to-work program, participants in the treated group declared receiving on average 119 euros more from work income than those in the control group (Column (2)), that is, about twice as much as the control group. Meanwhile, their DI benefits decrease by 105 euros (Column (3)), that is, a reduction of 9.6% compared to the control mean.[20] For two reasons, it is expected that the effects on wage income and DI benefits are not symmetrical. First, as explained in subsection 3.1.1, DI benefits replace 60% of lost income, which means that if someone exits DI and goes back to full-time employment, their wage income should be higher than their benefits. Second, when DI recipients work while on claim, their DI benefits are reduced, but the first 20% are exempted (more details in sub-section 3.1.1). For both these reasons, it is expected that DI beneficiaries who return to a full-time job or work while on claim will have higher earnings in total.

3.3.5 Effects on Health and Well-being

In this subsection, we explore the consequences of the programs for the health and well-being of participants. Our survey instrument designed for this project includes three sets of questions that allow us to build the most common indicators used in the literature in health and psychology. First, we ask twelve questions based on the short-form health

[20]The effect is slightly larger than when measured with the administrative registers (Table 4), which could reflect the fact that survey respondents work on average more hours per week and therefore incur a larger reduction in benefits.

survey validated by Ware, Kosinski, and Keller (1996). It includes questions on both physical and mental aspects in order to assess the impact of health on an individual's everyday life. Second, we ask survey participants to answer ten questions related to their self-esteem. These ten questions have been used since the seminal work of Rosenberg (1965) to measure both positive and negative feelings about oneself and to detect self-esteem problems. Third, we build on work by Schwarzer and Jerusalem (1995) and include ten questions to measure self-efficacy. The goal of this index is to observe how survey participants perceive their ability to cope with difficult demands in life.

Even though the primary goals of the return-to-work programs are not to improve participants' health or their self-perceptions, contacts with caseworkers and (potential) employers might have unintended consequences. Table 7 shows that SE does not seem to disproportionately affect the health, perceived self-esteem, or self-efficacy of program participants. We do not observe any significant effect for all three measures at any point in the follow-up period.

3.3.6 Suggested Mechanisms: Job Search Behavior and Vocational Training

In the previous subsections, we showed that participants in SE are more likely than those on regular rehabilitation to work while on claim, and as a result, rely less on DI benefits. We can think of at least three reasons why SE achieves higher part-time employment rates for DI recipients: (1) they receive more intense support from caseworkers; (2) they more actively look for a job; and (3) they dedicate less time to vocational training before attempting to go back to the labor market. Our survey will help test whether these hypothesized channels are met in practice. Our questionnaire helps us track the frequency of meetings between participants and their caseworkers, as well as the time that is dedicated to the job search or training.

Column (1) in Table 8 shows that during the first six months of participation in a return-to-work program, DI recipients in SE have had on average 4.8 more meetings with their caseworker compared to those in regular rehabilitation (that is, more than twice as many than in the control group, which had 3.8 meetings). Over time, the number of meetings with caseworkers tends to decrease, but it does so at a faster rate in the control group. Thus, when asked how many times they met with their caseworker at follow-up 18, participants in the treated group responded that they had 3.1 more meetings over the previous six months than their counterpart in the control group (mean of 1.9). These results indicate that SE does indeed offer more intensive guidance to DI recipients who are willing to re-enter the labor market.

Our questionnaire also allows us to track whether DI recipients are actively looking for a job and how much effort they commit to it. They are asked how many average hours they dedicate each week to seeking job offers, working on their CV and cover letters, preparing and going to interviews with potential employers, and other activities related to the job search (e.g., attending job fairs). We combine answers to these four categories and create a variable with the total number of hours dedicated to the job search. If a participant responds that they do not actively look for a job, the variable takes on the value 0. As such, it captures job search efforts at the extensive and intensive margins. Column (2) in Table 8 reveals that at each follow-up period, SE participants are more likely to declare that they are actively looking for a job. When it comes to how much efforts they put into it, we observe that at follow-up 12, participants in SE dedicate on average 2.5 hours more each week to their job search compared to those in regular rehabilitation (Column (3)). This effect fades away at follow-up 18, which most likely reflects the fact that many participants have already found a job and are no longer actively looking for employment.

We also ask survey respondents what their reservation wage is (net of social security contribution and income tax), as well as their preferred number of working hours when looking for a job. Interestingly, at follow-up 6 / 12, participants in SE reported that they preferred to work on average 2.4 / 3 hours less per week, that is, 10% / 12% lower than their control group counterparts (Column (4) in Table 8). They also declare that they would accept a lower wage of about 143 euros, that is, 10% lower than the control mean (Column (5), follow-up 12 in Table 8). We believe that these results could indicate another potential channel for the higher re-employment rate of SE participants. We hypothesize that individuals taking part in regular rehabilitation, which focuses on vocational training to regain work capacity, anticipate a return full-time employment. In contrast, our results seem to indicate that those in SE have lower expectations in terms of working time and wages, which might allow them to apply to a wider range of jobs.

Finally, our survey offers a way to track training attendance, even for short vocational trainings of only few hours or days. This is a truly positive feature compared to administrative registers, which only report on larger training periods for which NIHDI authorization is required and financial support offered. Column (6) in Table 8 shows that participants in SE have spent on average two weeks less on training at follow-up 12 and 18, this is about 50% less than in the control group. This result clearly reflects the different focus of the two programs, but also the fact that SE participants have more time available to look for a job and are not locked into a training that might delay their return to the labor market.

3.3.7 Heterogeneity analysis

In this last subsection, we explore the heterogeneous effects of the new SE program across chosen sub-groups of participants. We are particularly interested in the heterogeneity of results along six individual dimensions: location, gender, time on disability, blue/white-collar status, education, as well as the participants' belief in their capacity to overcome difficulties. In what follows, we highlight the reasons why we believe these distinctions are of interest for the research on DI and eventually present the estimated differences between groups. All subgroups are constructed around two mutually exclusive categories that encompass the entire sample of participants. We test for the difference between the two categories by adding an interaction term to equation (3.1), which becomes:

$$Y_i = \alpha + \beta Treat_i + \delta Categ_i + \zeta Treat_i * Categ_i + \gamma X_i + R_{region} + S_{month} + T_{year} + \epsilon_i \quad (3.2)$$

Figures 7, 8, 9 and 10 report the estimated difference ζ along six dimensions for the probability of exiting DI, the probability of working part-time, the probability of following a training, and the benefits amount.

Geographical location: We start by testing for differences in the effect of SE based on the region where participants lived in at the start of their return-to-work program. We distinguish between Flanders in the north of Belgium and Wallonia/Brussels in the south/center of the country. We are interested in this distinction because Flanders has a more dynamic job market, and according to Eurostat, an unemployment rate of 3.2% in 2019, compared to 7.2% in Wallonia and 12.6% in Brussels. Results in Figure 8 indeed show that participants in the new SE program were more likely to work part-time in Flanders, while there is no significant difference with VR in Wallonia/Brussels. We believe that this reflects the fact that it is harder for DI recipients to go back to work, even on a part-time basis, in a weak labor market where they compete with a large share of unemployed people.

Gender: We now look into the heterogeneity of the results by the gender of participants. We know from previous studies that the number of women on DI rolls has increased substantially in recent decades (Autor & Duggan, 2006), and particularly among young women after motherhood (Fontenay & Tojerow, 2020). We are therefore interested in the effects of the new SE program for this growing group of female DI recipients. Our estimates, using a gender dummy interacted with the treatment status, do not show statistically significant differences for the four outcomes considered. Despite large standard errors, however, we notice that men are 7 percentage points more likely to work part-time after entering the SE program than women. A better powered experiment might help reveal possible differences in the impact of SE across genders.

Time on disability: As explained in subsection 3.1.1, Belgian DI distinguishes between beneficiaries who have spent less than 12 months on DI rolls (short-term program) and those who have spent more (long-term program). We use this distinction to evaluate the effects of SE depending on the duration of the disability. Interestingly, we observe in Figure 9 that long-term beneficiaries who take part in the SE program are less likely than short-term beneficiaries to follow a training while on claim. While it was expected that because of the nature of the program, SE participants would be less likely to attend a training, we anticipated that long-term beneficiaries might need to regain human capital before attempting to return to the labor market. We can think of two reasons for this somewhat surprising result. First, long-term beneficiaries might have more stable mental health conditions, allowing them to more quickly re-enter the labor market without a need for training. Second, long-term beneficiaries have been away from the labor market longer and might be less willing to engage in long training that would further delay their return.

Blue/white collar, education level: We also explore how the effects of SE vary according to workers status (i.e., blue vs white collar) and their level of education (primary/secondary school vs higher education). The interaction effects with the two different dummies do not reveal statistically significant differences in the probability of exiting DI, the probability of working part-time, the probability of following a training while on claim, and the amount of benefits. We therefore conclude that SE works indifferently for blue-collar and white-collar workers, as well as for individuals who attended higher education or not.

Self-efficacy: Finally, we investigate the effects of the SE program based on the participants' belief, before the start of the return-to-work program, that they would be able to overcome difficulties and obstacles. To do so, we use the Generalized Self-Efficacy Scale (GSES) based on previous work by Schwarzer and Jerusalem (1995). The GSES includes ten questions that we report in table 9. We construct a self-efficacy score (ranging from 10 to 40) for all participants based on their answers to the dedicated in module in the baseline survey, that is before their entry in a return-to-work program. We then compute the z-score for each individual by subtracting from their score the sample mean and dividing by the standard deviation. Finally, we distinguish between those who have a low or high self-efficacy depending on whether they are below or above 0. The survey participants who obtain a higher score on the self-efficacy scale have stronger beliefs in their capacity to perform novel or difficult tasks, as well as to cope with adversity. Interestingly, when interacting the self-efficacy dummy with the treatment variable, we observe that those in the SE group who have a high self-efficacy score are 11.3 percentage point more likely to be working part-time than those who have a low score (Figure 8). We interpret this as a sign of the readiness of the participant to cope with the difficult task of returning to the labor market after a long time on disability. The GSES is an operative construct, which is highly correlated with subsequent behavior (Jones, Mandy, & Partridge, 2009; Schwarzer, 1992). As such, we believe that it could be used by doctors or caseworkers to asses the readiness of the candidates to participate in a return-to-work program, especially for SE which confronts participants to the labor market without prior training.

3.3.8 Impact of the COVID-19 Pandemic

Previous research on the consequences of economic shocks for beneficiaries of Active Labor Market Policies (ALMP) shows contrasting results. Barrera-Osorio, Kugler & Silliman (2021) find that the COVID-19 pandemic washed away the benefits of a job-training program in Colombia. Field, Linden, Malamud, Rubenson, & Wang (2019) also reveal that cohorts that graduate from vocational programs during economic downturns perform worse. However, others such as Beuermann, Bottan, Hoffmann, Jackson, & Vera Cossio (2021) find that beneficiaries of ALMP suffered fewer employment losses throughout the COVID-19 pandemic. It therefore remains unclear whether pre-pandemic programs might sustain their benefits throughout the turmoil that resulted from COVID-19.

NIHDI and its partners made sure that both return-to-work programs remained active during the pandemic and that caseworkers continued to meet participants virtually. This is not surprising given that Belgium is the country in the European Union with the highest share of teleworkers (above 50%) and a below average share of unemployed since the onset of the pandemic (Eurofound, 2020). For all these reasons, we believe that the impact of the COVID-19 pandemic was more moderate in Belgium, although obviously not null.

In a first exercise, we compare participants who entered the study early, from March 2018 to December 2018, with participants who entered at a later stage between January 2019 and December 2019. The rationale for this comparison is that the follow-up period of late entrants overlaps with the development of the COVID-19 pandemic, which started in March 2020 in Belgium. We plot in Figure 11 their probability to work part-time depending on whether they are early entrants (Mar.-Dec. 2018, left panel) or late entrants (Jan.-Dec. 2019, right panel). We notice a downward shift in the probability of working while on claim for late entrants ("2019 sample") in both groups. The gap between treatment and control participants at follow-up 18 is 12 percentage points for the "2018 sample" and declines to 8.2 percentage points for the "2019 sample." However, the relative gap compared to the control mean[21] is rather stable over time from 90% to

[21] The control mean at follow-up 18 is 13.2 percentage points for the "2018 sample" and 10.4 percentage

80%. This graphical exploration seems to suggest that even though late entrants in the study seem to have been affected by the COVID-19 pandemic, those who were randomly allocated to the SE program were still more likely to work while on claim than those in the VR program.

We now turn to more formal estimates of the effect of the COVID-19 pandemic. Since new participants entered the study continuously between March 2018 and December 2019, they have been impacted at different stages of their return-to-work program. For instance, those who entered in December 2019 had only three months pre-pandemic, while those who entered in January 2019 were hit toward the end of their 18-month follow-up. We leverage this unique feature of our study to estimate the impact of the COVID-19 pandemic on the trajectory of our study participants. We estimate the following model to track the effect of COVID-19 over time:

$$Y_{it} = \alpha + \beta Treat_i * Followup_t + \delta Treat_i * Followup_t * COVID_t + \gamma X_i + R_{region} + \epsilon_{it} \quad (3.3)$$

where the $Treat_i$ indicator is now interacted with a $Followup$ variable that tracks the number of months since the start of the return-to-work programs from zero to 18. The third term of equation 3.3 adds an interaction with a binary indicator $COVID$ which takes on a value 1 after March 2020. Thus, the vector of coefficients β captures the effects of being assigned to treatment over time before the pandemic, while the vector of coefficients δ captures the effects of the pandemic on the treated. Compared to previous estimations, we now pull together data for all participants during all follow-up periods (i.e., panel dataset). We also cluster standard errors are at the individual level.

Table 10 tracks the effects of the SE program on the probability of working while on claim from follow-up 6 to 18, and allows comparison to pre-pandemic impacts (first column) and impacts after COVID-19 hit Belgium in March 2020 (third column). We observe in column (1) that the probability of working part-time for participants in SE was already significantly higher at follow-up seven and increased over time, reaching 14 percentage points at follow-up 18. Column (2) reveals that the COVID-19 pandemic drastically reduced the positive effects for the treated in the medium-run (particularly between follow-ups nine and 13). However, we notice in column (3) that in the long-run, participants in SE whose follow-up overlapped with the COVID-19 pandemic were still 10.6 percentage points more likely to be working while on claim compared to those in regular rehabilitation. In other words, the COVID-19 pandemic delayed the return-to-work process of SE participants, but in the long-run, the effects of the program are still largely positive. We conclude that SE was effective in helping DI beneficiaries find and

points for the "2019 sample."

retain a job through the economic turmoil that resulted from COVID-19.

3.4 Robustness checks

This section offers two robustness checks that have become standard in the literature using randomized controlled trials (e.g. J. Cohen & Dupas, 2010; Fujiwara & Wantchekon, 2013). We start by showing that, despite the modest size of our sample, the results are very similar when using randomization inference instead of classical inference. In continuation, we account for the fact that we test the effect of the SE program on multiple outcomes and provide p-values accounting for the risk of false discovery.

Randomization Inference was first proposed by Fisher (1935) and further developed by Rosenbaum (2002) as an alternative for classical inference in a randomized experiment context. The main advantage of this procedure is providing inference with correct magnitude regardless of sample size. In addition, this test is nonparametric as it does not make distributional assumptions. Also known as permutation test, this method consists of reassigning the treatment and control status in the sample (in our case within the strata described in sub-section 3.2.1) and reestimating the parameter of interest (our β in equation 3.1) using this placebo assignment multiple times (we perform 1,000 random permutations).[22] This procedure gives p-values for the null hypothesis of zero treatment effect, which corresponds to the proportion of reestimated β that are larger (in absolute value) than the actual β. In Table 11, we report the p-values computed this way under the name "rand. inf. p-value" for all the outcomes considered in our previous estimations. One can see that these newly-computed p-values are very close in magnitude to those from the classical inference method used in our main analysis.

Multiple hypotheses testing: We also want to account for the fact that we are estimating the effects of the new SE program on 25 individual outcomes by adjusting p-values for multiple inference. In particular, we compute sharpened q-values following the procedure by Anderson (2008). This method controls for the false discovery rate, that is the expected proportion of rejections that are type I errors (i.e. false rejections). Table 11 shows the p-values computed this way under the name "sharpened q-values." As expected, the q-values of each test are adjusted upward to reduce the probability of a false rejection. However, one can observe that the effects measured for our main outcomes of interest, those related to the employment status of the participants, remain statistically significant at conventional levels.

[22] We use the Stata package "ritest" developed by Simon Heß (2017) to perform the randomization inference procedure.

We take all these results as a confirmation that, despite the limited size of our sample and the multiplicity of the outcomes tested, the effects measured for the new SE program are genuine and robust.

3.5 Cost-benefit Analysis

In this section, we perform a cost-benefit analysis of the two interventions to put in perspective the intention-to-treat effects measured previously. Given that the new SE program requires more intense supervision than regular VR and therefore supposes higher cost per individual participant, this is particularly important. As such, it is not obvious that a program should be preferred over another if its marginal benefits do not make up for its higher costs in a reasonable time frame.[23]

As explained in subsection 3.1.2, the NIHDI partners with regional employment agencies that offer rehabilitation services to DI recipients. The cost for each program participant is laid out in a cooperation agreement that was renewed in 2018 for the start of this study. NIHDI agrees to pay 4,800 euros per year for each participant in the SE program.[24] Over the course of the study, that is, 18 months, the cost was thus 7,200 euros for each SE participant. In comparison, NIHDI agrees to pay a flat fee for each participant in regular VR of 4,800 euros.[25] The difference between the two interventions was therefore 2,400 euros per individual participant over the 18-month study window. It certainly reflects SE's more intensive investment in human resources than VR. Those costs cover the salaries of caseworkers and their supervisors, as well as all necessary expenditures to perform their mission (e.g., office space, communication devices to stay in touch with program participants, transportation costs to meet prospective employers). It is therefore a comprehensive proxy of the individual cost if the programs were to be scaled up.

When it comes to the benefits of the program, we distinguish between two perspectives: (1) NIHDI budget and (2) the "society as a whole." First, from the perspective of NIHDI, benefits are measured by the reduction in DI benefits paid, which could be the result of individuals exiting DI completely or working while on claim. Second, from the perspective

[23] See for instance Crépon, Gurgand, Kamionka, and Lequien (2012) for an example of program whose positive effects are small relative to the cost of implementation.

[24] For the ease of calculation, we use the cost per participant in Flanders for the whole country. This is not a strong assumption since Flemish participants make up 70% of our sample. In addition, the cost in Brussels and Wallonia is slightly smaller, so our estimates are in fact the most conservative.

[25] The reason for the flat fee is that, contrary to SE, job coaching in regular rehabilitation is limited to six months after training has been completed. In addition, no support is provided while participants are enrolled in training.

of "society as a whole," the benefits encompass the value of the production generated by new jobs, as well as the savings for NIHDI's budget, which is financed by taxpayers. Following Fogelgren *et al.* (2021), we hypothesize that the production generated by new jobs is estimated by the wage cost.[26] This of course assumes that there are no displacement effects and that participants in one program do not crowd out jobs for participants in the other program, or for nonparticipants. Crépon, Duflo, Gurgand, Rathelot, & Zamora (2013) show in the context of unemployment insurance that this type of externality can drastically reduce the estimated benefits of a program, particularly in weak labor markets. In our particular context, we argue that this should not play a significant role since the Belgian labor market was rather strong during the experiment window, especially in Flanders where the unemployment rate was 3.2% in 2019, according to Eurostat. Tensions in the Flemish labor market are one reason that the employment agency was keen to implement activation policies for DI beneficiaries. For all these reasons, we assume that the benefits for the "society as a whole" are the production generated by the new jobs (valued at cost) and the budget savings for NIHDI.

In Table 12, we report the effects of the new SE program on DI benefits and wage income of participants (already shown in Table 6), as well as the combined effect for the "society as a whole," which corresponds to the sum of the absolute values of the two others. However, our sample of survey respondents, while perfectly balanced across programs, is not fully representative of the entire population of DI recipients with mental conditions in Belgium. In fact, in Table 2 we show that participants in the study are more likely to come from the north of the country (i.e., Flanders) and are equally likely to be a man or a woman, as well as a blue or white-collar worker. Population data from NIHDI's registers reveals instead that blue-collar workers and women are over-represented among DI recipients with mental conditions (59% and 61% respectively), while only one out of two beneficiaries comes from Flanders. To improve the representativeness of our study, we reweight our sample using the entropy balancing method by Hainmueller & Xu (2013). The results using this balanced sample are reported under the title "reweighted sample."

At this stage, we calculate how long the benefits of SE would need to last to make up for their higher cost, and we report the corresponding "catch-up time" in Table 12. From the perspective of the Social Security Administration, the reduction in monthly DI benefits is between 105 euros and 147 euros at follow-up 18, depending on whether we use the survey sample or the reweighted sample. If this effect were to remain stable beyond the 18-month window, the 2,400-euro gap between the cost of the two programs would be closed in 16 to 23 months. When considering gains for the "society as a whole," that is, the sum of the budget's savings and the value from the creation of jobs, one can see that

[26] Unfortunately, we only know the wage income of employees and not the total wage cost for their employers. This will therefore be a lower bound estimate for the value generated by the job.

the new SE program could make up for its higher cost in just 9 to 11 months.

There is of course no guarantee that the estimated effects will persist beyond the 18-month follow-up period. However, Figures 4 and 6 do not suggest that the control group is catching up with the treatment group and that the gap would close quickly. Thus, we argue that the estimated time that SE needs to make up for its higher cost (between 9 and 23 months depending on the chosen perspective) is relatively low compared to the potential future benefits.

3.6 Conclusion

The growing number of Disability Insurance (DI) recipients with mental health conditions, who often suffer ups and downs in the evolution of their illness, has blurred the line between those who are totally and permanently disabled and those who retain some work capacity or could recover it in the future. This trend motivates the implementation of active labor market programs to help DI beneficiaries return to work when their health allows it.

In this paper, we study the effects a new Supported Employment (SE) program introduced in March 2018 in Belgium. The program is characterized by a "work-first" approach with intense job coaching and follow-along support. Using a Randomized Control Trial, we compare the effects of this newly introduced program with regular Vocational Rehabilitation (VR) services, which have been in place for more than a decade in Belgium (i.e., control condition). Between March 2018 and December 2019, we recruited more than 660 DI recipients who suffer from mental illnesses and were willing to take part in a return-to-work program. Participants from across Belgium were randomly assigned to the new SE program or the regular VR program. We followed them for 18 months from the start of their program using both data from administrative registers and survey instruments designed for this research.

We find that compared to regular rehabilitation, SE increases the probability that DI recipients with mental conditions work while on claim and reduces their reliance on DI benefits. Specifically, we estimate that 18 months after the start of their return-to-work program, participants in the SE group are 9.5 percentage points more likely to be working part-time and receive 6% less in DI benefits than those in the control group. In addition, we observe that they find occupations in the private sector (and not in sheltered workshops).

We take advantage of our rich survey data to explore the channels through which SE achieves higher re-employment rates than regular rehabilitation. We find that participants in SE dedicate more time to the job search and less time to vocational training. In addition, DI recipients in SE report that they seek a job with fewer working hours and with lower pay. We hypothesize that SE participants form more realistic expectations of their capacity to work and as such are more likely to look for part-time jobs. In contrast, those in the control group who spend more time on vocational training might seek full work resumption, which is not always compatible with their current health status.

These findings should be of broad interest outside of Belgium since most OECD countries face rising disability rolls, especially individuals with mental health conditions. In this paper, we show that SE is successful in increasing the proportion of DI recipients who work while on claim, therefore reducing their reliance on benefits and easing the burden for the social security budget. The cost-benefit analysis reveals that the higher cost of SE per individual participant can be compensated within a reasonable time frame (of less than two years) if the observed effects were to remain beyond the 18-month follow-up window. This is of course an assumption that needs to be tested in future research.

Table 1: Characteristics of the Two Return-To-Work Programs

	Treatment	Control
Program:	Supported Employment (SE)	Vocational Rehabilitation (VR)
Launch:	2018	2009
Focus:	"Work first" approach with rapid job search	Human capital accumulation through training
Intensity:	1 caseworker for 20 beneficiaries (contact at least every 2 weeks)	1 caseworker for 100 beneficiaries
Duration:	Unlimited follow-along support	Max. 6 months after training completed

Table 2: Descriptive Statistics and Balancing Test

Panel A	Sample	Control	Treatment	Balancing test (T-C)
		Mean (SD)		Coeff. (SE)
Administrative data				
Female (0/1)	0.50	0.51	0.49	-0.02
	(0.50)	(0.50)	(0.50)	(0.04)
Age	40.47	41.02	40.03	-1.00
	(8.50)	(8.43)	(8.54)	(0.66)
Live in Flanders (0/1)	0.69	0.70	0.69	-0.01
	(0.46)	(0.46)	(0.46)	(0.04)
Months on disability (#)	44.11	44.32	43.94	-0.38
	(37.34)	(36.16)	(38.31)	(2.89)
Long-term disability (0/1)	0.83	0.84	0.83	-0.02
	(0.37)	(0.37)	(0.38)	(0.03)
Dependents - children or spouse (0/1)	0.20	0.22	0.18	-0.04
	(0.40)	(0.42)	(0.38)	(0.03)
Blue collar worker (0/1)	0.53	0.55	0.51	-0.05
	(0.50)	(0.50)	(0.50)	(0.04)
Daily benefits (euros)	47.24	48.00	46.63	-1.38
	(11.84)	(12.22)	(11.51)	(0.93)
Voluntary work (0/1)	0.16	0.15	0.16	0.01
	(0.37)	(0.36)	(0.37)	(0.03)
Baseline survey				
Belgian nationality (0/1)	0.87	0.87	0.88	0.01
	(0.33)	(0.34)	(0.33)	(0.03)
Higher education (0/1)	0.26	0.24	0.27	0.03
	(0.44)	(0.43)	(0.45)	(0.03)
Number of observations	667	298	369	667
Panel B				
F-test regression of treatment on all outcomes	0.79			
P-value	0.65			

*Notes: Panel A reports descriptive statistics before program assignment for the sample of individuals who entered the study between March 2018 and December 2019. Columns "Sample", "Control" and "Treatment" report the means (standard deviations in parentheses) of individual characteristics for the whole sample, the treatment and the control sub-samples, respectively. Column "Balancing test" reports the difference (standard errors in parentheses) between those assigned to treatment and those assigned to control for each outcome considered. Panel B reports an aggregate test for the equality of means across all variables. Data sources are from NIHDI administrative registers, as well as baseline survey administered to all participants before their entry in a return-to-work program. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 3: Response to Follow-up Surveys and Attrition Test

	Attrition (0/1)		
	Follow-up 6	Follow-up 12	Follow-up 18
Panel A			
Treatment (0/1)	-0.041	-0.063	0.009
	(0.037)	(0.039)	(0.039)
N	667	667	667
Attrition mean	0.360	0.448	0.508
Panel B			
Treatment (0/1)	-0.039	-0.056	0.010
	(0.037)	(0.038)	(0.038)
Female (0/1)	-0.156 ***	-0.101 **	-0.112 ***
	(0.037)	(0.039)	(0.040)
Age	-0.004 *	-0.001	-0.006 **
	(0.002)	(0.002)	(0.002)
Live in Flanders (0/1)	0.085 **	-0.008	0.023
	(0.039)	(0.042)	(0.042)
Months on disability (#)	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)
Long-term disability (0/1)	-0.046	-0.022	-0.013
	(0.052)	(0.056)	(0.057)
Dependents - children or spouse (0/1)	-0.045	-0.061	-0.065
	(0.055)	(0.057)	(0.056)
Blue collar worker (0/1)	0.057	0.087 **	0.098 **
	(0.040)	(0.041)	(0.042)
Disability benefits (euros)	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)
Voluntary work (0/1)	-0.070	-0.076	-0.061
	(0.052)	(0.054)	(0.056)
Belgian nationality (0/1)	-0.120 **	-0.074	-0.119 **
	(0.059)	(0.058)	(0.059)
Higher education (0/1)	-0.174 ***	-0.187 ***	-0.111 **
	(0.043)	(0.046)	(0.048)
N	667	667	667
Panel C			
Treatment (0/1)	0.034	0.212	0.213
	(0.281)	(0.293)	(0.296)
Baseline characteristics + Interactions with treatment			
N	667	667	667
Aggregate F-test	0.83	1.16	0.39
F test: p-value joint significance of interactions	0.62	0.31	0.97

*Notes: This table presents an analysis of survey attrition for the follow-up at 6, 12 and 18 months. The dependent variable is a dummy that takes on the value 1 if the person did not answer the follow-up survey. Panel A presents the difference in response rate between the treatment and control groups. Panel B presents an analysis on the type of people that were less likely to be surveyed. The covariates include the treatment indicator, as well as all the variables reported in Table 2. Panel C presents a test of whether the treatment affected the type of person who completed the follow-up surveys, in other words whether the treatment caused a sample composition bias. Most importantly, it reports the p-values for the joint significance of a full set of baseline characteristics interacted with the treatment indicator. Data sources are NIHDI administrative registers, as well as baseline survey administered to all participants before their entry in a return-to-work program. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 4: Administrative Registers - Status and Benefits

	(1) Disability status (0/1)	(2) Work while on claim (0/1)	(3) Training while on claim (0/1)	(4) Monthly DI benefits (euros)
Treat. 6 months	-0.019	0.029	-0.018	-64.129 **
	(0.014)	(0.023)	(0.013)	(28.952)
Obs.	667	667	667	667
Cont. mean	0.973	0.094	0.037	1181
Treat. 12 months	-0.011	0.061 **	-0.040 ***	-67.735 **
	(0.021)	(0.027)	(0.014)	(33.192)
Obs.	667	667	667	667
Cont. mean	0.930	0.121	0.050	1147
Treat. 18 months	-0.026	0.095 ***	-0.025 **	-69.967 **
	(0.021)	(0.027)	(0.011)	(35.470)
Obs.	667	667	667	667
Cont. mean	0.933	0.111	0.030	1165
Baseline cont.	YES	YES	YES	YES

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers. Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 5: Survey on Employment, Type of Contract and Sector of Activity

	(1) Paid work (0/1)	(2) Permanent contract (0/1)	(3) Temporary contract (0/1)	(4) Self-employed (0/1)	(5) Private sector (0/1)	(6) Public sector (0/1)	(7) Nonprofit sector (0/1)	(8) Sheltered work (0/1)
Treat. 6 months	0.055 **	0.035 *	0.031 **	-0.011	0.032	0.004	0.010	0.004
	(0.024)	(0.019)	(0.015)	(0.008)	(0.022)	(0.004)	(0.010)	(0.004)
Obs.	419	419	419	419	418	418	418	418
Cont. mean	0.044	0.022	0.011	0.011	0.039	0.000	0.006	0.000
Treat. 12 months	0.070 *	0.064 **	0.008	-0.002	0.054 *	-0.000	0.014 *	0.002
	(0.036)	(0.026)	(0.026)	(0.008)	(0.031)	(0.014)	(0.008)	(0.013)
Obs.	364	364	364	364	364	364	364	364
Cont. mean	0.104	0.039	0.058	0.006	0.071	0.019	0.000	0.013
Treat. 18 months	0.087 **	0.026	0.058 **	0.003	0.100 ***	0.001	0.009	-0.023
	(0.042)	(0.034)	(0.025)	(0.014)	(0.036)	(0.016)	(0.012)	(0.015)
Obs.	321	321	321	321	321	321	321	321
Cont. mean	0.130	0.089	0.027	0.014	0.075	0.021	0.007	0.027
Baseline cont.	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 6: Survey on Earnings

	(1) Total earnings (euros)	(2) Wage income (euros)	(3) DI benefits (euros)	(4) UI benefits (euros)
Treat. 6 months	17.670	62.411	-56.923	12.182
	(49.220)	(39.443)	(44.466)	(12.788)
Obs.	405	405	405	405
Cont. mean	1203.379	78.784	1117.124	7.471
Treat. 12 months	17.018	44.226	-30.372	3.164
	(52.683)	(53.599)	(52.539)	(17.010)
Obs.	349	349	349	349
Cont. mean	1243.051	164.934	1051.763	26.354
Treat. 18 months	29.336	119.394 **	-104.797 **	14.739
	(56.851)	(54.140)	(52.833)	(15.415)
Obs.	316	316	316	316
Cont. mean	1239.128	132.379	1093.132	13.617
Baseline cont.	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 7: Survey on Health and Well-being

	(1) Health related quality of life index	(2) Self-esteem index	(3) Self-Efficacy index
Treat. 6 months	0.016	0.036	0.021
	(0.052)	(0.080)	(0.082)
Obs.	407	394	394
Cont. mean	-0.071	-0.032	-0.020
Treat. 12 months	-0.030	-0.094	-0.028
	(0.061)	(0.091)	(0.091)
Obs.	352	345	345
Cont. mean	-0.014	0.116	0.045
Treat. 18 months	0.037	0.018	0.057
	(0.065)	(0.092)	(0.100)
Obs.	313	306	306
Cont. mean	-0.066	-0.008	0.002
Baseline cont.	YES	YES	YES

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 8: Survey on Job Search Behavior and Vocational Training

	(1) Nb. meetings caseworker	(2) Search job (0/1)	(3) Time job search (hours)	(4) Preferred hours	(5) Preferred wage (euros)	(6) Training nb. of weeks
Treat. 6 months	4.804 ***	0.238 ***	0.545	-2.391 **	-123.851 **	0.041
	(0.471)	(0.047)	(1.014)	(1.031)	(56.706)	(0.566)
Obs.	405	411	395	388	382	409
Cont. mean	3.783	0.350	4.618	24.008	1399.697	1.958
Treat. 12 months	3.938 ***	0.237 ***	2.525 **	-2.969 ***	-143.158 **	-1.887 **
	(0.510)	(0.050)	(0.999)	(1.072)	(55.351)	(0.781)
Obs.	350	354	349	344	329	351
Cont. mean	2.293	0.248	2.666	25.108	1441.353	3.748
Treat. 18 months	3.063 ***	0.159 ***	0.700	-1.672	-170.560 **	-1.948 **
	(0.616)	(0.053)	(0.723)	(1.198)	(67.322)	(0.875)
Obs.	319	320	317	307	301	318
Cont. mean	1.896	0.266	2.593	23.485	1454.008	4.049
Baseline cont.	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from a survey specifically designed for this study. Participants were asked to fill out the questionnaire sent every 6 months by email or post (depending on their preference). Significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Table 9: The General Self-Efficacy Scale

	Not at all true (=1)	Hardly true (=2)	Moderately true (=3)	Exactly true (=4)
1) I can always manage to solve difficult problems if I try hard enough.	o	o	o	o
2) If someone opposes me, I can find the means and ways to get what I want.	o	o	o	o
3) It is easy for me to stick to my aims and accomplish my goals.	o	o	o	o
4) I am confident that I could deal efficiently with unexpected events.	o	o	o	o
5) Thanks to my resourcefulness, I know how to handle unforeseen situations.	o	o	o	o
6) I can solve most problems if I invest the necessary effort.	o	o	o	o
7) I can remain calm when facing difficulties because I can rely on my coping abilities.	o	o	o	o
8) When I am confronted with a problem, I can usually find several solutions.	o	o	o	o
9) If I am in trouble, I can usually think of a solution.	o	o	o	o
10) I can usually handle whatever comes my way.	o	o	o	o

Notes: English version of the Generalized Self-Efficacy Scale based on Schwarzer and Jerusalem (1995). The total score is calculated by finding the sum of the all items. For the composite score therefore ranges between 10 and 40, with a higher score indicating more self-efficacy.

Table 10: Impact of COVID-19 Pandemic on Probability to Work While on Claim (0/1)

	(1) Treatment	(2) Treatment x COVID19	(3) Difference
Follow-up 6	0.034	-0.031	0.004
	(0.024)	(0.041)	(0.038)
Follow-up 7	0.056 **	-0.047	0.009
	(0.025)	(0.039)	(0.035)
Follow-up 8	0.072 ***	-0.063 *	0.010
	(0.027)	(0.037)	(0.031)
Follow-up 9	0.101 ***	-0.096 ***	0.005
	(0.030)	(0.036)	(0.028)
Follow-up 10	0.130 ***	-0.116 ***	0.015
	(0.032)	(0.038)	(0.028)
Follow-up 11	0.147 ***	-0.134 ***	0.013
	(0.034)	(0.038)	(0.026)
Follow-up 12	0.155 ***	-0.115 ***	0.040
	(0.037)	(0.042)	(0.027)
Follow-up 13	0.180 ***	-0.150 ***	0.030
	(0.041)	(0.044)	(0.025)
Follow-up 14	0.166 ***	-0.111 **	0.055 **
	(0.044)	(0.047)	(0.026)
Follow-up 15	0.152 ***	-0.082	0.070 ***
	(0.047)	(0.050)	(0.026)
Follow-up 16	0.142 ***	-0.070	0.072 ***
	(0.053)	(0.056)	(0.025)
Follow-up 17	0.121 **	-0.036	0.085 ***
	(0.058)	(0.061)	(0.025)
Follow-up 18	0.140 **	-0.034	0.106 ***
	(0.069)	(0.071)	(0.026)
Obs.	12,673		

*Notes: The table reports intention-to-treat effects at different follow-up period from 6 to 18 months since the start of the return-to-work program. Column (1) displays pre-pandemic effects of SE (β in equation (3.3)). Column (2) displays the interaction between the treatment indicator and a binary indicator "COVID-19" that takes on a value 1 from the start of the pandemic in March 2020 (δ in equation (3.3)) and should be interpreted as the effect of the pandemic on the treated. Column (3) reports the difference between the first two columns and should be interpreted as the effect of SE for participants impacted by the COVID-19 pandemic. Standard errors in parentheses are clustered at the individual level. The OLS regression controls for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years). Compared to previous estimations, the sample now pulls together all participants who entered the study between March 2018 and December 2019 at each follow-up period (i.e. a panel of 667 participants over 18 time periods). Data source is from NIHDI administrative registers. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Table 11: Robustness Checks

	(1) Disability status (0/1)	(2) Work while on claim (0/1)	(3) Training while on claim (0/1)	(4) Monthly DI benefits (euros)	(5) Paid work (0/1)	(6) Permanent contract (0/1)	(7) Temporary contract (0/1)	(8) Self-employed (0/1)	(9) Private sector (0/1)	(10) Public sector (0/1)	(11) Nonprofit sector (0/1)	(12) Sheltered work (0/1)
Treat. 6 months	-0.019	0.029	-0.018	-64.129 **	0.055 **	0.035 *	0.031 **	-0.011	0.032	0.004	0.010	0.004
SE	(0.014)	(0.023)	(0.013)	(28.952)	(0.024)	(0.019)	(0.015)	(0.008)	(0.022)	(0.004)	(0.010)	(0.004)
OLS p-value	0.194	0.219	0.170	0.027	0.026	0.064	0.045	0.152	0.144	0.320	0.291	0.320
Rand. inf. p-value	0.189	0.226	0.140	0.023	0.032	0.079	0.056	0.124	0.151	0.981	0.388	0.877
Sharpened q-value	0.324	0.331	0.318	0.128	0.128	0.181	0.141	0.309	0.309	0.434	0.434	0.434
Obs.	667	667	667	667	419	419	419	419	418	418	418	418
Treat. 12 months	-0.011	0.061 **	-0.040 ***	-67.735 **	0.070 *	0.064 **	0.008	-0.002	0.054 *	-0.000	0.014 *	0.002
SE	(0.021)	(0.027)	(0.014)	(33.192)	(0.036)	(0.026)	(0.026)	(0.008)	(0.031)	(0.014)	(0.008)	(0.013)
OLS p-value	0.582	0.023	0.004	0.042	0.050	0.015	0.753	0.806	0.083	0.975	0.085	0.863
Rand. inf. p-value	0.571	0.026	0.001	0.048	0.047	0.020	0.763	0.513	0.083	0.963	0.233	0.859
Sharpened q-value	0.698	0.053	0.036	0.072	0.079	0.045	0.760	0.760	0.117	0.880	0.117	0.760
Obs.	667	667	667	667	364	364	364	364	364	364	364	364
Treat. 18 months	-0.026	0.095 ***	-0.025 **	-69.967 **	0.087 **	0.026	0.058 **	0.003	0.100 ***	0.001	0.009	-0.023
SE	(0.021)	(0.027)	(0.011)	(35.470)	(0.042)	(0.034)	(0.025)	(0.014)	(0.036)	(0.016)	(0.012)	(0.015)
OLS p-value	0.216	0.001	0.023	0.049	0.040	0.454	0.020	0.829	0.006	0.935	0.468	0.141
Rand. inf. p-value	0.247	0.002	0.013	0.070	0.049	0.487	0.028	0.875	0.007	0.882	0.530	0.117
Sharpened q-value	0.231	0.007	0.067	0.085	0.085	0.415	0.067	0.564	0.036	0.598	0.415	0.165
Obs.	667	667	667	667	321	321	321	321	321	321	321	321
Baseline cont.	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. Under the name "Rand. inf. p-value," we report p-values from the randomization inference procedure developed by Heß for Stata (Heß, 2017) based on 1,000 random draws. The test statistic reports the proportion of placebo coefficients that are larger than the actual treatment effect. Under the name "Sharpened q-value," we report p-values adjusted for multiple inference testing using the procedure by Anderson (2008). The latter controls for the false discovery rate, that is the expected proportion of rejections that are type I errors. Significance levels are based on classical p-values reported under the name "OLS p-value": *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Table 12: Robustness Checks (cont.)

	(13) Total earnings (euros)	(14) Wage income (euros)	(15) DI benefits (euros)	(16) UI benefits (euros)	(17) Health related quality of life index	(18) Self-esteem index	(19) Self-Efficacy index	(20) Nb. meetings caseworker	(21) Search job (0/1)	(22) Time job search (hours)	(23) Preferred hours	(24) Preferred wage (euros)	(25) Training nb. of weeks
Treat. 6 months	17.670	62.411	-56.923	12.182	0.016	0.036	0.021	4.804 ***	0.238 ***	0.545	-2.391 **	-123.851 **	0.041
SE	(49.220)	(39.443)	(44.466)	(12.788)	(0.052)	(0.080)	(0.082)	(0.471)	(0.047)	(1.014)	(1.031)	(56.706)	(0.566)
OLS p-value	0.720	0.114	0.201	0.341	0.765	0.649	0.798	0.000	0.000	0.591	0.021	0.030	0.942
Rand. inf. p-value	0.709	0.106	0.189	0.400	0.763	0.657	0.803	0.000	0.000	0.575	0.022	0.015	0.942
Sharpened q-value	0.563	0.276	0.324	0.440	0.576	0.561	0.576	0.001	0.001	0.561	0.128	0.128	0.605
Obs.	405	405	405	405	407	394	394	405	411	395	388	382	409
Treat. 12 months	17.018	44.226	-30.372	3.164	-0.030	-0.094	-0.028	3.938 ***	0.237 ***	2.525 **	-2.969 ***	-143.158 **	-1.887 **
SE	(52.683)	(53.599)	(52.539)	(17.010)	(0.061)	(0.091)	(0.091)	(0.510)	(0.050)	(0.999)	(1.072)	(55.351)	(0.781)
OLS p-value	0.747	0.410	0.564	0.853	0.625	0.303	0.762	0.000	0.000	0.012	0.006	0.010	0.016
Rand. inf. p-value	0.719	0.391	0.545	0.871	0.617	0.297	0.763	0.000	0.000	0.012	0.005	0.008	0.007
Sharpened q-value	0.760	0.488	0.698	0.760	0.716	0.352	0.760	0.001	0.001	0.044	0.036	0.044	0.045
Obs.	349	349	349	349	352	345	345	350	354	349	344	329	351
Treat. 18 months	29.336	119.394 **	-104.797 **	14.739	0.037	0.018	0.057	3.063 ***	0.159 ***	0.700	-1.672	-170.560 **	-1.948 **
SE	(56.851)	(54.140)	(52.833)	(15.415)	(0.065)	(0.092)	(0.100)	(0.616)	(0.053)	(0.723)	(1.198)	(67.322)	(0.875)
OLS p-value	0.606	0.028	0.048	0.340	0.571	0.843	0.571	0.000	0.003	0.334	0.164	0.012	0.027
Rand. inf. p-value	0.591	0.036	0.053	0.393	0.583	0.857	0.581	0.000	0.002	0.312	0.159	0.011	0.014
Sharpened q-value	0.435	0.067	0.085	0.351	0.428	0.564	0.428	0.001	0.024	0.351	0.180	0.053	0.067
Obs.	316	316	316	316	313	306	306	319	320	317	307	301	318
Baseline cont.	NO	NO	NO	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO

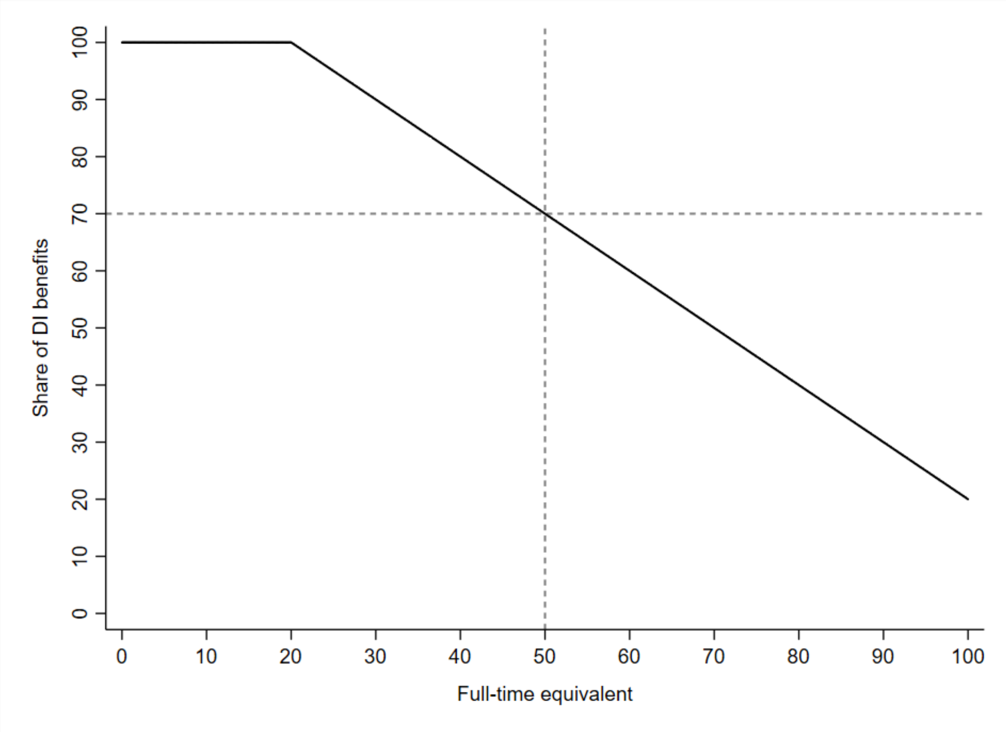
*Notes: The table reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions. Results are reported for three follow-up periods (6, 12, 18 months) after the individual filled the baseline questionnaire and started the return-to-work program. Standard errors in parentheses are robust to heteroskedasticity. Under the name "Rand. inf. p-value," we report p-values from the randomization inference procedure developed by Heß for Stata (Heß, 2017) based on 1,000 random draws. The test statistic reports the proportion of placebo coefficients that are larger than the actual treatment effect. Under the name "Sharpened q-value," we report p-values adjusted for multiple inference testing using the procedure by Anderson (2008). The latter controls for the false discovery rate, that is the expected proportion of rejections that are type I errors. Significance levels are based on classical p-values reported under the name "OLS p-value": *** p<0.01, ** p<0.05, * p<0.1*

Table 12: Cost-Benefit Analysis

Cost per participant (18 months)	SE = 7200 euros	VR = 4800 euros
Survey sample	**Effect size / SE (euros)**	**Catch up time (months)**
Reduction in DI benefits paid by Social Security	-104.80 (52.83)	23
Increase in wage income for DI recipients	119.39 (54.14)	20
Gains for "Society as a whole"	224.19 (90.96)	11
Reweighted sample **(female=61%, flanders=50%, blue collar=59%)**	**Effect size / SE (euros)**	**Catch up time (months)**
Reduction in DI benefits paid by Social Security	-147.39 (58.55)	16
Increase in wage income for DI recipients	116.61 (56.76)	21
Gains for "Society as a whole"	264.00 (97.63)	9

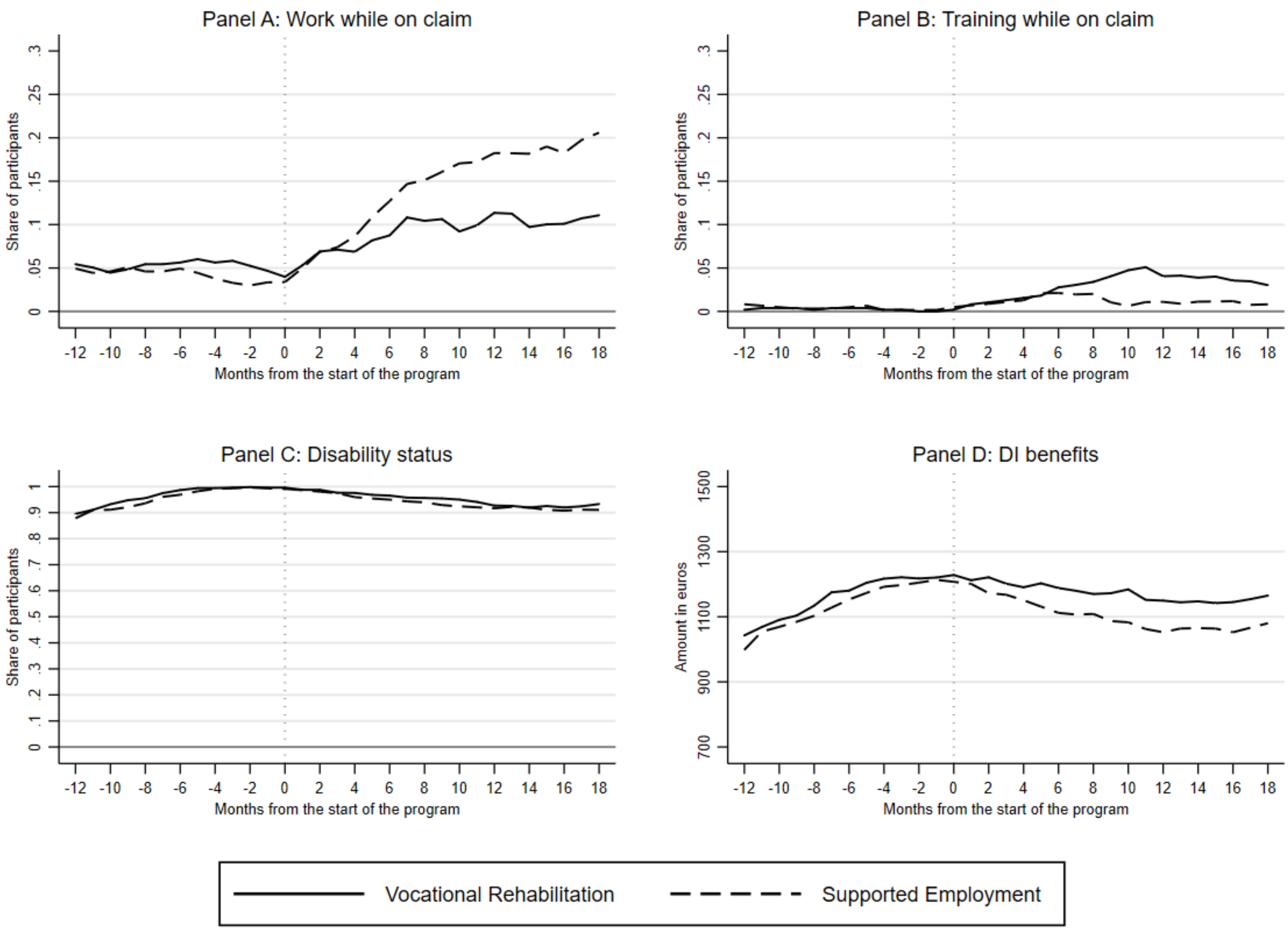
Notes: The cost per participant is laid out in a cooperation agreement between NIHDI and regional employment agencies who offer rehabilitation services. The table also reports intention-to-treat effects - coefficient β in equation (3.1) - from separate OLS regressions using survey answers at follow-up 18 months. The results for the "survey sample" are the same as in Table 6. We also use the entropy balancing method by Hainmueller & Xu (2013) to reweight the survey sample to known characteristics from the population of DI recipients with mental conditions in Belgium. The results using this balanced sample are reported under the title "reweighted sample." The outcome "gains for society as a whole" encompasses both the budget savings for NIHDI in the form of reduced benefits and the value of the production generated by the new jobs (estimated by the wage cost). It is therefore the sum of the absolute value of the two other outcomes. The "catch up time" corresponds to the number of months that the benefits of the SE program would need to last, beyond the 18-month follow-up period, to make up for its higher cost.

Figure 1: Reduction in Disability Insurance Benefits when Working while on Claim



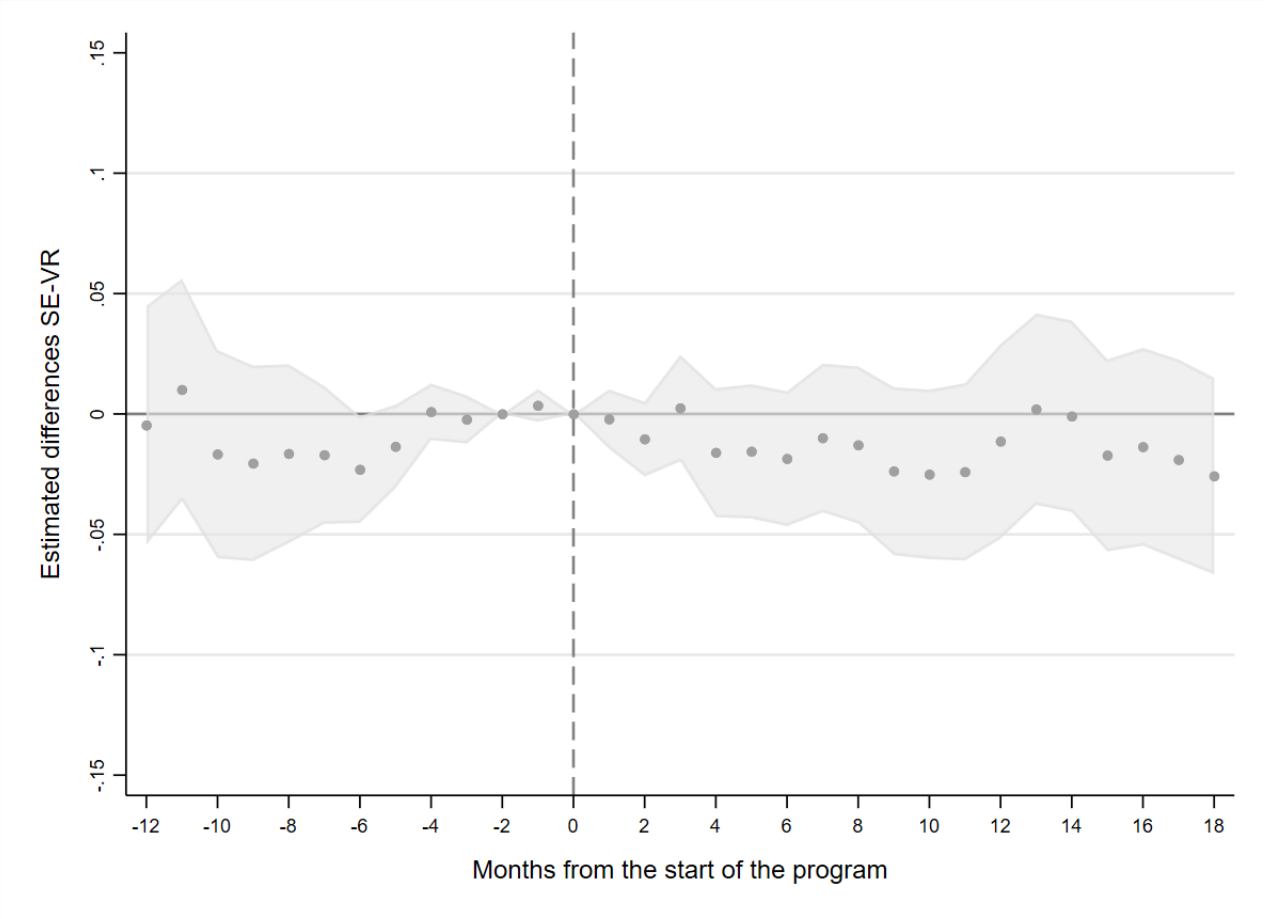
Notes: The rules are set in the Royal Decree implementing the law on compulsory insurance for medical care and cash benefits, consolidated on 14 July 1994, and amended in February 2018. Disability Insurance benefits are reduced by the amount of working time that exceeds 20% of a Full Time Equivalent (that is 38 hours a week in Belgium). DI recipients who work 20% (or less) of a FTE keep their full benefits. DI recipients who work half-time (50% of FTE) keep 70% of their benefits.

Figure 2: Comparison of Participants by Months since Randomly Allocated to a Return-to-work Program



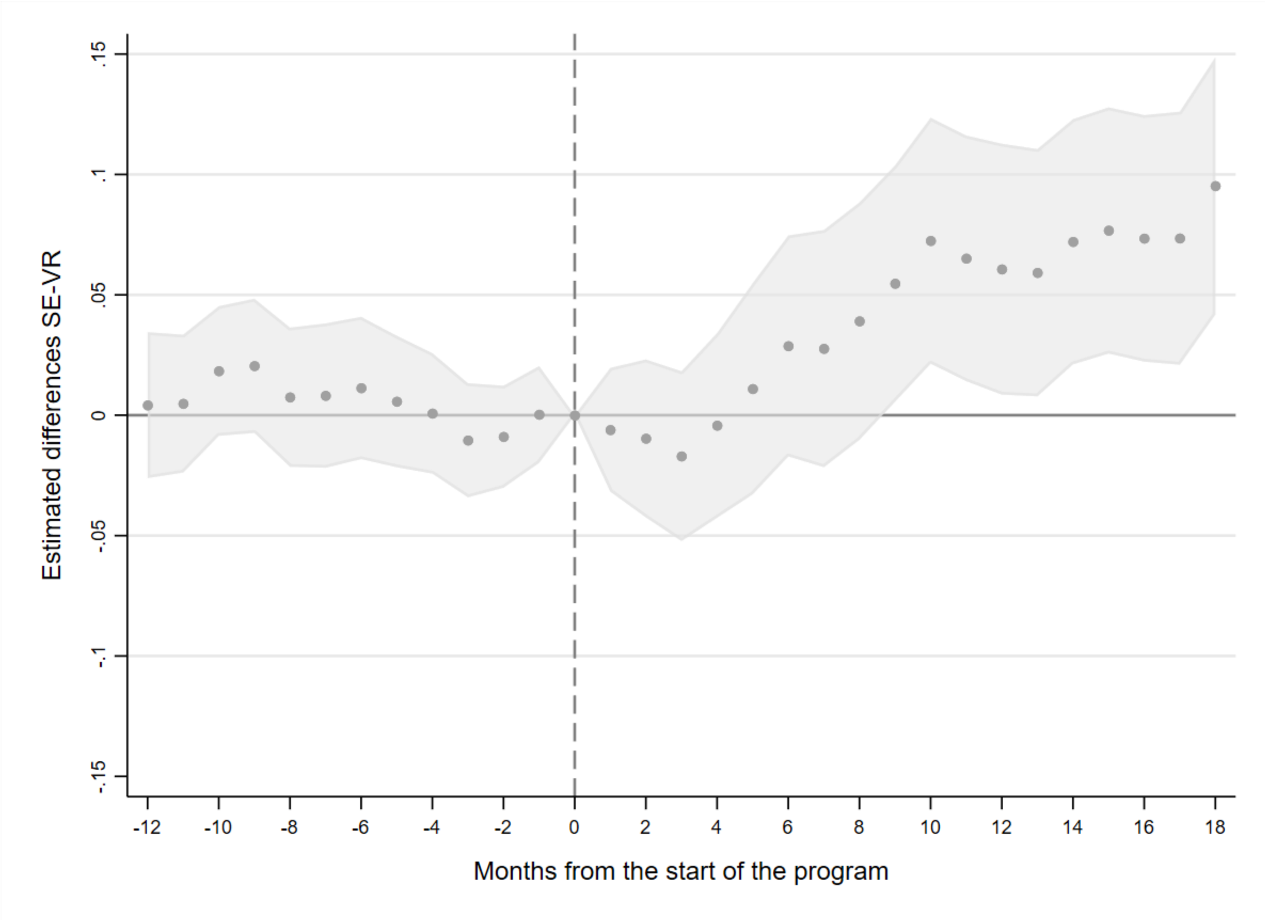
Notes: Horizontal axes show months since filling out baseline questionnaire and starting the return-to-work program. Vertical axes plot unconditional means for the share of study participants who work while on claim (Panel A), follow a training while on claim (Panel B), retain their disability status (Panel C), as well as the amount of monthly benefits they receive (Panel D). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 3: Intention-to-treat Effects on Disability Status (0/1)



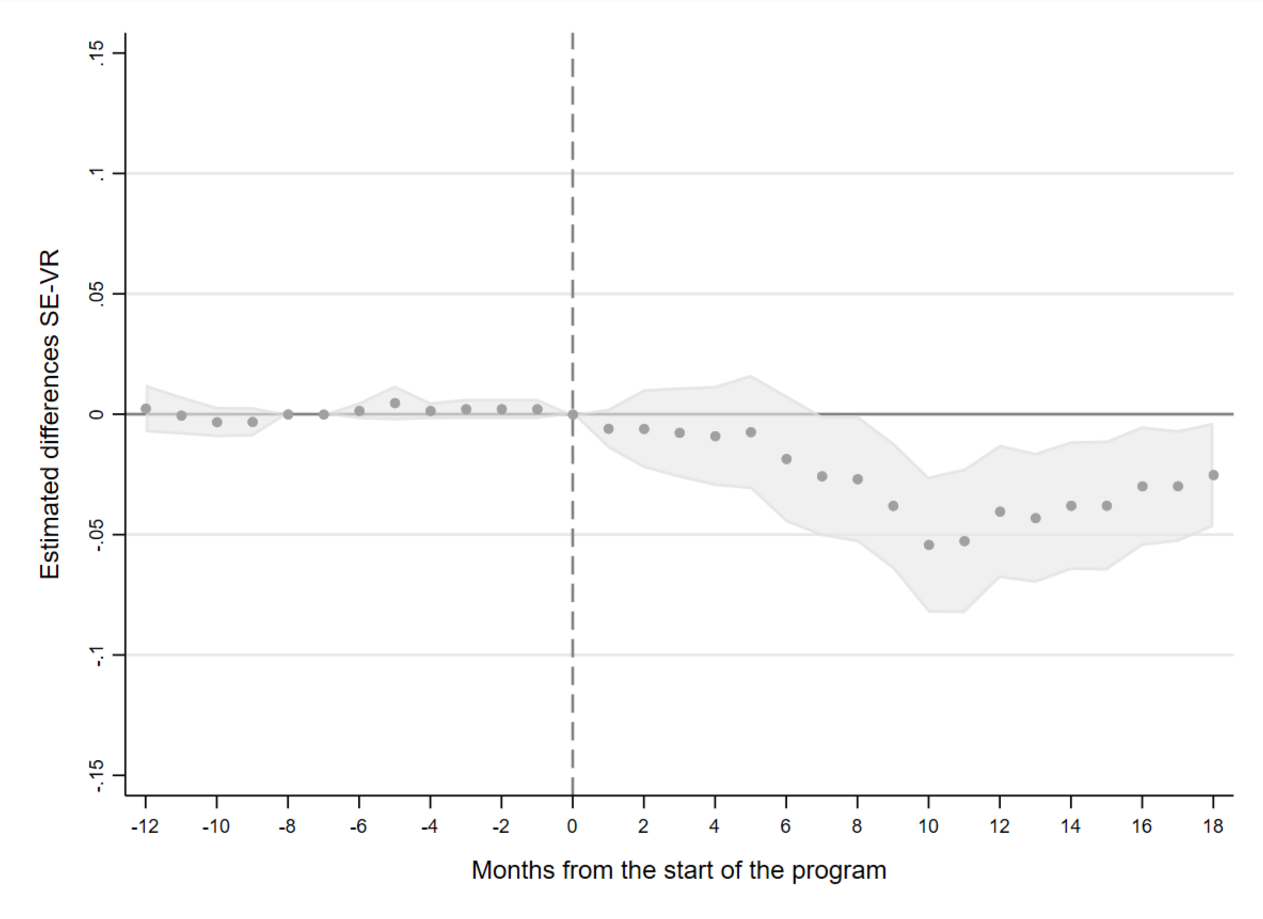
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (3.1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 4: Intention-to-treat Effects on Work while on Claim (0/1)



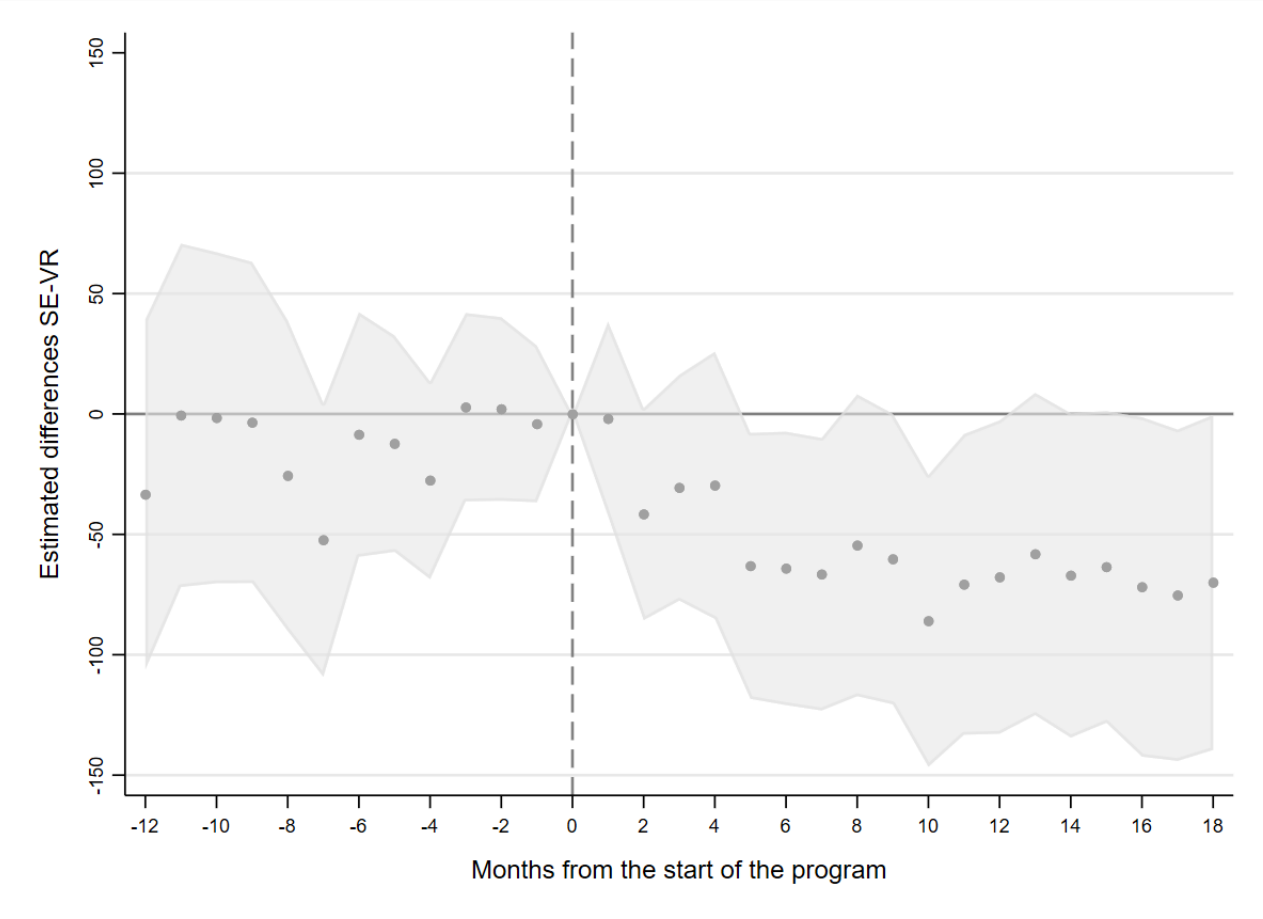
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (3.1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 5: Intention-to-treat Effects on Training while on Claim (0/1)



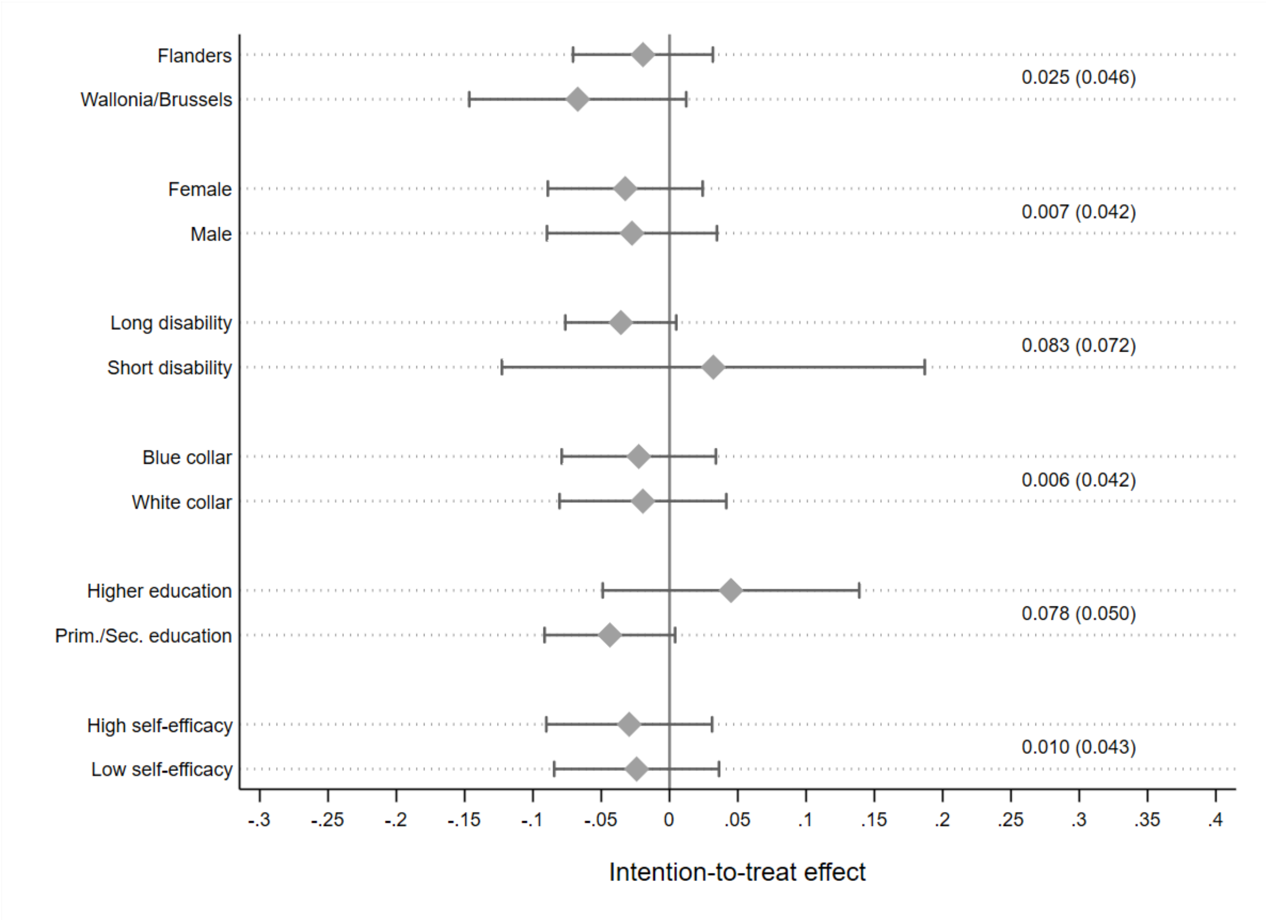
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (3.1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 6: Intention-to-treat Effects on Disability Benefits (euros)



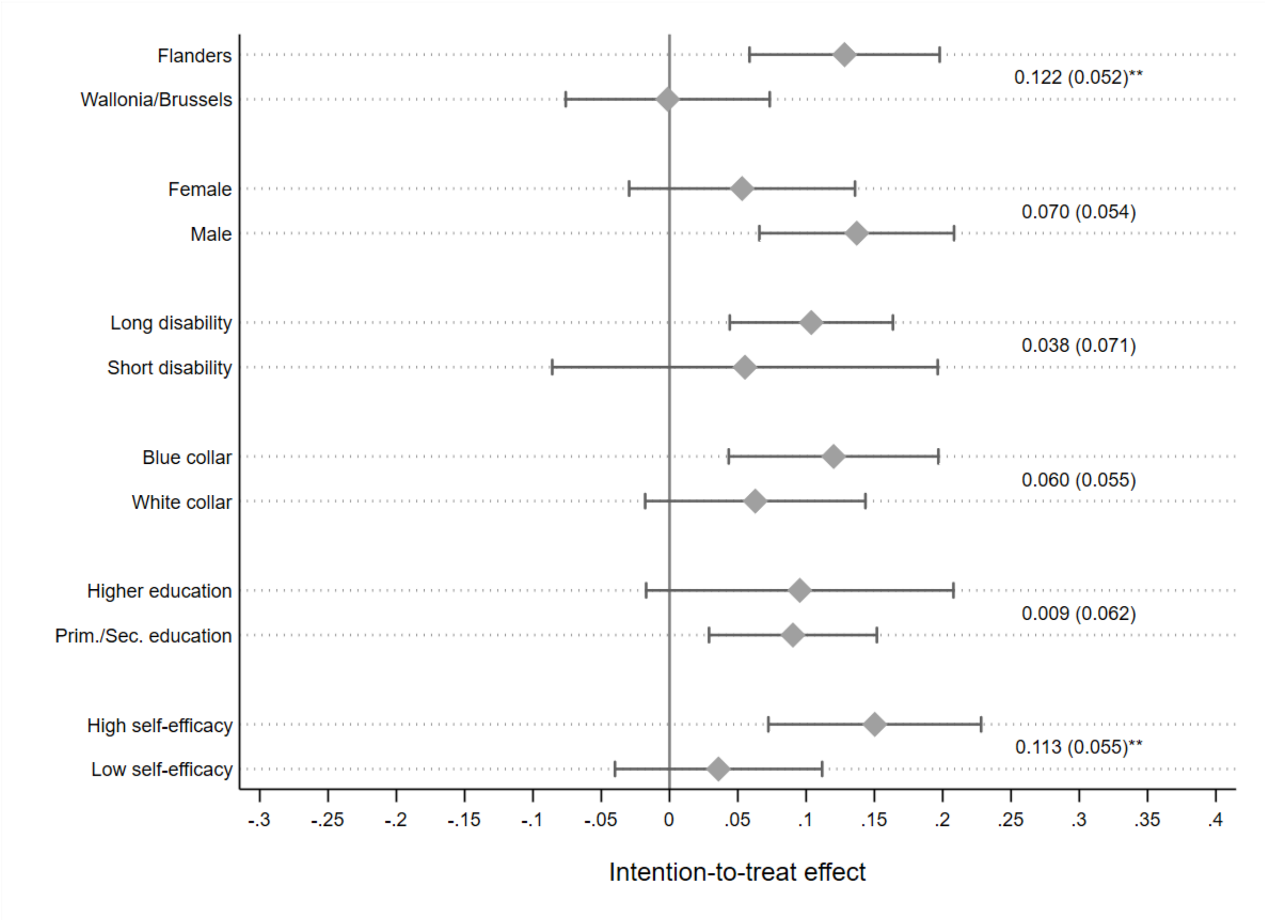
Notes: Horizontal axis shows months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots the difference between Supported Employment (SE) and Vocational Rehabilitation (VR). Each dot denotes the point estimate for intention-to-treat effect - coefficient β in equation (3.1) - at a given time horizon based on separate OLS regression. The shaded area denotes 95% confidence interval for the corresponding point estimate from heteroskedastic-robust standard errors. All estimations control for a vector of dummy variables corresponding to the four regions of the study, as well as a vector of beneficiaries' covariates that include the two stratification variables (dummies for gender and work experience in last two years), as well as the individuals' baseline value of the outcome variable (when available). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.

Figure 7: Heterogeneity Analysis: Effects on Disability Status (0/1)



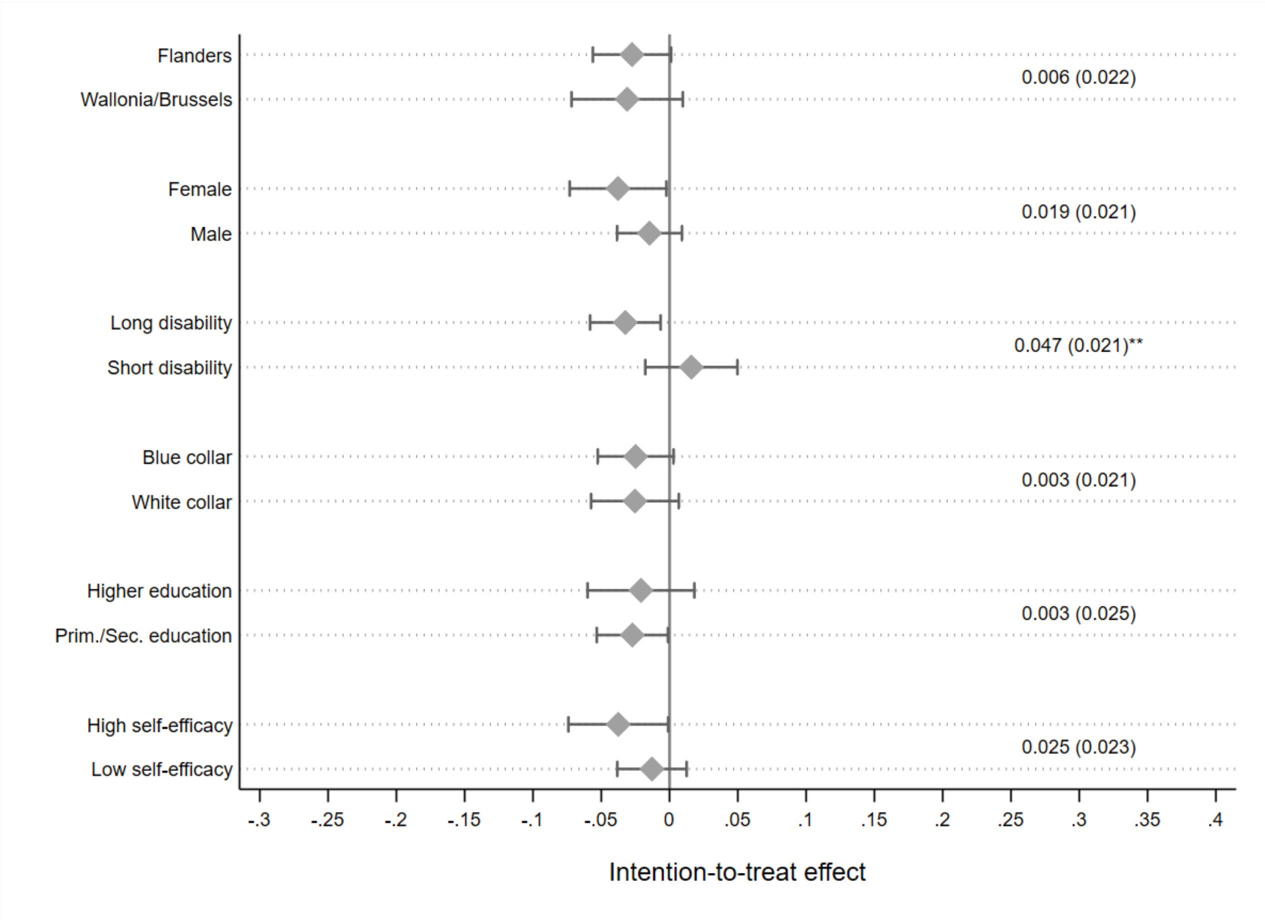
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (3.1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 8: Heterogeneity Analysis: Effects on Work while on Claim (0/1)



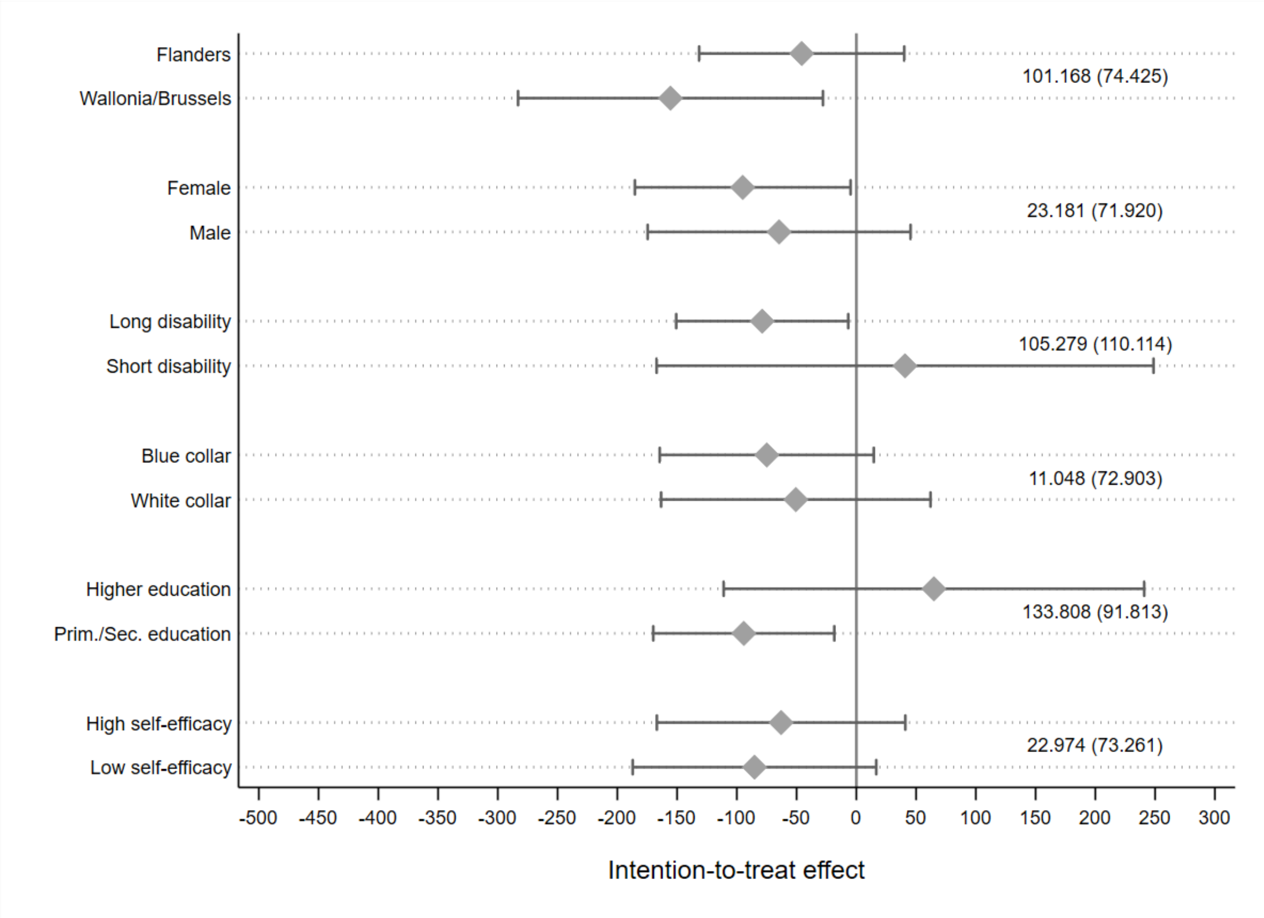
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (3.1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 9: Heterogeneity Analysis: Effects on Training while on Claim (0/1)



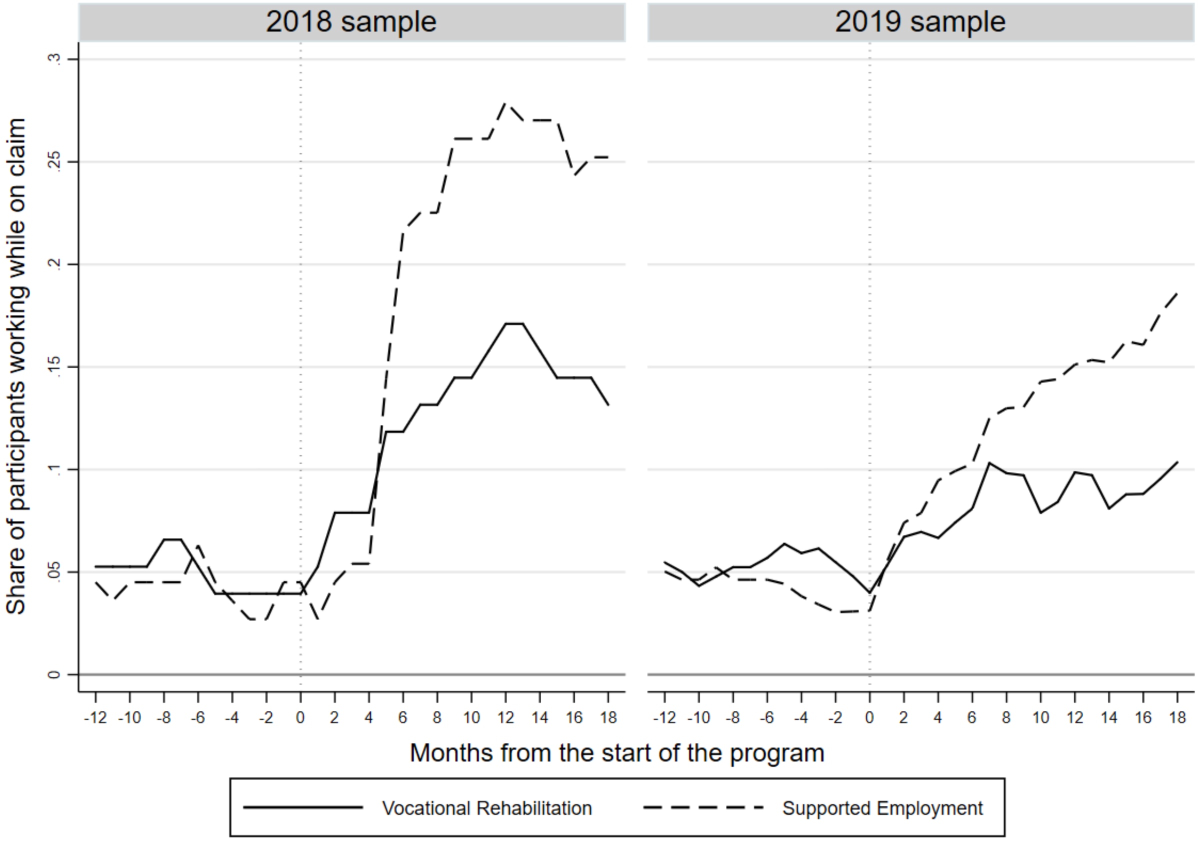
*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (3.1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 10: Heterogeneity Analysis: Effects on Disability Benefits (euros)



*Notes: The horizontal axis shows intention-to-treat estimates based on separate OLS regressions for the effects of Supported Employment (SE) compared to Vocational Rehabilitation (VR) 18 months after the start of the return-to-work program. The 95% confidence intervals are computed using heteroskedastic-robust standard errors. We also report on the graph the absolute value of the difference between the effects measured in the two groups (e.g. female vs male DI recipients), as well as the standard errors in parentheses (with significance levels displayed as *** p<0.01, ** p<0.05, * p<0.1). We use the same controls as in equation (3.1). Sample includes participants who entered the study between March 2018 and December 2019. Data source is from NIHDI administrative registers.*

Figure 11: Probability to Work while on Claim (Unconditional) for Participants who Started a Return-to-work Program in 2018 vs 2019



Notes: Horizontal axes show months since filling out baseline questionnaire and starting the return-to-work program. Vertical axis plots unconditional means for the share of study participants who work while on claim. Left panel includes only participants who entered the study between March 2018 and December 2018, while right panel includes only participants who entered the study between January 2019 and December 2019. Data source is from NIHDI administrative registers.

General Conclusion

This thesis covers three social insurance programs: maternity leave, paternity leave and Disability Insurance (DI). Each chapter discusses the design of a program, but also reveals significant spillover effects between them. This concluding section will summarize the most salient findings of each chapter and highlight their policy implications.

Policy design choices are not neutral: The first chapter sheds light on the maternity leave program and examines how the generosity of maternity leave allowance affects first time mothers' career trajectory and subsequent fertility decisions. It is of particular importance given that the level of compensation during maternity leave varies significantly across countries. Some, like Austria, France or Mexico, offer complete wage replacement, while others, like Canada, only cover 55% of lost income. In addition, many OECD countries have caps on the amount of benefits, which result in much lower replacement rates for high-earning women. The first chapter suggests that female entrepreneurship increases with the level of benefits provided during the maternity leave period. In fact, simulations for Belgium show that raising the benefit threshold to the 99^{th} percentile would contribute to reducing the gender gap in self-employment, while having positive spillover effects on the fertility of high-earning women.

Social insurance programs should not be studied in isolation: The second chapter shows how parenthood and parental gender affect the probability of experiencing work disability at a young age. The results suggest that gaps due to poor health and disability play an important role in mothers' labor market attachment. In this regard, the estimates reveal that about 1 out of 5 women who leave the labor market after having had children go on to claim DI benefits. At the same time, the second chapter demonstrates that the provision of paternity leave softens this child penalty, especially for first-time mothers. In fact, the study reveals substantial fiscal spillovers between two social security programs that are usually considered in isolation. The estimates show that those mothers who had a child with a father eligible for paternity leave receive on average 712 fewer euros in DI benefits, which represents a saving of 82 million euros to the Social Security budget. Back-of-the-envelope calculations suggest that spending on paternity leave could be more than compensated by the decade-long savings in mothers' DI benefits.

Randomized Controlled Trials (RCT) are an important addition to the applied economist's toolkit: The third chapter shows that Supported Employment (SE) is more successful than traditional vocational rehabilitation in increasing the proportion of DI recipients who work while on claim, therefore reducing their reliance on benefits and easing the burden for the social security budget. These findings should be of broad interest outside of Belgium, as most OECD countries face rising disability rolls, increasingly composed of people with mental health conditions. These results also highlight the importance of using RCTs for policy evaluation. In this specific case, because participants self-select into return-to-work programs, there was no comparable control group available to measure their impact. At the same time, there was no feasible natural experiment since the new program had never been tested in the Belgian context. To overcome this situation, researchers suggested to the National Institute for Health and Disability Insurance that it randomize access to this new program for a sample of participants, before considering scaling it up. Such an approach remains very much an exception rather than the norm in policy evaluation. The third chapter is therefore the result of a tight collaboration between researchers and civil servants at a public agency.

References

Adda, J., Dustmann, C., & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293-337.

Albanesi, S. (2016). Social insurance. In *The new palgrave dictionary of economics* (p. 1-7). London: Palgrave Macmillan UK.

Andersen, S. H. (2018). Paternity leave and the motherhood penalty: New causal evidence. *Journal of Marriage and Family*, *80*(5), 1125-1143.

Anderson, D. J., Binder, M., & Krause, K. (2002). The motherhood wage penalty: Which mothers pay it and why? *American Economic Review*, *92*(2), 354-358.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, *103*(484), 1481-1495.

Ando, M. (2017). How much should we trust regression-kink-design estimates? *Empirical Economics*, *53*(3), 1287-1322.

Andresen, M. E., & Nix, E. (2019). *What causes the child penalty? evidence from same sex couples and policy reforms* (Tech. Rep.). Statistics Norway, Research Department.

Andrew, A., Bandiera, O., Costa-Dias, M., & Landais, C. (2021). *Women and men at work* (Tech. Rep.).

Angelov, N., Johansson, P., & Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, *34*(3), 545-579.

Angelov, N., Johansson, P., & Lindahl, E. (2020). Sick of family responsibilities? *Empirical Economics*, *58*(2), 777-814.

Angrist, J. D., & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, *88*(3), 450-477.

Asai, Y. (2015). Parental leave reforms and the employment of new mothers: Quasi-experimental evidence from japan. *Labour Economics*, *36*, 72-83.

Autor, D. H., & Duggan, M. G. (2006). The growth in the social security disability rolls: A fiscal crisis unfolding. *Journal of Economic Perspectives*, *20*(3), 71-96.

Avdic, D., & Karimi, A. (2018). Modern family? paternity leave and marital stability. *American Economic Journal: Applied Economics*, *10*(4), 283-307.

Baggesgaard, Mertz, Ronchi, & Salvestrini. (2021). *Early exposure to entrepreneurship and the creation of female entrepreneurs.*

Bana, S., Bedard, K., & Rossin-Slater, M. (2020). The impacts of paid family leave benefits: Regression kink evidence from california administrative data. *Journal of Policy Analysis and Management*.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment*. *The Quarterly Journal of Economics*, *116*(1), 261-292.

Barrera-Osorio, F., Kugler, A. D., & Silliman, M. I. (2021). Job training through turmoil. *National Bureau of Economic Research Working Paper Series, No. 29565*.

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries.* NJ: Princeton University Press Princeton.

Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, *3*(1), S33-S58.

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, *2*(3), 228-55.

Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, *130*(2), 571-614.

Beuermann, D. W., Bottan, N. L., Hoffmann, B., Jackson, C. K., & Vera Cossio, D. A. (2021). Does education prevent job loss during downturns? evidence from exogenous school assignments and covid-19 in barbados. *National Bureau of Economic Research Working Paper Series, No. 29231*.

Bittmann, S. (2015). Ressources économiques des femmes et travail domestique des conjoints : quels effets pour quelles tâches? *Economie et Statistique*, 305-338.

Blanchflower, D. G., & Oswald, A. J. (1998). What makes an entrepreneur? *Journal of Labor Economics*, *16*(1), 26-60.

Boden, R. J. (1999). Flexible working hours, family responsibilities, and female self-employment: Gender differences in self-employment selection. *The American Journal of Economics and Sociology*, *58*(1), 71-83.

Boelmann, B., Raute, A., & Schönberg, U. (2020). Wind of change? cultural determinants of maternal labor supply. *Queen Mary University of London Working Papers*, *914*.

Borg, K., Hensing, G., & Alexanderson, K. (2001). Predictive factors for disability pension — an 11-year follow up of young persons on sick leave due to neck, shoulder, or back diagnoses. *Scandinavian Journal of Public Health*, *29*(2), 104-112.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *American Economic Review*, *79*(3), 482-503.

Bound, J., & Burkhauser, R. V. (1999). Economic analysis of transfer programs targeted on people with disabilities. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, p. 3417-3528). New York: Elsevier Science.

Broadway, B., & Kassenboehmer, S. C. (2019). Employment effects of job counseling for disability insurance recipients. *Melbourne Institute Working Paper*, *19*(18).

Bronars, S. G., & Grogger, J. (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review*, *84*(5), 1141-1156.

Bütikofer, A., Riise, J., & Skira, M. M. (2021). The impact of paid maternity leave on maternal health. *American Economic Journal: Economic Policy*, *13*(1), 67-105.

Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *The Review of Economics and Statistics*, *95*(3), 711-724.

Budig, M. J. (2006). Intersections on the road to self-employment: Gender, family and occupational class. *Social Forces*, *84*(4), 2223-2239.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326.

Card, D., Johnston, A., Leung, P., Mas, A., & Pei, Z. (2015a). The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in missouri, 2003-2013. *American Economic Review*, *105*(5), 126-30.

Card, D., Kluve, J., & Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894-931.

Card, D., Lee, D. S., Pei, Z., & Weber, A. (2015b). Inference on causal effects in a generalized regression kink design. *Econometrica*, *83*(6), 2453-2483.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234-261.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, *115*(531), 1449-1455.

Chatterji, P., & Markowitz, S. (2012). Family leave after childbirth and the mental health of new mothers. *J Ment Health Policy Econ*, *15*(2), 61-76.

Cheng, C.-Y., Fowles, E. R., & Walker, L. O. (2006). Continuing education module: postpartum maternal health care in the united states: a critical review. *The Journal of perinatal education*, *15*(3), 34-42.

Cohen, A., Dehejia, R., & Romanov, D. (2013). Financial incentives and fertility. *Review of Economics and Statistics*, *95*(1), 1-20.

Cohen, J., & Dupas, P. (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment*. *The Quarterly Journal of Economics*, *125*(1), 1-45.

Conde-Agudelo, A., Rosas-Bermudez, A., Castaño, F., & Norton, M. H. (2012). Effects of birth spacing on maternal, perinatal, infant, and child health: A systematic review of causal mechanisms. *Studies in Family Planning*, *43*(2), 93-114.

Congressional Research Service. (2018). *Trends in social security disability insurance enrollment* (Tech. Rep.).

Cools, S., Fiva, J. H., & Kirkebøen, L. J. (2015). Causal effects of paternity leave on children and parents. *The Scandinavian Journal of Economics*, *117*(3), 801-828.

Cortes, P., & Pan, J. (2020). Children and the remaining gender gaps in the labor market. *IZA Discussion Paper*, *13759*.

Cottier, L., Flückiger, Y., Kempeneers, P., & Lalive, R. (2018). Does job search assistance really raise employment? *IZA Discussion paper*(11766).

Craig, L. (2007). How employed mothers in australia find time for both market work and childcare. *Journal of Family and Economic Issues*, *28*(1), 69-87.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448-74.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment *. *The Quarterly Journal of Economics*, *128*(2), 531-580.

Crépon, B., Gurgand, M., Kamionka, T., & Lequien, L. (2012, Feb). *Is counseling welfare recipients cost-effective ? lessons from a random experiment.* Center for Research in Economics and Statistics.

Currie, J., & Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences*, *110*(30), 12265-12270.

Cygan-Rehm, K. (2016). Parental leave benefit and differential fertility responses: evidence from a german reform. *Journal of Population Economics*, *29*(1), 73-103.

Cygan-Rehm, K., Kuehnle, D., & Riphahn, R. T. (2018). Paid parental leave and families' living arrangements. *Labour Economics*, *53*, 182-197.

Dahl, G. B., & Gielen, A. C. (2020). Intergenerational spillovers in disability insurance. *American Economic Journal: Applied Economics*, *Forthcoming*.

Danzer, N., & Lavy, V. (2017). Paid parental leave and children's schooling outcomes. *The Economic Journal*, *128*(608), 81-117.

Dean, D., Pepper, J., Schmidt, R., & Stern, S. (2017). The effects of vocational rehabilitation services for people with mental illness. *Journal of Human Resources*, *52*(3), 826-858.

De Brouwer, O., Leduc, E., & Tojerow, I. (2019). The unexpected consequences of job search monitoring: Disability instead of employment? *IZA Discussion Paper*(12304).

De Brouwer, O., & Tojerow, I. (2018). Quels sont les déterminants des disparités géographiques de taux d'invalidité en belgique? *Revue Belge de Sécurité Sociale*, *2018*(1), 5-63.

Dillon, E. W., & Stanton, C. T. (2017). Self-employment dynamics and the returns to entrepreneurship. *National Bureau of Economic Research Working Paper Series*, *No. 23168*.

Dobkin, C., Finkelstein, A., Kluender, R., & Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, *108*(2), 308-52.

Drake, R. E., McHugo, G. J., Becker, D. R., Anthony, W. A., & Clark, R. E. (1996). The new hampshire study of supported employment for people with severe mental illness. *Journal of Consulting and Clinical Psychology*, *64*(2), 391-399.

Druedahl, J., Ejrnæs, M., & Jørgensen, T. H. (2019). Earmarked paternity leave and the relative income within couples. *Economics Letters*, *180*, 85-88.

Dunatchik, A., & Özcan, B. (2019). *Reducing mommy penalties with daddy quotas* (Tech. Rep.). Centre for Analysis of Social Exclusion, LSE. (CASE Papers)

Dustmann, C., & Schönberg, U. (2012). Expansions in maternity leave coverage and children's long-term outcomes. *American Economic Journal: Applied Economics*, *4*(3), 190-224.

Duvander, A.-Z., Lappegard, T., & Johansson, M. (2020). Impact of a reform towards shared parental leave on continued fertility in norway and sweden. *Population Research and Policy Review*, *Forthcoming*.

Ekberg, J., Eriksson, R., & Friebel, G. (2013). Parental leave — a policy evaluation of the swedish "daddy-month" reform. *Journal of Public Economics*, *97*(C), 131-143.

Eurofound. (2014). *Working conditions and job quality: Comparing sectors in europe* (Tech. Rep.).

Eurofound. (2020). Living, working and covid-19. *Publications Office of the European Union*, *COVID-19 series*.

European Commission, D. E. S. A. I. (2017). *Policy brief on women's entrepreneurship.* OECD Publishing.

Evans, D. S., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, *97*(4), 808-827.

Fadlon, I., & Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, *109*(9), 3162-91.

Fan, E., Liu, J.-T., & Chen, Y.-C. (2017). Is the quarter of birth endogenous? new evidence from taiwan, the us, and indonesia. *Oxford Bulletin of Economics and Statistics*, *79*(6), 1087-1124.

Farré, L., & González, L. (2019). Does paternity leave reduce fertility? *Journal of Public Economics*, *172*, 52-66.

Field, E. M., Linden, L. L., Malamud, O., Rubenson, D., & Wang, S.-Y. (2019). Does vocational education work? evidence from a randomized experiment in mongolia. *National Bureau of Economic Research Working Paper Series*, *No. 26092*.

Fisher, R. A. (1935). *The design of experiments.* Macmillan.

Flèche, S., Lepinteur, A., & Powdthavee, N. (2021). The importance of capital in closing the entrepreneurial gender gap: A longitudinal study of lottery wins. *Journal of*

Economic Behavior and Organization, *188*, 591-607.

Fogelgren, M., Ornstein, P., Rödin, M., & Skogman Thoursie, P. (2021). Is supported employment effective for young adults with disability pension? evidence from a swedish randomized evaluation. *Journal of Human Resources*, *(forthcoming)*.

Fontenay, S., & Tojerow, I. (2020). Work disability after motherhood and how paternity leave can help. *IZA Discussion Paper*(13756).

Fontenay, S., Tojerow, I., & Murphy, T. (2021). Child penalties across industries: Why job characteristics matter. *Applied Economics Letters*, *Forthcoming*.

French, E., & Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, *6*(2), 291-337.

Frey, W. D., Drake, R. E., Bond, G. R., Miller, A. L., Goldman, H. H., Salkever, D. S., & Holsenbeck, S. (2011). *Mental health treatment study: Final report to social security administration* (Tech. Rep.).

Fujiwara, T., & Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics*, *5*(4), 241-55.

Gangl, M., & Ziefle, A. (2009). Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in britain, germany, and the united states. *Demography*, *46*(2), 341-369.

Ganong, P., & Jäger, S. (2018). A permutation test for the regression kink design. *Journal of the American Statistical Association*, *113*(522), 494-504.

Gans, J. S., & Leigh, A. (2009). Born on the first of july: An (un)natural experiment in birth timing. *Journal of Public Economics*, *93*(1), 246-263.

Gelber, A., Moore, T. J., & Strand, A. (2017). The effect of disability insurance payments on beneficiaries' earnings. *American Economic Journal: Economic Policy*, *9*(3), 229-61.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics*, *37*(3), 447-456.

Georgellis, Y., & Wall, H. J. (2005). Gender differences in self-employment. *International Review of Applied Economics*, *19*(3), 321-342.

Gerards, R., & Theunissen, P. (2018). *Becoming a mompreneur: Parental leave policies and mothers' propensity for self-employment.* Maastricht University, unpublished manuscript.

Gerritse, E., Marcato, A., Plasman, R., & Tojerow, I. (2017). *Incapacité de travail: Quelles transitions et trajectoires après l'épisode d'incapacité?*

Geyer, J., & Welteke, C. (2021). Closing routes to retirement for women: How do they respond? *Journal of Human Resources*, *56*(1), 311-341.

Ginja, R., Jans, J., & Karimi, A. (2020). Parental leave benefits, household labor supply, and children's long-run outcomes. *Journal of Labor Economics*, *38*(1), 261-320.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091-1119.

González, L. (2013). The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply. *American Economic Journal: Economic Policy*, *5*(3), 160-88.

Gottlieb, J. D., Townsend, R. R., & Xu, T. (2021). Does career risk deter potential entrepreneurs? *Review of Financial Studies*, *Forthcoming*.

Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, *22*(3), 23-46.

Hainmueller, J., & Xu, Y. (2013). ebalance: A stata package for entropy balancing. *Journal of Statistical Software*, *54*(7), 1 - 18.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy*, *108*(3), 604-631.

Heß, S. (2017). Randomization inference with stata: A guide and software. *Stata Journal*, *17*(3), 630-651.

Heim, B. T., & Lurie, I. Z. (2010). The effect of self-employed health insurance subsidies on self-employment. *Journal of Public Economics*, *94*(11), 995-1007.

Hochschild, A., & Machung, A. (1990). *The second shift.* New York: Avon Books.

Holtz-Eakin, D., Joulfaian, D., & Rosen, H. S. (1994a). Entrepreneurial decisions and liquidity constraints. *The RAND Journal of Economics*, *25*(2), 334-347.

Holtz-Eakin, D., Joulfaian, D., & Rosen, H. S. (1994b). Sticking it out: Entrepreneurial survival and liquidity constraints. *Journal of Political Economy*, *102*(1), 53-75.

Hombert, J., Schoar, A., Sraer, D., & Thesmar, D. (2020). Can unemployment insurance spur entrepreneurial activity? evidence from france. *The Journal of Finance*, *75*(3), 1247-1285.

Hook, J. (2010). Gender inequality in the welfare state: Sex segregation in housework, 1965–2003. *American Journal of Sociology*, *115*(5), 1480-1523.

Hotchkiss, J. L., Pitts, M. M., & Walker, M. B. (2017). Impact of first birth career interruption on earnings: evidence from administrative data. *Applied Economics*, *49*(35), 3509-3522.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and u.s. economic growth. *Econometrica*, *87*(5), 1439-1474.

Hurst, E., & Lusardi, A. (2004). Liquidity constraints, household wealth, and entrepreneurship. *Journal of Political Economy*, *112*(2), 319-347.

IEFH. (2018). *Étude sur la dimension de genre du congé parental, du crédit-temps et de l'interruption de carrière* (Tech. Rep.).

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*(3), 933-959.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615-635.

INAMI. (2007). *Statistiques des indemnités - 2007.* https://www.riziv.fgov.be/fr/statistiques/indemnites/Pages/default.aspx.

ISSP. (2005). *International social survey programme: Work orientations iii - issp 2005.* GESIS Datenarchiv.

Jones, F., Mandy, A., & Partridge, C. (2009). Changing self-efficacy in individuals following a first time stroke: preliminary study of a novel self-management intervention. *Clinical Rehabilitation*, *23*(6), 522-533.

Kaplan, D. M. (2022). sivqr: Smoothed iv quantile regression. *Stata Journal*, *Forthcoming*.

Karimi, A. (2014). Effects of the timing of births on women's earnings - evidence from a natural experiment. *Institute for Evaluation of Labour Market and Education Policy Working Paper*, *2014*(17).

Kessler, R. C., Akiskal, H. S., Ames, M., Birnbaum, H., Greenberg, P., Hirschfeld, R. M., ... Wang, P. S. (2006). Prevalence and effects of mood disorders on work performance in a nationally representative sample of u.s. workers. *Am J Psychiatry*, *163*(9), 1561-8.

Kessler, R. C., Heeringa, S., Lakoma, M. D., Petukhova, M., Rupp, A. E., Schoenbaum, M., ... Zaslavsky, A. M. (2008). Individual and societal effects of mental disorders on earnings in the united states: results from the national comorbidity survey replication. *Am J Psychiatry*, *165*(6), 703-11.

Kirkwood, J. (2009). Motivational factors in a push-pull theory of entrepreneurship. *Gender in Management*, *24*(5), 346-364.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, *109*, 122-26.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181-209.

Kluve, J., & Schmitz, S. (2018). Back to work: Parental benefits and mothers' labor market outcomes in the medium run. *ILR Review*, *71*(1), 143-173.

Kotsadam, A., & Finseraas, H. (2011). The state intervenes in the battle of the sexes: Causal effects of paternity leave. *Social Science Research*, *40*(6), 1611-1622.

Lalive, R., Schlosser, A., Steinhauer, A., & Zweimüller, J. (2014). Parental leave and mothers' careers: The relative importance of job protection and cash benefits. *The Review of Economic Studies*, *81*(1), 219-265.

Landais, C. (2015). Assessing the welfare effects of unemployment benefits using the regression kink design. *American Economic Journal: Economic Policy*, *7*(4), 243-

78.
Laun, L., & Thoursie, P. S. (2014). Does privatisation of vocational rehabilitation improve labour market opportunities? evidence from a field experiment in sweden. *Journal of Health Economics*, *34*, 59-72.
Leduc, E., & Tojerow, I. (2020). Subsidizing domestic services: Job creation, employment formalisation and the unintended health costs. *IZA Discussion Paper*(13544).
Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281-355.
Liebman, J. B. (2015). Understanding the increase in disability insurance benefit receipt in the united states. *Journal of Economic Perspectives*, *29*(2), 123-50.
Lim, K., & Michelmore, K. (2018). The eitc and self-employment among married mothers. *Labour Economics*, *55*, 98-115.
Lindh, T., & Ohlsson, H. (1996). Self-employment and windfall gains: Evidence from the swedish lottery. *The Economic Journal*, *106*(439), 1515-1526.
Luciano, A., Drake, R. E., Bond, G. R., Becker, D. R., Carpenter-Song, E., Lord, S., ... Swanson, S. J. (2014). Evidence-based supported employment for people with severe mental illness: Past, current, and future research. *Journal of Vocational Rehabilitation*, *40*(1), 1-13.
Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, *107*(6), 1611-37.
Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review*, *103*(5), 1797-1829.
Malkova, O. (2018). Can maternity benefits have long-term effects on childbearing? evidence from soviet russia. *The Review of Economics and Statistics*, *100*(4), 691-703.
Markussen, S., & Røed, K. (2014). The impacts of vocational rehabilitation. *Labour Economics*, *31*, 1-13.
Markussen, S., & Røed, K. (2017). The gender gap in entrepreneurship – the role of peer effects. *Journal of Economic Behavior and Organization*, *134*, 356-373.
Marshall, T., Goldberg, R. W., Braude, L., Dougherty, R. H., Daniels, A. S., Ghose, S. S., ... Delphin-Rittmon, M. E. (2014). Supported employment: assessing the evidence. *Psychiatr Serv*, *65*(1), 16-23.
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.
Milligan, K. (2005). Subsidizing the stork: New evidence on tax incentives and fertility. *The Review of Economics and Statistics*, *87*(3), 539-555.
Modini, M., Tan, L., Brinchmann, B., Wang, M. J., Killackey, E., Glozier, N., ... Harvey,

S. B. (2016). Supported employment for people with severe mental illness: systematic review and meta-analysis of the international evidence. *Br J Psychiatry*, *209*(1), 14-22.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much?*. *The Quarterly Journal of Economics*, *122*(3), 1067-1101.

OECD. (2010). *Sickness, disability and work: Breaking the barriers.*

OECD. (2022). Labour force statistics : Summary tables. *OECD Employment and Labour Market Statistics (database).*

Olafsson, A., & Steingrimsdottir, H. (2020). How does daddy at home affect marital stability? *The Economic Journal*.

Olivetti, C., & Petrongolo, B. (2017). The economic consequences of family policies: Lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, *31*(1), 205-30.

Pailhé, A., & Solaz, A. (2008). Time with children: Do fathers and mothers replace each other when one parent is unemployed? *European Journal of Population*, *24*(2), 211-236.

Parker, K., & Wang, W. (2013). *Modern parenthood - roles of moms and dads converge as they balance work and family* (Tech. Rep.). Pew Research Center.

Parker, S. C. (2018). *The economics of entrepreneurship.* Cambridge University Press.

Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, *37*(4), 1009-1059.

Pepin, J. R., Sayer, L. C., & Casper, L. M. (2018). Marital status and mothers' time use: Childcare, housework, leisure, and sleep. *Demography*, *55*(1), 107-133.

Persson, P., & Rossin-Slater, M. (2019). When dad can stay home: Fathers' workplace flexibility and maternal health. *National Bureau of Economic Research Working Paper Series*, *No. 25902*.

Petrongolo, B., & Ronchi, M. (2020). Gender gaps and the structure of local labor markets. *Labour Economics*, *64*, 101819.

Raute, A. (2019). Can financial incentives reduce the baby gap? evidence from a reform in maternity leave benefits. *Journal of Public Economics*, *169*, 203-222.

Rege, M., & Solli, I. F. (2013). The impact of paternity leave on fathers' future earnings. *Demography*, *50*(6), 2255-2277.

Riphahn, R. T., & Wiynck, F. (2017). Fertility effects of child benefits. *Journal of Population Economics*, *30*(4), 1135-1184.

Rocha, V., & van Praag, M. (2020). Mind the gap: The role of gender in entrepreneurial career choice and social influence by founders. *Strategic Management Journal*, *41*(5), 841-866.

Rosen, S. (1981). The economics of superstars. *American Economic Review*, *71*(5), 845-858.

Rosenbaum, P. R. (2002). *Observational studies.* New York: Springer.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rossin-Slater, M. (2018). Maternity and family leave policy. In S. L. Averett, L. M. Argys, & S. D. Hoffman (Eds.), *The oxford handbook of women and the economy.* Oxford University Press.

Ryan, P. (1998). Random allocation of treatments in blocks. *Stata Technical Bulletin*, *41*, 43-46.

Saurel-Cubizolles, M.-J., Romito, P., Lelong, N., & Ancel, P. (2000). Women's health after childbirth: A longitudinal study in france and italy. *An International Journal of Obstetrics and Gynaecology*, *107*, 1202-1209.

Schwarzer, R. (1992). *Self-efficacy: Thought control of action.* Washington, DC, US: Hemisphere Publishing Corp.

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In . M. J. J. Weinman S. Wright (Ed.), *Measures in health psychology: A user's portfolio* (p. 35-37). Windsor, UK: NFER-NELSON.

Spiess, C. K., & Dunkelberg, A. (2009). The impact of child and maternal health indicators on female labor force participation after childbirth: Evidence for germany. *Journal of Comparative Family Studies*, *40*(1), 119-138.

Stichnoth, H. (2020). Short-run fertility effects of parental leave benefits: evidence from a structural model. *Empirical Economics*, *59*.

Sundström, M., & Duvander, A.-Z. E. (2002). Gender division of childcare and the sharing of parental leave among new parents in sweden. *European Sociological Review*, *18*(4), 433-447.

Tamm, M. (2019). Fathers' parental leave-taking, childcare involvement and labor market participation. *Labour Economics*.

Troske, K. R., & Voicu, A. (2013). The effect of the timing and spacing of births on the level of labor market involvement of married women. *Empirical Economics*, *45*(1), 483-521.

Truskinovsky, Y. (2021). The unintended consequences of informal child care subsidies for older women's retirement security. *Journal of Human Resources*, *Forthcoming*.

Tudor, S. (2020). Financial incentives, fertility and early life child outcomes. *Labour Economics*, *64*, 101839.

Ugreninov, E. (2013). Can family policy reduce mothers' sick leave absence? a causal analysis of the norwegian paternity leave reform. *Journal of Family and Economic Issues*, *34*(4), 435-446.

Visier, L. (1998). Sheltered employment for persons with disabilities. *Int'l Lab. Rev.*, *137*, 347.

Wallman, T., Wedel, H., Palmer, E., Rosengren, A., Johansson, S., Eriksson, H., &

Svärdsudd, K. (2009). Sick-leave track record and other potential predictors of a disability pension. a population based study of 8,218 men and women followed for 16 years. *BMC Public Health*, *9*(1), 104.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220-233.

Wellington, A. J. (2006). Self-employment: the new solution for balancing family and career? *Labour Economics*, *13*(3), 357-386.

Wewiorski, N. J., & Fabian, E. S. (2004). Association between demographic and diagnostic factors and employment outcomes for people with psychiatric disabilities: A synthesis of recent research. *Mental Health Services Research*, *6*(1), 9-21.