ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics & Business Economics

The effect of the 1974 Swedish parental leave reform on the gender gap in long-term mental health

Name student: Stijn Rotman

Student ID number: 447997

Supervisor: Lisa Voois

Second Assessor: xxx

Date final version: 15-07-2020

Abstract

Female life expectancy at birth is higher than men's, even though women report worse physical and mental health than men. This phenomenon has been extensively studied and tried to counteract with social policy. This thesis evaluates one such policy: the 1974 parental leave reform in Sweden that equalised paternity leave and maternity leave. The impact of this policy on the difference in mental health between men and women has been studied, using data form the Survey of Health, Ageing and Retirement in Europe. Two triple differences models have been estimated comparing mental health for Denmark and Sweden before and after the policy implementation for women and men separately. The resulting Average Treatment effects on Treated men and women have been used to evaluate the change in the gender gap in mental health. No significant change in the gender gap in mental health has been concluded, this result has been found to be generalizable. While other benefits of paternity leave have been demonstrated and should be considered by policymakers, it should also be considered that the long-term benefits regarding gender inequality in mental health are low or absent.

This thesis uses data from SHARE Waves 1 and 2 (DOIs: 10.6103/SHARE.w1.700, 10.6103/SHARE.w2.700), see Börsch-Supan et al. (2013) for methodological details.

The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see www.share-project.org).

Table of Contents

The Effect of the 1974 Swedish Parental Leave Reform on the Gender Gap in Long-term Mental Health

Morbidity and Mortality

Female life expectancy at birth is higher than men's, even though women report worse physical and mental health than men (Idler & Benyamini, 1997; Macintyre et al. 1999; Wingard et al. 1989). This paradox was initially identified by Nathanson (1975), who offered three possible explanations:

- Women do not suffer from worse health than men, they *report* more illness, because they feel more comfortable exhibiting their pain and symptoms, this view is elaborated on by Jaworska and Ryan (2018) and Sullivan et al. (2014). Both studies identify masculine stereotypes as hindering factors for seeking help in men with mental health problems.
- Women's traditional role of caregiving is more compatible with *sick role behaviour* (i.e. the restrictions in activities and submission to care), leading to a higher reporting of illness. Gove and Tudor (1973) called this explanation the 'fixed role hypothesis'. They argue that the reason women adapt to the sick role more easily, is that it is easier for them to put off their housework for a few days, while men cannot call in sick in their job too often, because that could lead to losing their jobs.
- Women have worse physical and mental health because their role as caregiver is a more stressful one. In sociological literature, this approach is known as the 'nurturant hypothesis'. Marcus and Seeman (1981) state that this hypothesis has an inherently different approach than the fixed role hypothesis by focusing on 'real illness' instead of self-reported illness. However, it was argued by Gove (1984) that these hypotheses complement each other:

> Evidently, at least part of the gender differences in physical and mental illness are due to the fact that the roles of men tend to have more structure than the roles of women, while women are more likely than men to occupy a nurturant role. (p. 82)

Another explanation of these gender differences in self-reported health is found by Case and Paxson (2005), who argue that the different chronic conditions men and women face can explain the differences in self-assessed health completely. Their epidemiological approach contradicts the first two explanations from Nathanson (1975) as well as the conclusions from Idler (2003), who found that women are more accurate when self-reporting health, and Spiers et al.

(2003), who state that less serious conditions are more likely to be taken into account by women than by men in the self-assessment of health. As for the gender gap in mortality, Case and Paxson (2005) found that men suffer from more severe forms of smoking-related diseases than women, suggesting higher tobacco usage among men as an explanation. Their hypothesis is in line with earlier research on this subject (Pampel, 2002; Valkonen & Van Poppel, 1997).

A Closer Look at Mental Health

Different measures of mental health, for instance suicidality, mental illness and depression, lead to different magnitudes of gender differences in mental health. The suicide rates of men are higher than those of women, reluctance in seeking professional help when suffering from depression being an important factor (Sullivan et al., 2014). Women are found to suffer from mental illness and depression more often (Macintyre et al., 1996; Van de Velde et al., 2010). To compare people with depressive symptoms in different European countries, the EURO-D scale was developed by Prince et al. (1999). The EURO-D scale is a 12-point scale that increases by 1 for each of the following depressive symptoms: pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, and tearfulness. A higher score points to worse mental health, with a score of 4 or higher indicating depression. Several factors lead to a higher score on this scale for women than for men: Buber and Engelhardt (2011) identify health and living conditions as predictors for depressive symptoms, finding these are worse for women and can also have different effects among men and women. Calvó-Perxas et al. (2016) also find an association between depression and pain, stating that women report pain and depression more often than men, as well as the combination of the two. These factors are found by Acciai and Hardy (2017) too, who also rule out different reporting behaviour between men and women as a reason for the higher incidence of depressive symptoms.

The nurturant hypothesis can help to further understand what drives the gender gap in depressive symptoms and EURO-D score. This idea of the female role as caregiver that decreases mental health is supported by Oakley (2018), who found that loneliness due to housework is a frequent complaint among housewives. Manuel and colleagues (2012) found that motherhood increases the chance of depressive symptoms, while emotional support from the partner and instrumental support are factors that partially mitigate this effect. Hank and Jürges (2007) found that the division of housework among older people is determined in equal parts by the characteristics of an older couple and their countries' score on the United Nations' gender empowerment measure.

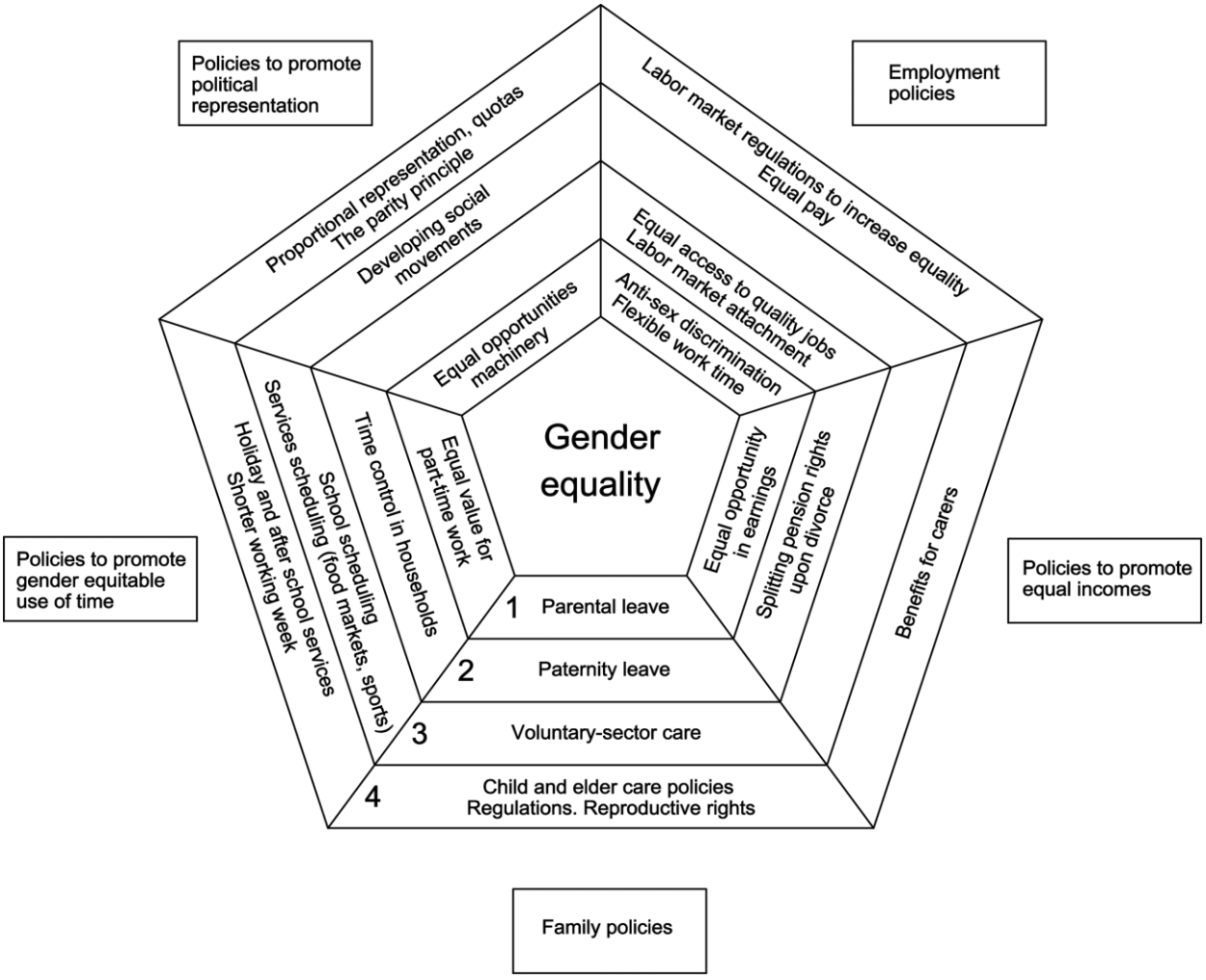
Closing the Gap with Gender Policy

Research about the determinants of the gender differences in physical and mental health is still ongoing, for example by Uccheddu and colleagues (2019), who state that better understanding the factors driving the gender gap in health helps to create gender policy that is more productive at reducing this inequality. Their research focused on the different effects of socio-economic status on health between men and women, measuring health using the 'Frailty Index', a measure aimed at the accurate indication of older individuals' overall health. They found that the effect of socio-economic status on women's health is smaller in defamilialized and decommodified countries. That is, countries with better welfare policy where women are less dependent on their families. This can be achieved by childcare and parental policies. These findings are in line with a study by Boye (2011) that showed that work-family conflict is a hindering factor in the association between paid working hours and higher levels of well-being for women. This suggests that, to increase well-being among women, more hours of paid work should be accompanied by a decrease in housework hours. A direct link between housework in combination with paid work and depression was found by Hoshino and colleagues (2016), who found that the combination of housework and paid work is a stress factor that impacts depression.

The effects of different aspects of gender policy on health are reviewed in a meta-analysis by Palència et al. (2017). This study uses the distinction of gender policy as introduced by Pascall and Lewis (2004). They distinguish five components of developmental gender regimes: paid work, care, income, time, and voice. These social determinants can be interpreted as follows: paid work means an equal access to jobs and equal pay and flexibility, care means the equal distribution of raising children and doing housework, income means independence from men in earning income, via own jobs and pensions, time means the flexibility to distribute the quantities of time spent in paid work, unpaid work and leisure, and voice means female representation in governments and companies, as well as having an equal say in major decisions as their male counterparts. These components can be implemented in four different levels of society: individual, household, community, and society. The social determinants and levels of society coincide, stemming from the researchers' goal of creating a holistic model. This model was slightly modified by Borrell et al. (2014), changing the social determinants of health to types of policies, maintaining the same societal levels of policy intervention (figure 1). In my thesis, I will study a specific policy that could be placed between the first and second quadrant of family policy: a non-mandatory paternal leave reform. Fathers were not obligated to take time off work but gained the opportunity to do so. If they pleased to not take time off work, they could give their time to their partners.

Figure 1

Policies for Gender Equality



Note. Modified from Pascall and Lewis (2004). Levels of policy intervention: 1, individual; 2, household; 3, civic/community; 4, society/politic.

Fontenay and Tojerow (2020) found that an increase in paternity leave is associated with a decrease in the chance for women to enter disability, likely due to the more equal division of housework among mothers and fathers. This effect is still observed after a period of 12 years. However, they did not find positive effects on female mental health. Andersen (2018) studied the effects of paternal leave reforms in Denmark, concluding that paternity leave reduced the wage gap within households and overall increased total household wages. Månsdotter et al. (2007) demonstrated that fathers that took paternity leave had a significantly lower risk of mortality than

fathers that did not take paternity leave. Månsdotter and Lundin (2010) showed that, while taking paternity leave and having a lower level of masculinity are both associated with lower risk of mortality, the causal pathways of these effects are separate. As for maternity leave, similar positive effects were found (Avendano et al, 2015; Dagher et al., 2014; Ruhm, 2011; Staehelin et al., 2007). This gives rise to the hypothesis that paternity leave has other long-term positive effects because of a redistribution of housework.

Parental Leave in Sweden

I will focus on the gender-neutral paid parental leave benefit introduced in 1974 in Sweden (Swedish Government Ministry of Employment, 2019). This reform ensured the same rights for mothers and fathers in terms of family and work by giving men the same amount and flexibility in parental leave as women. Parents receive a fixed number of days parental leave per child, the parental leave is transferable between parents. In later reforms, implemented in 1995, 2002, and 2016, more days were added, and a fixed number of months became non-transferable between parents (Swedish Government Ministry of Employment, 2019). As shown in table 1, men took up to two weeks parental leave on average in the first years succeeding the parental leave reform. Data regarding the percentage of men that took parental leave is not available. However, due to the number of days men went on parental leave on average, high participation can be assumed.

Table 1

Days of Paternal Leave per Child

Year	Live Births[ab]	Parental Leave[c]			
		Total Days[a]	Male Days (%)	Male Days[a]	Male Days per Birth
1974	110	19017	0	0	0
1980	97	27020	5	1351	14
1985	98	33193	6	1992	20
1990	124	48292	7	3380	27
1995	103	47026	10	4703	45
2000	90	35661	12	4279	47
2005	101	42659	20	8532	84
2010	116	49719	23	11435	99

[a]Values in thousands [b]Data obtained from Statistics Sweden (2020). [c]Data obtained from Statistics Sweden (2018).

Sweden is seen as archetypical for the dual-earner/dual-carer model in studies comparing gender regimes and family policy as a whole (Burstrom et al., 2010; Fritzell et al., 2012; Palència et al., 2017), being a counterpart to Britain for the market-oriented model and Italy for the traditional model. In the traditional model, policies are aimed at a traditional distribution of work, where men are employed while women stay at home as caregivers. The market-oriented model does not have many policies supporting households, with a free market to distribute resources instead. In the dual-earner/dual-carer model, many policies are aimed at female employment and male participation in the household. I will concentrate on the long-run effects of this specific policy in Sweden, comparing the gender gap for people who did not benefit from the policy to the gender gap in mental health for people who did benefit from the policy.

The research question I will study is “To what extent has the Swedish parental leave reform of 1974 contributed to the decrease of the gender gap in mental health in the long run?” The null hypothesis is that the Swedish 1974 parental leave reform had no significant effect on the gender gap in mental health in the long run. The alternative hypothesis is that the Swedish 1974 parental leave reform significantly decreased the gender gap in mental health in the long run. As stated earlier, there is extensive research on the effects of childcare policy on the gender gap in physical health in both the long and the short run and the research on the gender gap in mental health in the short run. My focus brings a new perspective of the effects of paternity leave on inequality between men and women in the long run to the scientific field of study. Furthermore, it broadens the comprehension of the effects of paternity leave for policymakers, aiding them to create policy that is more effective at reducing gender inequality.

This section introduced the relevant academic concepts and theories regarding gender differences in mental health and policies to counteract this inequality. A research question has been defined and relevant hypotheses are posed. The subsequent sections, Data & Methodology, Results, Conclusion, and Discussion will attempt to answer the research question. To start with, the Data & Methodology section will describe the methods used to collect and prepare the data. In addition, it will describe the research methodology used to analyse the data. Then, the Results section will report the implementation of the methodology as well as its outcomes, focussing on significance and magnitude of the effect. Furthermore, the Conclusion interprets the obtained results in relation to the null hypothesis and alternative hypothesis to pose an answer to the research question. Finally, the Discussion will reflect critically on both the internal and external validity of this thesis by describing the limitations and assumptions of the research methodology and assessing whether the obtained results will be homogeneous over place and time.

Data & Methodology

Data

This section will describe the dataset used to analyse the effect of the 1974 Swedish parental leave reform and explain the relevant scales and variables present in this dataset. To study the research question, I will use data from the Study of Health, Ageing and Retirement in Europe (SHARE) database. Data for the SHARE database is collected in seven waves from 2004 to 2017. The first two waves (2004-2007) will be used, to exclude possible effects on mental health from the 2008 Great Recession. This database contains interviews of people aged 50 or older, giving insight in their "physical and mental health, economic and non-economic activities, income and wealth, transfers of time and money within and outside the family as well as life satisfaction and well-being." (Börsch-Supan et al., 2013, p. 992). The outcome variables of these surveys I am interested in are EURO-D score and EURO-D caseness. The EURO-D scale is explained in the Introduction. EURO-D caseness is defined as a EURO-D score of 4 or higher it is the indication of being depressed.

SHARE data regarding demographics, children, behavioural risks, and mental health indices were combined into one dataset. Observations with zero children were dropped as well as observations that both contained one or more children before as well as one or more children after 1974, the year the new policy was implemented. The control country, Denmark (see Methods section), implemented paternity leave in 1989 (Andersen, 2018). Every person that benefitted from this reform will not be considered either. Individuals that were not born in the country of interview are also excluded from analysis, because it cannot be determined whether they benefitted from the parental leave reform. Only observations of Swedish parents who benefitted from paternity leave with all their children, Swedish parents who benefitted from paternity leave with none of their children, Danish people who did not benefit from paternity leave with any child but would have benefitted with all their children had they lived in Sweden, and Danish people who did not benefit from paternity leave with any child even if they had lived in Sweden remain. The resulting values for EURO-D score and EURO-D caseness are summarized by country and gender in table 2. From this table, it can be concluded that the differences in mental health between men and women are highly significant and substantial both in Sweden and Denmark.

Table 2

Descriptive statistics by country and gender

Variable	Sweden			Denmark		
	Men	Women	Difference	Men	Women	Difference
EURO-D score	1.550	2.260	-0.710***	1.524	2.053	-0.530***
EURO-D caseness	0.112	0.224	-0.112***	0.118	0.207	-0.090***
Number of observations	1270	918		1586	1182	

Note. Difference = Men – Women.

*p < .1. **p < .05. ***p < .01

Three categories of control variables will be in place in the models (see Methods section Control Variables): age at interview, age at birth of first child, and education. Age at interview is available in the SHARE data. Age at birth of first child is obtained by subtracting the birthyear of the individual from the birthyear of the first child. Education is measured by the International Standard Classification of Education (ISCED) score provided in the SHARE data. It was developed by UNESCO to compare the level of education among individuals and countries in six levels (2006). Level 1 stands for primary education or the first stage of basic education, level 2 stands for lower secondary or the second stage of basic education, level 3 stands for (upper) secondary education, level 4 stands for post-secondary non-tertiary education, level 5 stands for the first stage of tertiary education, and level 6 stands for the second stage of tertiary education. These categories are fit into three categories: primary education (levels 1 and 2), secondary education (levels 3 and 4), and tertiary education (levels 5 and 6). The following section will introduce a framework to formally assess the hypotheses. This assessment will aid in finding an answer to the research question in later sections.

Methodology

The preceding section stated the research question and hypotheses of this thesis. The preceding subsection described the steps taken in preparing the dataset for analysis of the hypotheses and research question. This subsection will first provide a framework for evaluating the impact of the 1974 parental leave reform on two outcome variables: EURO-D score and EURO-D caseness. Then, it will specify the control group and control variables.

EURO-D Score

In order to investigate the impact of the 1974 parental leave reform on the gender gap in EURO-D score, the following set of equations needs to be evaluated:

$$\Delta Gendergap = ATT^F - ATT^M \quad (1.1)$$

$$ATT^F = E\left(EUROD_i^F(1)\middle|S, t=1\right) - E\left(EUROD_i^F(0)\middle|S, t=1\right) \quad (1.2)$$

$$ATT^M = E\left(EUROD_i^M(1)\middle|S, t=1\right) - E\left(EUROD_i^M(0)\middle|S, t=1\right) \quad (1.3)$$

Where ∆Gendergap stands for the change in the difference of EURO-D score for men and women that is due to the implementation of the new law in Sweden. A significant negative value for ∆Gendergap would indicate a decrease of the gender gap in EURO-D score. ATT^F and ATT^M are the average treatment effects on the treated for women and men, respectively. In other words, the difference between the EURO-D score for Swedish women (men) under the new law, $\mathrm{E}\left(\mathrm{EUROD}_\mathrm{i}^{F(M)}(1)\middle|\mathrm{S}, \mathrm{t}=1\right)$ and the EURO-D score for Swedish women (men) had the new law not been implemented, $\mathrm{E}\left(\mathrm{EUROD}_\mathrm{i}^{F(M)}(0)\middle|\mathrm{S}, \mathrm{t}=1\right)$. The first terms of (1.2) and (1.3) are observed. The second terms are counterfactual, thus require to be estimated. Mere with-and-without (where a control country poses as a counterfactual) or before-after (where the period before the implementation of the new law poses as a counterfactual) comparisons do not suffice, because they suffer from selection bias. Rather, the triple differences (DDD) model in (1.4) will be estimated. This model contains two countries: Sweden and not-Sweden, two genders: male and female, and two time periods: before the implementation of the new policy and after the implementation of the new policy. The second time period ends in the year a parental leave policy is implemented in not-Sweden.

$$EUROD_{it} = \alpha + \beta S_i + \gamma t + \delta F_i + \zeta S_i F_i + \eta F_i t + \theta S_i t + \kappa S_i t F_i + \lambda X_{ti} + \nu X_{ti} F_i + \varepsilon_{it} \quad (1.4)$$

Where parameter α is the baseline, β captures the initial difference between the control country and Sweden, S is a dummy that indicates if an individual is Swedish, γ captures the time effect, t stands for the time period, δ captures the baseline difference between men and women, F is a dummy that indicates if an individual is female, ζ tests if the difference between Swedish men and women is equal to the difference between men and women from the control country, η tests if the time effect for women differs from that of men, θ captures the treatment effect for Swedish

men by testing if the time effect for Sweden differs from that of the control country, κ tests if the treatment effect differs between men and women, λ is a vector of the effects of country-specific time-varying factors, X is a vector of these country-specific time-varying variables, ν is a vector of tests if the country-specific time-varying factors have different effects between men and women, and ε_{it} is the error term. Since the treatment effect for men is captured by θ and κ captures the difference in treatment effect between men and women, the treatment effect for women is found by adding θ and κ. The value for ΔGendergap is equal to κ.

EURO-D Caseness

Using EURO-D score as the outcome variable has upsides as well as downsides. A problem with EURO-D score as outcome variable is that the linear interpretation of the sum of dummy variables assumes that every point increase in EURO-D score captures the same decrease in mental health. It is unsure whether this assumption holds since a change in EURO-D score from 2 to 1 might have other implications for mental health gains than a change from 6 to 5. An argument for using EURO-D score instead of EURO-D caseness is that it captures mental health gains at every level of mental health. Analysing only the impact on EURO-D caseness would find the gains for people who went from a score of 4 to 3 but disregards mental health gains for people who went from 2 to 1 or 6 to 5, since their EURO-D caseness did not change. Analysing both EURO-D score and EURO-D caseness leads to an understanding of the impact of the 1974 parental leave reform on the whole spectrum of mental health as well as the specific point of change from 'depressed' to 'not depressed'. The analysis of EURO-D caseness is done in a framework of formulas that is similar to that of EURO-D score:

$$\Delta Gendergap = ATT^F - ATT^M \tag{2.1}$$

$$ATT^F = E\left(Caseness_i^F(1)\middle|S, t = 1\right) - E\left(Caseness_i^F(0)\middle|S, t = 1\right) \tag{2.2}$$

$$ATT^M = E\left(Caseness_i^M(1)\middle|S, t = 1\right) - E\left(Caseness_i^M(0)\middle|S, t = 1\right) \tag{2.3}$$

$$\begin{gathered} Pr(Caseness_{it} = 1|S_i, t, F_i, X_{ti}) = \\ \alpha + \beta S_i + \gamma t + \delta F_i + \zeta S_i F_i + \eta F_i t + \theta S_i t + \kappa S_i t F_i + \lambda X_{ti} + \nu X_{ti} F_i + \varepsilon_{it} \end{gathered} \tag{2.4}$$

The difference between this framework and the framework used for the evaluation of EURO-D score lies mainly in the different nature of the outcome variables. Since EURO-D

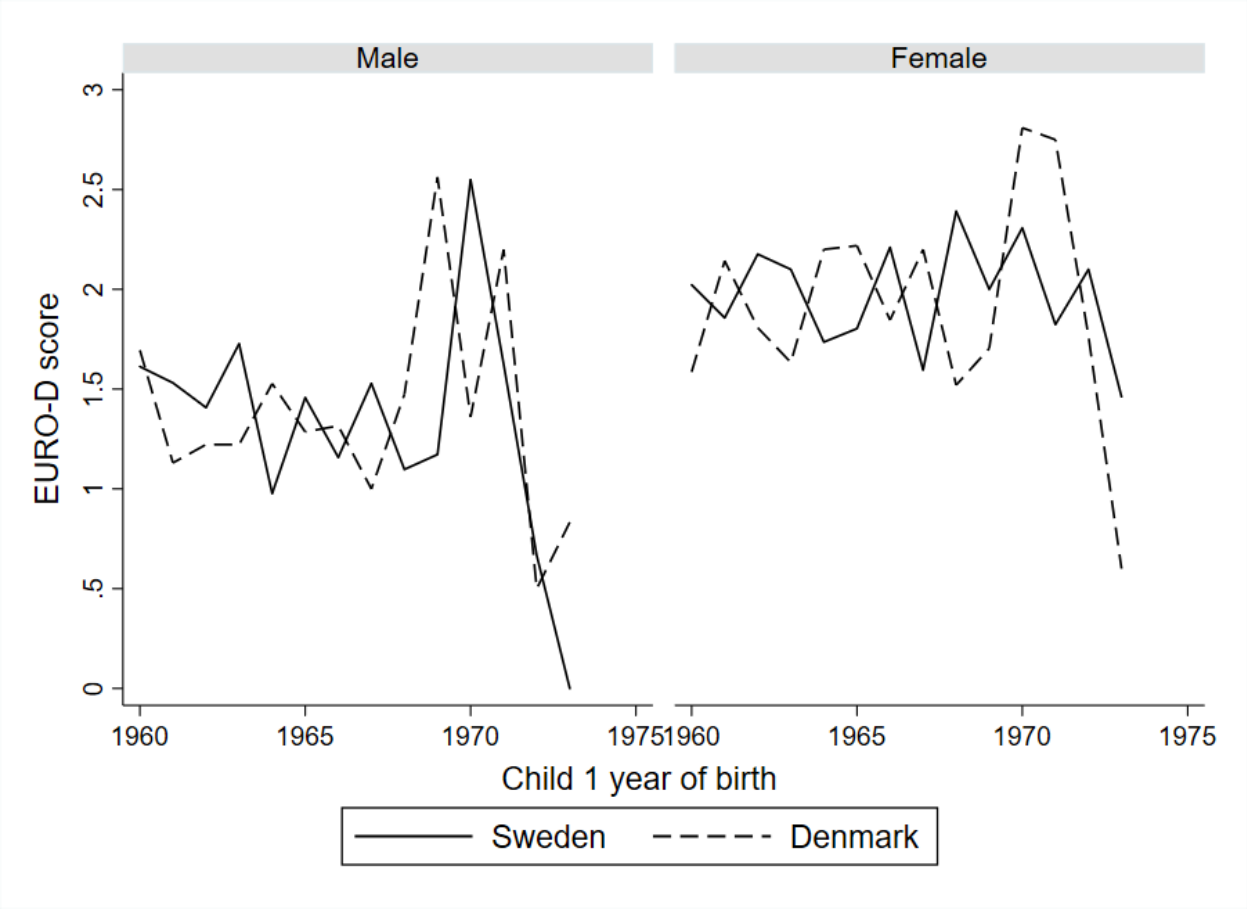
caseness is a dummy variable (someone is either depressed or not depressed), the DDD model estimates the chance of EURO-D caseness being equal to 1, conditional on nationality, time, gender, and control variables. The interpretation of the coefficients is similar: θ captures the effect of the 1974 policy on the chance of being depressed for men, κ tests if the effect is equal for men and women, its value is equal to ΔGendergap. The effect of the 1974 policy on the chance of being depressed for women is found by adding θ and κ.

Control Country

To minimize unobserved country-specific time-varying factors, the control country should be a country that resembles Sweden in the trends of EURO-D score and EURO-D caseness in pre-intervention years as closely as possible. Since the DDD models of (1.4) and (2.4) contain only two time periods, there are no pre-intervention trends in these models. For the assessment of the parallel trends assumption, the dataset will be divided into separate years based on the birth of the first child. Note that the exact values for mental health do not have to correspond, merely their trends. A more sophisticated way of finding a control country is the synthetic control method (Abadie et al., 2010), described in Appendix A, accompanied by an explanation why this method turned out not to lead to a meaningful conclusion. Countries that were weighted in the creation of synthetic Sweden (table A2) will be considered as control countries. Note that, as stated earlier, the time indicator is not mental health in a specific year, but rather the birthyear of the individual's first child. Since the treatment effects are estimated for men and women separately, the trend of the control country should resemble the trend of Sweden as closely as possible for both men and women. Figure 2 compares the trends of Sweden and Denmark for EURO-D score and EURO-D caseness. Comparisons of Sweden and the other potential control countries are shown in appendix B. While none of the potential control countries have the same trend in EURO-D score as Sweden the trends seems to run more parallel with Denmark than the other countries. The minimal contribution of Denmark in synthetic Sweden does not pose a problem for the validity of Denmark as a control country since the comparison method is different. The synthetic control method tries to match the level of the outcome variable as closely as possible by giving weights to potential control countries to resemble predictors of the outcome variable as closely as possible. This process results in a synthetic control group with a similar trend in the outcome variable as a result. However, corresponding levels of the outcome variable are not necessary in a DDD model, parallel trends are sufficient. In conclusion, while it does not provide a perfect match, Denmark will be used as control country.

Figure 2

Comparison of trends in EURO-D score: Sweden vs. Denmark



Control Variables

Now that the models and control country are defined, the last important affair of the methodology is the selection of control variables. Since the goal of the models is not to create a framework that explains the influencing factors of depressive symptoms and depression as much as possible, not all variables would qualify as control variables. The baseline difference in mental health between Denmark and Sweden is captured by α and time-varying influencing factors are also accounted for, by γ, provided that the time effect is constant among Denmark and Sweden. Only factors that differ over time in different ways between Denmark and Sweden need to be controlled for. However, while this condition is necessary to qualify as control variable, it is not sufficient. It is also important that the DDD model does not include 'mechanisms': variables that influence mental health and are caused by the parental leave reform. Including mechanisms as

control variables is harmful to the internal validity of the model because it takes away part of the effect of the policy implementation on mental health. I will compare earlier papers studying the effect of parental leave on mental health (Avendano et al., 2015; Dagher et al., 2014;Fontenay & Tojerow, 2020) with respect to the control variables they add and evaluate them as possible control variables, provided they are available in the SHARE data. The result of this comparison is shown in table 3.

Table 3

Common control variables for the effect of parental leave on mental health

Control variables	Avendano et al.	Dagher et al.	Fontenay & Tojerow
Age at childbirth	x	x	x
Education	x	x	
Marital status	x	x	
Number of children	x		x
Age	x		
Limitations with activities of daily living	x		
Limitations with instrumental activities of daily living	x		
Limitations with activities	x		
Smoking	x		
Drinking	x		
Race		x	
Income		x	
Occupation			

Common control variables are age at childbirth, education, marital status, and number of children. These variables will be considered as possible control variables, along with the age at interview. Assessing which potential control variables are mechanisms is hard to formalize. I will use earlier research to determine which variables are mechanisms and which variables are not. Only variables that have not been studied mechanisms of parental leave reforms will be added to (1.4) and (2.4) as control variables.

Avdic and Karimi (2018) find an effect of paternity leave on marital stability. Their findings suggest that marital status is a mechanism of paternity leave, it will therefore not be added as a control variable. No studies evaluating the impact of paternity leave on education have been found. Since education is often determined before one has children, the level of education is likely

not to be a mechanism, it will therefore be added as a control variable to the DDD models. The effect of paternity leave on fertility is studied by Duvander and Andersson (2006). Their findings suggest an increase of childbearing for parents who benefitted from paternity leave. This makes clear that the number of children is a mechanism of paternity leave, it will therefore not be included in the DDD models. Månsdotter and colleagues (2007) find a positive effect of the implementation of paternity leave on life expectancy for both men and women. On average, this leads to a higher life expectancy for people who benefitted from paternity leave. This could lead to the conclusion that age at interview is a mechanism of paternity leave and should be omitted from the DDD models. However, the effect of omitting the age at interview might bias the treatment effects more than the effect of including the possible mechanism in the models. The rationale behind this is that the different ages at interview of people in time period 1 than people in time period 0 differ by design of the two groups. It is only logical that two groups that are selected on the moment they had their first child differ in age. Therefore, age at interview will be added as a control variable. No study assessing the effect of paternity leave on age at childbirth has been found. Therefore, it will be added as a control variable as well.

This section described the collected data in the SHARE survey and the variables used as outcome variables. Then, two DDD regression models were stated to evaluate ΔGendergap for EURO-D score and EURO-D caseness. Furthermore, the control country was specified to be Denmark. An assessment has been done to find possible variables that need to be controlled for. The control variables included in the models will be level of education, age at interview and age at childbirth. Next, the Results sections will implement the DDD models and perform a graphical analysis to aid in finding an answer to the research question. Then, the Conclusion & Discussion will interpret the results and comment on the assumptions, internal validity, and external validity of the analysis.

Results

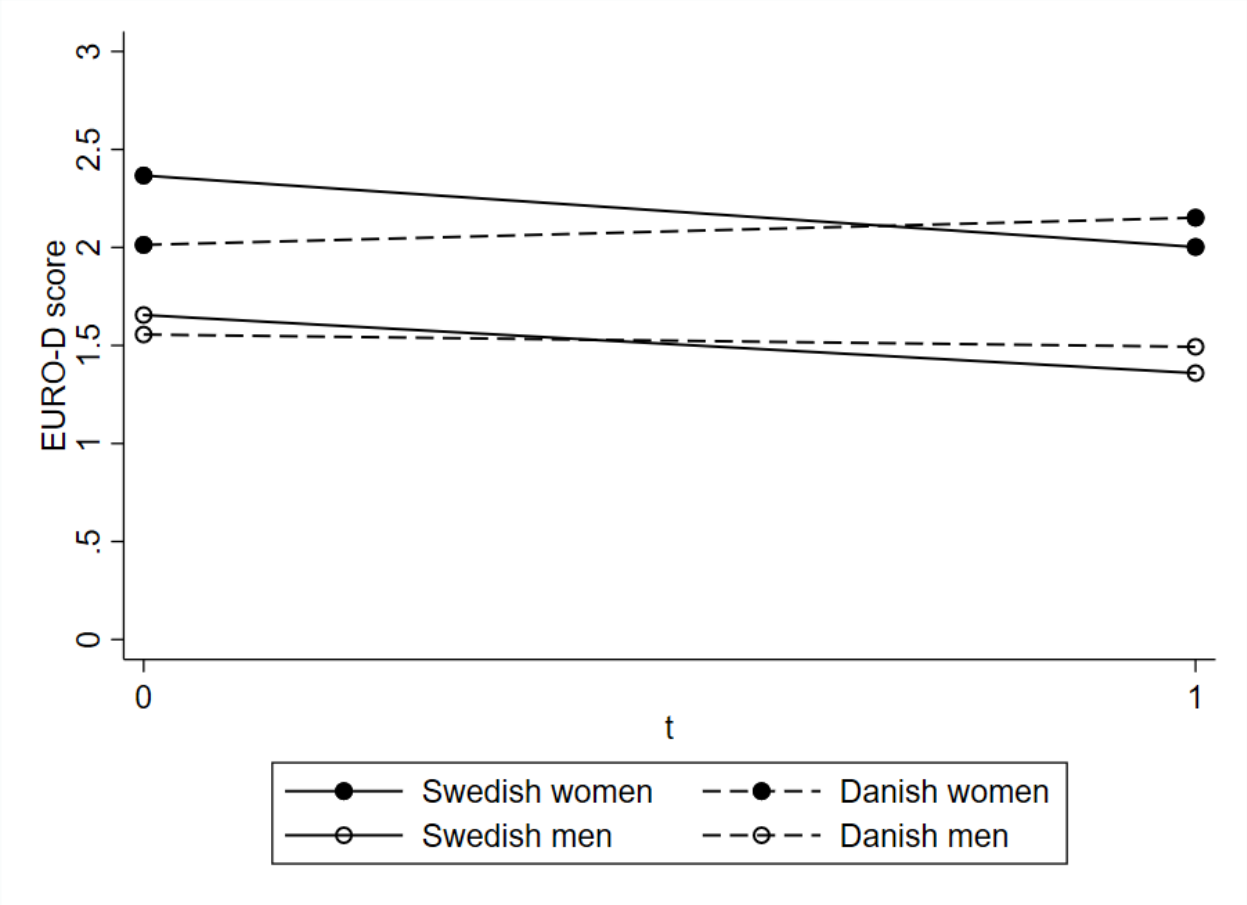
The Data & Methodology section provided two DDD models to evaluate the impact of the 1974 parental leave reform in Sweden on EURO-D score and EURO-D caseness. It also provided a description of the variables used in the models. This section will implement the model and provide a graphical and statistical analysis to pave the way for the assessment of the hypotheses and the answer to the research question in the Conclusion.

EURO-D Score

Figure 3 plots the mean values for EURO-D score at times 0 and 1 (before 1974 and 1974 onwards) for Sweden and Denmark by gender. The treatment effects can be estimated by starting the line of Denmark at the point of Sweden and measuring the difference between the resulting value at time period 1 and the actual value of Sweden at time period 1. From this figure, it can be inferred that it is likely that there are treatment effects for both men and women, because the lines of Denmark go up or stay roughly constant and the lines of Sweden go down. The implementation of (1.4) leads to more precise estimates and provides standard errors.

Figure 3

EURO-D scores by country, time period, and gender



The results of the implementation of (1.4) are summarised in table 6. As presented in the table, the average treatment effect on treated men equals -0.116, this effect is not statistically significant. The average treatment effect on treated women equals -0.186. Since the coefficient of the gender gap is not significant, it cannot be rejected that the average treatment effect on

treated women is equal to the average treatment effect on treated men. Therefore, it cannot be rejected that the average treatment effect on treated women is equal to zero.

Table 6

DDD model estimating the treatment effect and testing if the treatment effect differs by gender for EURO-D score

Variable	Coefficient	SE	95% CI		p
Intercept	2.762	0.869	1.058	4.466	0.001***
Sweden	-0.085	0.150	-0.379	0.210	0.574
Women	2.349	1.282	-0.166	4.863	0.067*
Time	-0.120	0.246	-0.603	0.362	0.625
Sweden*Women	0.042	0.211	-0.372	0.456	0.843
Women*Time	0.061	0.342	-0.609	0.732	0.857
Treatment men	-0.116	0.192	-0.492	0.260	0.545
Gendergap	-0.070	0.288	-0.636	0.496	0.808
Age at survey	-0.013	0.014	-0.040	0.014	0.344
Women*Age at survey	-0.018	0.020	-0.056	0.020	0.357
Age at birth	0.020	0.015	-0.009	0.049	0.180
Women*age at birth	-0.019	0.021	-0.061	0.022	0.363
Highest education					
Primary	-0.897	0.472	-1.822	0.027	0.057*
Secondary	-0.978	0.470	-1.898	-0.057	0.037**
Tertiary	-1.151	0.471	-2.075	-0.227	0.015**
Women*Highest education					
Primary	-0.075	0.762	-1.569	1.419	0.922
Secondary	-0.288	0.760	-1.780	1.203	0.705
Tertiary	-0.094	0.761	-1.586	1.398	0.902
Observations	3012				

*p < .1. **p < .05. ***p < .01

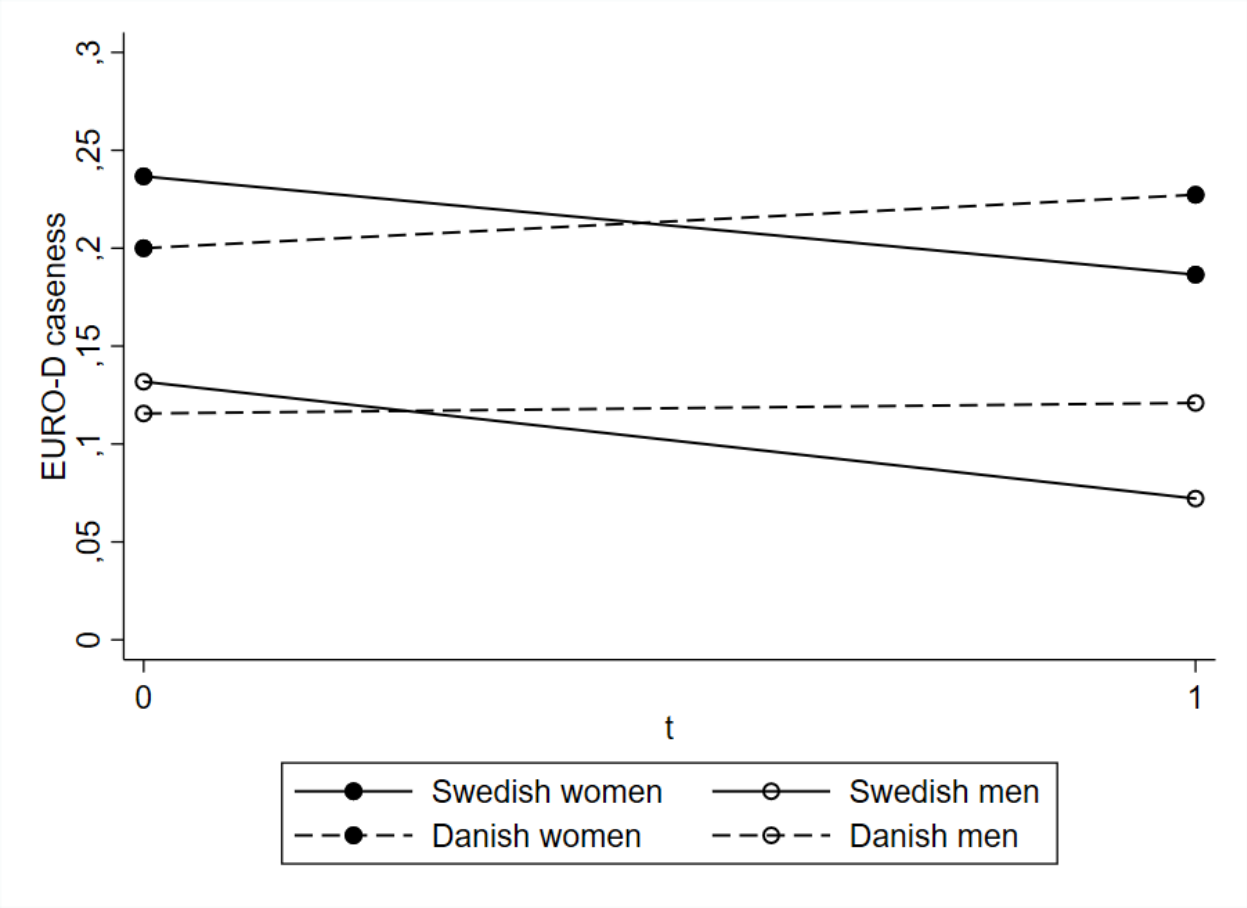
EURO-D Caseness

Figure 4 plots the mean values for EURO-D caseness at times 0 and 1 for Denmark and Sweden by gender. As with figure 3, an indication of the treatment effect can be found by drawing the trend of Denmark with Sweden at time period 0 as starting point. Comparing figure 3 and figure 4, it seems likely that the treatment effects are larger for men as well as for women for EURO-D caseness than for EURO-D score, because the trends for Denmark have a stronger negative slope and the trends for Sweden have a stronger positive slope. With EURO-D caseness

as well as with EURO-D score, the effect seems stronger for women than for men. However, more formal analysis by implementing the model in (2.4) is needed to make such comparisons valid.

Figure 4

EURO-D caseness by country, time period, and gender



The results of the implementation of (2.4) are summarised in table 7. As presented in the table, the average treatment effect on treated men equals -0.044, a decrease in the chance of being depressed of 4.4 percentage points. This effect is not statistically significant. The average treatment effect on treated women equals -0.016, a decrease in the chance of being depressed of 1.6 percentage points. Since the coefficient of the gender gap is not significant, it cannot be rejected that the average treatment effect on treated women is equal to the average treatment effect on treated men. Therefore, it cannot be rejected that the average treatment effect on treated women is equal to zero.

Table 7

DDD model estimating the treatment effect and testing if the treatment effect differs by gender for EURO-D caseness

Variable	Coefficient	SE	95% CI		p
Intercept	0.386	0.179	0.035	0.736	0.031**
Sweden	-0.007	0.026	-0.057	0.044	0.793
Women	0.311	0.280	-0.239	0.861	0.267
Time	-0.041	0.042	-0.124	0.041	0.327
Sweden*Women	-0.022	0.040	-0.100	0.057	0.585
Women*Time	0.039	0.065	-0.088	0.166	0.549
Treatment men	-0.044	0.033	-0.109	0.021	0.184
Gendergap	0.030	0.055	-0.078	0.138	0.587
Age at survey	-0.005	0.003	-0.010	0.000	0.051*
Women*Age at survey	0.001	0.004	-0.007	0.008	0.828
Age at birth	0.005	0.003	-0.001	0.011	0.085*
Women*age at birth	-0.005	0.004	-0.014	0.003	0.213
Highest level of education					
Primary	-0.072	0.125	-0.318	0.174	0.567
Secondary	-0.099	0.125	-0.344	0.146	0.430
Tertiary	-0.100	0.125	-0.345	0.146	0.427
Women*Highest education					
Primary	-0.138	0.203	-0.536	0.260	0.497
Secondary	-0.136	0.203	-0.534	0.261	0.501
Tertiary	-0.134	0.203	-0.532	0.263	0.509
Observations	3012				

*p < .1. **p < .05. ***p < .01

∆Gendergap

The preceding subsections described the results of the DDD models in (1.4) and (2.4), this subsection will interpret these results in terms of the change in the gender gap in EURO-D score and EURO-D caseness due to the 1974 parental leave reform. Table 8 interprets the results of tables 6 and 7 in terms of the average treatment effects on treated men and women and of ∆Gendergap. This table shows that the gender gap in terms of EURO-D score decreased, while it increased for EURO-D caseness. However, neither of these results are significant. The row labeled '∆Gendergap (relative)' is calculated by dividing the absolute changes in gender gap by the initial gender gaps of EURO-D score and EURO-D caseness. The large relative increase in

gender gap in EURO-D caseness stands out. However, since the absolute values of ΔGendergap are not significant, the relative values are not significant either.

Table 8

Comparison of outcomes of EURO-D score and EURO-D caseness

Outcome	EURO-D score	EURO-D caseness
Initial Gendergap	0.710	0.112
ATT (men)	-0.116	-0.044
ATT (women)	-0.186	-0.016
ΔGendergap (absolute)	-0.070	0.030
ΔGendergap (relative)	-9.9%	26.8%

*p < .1. **p < .05. ***p < .01

This section reported the results of the implementation of the frameworks that were described in the Data & Methodology section, testing for the difference in average treatment effect on the treated between men and women for two measures of mental health: EURO-D score and EURO-D caseness and interpreting the treatment effects in the context of the change in gender gap for these outcome variables. No statistically significant average treatment effects on treated men were found, the average treatment effects on treated women did not differ significantly from that of treated men. The following section will draw a conclusion with respect to the hypotheses to state an answer to the research question.

Conclusion & Discussion

This thesis studied the research question “To what extent has the Swedish parental leave reform of 1974 contributed to the decrease of the gender gap in mental health in the long run?” In the Introduction, an alternative was posed to the null hypothesis that the Swedish 1974 parental leave reform had no significant effect on the gender gap in mental health in the long run: the alternative hypothesis is that the Swedish 1974 parental leave reform significantly decreased the gender gap in mental health in the long run. The Data & Methodology section described the contents of the SHARE dataset, the data used to analyse the hypotheses and research question. This section also provided a framework to formally study the effect of the 1974 Swedish parental leave reform on ΔGendergap, the change in the gender gap in mental health because of the new policy, measured in EURO-D score and EURO-D caseness. The Results section reported EURO-D score and EURO-D caseness separately. In this section, these separate sections will be used to form a joint conclusion, since the research question concerned only one outcome: mental

health. The separate analysis of mental health in two outcome variables in different models should not lead to different conclusions regarding the hypotheses and research question.

While the results of ΔGendergap in terms of EURO-D score and EURO-D caseness are contradictory, both are insignificant. Therefore, I conclude that the difference in mental health for women and men was not impacted by the implementation of the 1974 parental leave reform in Sweden. In other words, there are insufficient grounds to reject the null hypothesis that the Swedish 1974 parental leave reform had no significant effect on the gender gap in mental health in the long run. The answer to the research question and the conclusion of this thesis is that the implementation of the 1974 parental leave reform in Sweden had no significant effect on the difference in mental health between men and women in the long run. The following section will elaborate on the internal and external validity of this result as well as its policy implications.

The most important assumption of the DDD models is the parallel trends assumption. This assumption relaxes the assumption of a with-and-without comparison that Sweden would have the same outcome as Denmark had Sweden not implemented the new policy, as well as the assumption of a before-after comparison that the outcome at time period 1 would be the same as the outcome at time period 0 had the new policy not been implemented at time period 1. The parallel trends assumption states that the outcomes of Sweden and Denmark would have changed with the same magnitude had the new policy not been implemented in Sweden. Because the DDD models contain two time periods, the parallel trends assumption cannot be formally assessed (otherwise, placebo tests putting the intervention year at earlier dates could have been performed). However, it has been attempted to check this assumption in figure 2. It can be concluded from this figure that the trends of Denmark and Sweden do not run perfectly parallel. As has been stated in the Data & Methodology section, Denmark does not provide a perfect match. However, sufficient similarity in trends has been assumed for further analysis.

When evaluating the impact of a change in behaviour or policy, it is important to be aware of the bad control problem. Not every variable qualifies as a good control variable and adding unnecessary control variables to a model can be harmful to its validity. Not all variation in the values of a potential control variable is necessarily due to the effect of the independent variable on the potential control variable. Consequently, a potential control variable could be a mechanism and endogenous at the same time. Since it is both impossible to distinguish the endogenous effect from the mechanism and to control for a variable only partially, some bias in the models is inevitable. Modern econometrics has not yet found a solution to this problem.

The primary data source for this study was the SHARE dataset. In the data resource profile, Börsch-Supan and colleagues (2013) pose that, while the response rate is higher than similar studies, the response rate is still quite low (about 62% on average for wave 1 and about 54% for wave 2). In combination with the moderate attrition levels, this could lead to selection bias if unit nonresponse and attrition are non-random, especially when evaluating social policy. The found results are robust on including cross-sectional and longitudinal weights, indicating the sample of respondents to be representative of the population. However, the format of the survey leads to the limitation that no analysis on the effects of paternity leave can be done directly. Since no questions regarding parental leave were present in the survey, paternity leave is inferred based on the combination of the country and year of childbirth. Therefore, different effects for different durations of paternity leave cannot be estimated separately.

The estimation of the average treatment effects on treated men and women for EURO-D caseness was done in a linear probability model. This is not an ideal method, since it does not account for the constraints that probabilities cannot be smaller than zero or larger than one (Cameron & Trivedi, 2005). Even though the estimated model did not lead to any predicted values smaller than zero or larger than one, estimating a logit or a probit model would still lead to a stronger internal validity. Ai and Norton (2003) reveal the problems regarding the interpretation of nonlinear models with interaction terms. However, the suggested method for finding treatment effects in a nonlinear difference-in-differences model by Puhani (2012) is also applicable to nonlinear DDD models. The reason for performing a linear DDD instead of a logit or probit is the inaccuracy of regularly calculated standard errors in nonlinear regression models with interaction terms (Ai & Norton, 2003) the solution of Norton and colleagues (2004) for calculating standard errors is too specific to be applicable to a DDD model. Puhani (2012) suggests using the delta method or bootstrapping to obtain the correct standard errors. However, both are outside the scope of this thesis.

The difference-in-differences method is held in high regard in terms of internal validity. However, due to the imperfect assessment of the parallel trends analysis, the bad control problem, and the nature of the survey data, a degree of uncertainty remains regarding the internal validity of the conclusion. Generalising the conclusion of this thesis to a policy advice requires external validity, which is impacted by the level of internal validity. In summary, the finding that the parental leave reform did not lead to a significant reduction in mental health needs to be internally valid as well as externally valid to give meaningful policy advice.

External validity seems probable given the similarity of the effects to the research by Fontanay and Tojerow (2020), which found low impact on mental health in an analysis of the implementation of paternal leave in Belgium. The reforms of the parental leave in Sweden in later periods (the first reform to parental leave was in 1995; Swedish Government Ministry of Employment, 2019) suggest that the long-term results of the 1974 policy were not found sufficient at other times in Sweden either. This leads to the conclusion that the low or absent effect on the gender gap in mental health in the long run should be taken into consideration by policymakers. This is not a general advice against the implementation of paternal leave or gender equality oriented parental leave. Other positive impacts of these policies have been studied in former research that has been exemplified in this thesis. These should be considered as well.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493-505. https://doi.org/10/bg2f7t

Acciai, F., & Hardy, M. (2017). Depression in later life: A closer look at the gender gap. *Social Science Research, 68*, 163-175. https://doi.org/10/gcmzpc

Ai, C., & Norton, E.C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123-129. https:??doi.org/10/bdd78k

Andersen, S.H. (2018). Paternity leave and the motherhood penalty: New causal evidence. *Journal of Marriage and Family, 80,* 1125-1143. https://doi.org/10/gd6hjc

Avdic, D., & Karimi, A. (2018). Modern family? Paternity leave and marital stability. *American Economic Journal: Applied Economics, 10*(4), 238-307. https://doi.org/10/gfj6z5

Avendano, M., Berkman, L.F., Brugiavini, A., & Pasini, G. (2015). The long-run effect of maternity leave benefits on mental health – Evidence from European countries. *Social Science & Medicine, 132*, 45-53. https://doi.org/10/f69r8r

Borrell, C., Palència, L., Muntaner, C., Urquía, M., Malmusi, D., & O'Campo, P. (2014). Influence of macrosocial policies on women's health and gender inequalities in health. *Epidemiologic Reviews, 36*, 31-48. https://doi.org/10/f5m76m

Börsch-Supan, A. (2019). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1* (7.0.0) [Data set]. SHARE-ERIC. https://doi.org/10/dvkk

Börsch-Supan, A. (2019). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2* (7.0.0) [Data set]. SHARE-ERIC. https://doi.org/10/dvkm

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology, 42*(4), 992-1001. https://doi.org/10/f5dndk

Boye, K. (2011). Work and well-being in a comparative perspective: The role of family policy. *European Sociological Review*, *7*(1), 16-30. https://doi.org/10/fmfgd5

Buber, I., & Engelhardt, H. (2011). The association between age and depressive symptoms among older men and women in Europe. Findings from SHARE. *Comparative Population Studies – Zeitschrift für Bevölkerungswissenschaft, 36(*1), 103-126. https://doi.org/ 10.4232/10.CPoS-2011-02en

Burstrom, B., Whitehead, M., Clayton, S., Fritzell, S., Vannoni, F., & Costa, G. (2010). Health inequalities between lone and couple mothers and policy under different welfare regimes – The example of Italy, Sweden and Britain. *Social Science & Medicine, 70*. 912-920. https://doi.org/10/fshfpk

Calvó-Perxas, L., Vilalta-Franch, J., Turró-Garriga, O., López-Pousa, S., & Garre-Olmo, J. (2016). Gender differences in depression and pain: A two year follow-up study of the Survey of Health, Ageing and Retirement in Europe. *Journal of Affective Disorders, 193*, 157-164. https://doi.org/10/f78hqb

Cameron, C.A., & Trivedi, P.K. (2005). *Microeconometrics: Methods and applications.* Cambridge University Press.

Case, A., & Paxson, C. (2005). Sex differences in morbidity and mortality. *Demography, 42*, 189-214. https://doi.org/10/cjpfn8

Dagher, R.K., McGovern, P.M., Dowd, B.E. (2014). Maternity leave duration and postpartum mental and physical health: implications for leave policies. *Journal of Health Politics, Policy and Law, 39*(2), 369-416. https://doi.org/10/dwrq

Duvander, A.Z., & Andersson, G. (2006). Gender equality and fertility in Sweden; A study on the impact of the father's uptake of parental leave on continued childbearing. *Marriage & Family Review, 39*(1-2), 121-142. https://doi.org/10/bvfxmk

Fontenay, S., & Tojerow, I. (2020). The “other" child penalty: Work disability after motherhood and how paternity leave can help. [not published in journal].

Fritzell, S., Vannonni, F., Whitehead, M., Burström, B., Costa, G., Clayton, S., & Fritzell, J. (2012). Does non-employment contribute to the health disadvantage among lone mothers in Britain, Italy and Sweden? Synergy effects and the meaning of family policy. *Health & Place, 18*. 199-208. https://doi.org/10/b85xqz

Gove, W.R. (1984). Gender differences in mental and physical illness: The effects of fixed roles and nurturant roles. *Social Science & Medicine, 19*(2). 77-84. https://doi.org/10/fnh6bq

Gove, W.R., & Tudor, J.F. (1973). Adult sex roles and mental illness. *American Journal of Sociology, 78*(4), 812-835. https://www.jstor.org/stable/2776605

Hank, K., & Jürges, H. (2007). Gender and the division of household labor in older couples: A European perspective. *Journal of Family Issues, 28*(3), 399-421. https://doi.org/10.1177/0192513X06296427

Hoshino, A., Amano, S., Suzuki, K., & Suwa, M. (2016). Relationships between depression and stress factors in housework and paid work among Japanese women. *Hong Kong Journal of Occupational Therapy*, *27*(1), 35–41. https://doi.org/10/dvkn

Idler, E.L. (2003). Discussion: Gender differences in self-rated health, in mortality, and in the relationship between the two. *The Gerontologist*, 372-375. https://doi.org/10/b75crf

Idler, E.L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behaviour*, 21-37. https://doi.org/10/b9ssnt

Jaworska S., & Ryan, K. (2018). Gender and the language of pain in chronic and terminal illness: A corpus-based discourse analysis of patients' narratives. *Social Science & Medicine, 215,* 107-114. https://doi.org/10/gfm5nx

Macintyre, S., Ford, G., & Hunt, K. (1999). Do women 'over-report' morbidity? Men's and women's responses to structured prompting on a standard question on long standing illness. *Social Science & Medicine, 48*(1), 89-98. https://doi.org/10/bm7pbd

Macintyre, S., Hunt, K., & Sweeting, H. (1996). Gender differences in health: Are things really as simple as they seem? *Social Science & Medicine, 42*(4), 617-624. http://refhub.elsevier.com/S0049-089X(17)30116-3/sref26

Månsdotter, A., Lindholm, L., & Winkvist, A. (2007). Paternity leave in Sweden; Costs, savings and health gains. *Health Policy, 82*(1), 102-115. https://doi.org/10/ckcjhm

Månsdotter, A., & Lundin, A. (2010). How do masculinity, paternity leave, and mortality associate? A study of fathers in the Swedish parental & child cohort of 1988/89. *Social Science & Medicine, 75*, 576-583. https://doi.org/10/fkxbds

Manuel, J.I., Martinson, M.L., Bledsoe-Mansori, S.E., & Bellamy, J.L. (2012). The influence of stress and social support on depressive symptoms in mothers with young children. *Social Science & Medicine, 75*, 2013-2020.

Marcus, A.C., & Seeman, T.E. (1981). Sex differences in health status: A reexamination of the nurturant role hypothesis. *American Sociological Review, 46*(1), 119-123. https://doi.org/10/c9sxkh

Nathanson, C. A. (1975). Illness and the feminine role: A theoretical review. *Social Science & Medicine, 9*(2), 57-62. https://doi.org/10/fmq8cx

Norton, E.C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal, 4*(2), 154-167. https://doi.org/10/d23w

Oakley, A. (2018). *The sociology of housework (reissue)*. Bristol University Press. https://doi.org/10/dvkr

Palència, L., Moortel, D. de, Artazcoz, L., Salvador-Piedrafita, M., Puig-Barrachina, V., Hagqvist, E., Pérez, G., Ruiz, M.E., Trujillo-Alemán, S., Vanroelen, C., Malmusi, D., & Borrell, C. (2017). Gender policies and gender inequalities in health in Europe: Results of

the SOPHIE project. *International Journal of Health Services, 47*(1), 61-82. https://doi.org/10/f3rxh3

Pampel, F.C. (2002). Cigarette use and the narrowing sex differentials in mortality. *Population and Development Review, 28*(1), 77-104. https://doi.org/10/bpzmsk

Pascall, G., & Lewis, J. (2004). Emerging gender regimes and policies for gender equality in a wider Europe. *Journal of Social Policy, 33*(3), 373-394. https://doi.org/10/d4rbrn

Prince, M. J., Reischies, F., Beekman, A. T. F., Fuhrer, R., Jonker, C., Kivela, S. L., Lawlor, B., Lobo A., Magnusson, H., Fichter, M. M., Van Oyen, H., Roelands, M., Skoog, I., Turrina, C., & Copeland, J. R. (1999). Development of the EURO-D scale: A European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry, 174*(4), 330-338. https://doi.org/10/bw2546

Puhani, P.A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear “difference-in-differences” models. *Economics Letters, 115*(1), 85-87. https://doi.org/10/fkkr4c

Ruhm, C.J., (2011). Policies to assist parents with young children. *The Future of Children, 21*(2), 37-68. https://doi.org/10/dwrr

Spiers, N., Jagger, C., Clarke, M., & Arthur, A. (2003). Are gender differences in the relationship between self-rated health and mortality enduring? Results from three birth cohorts in Melton Mowbray, United Kingdom. *The Gerontologist*, 406-411. https://doi.org/10/bp65pz

Staehelin, K., Bertea, P.C., & Stutz, E.Z. (2007). Length of maternity leave and health of mother and child; A review. *International Journal of Public Health, 52*(4), 202-209. https://doi.org/10/c85pcx

Statistics Sweden. (2018). *Women and men in Sweden 2018.* https://www.scb.se/en/finding-statistics/statistics-by-subject-area/living-conditions/gender-statistics/gender-statistics/pong/publications/women-and-men-in-sweden2.-facts-and-figures-2018/

Statistics Sweden. (2020). *Summary of population statistics 1960-2019*. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/yearly-statistics--the-whole-country/summary-of-population-statistics/

Sullivan, L., Camic, P.M., & Brown, J.S.L. (2014). Masculinity, alexithymia, and fear of intimacy as predictors of UK men’s attitudes towards seeking professional psychological help. *British Journal of Health Psychology, 20*(1), 194-211. https://doi.org/10/dvmv

Swedish Government Ministry of Employment. (2019). *Feminist government*. https://tinyurl.com/yaos39bb (pdf file)

Uccheddu, D., Gauthier, A.H., Steverink, N., & Emery, T. (2019). Gender and socioeconomic inequalities in health at older ages across different European welfare clusters: Evidence from SHARE data, 2004–2015. *European Sociological Review, 35*(3), 346-362. https://doi.org/10/c3h8

United Nations Educational, Scientific and Cultural Organization. (2006). International Standard Classification of Education: ISCED. http://uis.unesco.org/en/topic/international-standard-classification-education-isced

Valkonen, T., & Van Poppel, F. (1997). The contribution of smoking to sex differences in life expectancy: Four Nordic countries and The Netherlands 1970–1989. *European Journal of Public Health, 7*(3), 302-310. https://doi.org/10/ct56qv

Van de Velde, S., Bracke, P., & Levecque, K. (2010). Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science & Medicine, 71*, 305-313. http://refhub.elsevier.com/S0049-089X(17)30116-3/sref46

Wingard, D.L., Cohn, B.A., Kaplan, G.A., Cirillo, P.M., & Cohen, R.D. (1989). Sex differentials in morbidity and mortality risks examined by age and cause in the same cohort. *American Journal of Epidemiology, 130*(3), 601-610. https://doi.org/10/dvmg

Appendix A: Implementation of the Synthetic Control Method

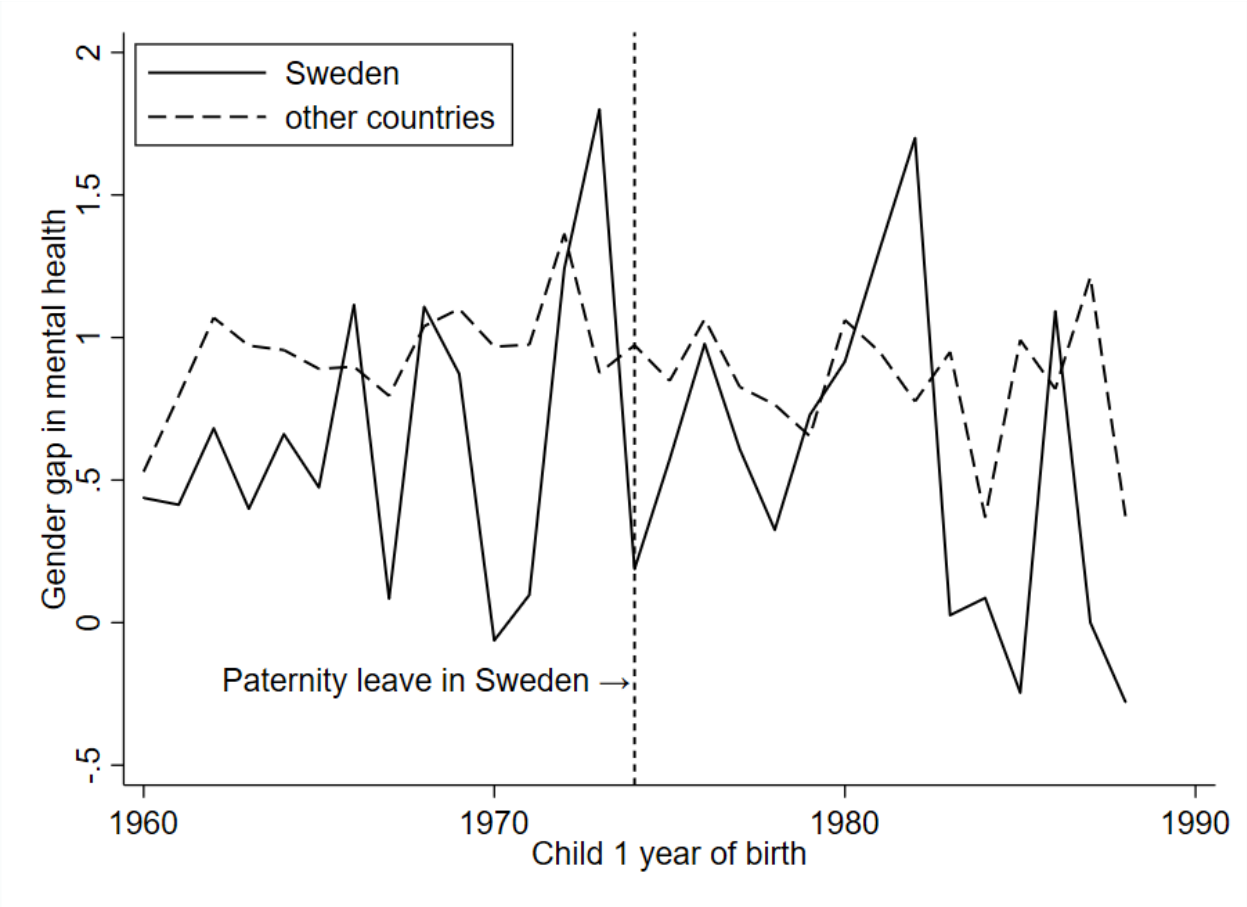
Data & Methodology

A panel data set of aggregate data per country and year of birth of the first child will be constructed, only focusing on individuals who had all their children either before or after the Swedish parental leave reform of 1974. The year of birth of the first child will function as the indicator of the year in the panel data set. Individuals who had some children before the reform and some children after the reform are excluded to reduce possible spillover effects. The birthyear of the first child is taken as indicator for year because the situation during the birth of a first child is more important for an individual's mental health than the situation during the birth of other children (Avendano et al., 2015).

The panel data set will be used to implement the synthetic control method designed by Abadie et al. (2010). This method is more sophisticated in creating a counterfactual than a difference-in-differences model, because a combination of other countries may give a better comparison group than any one country alone. The synthetic control method also formalizes the selection of comparable countries. This method creates a counterfactual by giving other countries than Sweden different weights in such a manner that the weighted averages of influencing factors of the gender gap in mental health resemble those of Sweden as closely as possible, creating a 'synthetic Sweden'. Only countries that did not implement a paternity leave policy of which the people in the sample benefitted are eligible for this group. Sweden was the first country to implement a paternity leave policy (Swedish Government Ministry of Employment, 2019), of the other countries in the dataset, Austria, Germany, the Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland, Belgium, Israel, the Czech Republic, Poland and Ireland, Denmark was the first to implement a paternity leave policy. The parental leave policy in Denmark was implemented in 1989 (Andersen, 2018), therefore all data from 1989 onwards will not be taken into consideration. A comparison of the gender gap in mental health between men and women for Sweden and the other countries is provided in figure A1. Clearly, the trends are too different to estimate the effect of paternal leave on the gender gap in mental health.

Figure A1

Trends of the Gender Gap in EURO-D Score: Sweden Compared to Other Countries



Estimating the values of the gender gap in Sweden before the implementation of paternity leave and giving the other countries weights to match the pre-intervention values would result in a better control group with a more similar trend for the gender gap in mental health preceding the Swedish policy change. The variables that will be used to construct synthetic Sweden are shown in table A1. To further increase the estimation power of the line, the countries will also be weighted on their outcome values for the years 1973, 1968, and 1963 (the year before the implementation of paternity leave in Sweden and two five-year lags)

Table A1

Linear Regression of the Gender Gap in EURO-D Score in Sweden from 1958 to 1973

Variable	Coefficient
Ln(total household income)	-4.670***
Age	-0.280***
Number of chronic diseases	-0.850***
Number of grandchildren	-0.637*
Constant	67.970*

*p < .1. **p < .05. ***p < .01

Results

Implementation of the synthetic control method resulted in the weights shown in table A2, the resulting means of the predictors of the gender gap in mental health are summarized in table A3. Figure A2 shows the resulting comparison of the trends of Sweden and synthetic Sweden.

Table A2

Country Weights in Synthetic Sweden

Country	Weight	Country	Weight
Austria	0.167	Greece	-
Germany	-	Switzerland	-
Netherlands	0.436	Belgium	0.315
Spain	-	Israel	-
Italy	-	Czech Republic	-
France	-	Poland	-
Denmark	0.082	Ireland	-

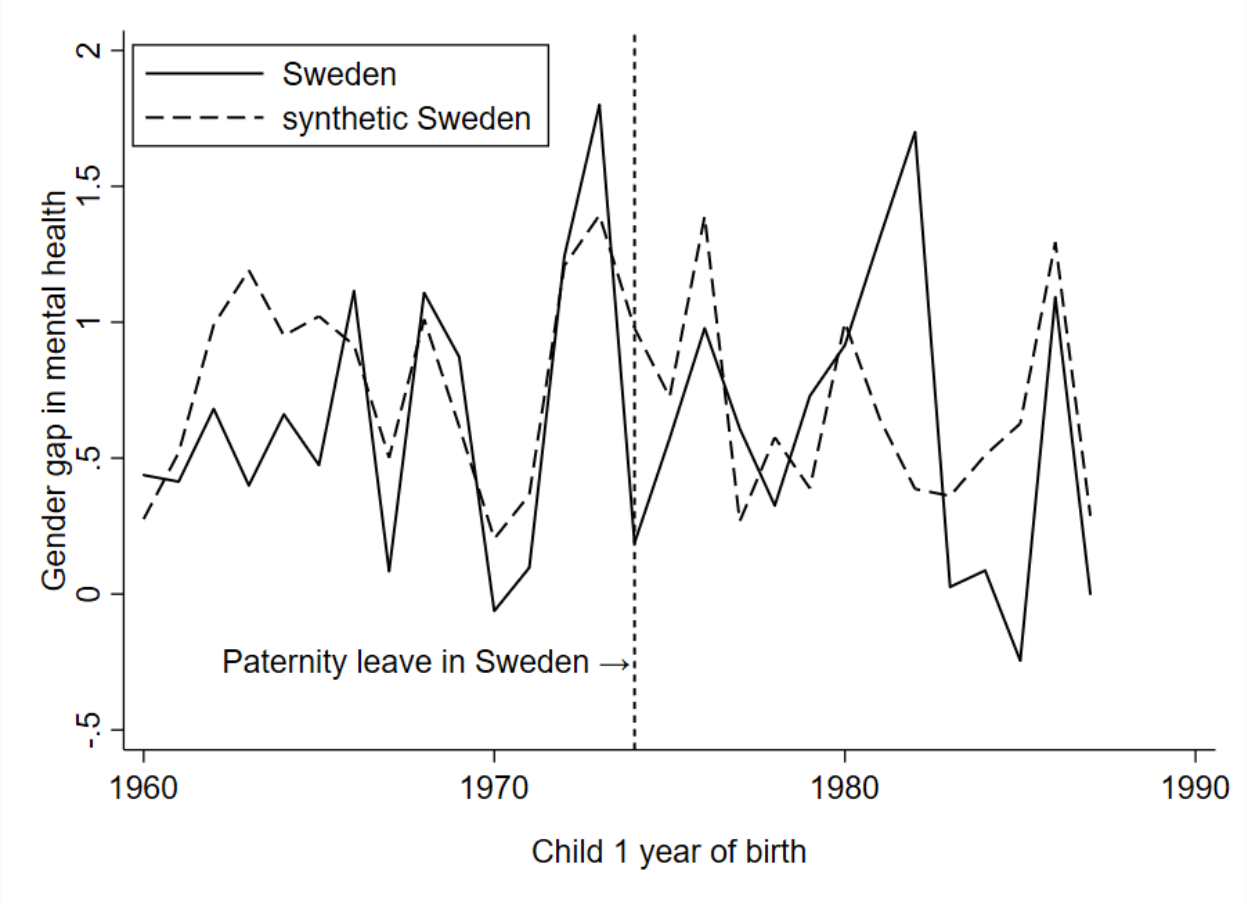
Table A3

Gender Gap Predictor Means

Variable	Sweden	
	Real	Synthetic
Ln(total household income)	10.448	10.354
Age	64.770	64.911
Number of chronic diseases	1.534	1.535
Number of grandchildren	3.130	2.796
Gender gap in mental health 1973	1.800	1.394
Gender gap in mental health 1968	1.108	1.011
Gender gap in mental health 1963	0.399	1.191

Figure A2

Trends of the Gender Gap in EURO-D Score: Sweden Compared to Synthetic Sweden



Conclusion

Even though the trend of synthetic Sweden resembles the trend of Sweden more closely than the trend of all control countries combined does, it still diverges from the trend of Sweden too much to conclude a decrease in the gender gap in mental health in Sweden for individuals who benefitted from paternity leave. No conclusions can be drawn from the results section of this appendix. The rest of this section will find an explanation why no sufficient results could be produced.

It is necessary for the implementation of synthetic controls that the microdataset is fit into an aggregate dataset. While this is legitimate and even suggested by the designers of the synthetic control method (Abadie et al., 2010), it leads to problems when the sample size of the microdata is low and outliers are present. As shown in table A4, the sample size is very low for the years preceding the policy change. This is due to the exclusion of parents who had their first child before 1974 and one or more children in 1974 or later. Inclusion of these observations would lead to a better handling of outliers, as seen in the synthetic Sweden trend (figure A2), which follows the direction of actual Sweden with less extreme peaks. However, it could lead to bias in the results, since the distinction between parents that did and did not receive paternal leave becomes less clear.

Table A4

Number of Swedish Observations per Birthyear of Firstborn Child

Year	No Spillover Children		Whole Sample	
	Real	Synthetic	Real	Synthetic
1965	102	292	133	363
1966	82	267	121	348
1967	109	296	159	403
1968	93	258	157	384
1969	59	275	131	425
1970	50	201	134	388
1971	26	150	129	421
1972	17	111	114	435
1973	19	83	101	358
1974	80	379	85	387
1975	95	351	101	355

Another solution for increasing the sample size, thus producing a better handling of outliers, would be the division of the observations into arbitrary time periods. This is already done in the difference-in-differences regression: only two time periods exist, before and after the intervention. The problem in decreasing the number of time periods is mainly the oversimplification of the matching process: a synthetic control unit based on only one pre-treatment period would produce an unconvincing counterfactual. A balance needs to be found between using enough time periods to accurately estimate the trend and having enough observations per time period to correct outliers to increase the chance of obtaining a counterfactual fitting the pre-intervention years well.

Appendix B: Comparisons of Pre-Intervention EURO-D Score Trends

Figure B1

Comparison of trends in EURO-D score: Sweden vs. Austria

Figure B2

Comparison of trends in EURO-D score: Sweden vs. Belgium

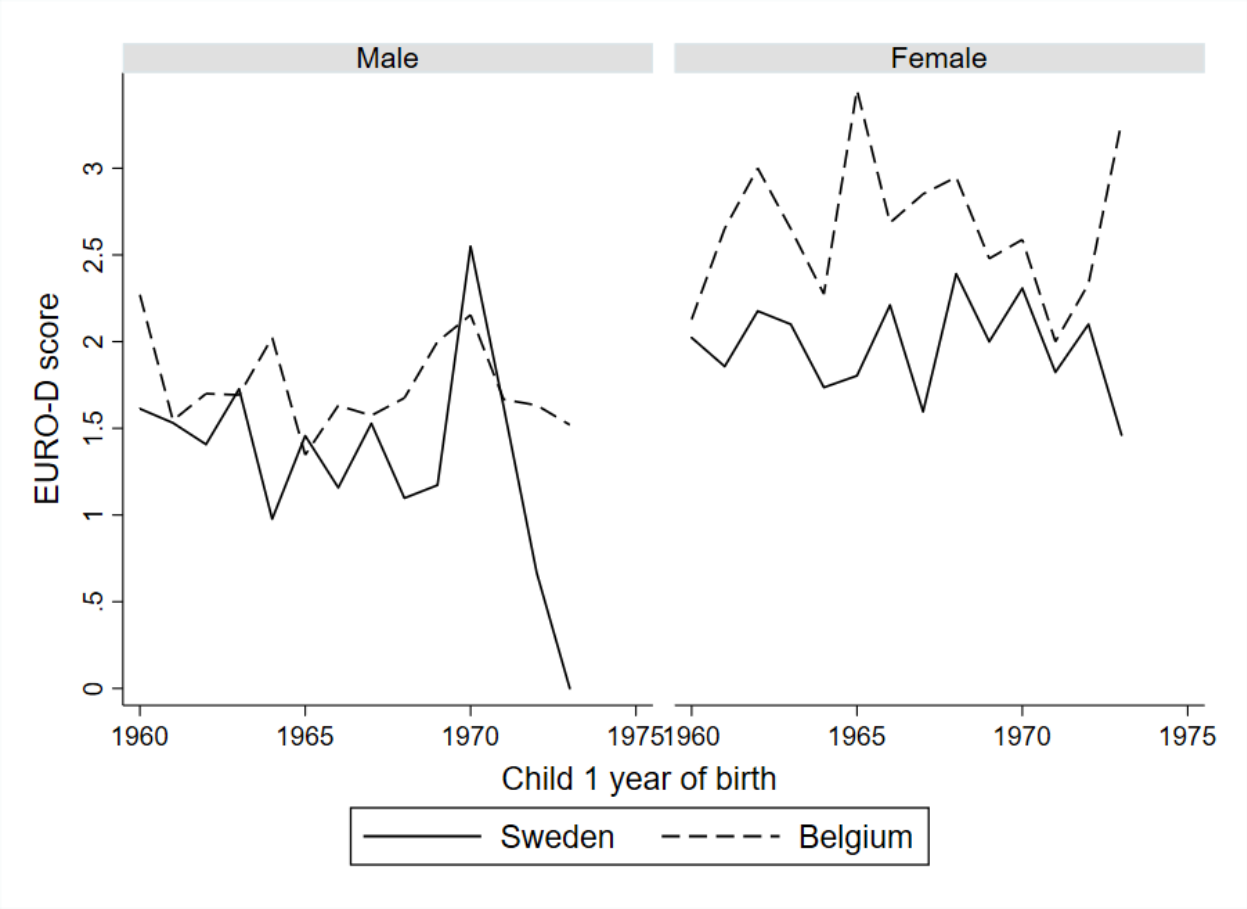


Figure B3

Comparison of trends in EURO-D score: Sweden vs. the Netherlands