# Supplemental File 1

**Table 1: Table of Policies**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1995 | | | | 2008 | | | |
|  | Duration (per year) | WRR | Length of service required | Other Conditions | Duration (per year) | WFF | Length of service required | Other Conditions |
| Austria | 1 week | 100% | none | none | 1 week | 100% | none | none |
| Belgium | no paid leave | no paid leave | no paid leave | no paid leave | up to 1 year | flat rate | 1 year | household member or 2nd degree relative, severe illness |
| Czech Republic | 9 days | 60% | none | household member,  severe illness | 9 days | 60% | none | household member,  severe illness |
| Denmark | no paid leave | no paid leave | no paid leave | no paid leave | up to 6 months | flat rate | none | severe illness, or serious reduced function |
| France | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |
| Germany | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |
| Greece | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |
| Ireland | no paid leave | no paid leave | no paid leave | no paid leave | up to 2 years | flat rate | 156 weeks of contributions | serious reduced function |
| Italy | up to 36 days | 100% | none | cohabiting 2nd degree relative,  serious reduced function | up to 36 days | 100% | none | cohabiting 2nd degree relative,  serious reduced function |
| Netherlands | indefinite/short-term | 100% | none | none | 10 days | 70% | none | 1st degree relative |
| Poland | 14 days | 80% | none | 2nd degree relative | 14 days | 80% | none | 2nd degree relative |
| Spain | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |
| Sweden | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |
| Switzerland | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave | no paid leave |

Table 2: Characteristics of Included Countries

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Sample size** | **Mean age in 1990 (SD)** | **Mean proportion working pre-policy (SD)** | **Mean proportion working post-policy (SD)** | **Mean difference in proportion working pre/post policy** | **Treated** |
| **Belgium, Women** | 1544 | 48.9(10.4) | 0.41(0.03) | 0.28(0.05) | -0.14 | 1 |
| *France, Women* | 1382 | 48.6(10.5) | 0.51(0.03) | 0.35(0.05) | -0.15 | 0 |
| *Greece, Women* | 1651 | 47.3(10.5) | 0.33(0.03) | 0.22(0.03) | -0.11 | 0 |
| *Italy, Women* | 1349 | 47.8(9.2) | 0.32(0.03) | 0.19(0.04) | -0.13 | 0 |
| **Belgium, Men** | 1270 | 48.4(9.8) | 0.72(0.06) | 0.43(0.1) | -0.29 | 1 |
| *France, Men* | 1062 | 47.9(9.8) | 0.71(0.06) | 0.44(0.08) | -0.27 | 0 |
| *Spain, Men* | 996 | 50(10) | 0.72(0.07) | 0.43(0.09) | -0.29 | 0 |
| *Switzerland, Men* | 563 | 48.5(9.5) | 0.84(0.06) | 0.57(0.08) | -0.27 | 0 |
| **Denmark, Women** | 1139 | 47.4(10.6) | 0.66(0.06) | 0.47(0.05) | -0.19 | 1 |
| *Austria, Women* | 583 | 49.9(9.5) | 0.38(0.08) | 0.16(0.03) | -0.22 | 0 |
| *Sweden, Women* | 1074 | 49.6(9.7) | 0.7(0.08) | 0.45(0.05) | -0.25 | 0 |
| **Denmark, Men** | 952 | 46.8(9.4) | 0.79(0.07) | 0.54(0.06) | -0.25 | 1 |
| *Greece, Men* | 1320 | 47.9(9.2) | 0.73(0.08) | 0.46(0.05) | -0.27 | 0 |
| *Sweden, Men* | 873 | 50(9.2) | 0.76(0.09) | 0.47(0.06) | -0.29 | 0 |

Table 3: All Regression Results\*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Belgium Women** | | |  | **Belgium Men** | | |
|  | Estimate | 95% CI |  |  | Estimate | 95% CI |
| (Intercept) | 0.528 | 0.499;0.557 |  | (Intercept) | 1.06 | 1.038;1.082 |
| policy.introduced | 0.006 | -0.002;0.014 |  | policy.introduced | 0.002 | -0.010;0.014 |
| treated | 0.113 | 0.086;0.140 |  | treated | -0.127 | -0.151;-0.103 |
| **Interaction\*\*** | -0.012 | -0.028;0.004 |  | **Interaction\*\*** | -0.019 | -0.037;-0.001 |
| countryFrance | 0.185 | 0.161;0.209 |  | countryFrance | -0.146 | -0.170;-0.122 |
| countryGreece | 0.010 | -0.015;0.035 |  | countrySpain | -0.088 | -0.112;-0.064 |
| year1 | 0.041 | 0.021;0.061 |  | year1 | -0.048 | -0.072;-0.024 |
| year2 | 0.084 | 0.062;0.106 |  | year2 | -0.057 | -0.081;-0.033 |
| age\_cat40.49\*\*\* | -0.058 | -0.082;-0.034 |  | age\_cat40.49\*\*\* | 0.004 | -0.012;0.020 |
| age\_cat50.59 | -0.229 | -0.262;-0.196 |  | age\_cat50.59 | -0.104 | -0.126;-0.082 |
| age\_cat60.69 | -0.560 | -0.595;-0.525 |  | age\_cat60.69 | -0.645 | -0.670;-0.620 |
| age\_cat70+ | -0.665 | -0.700;-0.63 |  | age\_cat70+ | -0.879 | -0.904;-0.854 |
|  | | | | | | |
| **Denmark Women** | | |  | **Denmark Men** | | |
|  | Estimate | 95% CI |  |  | Estimate | 95% CI |
| (Intercept) | 0.963 | 0.936;0.990 |  | (Intercept) | 1.003 | 0.981;1.025 |
| policy.introduced | -0.004 | -0.020;0.012 |  | policy.introduced | -0.008 | -0.022;0.006 |
| treated | -0.064 | -0.088;-0.040 |  | treated | -0.031 | -0.051;-0.011 |
| **Interaction\*\*** | 0.008 | -0.014;0.030 |  | **Interaction\*\*** | 0.010 | -0.012;0.032 |
| countryAustria | -0.293 | -0.320;-0.266 |  | countryGreece | -0.071 | -0.091;-0.051 |
| year1 | -0.027 | -0.054;0.000 |  | year1 | -0.041 | -0.065;-0.017 |
| year2 | 0.024 | -0.005;0.053 |  | year2 | -0.029 | -0.054;-0.004 |
| age\_cat40.49\*\*\* | -0.021 | -0.045;0.003 |  | age\_cat40.49\*\*\* | 0.006 | -0.010;0.022 |
| age\_cat50.59 | -0.127 | -0.160;-0.094 |  | age\_cat50.59 | -0.061 | -0.085;-0.037 |
| age\_cat60.69 | -0.593 | -0.630;-0.556 |  | age\_cat60.69 | -0.525 | -0.554;-0.496 |
| age\_cat70+ | -0.850 | -0.887;-0.813 |  | age\_cat70+ | -0.892 | -0.923;-0.861 |
| \*Results from linear regression model for change in proportion working  \*\*treatment effect  \*\*\*reference category: age 20-39 | | | | | | |

# Supplemental File 2: Control Country Selection Procedure

**Method**

1. Functional form for year was determined separately for each treated country and gender. First, proportion working was regressed on indicator variables for year, and marginal predictions were plotted. Second, proportion working was regressed on linear variable for year, F-tests were performed, and marginal predictions were plotted. Third, proportion working was regressed on linear and quadratic variables for year, F-tests were performed, and marginal predictions were plotted. Functional form was determined by F-test results and marginal plots. All analyses were restricted to pre-policy period.
2. Regression model with the treated country and all eligible control countries was fit using the appropriate functional form for year. Terms for country, age group, year, and country\*year interaction were included, with treated country as the referent.
3. An F-test for joint significance of country\*year interaction terms was performed. If p<0.05, the country with the lowest p-value was removed. If two countries had the same p-value, the country with the largest absolute value for interaction term was removed.
4. Steps 2 & 3 were repeated until chunk test p-value ≥ 0.05; marginal predicted proportion working by year were plotted by country.
5. Marginal prediction plots were visually inspected, and countries whose pre-policy trends did not appear parallel were removed.

**Results**

The results are presented for each treatment strata in terms of the order in which candidate control countries were eliminated and the value of the chunk test p-value. The countries highlighted in blue are the selected controls in main analysis.

**Belgium Men**

Time trend was modeled using a linear term for year. The control countries selected were Switzerland, Spain and France.

1. Poland (p<0.0001)
2. Italy (p<0.0001)
3. Greece (p=0.0001)
4. Sweden (p=0.0011)
5. Czech Republic (p=0.0067)
6. **Austria** (p=0.1063; Control by p-value only)
7. **Germany** (p= 0.1759; Control by p-value only)
8. **Switzerland** (Control by p-value and visual inspection)
9. **Spain** (Control by p-value and visual inspection)
10. **France** (Control by p-value and visual inspection)

|  |  |
| --- | --- |
| *P-Value Only* | *P-Value and Visual Inspection* |
|  |  |

**Belgium Women**

Time trend was modeled using a linear term for year. The control countries selected were Greece, Italy and France.

1. Poland (p<0.0001)
2. Czech Republic (p<0.0001)
3. Spain (p<0.0001)
4. Sweden (p=0.0001)
5. **Switzerland** (p=0.0671; Control by p-value only)
6. **Austria** (p=0.1728 Control by p-value only)
7. **Germany** (p=0.2776; Control by p-value only)
8. **Greece** (Control by p-value and visual inspection)
9. **Italy** (Control by p-value and visual inspection)
10. **France** (Control by p-value and visual inspection)

|  |  |
| --- | --- |
| *P-Value Only* | *P-Value and Visual Inspection* |
|  |  |

**Denmark Men**

Time trend was modeled using a linear term for year. Only Greece was found to be an appropriate control.

1. Poland (p<0.0001)
2. Italy (p<0.0001)
3. Austria (p<0.0001)
4. Germany (p=0.0017)
5. France (p=0.0207)
6. **Switzerland** (p=0.1025; Control by p-value only)
7. **Spain** (p=0.1172; Control by p-value only)
8. **Czech Republic** (p=0.3479; Control by p-value only)
9. **Greece** (Control by p-value and visual inspection)
10. **Sweden** (Control by p-value and visual inspection)

|  |  |
| --- | --- |
| *P-Value Only* | *P-Value and Visual Inspection* |
|  |  |

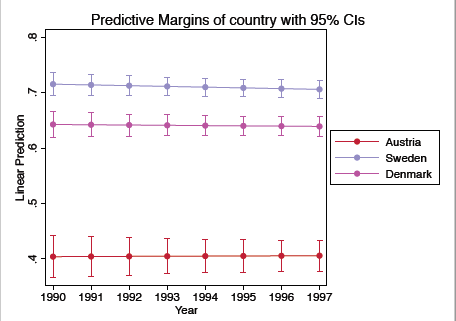
**Denmark Women**

Time trend was modeled using a linear term for year. F-tests supported the inclusion of a quadratic term for year; however, this may be due to overfitting of a relatively flat time trend. For main analysis, Austria was the best control. For analysis including a quadratic term, all candidate controls were insufficient by F-test p-value criterion.

***Linear***

1. Spain (p<0.0001)
2. Poland (p<0.0001)
3. Switzerland (p<0.0001)
4. Italy (p<0.0001)
5. Greece (p<0.0001)
6. Czech Republic (p<0.0001)
7. France (p=0.0001)
8. Germany (p=0.0198)
9. **Austria** (Control by p-value only/p-value and visual inspection)
10. **Sweden** (Control by p-value only/p-value and visual inspection)

*P-value only/P-Value and Visual Inspection*



***Model with quadratic time trend***

1. France (p<0.0001)
2. Spain (p<0.0001)
3. Switzerland (p<0.0001)
4. Greece (p<0.0001)
5. Austria (p<0.0001)
6. Italy (p<0.0001, but linear interaction p=0.054 and quadratic interaction p=0.873; Control by interaction p-value only)
7. **Poland** (p<0.0001, but linear interaction p=0.154 and quadratic interaction p=0.592; Control by interaction p-value only)
8. **Czech Republic** (p<0.0001, but linear interaction p=0.985, quadratic interaction p=0.144; Control by interaction p-value only)
9. **Germany** (p=0.0271, but linear interaction p=0.319, quadratic interaction p=0.936; Control by interaction p-value and visual inspection)
10. **Sweden** (Control by F-test p-value only/p-value and visual inspection)

|  |  |
| --- | --- |
| *P-Value (Interaction, not Chunk) Only* | *P-Value (Interaction, not Chunk) and Visual Inspection* |
|  |  |

# Supplemental File 3: Additional Sensitivity Analyses

Different controls

For each treated country, control countries were sequentially added according to the evidence that the trend was parallel to the treated country, as explained in Supplemental File 1. While the final decision for controls also relied on visual inspection, countries with higher p-values on the coefficient for the interaction term were considered better controls because there was less statistical evidence that their trend in the proportion working (i.e. the outcome) was different than the trend in the treated country before the policy change. The effect estimates remained robust across the removal/addition of different control countries, and only began to shift as the quality of the control group worsened (Figure 1).

*Figure 1: Results of Sensitivity analysis using different control groups*

|  |  |
| --- | --- |
|  |  |
|  |  |

Non-linearity in the pre-policy time trend

Since we found evidence of non-linearity among women in Denmark, we completed alternate analyses using only Sweden as a control because it best matched the non-linear time trend observed in Supplemental File 1 (Table 1).

*Table 1: Alternate analysis for Denmark Women*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | SE | 95% CI |
| primary | 0.008 | 0.011 | -0.014;0.03 |
| Sweden only | 0.014 | 0.012 | -0.01;0.038 |

Inclusion of country-level variables and alternate analyses

**m2:** We included an indicator marking whether a subject was eligible for retirement in each year according to each country’s retirement rules. This information was available in the SHARE Job Episodes Panel.

**m3:** We restricted the sample to those reporting not being retired using the 0/1 retirement indicator provided by SHARE.

**m4:** Additional country-level information that influence the proportion working, and may also influence a country’s decision to change their policy were included. These included per capita national income, gross domestic product per capita (purchasing power parity) and the average national labour force participation rates, which were extracted from the World Bank’s World Development Indicators and Global Development Finance databases[1] (Table 2).

**m5:** We excluded those who reported being retired before the policy change year. This was based on the 0/1 retirement indicator. In cases where individuals reported being retired, working again, then being retired again, the first year of retirement was used.

**m6:** We excluded France as a control for Belgium

Table 2: Alternate analyses

Belgium Females

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | SE | 95% CI |
| primary | -0.012 | 0.008 | -0.028;0.005 |
| m2 | -0.013 | 0.008 | -0.029;0.003 |
| m3 | -0.015 | 0.009 | -0.033;0.003 |
| m4 | -0.022 | 0.009 | -0.039;-0.004 |
| m5 | 0.018 | 0.023 | -0.027;0.063 |
| m6 | -0.025 | .009 | -0.043;-0.007 |
|  |  |  |  |

Belgium Males

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | SE | 95% CI |
| primary | -0.019 | 0.009 | -0.037;-0.001 |
| m2 | -0.031 | 0.009 | -0.05;-0.013 |
| m3 | -0.033 | 0.012 | -0.056;-0.01 |
| m4 | -0.016 | 0.01 | -0.034;0.003 |
| m5 | -0.037 | 0.017 | -0.07;-0.004 |
| m6 | -0.021 | 0.011 | -0.014;0.03 |

Denmark Females

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | SE | 95% CI |
| primary | 0.008 | 0.011 | -0.014;0.03 |
| m2 | 0.004 | 0.011 | -0.017;0.025 |
| m3 | 0.014 | 0.013 | -0.012;0.041 |
| m4 | 0.013 | 0.011 | -0.009;0.036 |
| m5 | -0.019 | 0.28 | -0.074;0.036 |

Denmark Males

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | SE | 95% CI |
| primary | 0.01 | 0.011 | -0.011;0.032 |
| m2 | 0.021 | 0.01 | 0.001;0.042 |
| m3 | 0.016 | 0.01 | -0.005;0.036 |
| m4 | -0.003 | 0.013 | -0.028;0.022 |
| m5 | -0.043 | 0.026 | -0.094;0.008 |

The estimates were consistent with the main analysis except for males in Denmark when we restrict the sample based on retirement. With this restriction, we observe an increase in the proportion working of 0.021 (95% CI 0.001;0.042). Further exploration of this, however, similarly suggests that we cannot conclude the positive increase is due to the policy change because the effect was apparent before the actual policy change (Table 3).

*Table 3: Additional analysis for Denmark Men*

|  |  |  |
| --- | --- | --- |
|  | Estimate | SE |
| leadeffect | 0.02 | 0.01 |
| primary | 0.016 | 0.01 |
|  |  |  |

Sampling weights

We redid the primary analysis with the cross-sectional and longitudinal individual-level sampling weights provided by SHARE. Whether to include sampling weights is controversial for methods that aim to identify a causal effect because the aim is not usually to be representative, but rather the identify the effect. This is contrast to, for example, descriptive analyses where the goal may be to describe characteristics of the population and the sample is less of interest. However, as people differ on this, this analysis considers whether including the weights affects our results.

SHARE provides two types of weights for individuals - longitudinal and cross-sectional. The weights are full described elsewhere[2]. Longitudinal weights allow for accounting for attrition across SHARE waves while cross-sectional weights reconstruct the population for a given wave. We applied both types of weights (the longitudinal weights for waves 2-3 and the cross-sectional weights for wave 3) and re-estimated our main model. Inclusion of these weights did not affect our results (Table 4).

cs.weights = cross-sectional weights long.weights = longitudinal weights

*Table 4: Inclusion of sampling weights*

*Belgium Women*

|  |  |
| --- | --- |
|  | Estimate[95% CI] |
| primary | -0.0116[-0.028;0.004] |
| long.weights | -0.0107[-0.029;0.006] |
| cs.weights | -0.0081[-0.028;0.005] |

*Belgium Men*

|  |  |
| --- | --- |
|  | Estimate[95% CI] |
| primary | -0.019[-0.037;-0.001] |
| long.weights | -0.0155[-0.039;0.001] |
| cs.weights | -0.0225[-0.038;0] |

*Denmark Women*

|  |  |
| --- | --- |
|  | Estimate[95% CI] |
| primary | 0.0077[-0.014;0.03] |
| long.weights | -0.0019[-0.016;0.031] |
| cs.weights | 0.0047[-0.014;0.03] |

*Denmark Men*

|  |  |
| --- | --- |
|  | Estimate[95% CI] |
| primary | 0.0102[-0.011;0.032] |
| long.weights | 0.004[-0.012;0.033] |
| cs.weights | 0.0021[-0.011;0.031] |

**References**

1. World Bank (2017). <http://data.worldbank.org/data-catalog/world-development-indicators>.
2. SHARE (March 31, 2017). Release Guide 6.0.0. <http://www.share-project.org/fileadmin/pdf_documentation/SHARE_release_guide_6-0-0.pdf>.