

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## The Effect of Concussion on Salary and Employment-A Population-Based Event Time Study using a Quasi- Experimental Design

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038161
Article Type:	Original research
Date Submitted by the Author:	02-Mar-2020
Complete List of Authors:	Fallesen, Peter; Stockholm University, Swedish Institute for Social Research Campos, Benito; Heidelberg University Hospital Department of Neurosurgery, ;
Keywords:	Trauma management < ORTHOPAEDIC & TRAUMA SURGERY, HEALTH ECONOMICS, Neurosurgery < SURGERY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 1 **The Effect of Concussion on Salary and Employment-A Population-Based Event Time**  
5 2 **Study using a Quasi-Experimental Design**  
6  
7  
8 3

9  
10 4 <sup>1,2</sup>Peter Fallesen, PhD, <sup>3,4</sup>Benito Campos, M.D.  
11  
12 5

13  
14  
15 6 <sup>1</sup>Swedish Institute for Social Research, Stockholm University, SE-106 91 Stockholm, Sweden

16 7 <sup>2</sup>ROCKWOOL Foundation, Ny Kongensgade 6, DK-1472 Copenhagen C, Denmark

17 8 <sup>3</sup>Department of Neurosurgery, University of Heidelberg, INF400, 69120 Heidelberg, Germany

18 9 <sup>4</sup>Department of Health Tech, Technical University of Denmark, DK-2800 Kongens Lyngby,  
19 10 Denmark  
20  
21  
22 11

23  
24 12 **Address of Correspondence:** Peter Fallesen, Swedish Institute for Social Research,  
25 13 Stockholm University, SE-106 91 Stockholm, Sweden, [peter.fallesen@sofi.su.se](mailto:peter.fallesen@sofi.su.se),  
26 14 +4561333184  
27  
28 15

29 16 **Keywords:** administrative data; concussion; income; productivity  
30  
31

32 17 **Abstract word count: 196**  
33

34 18 **Text word count: 3025 (excluding references and figure/table legends)**  
35

36 19 **Number of references: 20**  
37

38 20 **Number of tables and/or figures: 2 tables and 4 figures**  
39  
40

41 21 **Number of videos: 0**  
42  
43 22  
44 23  
45  
46 24  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 25 **Abstract**

26 **Objective:** Concussions are the most frequent traumatic brain injuries. Yet, the socio-economic  
27 impact of concussions remains unclear. We study the socio-economic effect of concussions on  
28 working age adults on a population scale.

29 **Design:** Our population-based, event time study uses administrative data as well as hospital  
30 and emergency room records for the population of Denmark.

31 **Setting:** We study all Danish patients, aged 20-59 y, who were treated at a public hospital or  
32 emergency room between 2003-2017 after suffering a concussion without other intracranial or  
33 extracranial injuries (n=55,424 unique individuals) with no prior diagnosis of intra- or  
34 extracranial injury within the past ten years leading up to the incident.

35 **Primary and Secondary Outcome Measures:** As primary endpoint, we investigate the mean  
36 effect of concussion on annual salaried income within a five-year period after trauma. In an  
37 exploratory analysis, we study whether the potential impact of concussion on annual salaried  
38 income is driven by patient age, education or economic cycle.

39 **Results:** Concussion was associated with an average change in annual salary income of -  
40 1,223€ (95% CI, -1,540€; -905€, p<.001) corresponding to a salary change of -4.1 % (95% CI,  
41 -5.2 %; -3.1 %). People between 30-39 y and those without high school degrees suffered the  
42 largest salary decreases. Affected individuals leaving the workforce drove the main part of the  
43 decrease. Absolute annual effect sizes were countercyclical to the unemployment rate.

44 **Conclusions:** Concussions have a large and long-lasting impact on salary and employment of  
45 working-age adults on a nationwide scale.

1  
2  
3  
4 47 **Strengths and limitations of this study**

- 5  
6 48 - We use natural experiments to obtain plausible causal effects between concussion and  
7  
8 49 salary/employment.  
9  
10  
11 50 - We study a large, population-based sample with multiple data layers.  
12  
13 51 - We study how economic cycles affect our outcome measures.  
14  
15 52

16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 53 Introduction

54 Concussions are by far the most frequently occurring intracranial injuries affecting  
55 approximately 1 in 219 Danes every year.<sup>10</sup> Immediate symptoms may last for days or weeks,  
56 and 10-15% of patients diagnosed with concussion suffer from long-term symptoms such as  
57 headache, fatigue, and intolerance to stress. Clinical practice has encouraged patients to  
58 restrict social, mental, and physical activity in the weeks following a concussion (see<sup>14</sup> for  
59 review), although prolonged inactivity may prolong symptoms.<sup>4,8,12</sup> Thus, symptoms,  
60 comorbidities, and suggested treatment are associated with short- to long-term absence from  
61 work and lower productivity.

62 Yet, the causal effect of concussion on economic burdens for individuals and society through  
63 decreased labor market activity has not been identified. First, concussion is a sudden incidence  
64 and thus not amenable to prospective study nor randomization. Cohort and case-control  
65 studies<sup>1,3,7,16,17</sup> provide some valuable evidence on employment and labor market outcomes  
66 among those who suffered concussions but are prone to selection bias. Individuals at high risk  
67 of concussions may differ on unobserved characteristics (e.g., risk aversion, routine activities)  
68 from those at low risk. People who are more likely to suffer concussions may also, on average,  
69 have more precarious or unstable employment trajectories prior to the incident, which may  
70 further bias prospective studies. Given the high incidence of concussion, even small losses of  
71 productivity and discrete drops in employment would have a significant socioeconomic impact  
72 and thus, it would require large patient cohorts with suitable controls to grasp the full  
73 socioeconomic impact of concussions. Thus, absent the possibility of randomization, using a  
74 natural or quasi-experimental design is the only likely option to parse out the *causal* effect of

1  
2  
3  
4 75 concussions on labor market activity.<sup>9</sup>  
5  
6

7 76 We examine how concussions affect salary and employment of working age individuals in  
8  
9 77 Denmark, a representative north-European industrial nation with a strong welfare state and a  
10  
11 78 flexible labor market. We use administrative longitudinal data linked to hospital and emergency  
12  
13 79 room diagnostic data on all Danes, who received a primary diagnosis of concussion between  
14  
15 80 2003 and 2017. To address the problem of unmeasured bias between those that do and do not  
16  
17 81 experience a concussion, we use a quasi-experimental event-study approach<sup>5,6</sup> where we  
18  
19 82 compare similar individuals, who experienced their concussions at different time points. Under  
20  
21 83 mild assumptions of parallel trends in wage progression prior to concussion and random timing  
22  
23 84 of concussion event within a five-year time frame, the approach recovers a robust estimation  
24  
25 85 of the effect of concussion on annual salary and employment status.  
26  
27  
28  
29

30 86  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54



## 87 **Material & Methods**

### 88 *Data Sources and Sample Construction*

89 Our concussion data originates from the Danish National Patient Registry (DNPR) (see<sup>11</sup> for  
90 description). DNPR is published annually and holds information on all hospitalizations at public  
91 hospitals in Denmark since 1977, and on all emergency room visits and outpatient treatments  
92 at public hospitals since 1994. Our combined exposure and control cohort includes all Danes  
93 aged 20-59 y, who received a primary diagnosis of concussion (ICD-10 code S06.0, ICD-8 code  
94 N850) between 2003 and 2017 and did not sustain any kind of additional intracranial or  
95 extracranial injury. Individuals who regularly engage in activities associated with a high risk of  
96 sustaining multiple concussions may differ from the average concussion patient and would  
97 likely be over-represented in the exposure sample. To avoid such potential bias, we exclude all  
98 individuals, who were diagnosed with any kind of brain trauma during a ten-year period prior to  
99 the concussion event. Altogether, we study a cohort of 55,424 individuals. Only attrition is  
100 through mortality and out-migration, and out-migrated or deceased patients with missing spells  
101 in the follow up period is excluded in those periods.

102 As a measure of productivity, we use price-index deflated annual salaried income. Salary  
103 information comes from Statistics Denmark's Income Statistics database. The database  
104 includes all declared annual incomes including income from self-employment. The Danish Tax  
105 Authorities supply the data to Statistics Denmark. Overall accuracy is considered very good.<sup>15</sup>  
106 Table 1 reports number of observations for the samples and number of observations with  
107 missing salary information. As evident, only between 0.01 to 0.02 percent of observations  
108 across exposure and control groups have missing salary information. We disregard these

1  
2  
3  
4 109 observations in the main analysis. Through social security numbers we link information on  
5  
6 110 salaried income to records on diagnosed concussions. Further, we obtain information on high  
7  
8  
9 111 school or equivalent level degree at time of concussion using the Danish Education Database.  
10  
11 112 From the Danish Population Database, we obtain demographic information on age and gender  
12  
13 113 for all respondents. Since the data used in the study come from de-identified administrative  
14  
15  
16 114 registers that Statistics Denmark makes available for research purposes for approved  
17  
18 115 institutions, no approval from an ethics committee was needed to carry out the study. The  
19  
20 116 research was carried out as part of project no. 706630 approved by Statistics Denmark.  
21  
22

### 23 117 24 25 118 *Quasi-experimental design*

26  
27  
28 119 We use a quasi-experimental, difference-in-differences event time approach previously  
29  
30 120 described in a health setting by Dobkin et al.<sup>5</sup> We compare two groups of individuals from the  
31  
32 121 same cohort, where both groups experience concussions, but at two different time points  
33  
34  
35 122 ( $t_c, t_c + \Delta$ ). Specifically, we sample all 55,496 individuals into six different subgroups: i) The  
36  
37 123 *exposure group*, which includes all patients, who suffered their concussion during the period  
38  
39 124 2003-2012 ( $n=37,848$ ) and ii) five *control groups*, which comprise patients who experienced  
40  
41 125 their concussions  $\Delta = \{1$  ( $n=34,551$ ),  $2$  ( $n=31,851$ ),  $3$  ( $n=29,922$ ),  $4$  ( $n=28,530$ ), and  $5$   
42  
43  
44 126 ( $n=27,421\})$  years later than the exposure group and did not experience any kind of brain injury  
45  
46 127 in the  $10 + \Delta$  years before the concussion event (note that the design allows individuals to both  
47  
48 128 be part of the exposure and control group). Our model is built on the assumption that the exact  
49  
50  
51 129 timing of a concussion is random for small enough sizes of  $\Delta$ , and on the additional assumption  
52  
53 130 that the exposure and the control groups would have displayed parallel trends in salary if the  
54

1  
2  
3  
4 131 control group had not suffered a concussion at  $t_c$ . Table 1 show the number of patients in the  
5  
6  
7 132 exposure group and the five control groups for each year relative to exposure group's  
8  
9 133 concussion incident. Using multiple comparison groups allow us the gage the validity of the  
10  
11 134 assumption that the exact timing of a concussion is random for small enough sizes of  $\Delta$

12  
13  
14 135 To estimate the impact of concussion on labor market outcomes, we focus on the change in  
15  
16 136 annual salary as our primary outcome, and, in further exploratory analyses, study additional  
17  
18 137 outcomes such as income from health-related benefits, income from welfare benefits, and  
19  
20  
21 138 employment rates. Our data is nested within a three-level structure: Exposure or control group  
22  
23 139  $g$ , which includes individuals  $i$ , at times to exposure-groups concussion incident  $t$ . First, we  
24  
25 140 estimate a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3,$   
26  
27 141  $4, \text{ and } 5\}$  using ordinary least squares:

$$\begin{aligned}
 & \text{Salary}_{git} = \beta_0 + \gamma \text{Exposure}_g + \theta \text{Post}_t + \delta \text{Post} \times \text{Exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=20}^{59+\Delta} I(Age) \eta_{Age} \\
 & + \sum_{Year=1999}^{2012} I(Year) \eta_{Year} + \epsilon_{git} \quad (1)
 \end{aligned}$$

28  
29  
30 142  
31  
32 143  
33  
34  
35 144 where  $\text{Salary}_{git}$  measures annual salaried income deflated to 2015-level;  $\text{Exposure}_g$  indicates  
36  
37 145 whether the observation belongs to the exposure or control group;  $\text{Post}_t$  captures the period  
38  
39 146 after the exposure group's concussion occurred;  $\text{Post}_t \times \text{Exposure}_{git}$  captures the effect  
40  
41  
42 147 concussion, measured as share of year  $t \geq 0$  affected by concussion;  $\mathbf{X}_i$  is a set of covariates  
43  
44 148 that includes a high school indicator and a gender dummy;  $\epsilon_{git}$  is the error-term; and the two  
45  
46 149 last sets of indicator variables  $I(Age)$  and  $I(Year)$  capture age and incident year (for control  
47  
48  
49 150 group, the year indexed against). Under an assumption of parallel trends in salaried earnings  
50  
51 151 between exposure and control groups had the exposure concussion not occurred,  $\delta$  then  
52  
53 152 captures the annual causal effect of concussion on salary for people exposed to concussions.

1  
2  
3  
4 153 For additional exploratory analyses, we also estimate separate models across gender,  
5  
6 154 educational level, and age, as well as across the salary distribution (see Supplemental  
7  
8  
9 155 Methods, Supplemental Digital Content 1, for further details).  
10  
11  
12 156

### 14 157 *Standard Protocol Approvals, Registrations, and Patient Consents*

15  
16 158 Since the data used in the study come from de-identified administrative registers that Statistics  
17  
18  
19 159 Denmark makes available for research purposes for approved institutions, no approval from an  
20  
21 160 ethics committee was needed to carry out the study. The research was carried out as part of  
22  
23 161 project no. 706630 approved by Statistics Denmark.  
24  
25

### 28 163 *Patient and Public Involvement*

29  
30 164 There was no involvement from patients or members of the public in the design, or conduct,  
31  
32 165 or reporting, or dissemination plans of the research.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Results

### *Concussion leads to long-term loss in salaried income*

Compared to people who experienced their concussion one to five years after the exposure group, concussions had a sizeable effect on salaried income. Comparing to people who experienced a concussion one year after the exposure group, we estimate the loss in salaried income to be 423€ (95% CI: -9129€; 73€,  $p=.095$ ; Table 2), corresponding to a salary decrease of 1.5 % (95% CI: -.3.0 %; 3.2 %; Figure 1). Comparing to the control group, who suffered their concussions five years after the exposure group, annual salary of the exposure group was 1,243€ (95% CI: 1,564€; 922€,  $p<.001$ ) lower than annual salary of the control group  $\Delta=5$ , corresponding to a salary decrease of 4.2 % (95% CI: 3.1 %; 5.3 %; Figure 1). Normalized wage progression for the control groups, who suffered a concussion 1, 2, 3, 4, and 5 years after the exposure group, showed similar trends and similar levels pre-exposure, indicating that the parallel wage trends assumption was met (Figure 2 and table S1, Figures S1 in Supplemental Digital Content 2). Thus, there is a sizeable effect of concussion on salary, and the effect only fully reveal itself after the first year since exposure incident.

We hypothesized that the salary decrease caused by concussion resulted from a combination of lower salary and exit from the labor market, either through short- or long-term absence/unemployment. In an exploratory analysis, we tested whether labor force exit drove the full effect of concussion on salary (Figure 3). Compared to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group, a concussion was associated with 2.6% (95% CI: 3.0 %; 2.2 %,  $p <.001$ ) increase in the risk of receiving € 0 in annual salary, or, in other words, effectively exiting employment. From the cumulative density function in Figure

1  
2  
3  
4 188 3 we infer that exit from employment among the bottom half of the salary distribution drove the  
5  
6  
7 189 effect of concussion on salary.  
8

9 190  
10  
11  
12 191 *Long-term loss in salaried income stems from exit from the labor market*

13  
14 192 To further examine whether exit from the labor market was caused either through short- or long-  
15  
16 193 term absence/unemployment, we estimated a dynamic model using the control group  $\Delta=5$ ,  
17  
18 194 which suffers a concussion five years after the exposure group. Following the concussion  
19  
20  
21 195 incidence, sick leave benefits payments were higher in the exposure group for the first two  
22  
23 196 years following the incident. Concussed individuals were to some extent compensated for their  
24  
25 197 salaried income loss through sick leave benefits in the first few years following a concussion,  
26  
27  
28 198 but the compensation ended while the salary drop persisted. Further, employment in the  
29  
30 199 exposure group remained lower than in the control group  $\Delta=5$  and remained so for the entire  
31  
32 200 post-exposure period (see table S2, Supplemental Digital Content 2 for further details). Total  
33  
34  
35 201 income decline was lower than the salary decline through the five years, which indicates that  
36  
37 202 some form of public benefits covered part of the salary loss (see Figure S2, Supplemental  
38  
39 203 Digital Content 2 for further details).  
40

41  
42 204 These results suggest that the impact of concussions on salary largely stems from affected  
43  
44 205 individuals leaving the labor force completely, likely instead sustaining themselves through  
45  
46 206 early retirement, disability pensions, self-sufficiency, or other income sources.  
47  
48

49 207  
50  
51 208 *Younger patients without high school degree drove the effect of concussion on loss income*

52  
53  
54 209 The exposure group and all control groups differed slightly in terms of average patient age,  
55

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

210 male to female ratio, and for control group  $\Delta=5$ , in the frequency of individuals with at least a  
211 high school degree (see Table S3, Supplemental Digital Content 2 for further details). To ensure  
212 that differences in gender, education, or age did not influence our results, we subdivided our  
213 exposure group into subgroups based on gender, education status, and age at time of  
214 concussion. We then estimated the impact of concussion on salary and employment across all  
215 values of  $\Delta$  and for all subgroups (see, Figures S3-S8, Supplemental Digital Content 2 for  
216 further details). Patients between age 30-39 and those without a high school degree  
217 experienced the largest absolute and relative declines in salary.

218 Finally, we addressed the role of timing of concussion across different years. Given that per  
219 design our exposure group always suffered their concussion earlier than the control groups do,  
220 changing labor market conditions could moderate effects. Part of our sample suffered their  
221 concussion during or just prior to the Great Recession in 2009-2010, which arguably presented  
222 the largest shock to both the global and local economy since the Great Depression in the 1930s.  
223 In Denmark, the great recession was preceded by a series of years of economic growth, low  
224 unemployment, and increasing salaries (see Figure S8, Supplemental Digital Content 2 for  
225 salary development from 1994 to 2017). We estimated the impact of concussion on salary  
226 separately for each year from 2003-2012 and plotted the estimate against the percent of full-  
227 time unemployment in the Danish labor force (Figure 4). Suffering a concussion during an  
228 economic boom had a substantially higher impact on salary than doing so during a recession  
229 when comparing to control groups who suffered concussions two to five years later than  
230 exposure group.

## Discussion

The impact of concussion on employment and salary remains understudied. In a systematic review of four studies on the association between mild TBI and return to work, Cancelliere and colleagues<sup>1</sup> found that most workers return to work within 3-6 months of suffering a mild TBI, but that the long-term impact (more than one year after concussion) was not studied. In addition, studies included small to medium sample sizes, varied measures of return to work, and employed both case-control and cohort designs. Using an inception cohort study design, Theadom and colleagues<sup>16</sup> collected follow up data four years after mild TBI incidents on 245 New Zealanders who were employed prior to incident. They found a 3.6 percent productivity decline among those who suffered a concussion, compared to a 2.3 population average decline. The group suffering mild TBI also reported more difficulties carrying out work-related tasks. In a related study,<sup>16</sup> Theadom and colleagues further found that the mild TBI group had persisting cognitive symptoms four years after suffering their concussion compared to an age-sex matched control group. Also using a case-control design and data from Taiwan, Chu and colleagues<sup>3</sup> found that one month after incident, 26 percent of patients had still not managed to return to work, and a large share of those who did return scored below full-time employment on a work quality index. Only one other study by Graff and colleagues<sup>7</sup> include a large patient cohort (n=19,732). Using case-control they found an association between concussion and failing to return to work of 1.54 odds ratio, but also found that exposed individuals had lower labor market attachment and was more likely to receive health related benefits pre-incident compared to the control group.

In the present study, we overcame some of the obstacles faced by previous work on the impact



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

254 of concussions on labor market outcomes by including a vast cohort of patients and exploiting  
255 a quasi-experimental design that allow us to plausibly account for unobserved difference  
256 between exposure and control group. In addition, salary and employment data reported here  
257 were compiled routinely through third-party reporting and were mandatory for all subjects, thus  
258 giving a complete and comprehensive picture of the economic impact of concussion on a  
259 nationwide scale.

260 Altogether, we showed that Danes between 20-59 year of age, who suffered a concussion  
261 during the period 2003-2012 experienced average salary losses of 4.1%. The impact of  
262 concussions on salary already materialized one year after the incident and remained sizeable  
263 for at least five years. This result is in line with an implementation period in which the impact of  
264 concussion on wages fully develops. First, concussions occur at some point during the year,  
265 thereby not affecting already earned salary that year. Second, in Denmark, most employees  
266 are entitled to receive their salary during sickness absence for an open ended, but not indefinite  
267 period.

268 If we assume that people return to their expected salary levels after a five-year recovery period  
269 (a very conservative assumption that is not supported by our data), the mere net annual salary  
270 loss in our sample would be approximately €23,000,000 measured in 2015-value. That would  
271 neither include hospital charges, medical costs for the treatment of concussion, the foregone  
272 tax from income, and the increased need for welfare spending, nor would it account for the  
273 large group of individuals who never seeks treatment<sup>13</sup> or receive their diagnosis from their  
274 general practitioner rather than in a hospital or emergency room, and thus escape our study.  
275 Thus, total public costs are likely substantially higher.

1  
2  
3  
4 276 In addition, both in absolute and relative terms, the early peak-working aged individuals (30-39  
5  
6  
7 277 y) and the less-educated individuals in our cohort seemed to be most affected after suffering a  
8  
9 278 concussion. These findings might have an additional and yet unmeasured social impact,  
10  
11 279 especially if our results are transferrable to other nations with a less established welfare state  
12  
13 280 and a less flexible labor market. In such countries, the impact on the young and less-educated  
14  
15  
16 281 individuals suffering a concussion and thus on society might be accentuated.

17  
18 282 Comparing our hospital incidence rates to more complete canvases of incidences<sup>2</sup>, it seems  
19  
20  
21 283 likely that the actual cost in the population is more than twice as large as what we estimate. If  
22  
23 284 we were to consider the average concussion incidence rates for six other advanced European  
24  
25 285 countries that are somewhat comparable to Denmark (Norway, Finland, Germany, Netherlands,  
26  
27 286 England and France) and under the assumption that concussion have a similar impact on  
28  
29  
30 287 earnings in these countries, the net annual salary loss would be approximately €1,099,400,000  
31  
32 288 measured in 2015-value. While our study likely underestimates the total socioeconomic impact  
33  
34 289 of concussion, it suggests that concussions has a large economic impact on a nationwide scale  
35  
36  
37 290 and on productivity and income at the patient level.

## 38 39 291 40 41 42 292 **CONCLUSION**

43  
44  
45 293 Using timing of concussion as a natural experiment, we provide first plausible causal estimates  
46  
47 294 of the effect of concussion on salary and employment. Our results show that concussion has a  
48  
49 295 large and long-term negative causal impact on salary and employment. People between 30-39  
50  
51 296 y and those without high school degrees suffered the largest salary decreases.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

298 **Disclosures:**

299 The authors report no conflict of interest.

300

301 **Acknowledgements:**

302 NA

303

304 **Funding:**

305 This work was supported by the Swedish Research Council for Health (PF), Working Life and  
306 Welfare (Grant no. 2016-07099) with additional funding from the ROCKWOOL Foundation (PF).

307 The research was carried out independently of the funders.

308

309 **Authors contributions:** P.F. and B.C. conceived of the presented idea, P.F. performed the  
310 computations. P.F. and B.C. verified the statistical methods. P.F. and B.C. discussed the results  
311 and wrote the manuscript. The corresponding author confirms that he had full access to all the  
312 data in the study and had final responsibility for the decision to submit for publication.

313

314 **Data Availability Statement**

315 The data used in this study has been made available through a trusted third party, Statistics  
316 Denmark. Due to privacy concerns the data cannot be made available outside the hosted  
317 research servers at Statistics Denmark. University-based and private Danish scientific  
318 organizations can be authorized to work with data within Statistics Denmark. Such organization  
319 can provide access to individual scientists inside and outside of Denmark. Requests for data

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

320 may be sent to Statistics Denmark:  
321 <http://www.dst.dk/en/OmDS/organisation/TelefonbogOrg.aspx?kontor=13&tlfbogsort=sektion>  
322 or the Danish Data Protection Agency: [https://www.datatilsynet.dk/english/the-danish-data-](https://www.datatilsynet.dk/english/the-danish-data-protection-agency/contact/)  
323 [protection-agency/contact/](https://www.datatilsynet.dk/english/the-danish-data-protection-agency/contact/). The authors document and make available all code needed to  
324 reproduce the findings in the study.

For peer review only

## References

1. Cancelliere C, Kristman VL, Cassidy JD, Hincapié CA, Côté P, Boyle E, et al: Systematic Review of Return to Work After Mild Traumatic Brain Injury: Results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. **Archives of Physical Medicine and Rehabilitation** **95**:S201–S209, 2014
2. Cassidy JD, Carroll LJ, Peloso PM, Borg J, Von Holst H, Holm L, et al: Incidence, risk factors and prevention of mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on mild traumatic brain injury. **Journal of Rehabilitation Medicine** **36**:28–60, 2004
3. Chu SY, Tsai YH, Xiao SH, Huang SJ, Yang CC: Quality of return to work in patients with mild traumatic brain injury: a prospective investigation of associations among post-concussion symptoms, neuropsychological functions, working status and stability. **Brain Injury** **31**:1674–1682, 2017
4. DiFazio M, Silverberg ND, Kirkwood MW, Bernier R, Iverson GL: Prolonged Activity Restriction After Concussion: Are We Worsening Outcomes? **Clinical pediatrics** **55**:443–51, 2016
5. Dobkin C, Finkelstein A, Kluender R, Notowidigdo MJ: The Economic Consequences of Hospital Admissions. **American Economic Review** **108**:308–352, 2018
6. Fadlon I, Nielsen TH: **Household Responses to Severe Health Shocks and the Design of Social Insurance**. Cambridge, MA, 2015
7. Graff HJ, Siersma V, Møller A, Kragstrup J, Andersen LL, Egerod I, et al: Labour market attachment after mild traumatic brain injury: nationwide cohort study with 5-year register follow-up in Denmark. **BMJ open** **9**:e026104, 2019
8. Lal A, Kolakowsky-Hayner SA, Ghajar J, Balamane M: The Effect of Physical Exercise After a Concussion: A Systematic Review and Meta-analysis. **The American journal of sports medicine** **46**:743–752, 2018
9. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al: Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. **J Epidemiol Community Health**:2019
10. Pinner M, Jensen R, Birket-Smith M, Gade A, Riis JO, others: **Konsensusrapport Om Commotio Cerebri (Hjernerystelse)**. Videnscenter for Hjerneskaade, 2003
11. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT: The Danish National Patient Registry: a review of content, data quality, and research potential. **Clinical epidemiology** **7**:449–90, 2015
12. Schneider KJ, Leddy JJ, Guskiewicz KM, Seifert T, McCrea M, Silverberg ND, et al: Rest and treatment/rehabilitation following sport-related concussion: a systematic review. **British journal of sports medicine** **51**:930–934, 2017
13. Setnik L, Bazarian JJ: The characteristics of patients who do not seek medical treatment for traumatic brain injury. **Brain Injury** **21**:1–9, 2007
14. Silverberg ND, Iverson GL: Is rest after concussion “the best medicine”? recommendations for activity resumption following concussion in athletes, civilians, and military service members. **The Journal of head trauma rehabilitation** **28**:250–9, 2013
15. Statistics Denmark: Documentation of statistics: Income Statistics - Statistics Denmark.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

2015

Available:

<https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/income-statistics>. Accessed 20 August 2018

16. Theadom A, Barker-Collo S, Jones K, Kahan M, Te Ao B, McPherson K, et al: Work Limitations 4 Years After Mild Traumatic Brain Injury: A Cohort Study. **Archives of Physical Medicine and Rehabilitation** **98**:1560–1566, 2017
17. Theadom A, Starkey N, Barker-Collo S, Jones K, Ameratunga S, Feigin V, et al: Population-based cohort study of the impacts of mild traumatic brain injury in adults four years post-injury. **PLOS ONE** **13**:e0191655, 2018
18. Faul M, Coronado V. Epidemiology of traumatic brain injury. **Handbook of clinical neurology** **127**:3-13, 2015
19. Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic Brain Injury–Related Emergency Department Visits, Hospitalizations, and Deaths — United States, 2007 and 2013. **MMWR Surveillance Summaries** **66**:1-16, 2017
20. Chernozhukov V, Fernández-Val I, Melly B. Inference on Counterfactual Distributions. **Econometrica** **81**:2205-2268, 2013

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 386 **Figure Legends**

### 387 **Figure 1. Estimated effect of concussions in percentage on salary for the exposure group** 388 **measured against each control group**

389 Figure shows the percentage change in salary experienced by the exposure group following  
390 their concussions compared to the expected trajectory absent the concussion (calculated from  
391 the control groups) with 95 % confidence intervals. See table 1 for separate p-values for each  
392 estimate.

### 394 **Figure 2. Salary development for exposure and control groups across time of exposure**

395 Figure shows the salary trajectories for the exposure group (black) who suffers concussion at  
396 year zero against normalized wage trajectories for the control groups who suffer their  
397 concussions one to five years later.  $\Delta$  indicates the number of years between exposure and  
398 control incident. Table shows that there are no significant differences in the normalized salary  
399 levels for exposure and control group prior to exposure incident (see Figure S1, Supplemental  
400 Digital Content 2 for unnormalized salary trajectories).

### 402 **Figure 3. The cumulative density function (cdf) for salary post-treatment among the** 403 **treatment group and their counterfactual, and the difference between the two cdfs** 404 **expressed as the effect of concussion on the probability of earning below that salary-** 405 **level following exposure event.**

406 The figure shows the observed cumulative salary distribution following a concussion (red) and

1  
2  
3  
4 407 the expected counterfactual salary distribution absent the concussion (blue). The black line  
5  
6  
7 408 shows the difference between the observed and the counterfactual distribution, and the grey  
8  
9 409 dash lines show the 95 % confidence interval. The closed to constant decline of the difference  
10  
11 410 between the two distributions as the salary increase indicates that the main part of the effect of  
12  
13 411 concussions on salary are driven by people having a salary equal to zero.

15  
16  
17 412 **Figure 4. Effect of concussion on salary across incident years and control groups**  
18  
19 413 **together with the percentage fulltime unemployed of the labor force.**

20  
21  
22 414 Figure shows annual estimates of concussion against each control group separately mapped  
23  
24 415 against the share of the labor force that is full time unemployed. 95 % confidence intervals. The  
25  
26 416 estimates for the effect of concussion on salary almost uniformly increase in absolute  
27  
28 417 magnitude when unemployment decreases, and decrease when unemployment increase,  
29  
30 418 indicating that the effect of concussion on salary is countercyclical to the economic cycle.

31  
32  
33  
34 419  
35  
36 420 **Supplemental Digital Content titles & legends**

37  
38  
39 421 **Supplemental Digital Content [#1].** Text file. Supplemental materials and methods. This file  
40  
41 422 contains further details on our quasi-experimental, difference-in-differences event time  
42  
43 423 approach.

44  
45  
46 424  
47  
48 425 **Supplemental Digital Content [#2].** Table. Supplemental results Table S1: Test of parallel  
49  
50 426 trends assumption pre-exposure incident against each control group separately using eq. S3 in  
51  
52 427 supplementary methods. Separate exposure dummies for all time periods (except the year prior



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

to exposure, which serves as reference period).

**Supplemental Digital Content [#2].** Table. Supplemental results Table S2: Effect of concussion on different labor market outcome parameters using separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period): In this exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group. Outcomes include annual salaried income (annual salary), total annual income (total income), annual sick leave benefits received (sick leave benefits) as well as a binary indicator of employment (probability of employment). Monetary outcomes were measured at 2015-level in € 1,000.

**Supplemental Digital Content [#2].** Table. Supplemental results Table S3: Demographic factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years leading up to the concussion event in each of the groups. Factors include patient age (in years), share of sample female (1=100% female), and share of individuals with at least a high school degree (1=100%).

**Supplemental Digital Content [#2].** Figure. Supplemental results Figure S1: Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels

**Supplemental Digital Content [#2].** Figure. Supplemental results Figure S2: The Cumulative Density Function (CDF) for Total Income Post-Treatment among the Treatment Group and

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

450 Their Counterfactual, and the Difference between the Two CDFs Expressed as the Effect of  
451 Concussion on the Probability of Total Income Below that Income-Level following Exposure  
452 Event.

453  
454 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S3: Percentage  
455 Effect of Concussion on Relative Salary Across Age Groups.

456  
457 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S4: Percentage  
458 Effect of Concussion on Relative Salary Across High School Completion.

459  
460 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S5: Percentage  
461 Effect of Concussion on Relative Salary Across Gender.

462  
463 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S6: Effect of  
464 Concussion on Absolute Salary in 1K Euro Across Age groups.

465  
466 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S7: Effect of  
467 Concussion on Absolute Salary in 1K Euro Across Education.

468  
469 **Supplemental Digital Content [#2].** Figure. Supplemental results Figure S8: Effect of  
470 Concussion on Absolute Salary in 1K Euro Across Gender.

1  
2  
3  
4  
5  
6  
7  
8

**Table 1. Number of observations for exposure and control groups across time since exposure and number of observations with missing salary information**

473

9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

Years until exposure	Exposure					
	group	Control $\Delta=1$	Control $\Delta=2$	Control $\Delta=3$	Control $\Delta=4$	Control $\Delta=5$
-4	36,804	33,681	31,112	29,190	27,859	26,794
-3	36,978	33,834	31,245	29,366	27,973	26,907
-2	37,195	34,003	31,407	29,501	28,146	27,031
-1	37,449	34,224	31,582	29,687	28,288	27,220
0	37,848	34,551	31,851	29,922	28,530	27,421
1	37,467		31,755	29,832	28,433	27,337
2	36,940			29,807	28,421	27,295
3	36,484				28,421	27,304
4	36,084					27,314
Total observations	333,249	170,293	188,952	207,305	226,071	244,623
Observations with missing salary	81	32	31	44	35	29

474

Control groups have not suffered a concussion in  $10+\Delta$  years before incident, exposure group has not suffered concussion the 10 years before exposure incident.

475  
476  
47734  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 2. Estimated effect of concussion on salary of exposure group compared to control groups that suffered their concussion  $\Delta = 1, 2, 3, 4, 5$  y after the exposure group's concussion event, measured at 2015-level.  $N_{\text{Exposure}}: 37,848$**

	Estimated salary effect ( $\delta$ )	95 % CI	$p$	$N_{\text{Control}}$
$\Delta = 1$ y	-423€	(-919€; 73€)	.095	34,551
$\Delta = 2$ y	-825€	(-1,108€; -543€)	<.001	31,851
$\Delta = 3$ y	-1,019€	(-1,331€; -707€)	<.001	29,922
$\Delta = 4$ y	-1,126€	(-1,446€; -805€)	<.001	28,530
$\Delta = 5$ y	-1,243€	(-1,564€; -922€)	<.001	27,421

Results obtained from estimations following Eq. (1). Models include controls for high school diploma, gender, age, and observation year. Results obtained using reghdfe in Stata.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

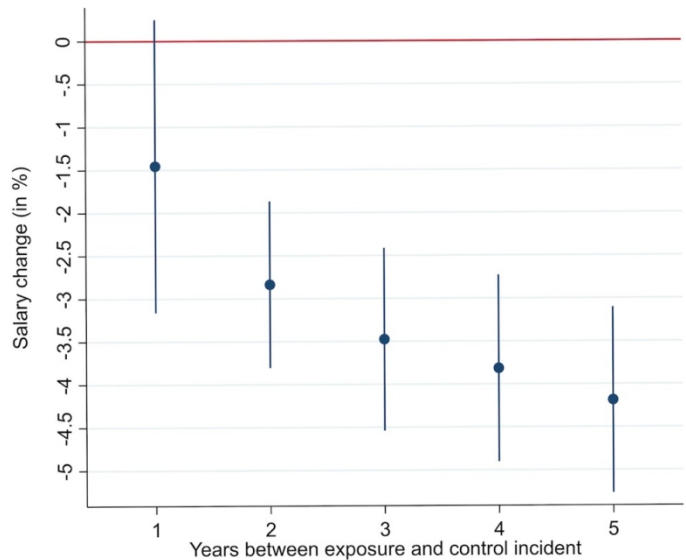


Figure 1

338x190mm (108 x 108 DPI)

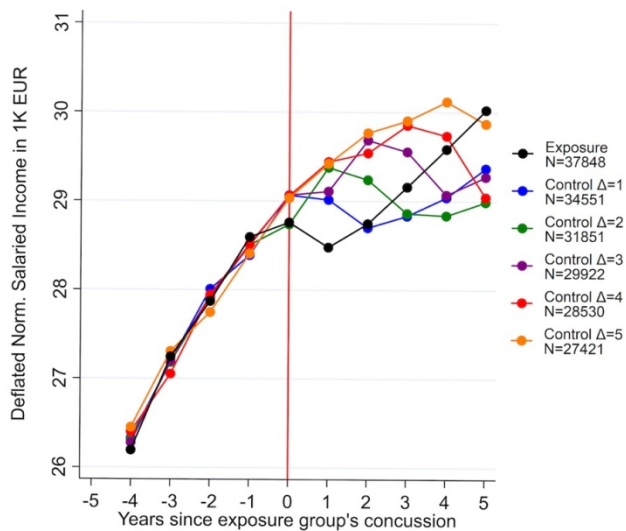


Figure 2

338x190mm (108 x 108 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

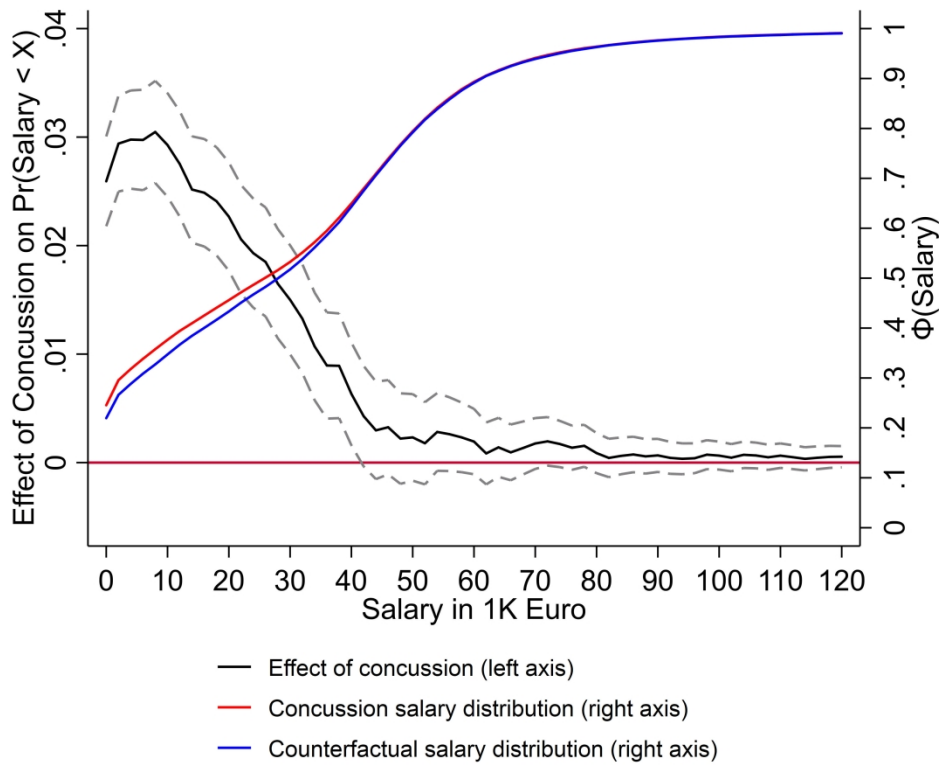


Figure 3

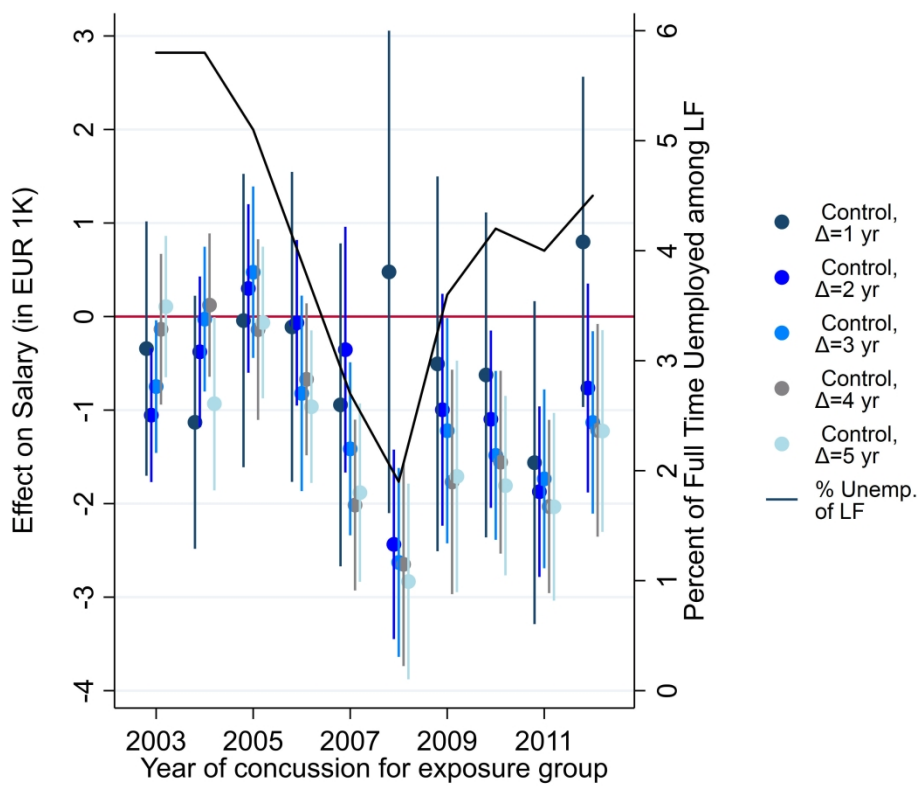


Figure 4



## SUPPLEMENTAL MATERIALS AND METHODS

Our quasi-experimental, difference-in-differences event time approach compares two groups of individuals from the same cohort, where both groups experience concussions, but at two different time points ( $t_c, t_c + \Delta$ ). For the simple situation where we have three periods ( $t=0, 1, 2$ ) and the exposure group (T) experiences their concussion at the start of period 1 ( $t_c=1$ ), and the control group (C) at the start of period 2 ( $t_c + \Delta=2$ ), the effect of concussion on salary (Y) is:

$$\Delta = (Y_1^T - Y_1^C) - (Y_0^T - Y_0^C)$$

The effect of concussion on salary in  $t=1$  is estimated by comparing the average difference in salary between exposure and control groups for the post-concussion period  $t=1$  ( $Y_1^T - Y_1^C$ ) to the average difference in salary for the pre-concussion, or baseline, interval  $t=0$  ( $Y_0^T - Y_0^C$ ). Assuming the exact timing of a concussion is random for small enough sizes of  $\Delta$ , and under the additional assumption that the exposure group would have had parallel trends in salary as the control group absent suffering concussion at  $t_c$ ,  $\delta$  captures the causal effect of concussion among those who suffer concussions – also known as the average effect on the treated (AT). The AT does not capture how concussions would affect a random person. The AT captures how concussions causally affect those who suffer concussions.

For our study, the parallel trends assumption states that exposure and control groups have parallel developments in salary leading up to the exposure group's concussion and the exposure and control groups would have further exhibited parallel salary trajectories if the concussion had not occurred. To test the parallel trends assumption, we estimate a dynamic version of the model specification (shown in supplementary table S1), which explicitly allows us to test whether the parallel trend assumption for our sample is probable.

To validate that the timing of concussion is random with our study period, we present estimates for effect of exposure across different periods between exposure and control incident ( $\Delta$ ). Most recorded concussions outside contact sports and military engagements stem from unforeseen events, such as falls or striking/being struck by an object<sup>18,19</sup>, so assuming random timing is likely valid. People who regular engage in activities that result in high risk of multiple concussions may be different than the average concussion patient and would be more likely to end up in the exposure sample than in the control sample, which could induce bias. To avoid such potential bias, we restrict our sample to individuals without prior diagnoses for intracranial injuries ten years prior to exposure.

At  $t=-1$ , i.e. one year before the exposure group suffered a concussion, the control groups were slightly smaller than the exposure group, and two control groups ( $\Delta=4$  and 5) differed slightly but significantly in terms of average patient age ( $p < .001$ ; supplementary table S2), male to female ratio ( $p < .001$ ), and for control group  $\Delta=5$ , in the frequency of individuals with at least a high school degree ( $p < .001$ ). However, the differences are numerically small. To test that composition differences between exposure and control do not drive our results, we provide separate results for individuals with and without high school degree, for males and females, and for different age groups across all different values of  $\Delta$ .

Further, our design inherently leads to the possibility of timing issues—our exposure group always suffers their concussion earlier (in terms of calendar time and age) than the control groups do. If the labor market is constantly improving or worsening during the period we consider, this could substantially influence our results. Therefore, we also estimate separate models across exposure incident year and control group. Estimating separate models allow us the added benefit of being able to examine whether the business cycle influences the effect of concussions on salary.

### Statistical model

To estimate the impact of concussion on salary, we define the following variables: Exposure or control group  $g$ , which includes individuals  $i$ , at times  $t$  exposure-groups concussion incident  $t$ . First, we estimate a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  using ordinary least squares:

$$\text{Salary}_{git} = \beta_0 + \gamma \text{exposure}_g + \theta \text{post}_t + \delta \text{post}_t \times \text{exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{\text{Age}=26}^{48+\Delta} I(\text{Age}) \eta_{\text{age}} + \sum_{\text{year}=1999}^{2012} I(\text{year}) \eta_{\text{year}} + \epsilon_{git} \quad (\text{S1})$$

where  $\text{Salary}_{git}$  measures annual salaried income deflated to 2015-level,  $\text{exposure}_g$  indicates whether the observation belongs to the exposure or control group,  $\text{post}_t$  captures the period after the exposure group's concussion occurred, and  $\text{post}_t \times \text{exposure}_{git}$  captures the effect concussion, measured as share of year  $t \geq 0$  affected by concussion. In this way, someone who suffers a concussion July 1 has  $\text{post}_t \times \text{exposure}_{git} = 0.5$  for  $t = 0$  and  $\text{post}_t \times \text{exposure}_{git} = 1$  for  $t > 0$ .  $\mathbf{X}_i$  is a set of covariates that includes a high school indicator and a gender dummy,  $\epsilon_{git}$  is the error-term, and the two last sets of indicator variables  $I(\text{Age})$  and  $I(\text{Year})$  capture age and incident year levels (control group indexed against incident year). Under the parallel trends assumption,  $\delta$  then captures the annual effect of concussion on salary. In eq. 1,  $\text{exposure}_g$  normalizes any pre-exposure differences between the exposure and control group, thereby creating a joint baseline pre-exposure.

We estimate robust individual-level clustered standard errors to account for the possibility that individuals enter the data twice both as control (0) and exposure (1) individuals ( $g=\{0,1\}$ ), and that they are observed for multiple periods ( $t=\{-4, \dots, \Delta-1\}$ ). To calculate the relative salary decrease after concussion, we exploit the parallel trends assumption to generate the expected counterfactual salary level, i.e. had the concussion not occurred, and calculate the decline expressed in percentage as: % change =  $\delta / E(\widehat{Salary}_{git} | g = 1, post_t = 1, post_t \times exposure_{git} = 0)$ . In this way, we provide both absolute estimates measured in 1K Euro, as well as percentage change.

We expect  $\delta$  from eq. (1) to likely be negative. Yet, a decrease in annual salary can arrive through two different channels. Concussions may affect salary through either decreasing income among those employed or by reducing the number of individuals who are employed and earning any salary at. To parse out which of the two channels is driving the results, we examine how concussion affects the salary distribution among the exposure group following. Following Chernozhukov et al.<sup>20</sup> we estimate a series of regressions across the whole salary distribution, where, for a finite set of points, we predict how concussion affects the likelihood of having earnings on the left side of each finite point, as follows:

$$\sum_{j=0}^{\max(Salary)} p_j = \beta_{0j} + \delta_j post_t \times exposure_{git} + \theta post_t + \gamma_j exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age,j} + \sum_{year=1999}^{2012} I(year) \eta_{year,j} + \epsilon_{git,j} \quad (S2)$$

where  $p_j = \Pr(Salary_{git} \leq j)$  and  $j$  is the interval from 0 to  $\max(Salary)$ . Across the salary distribution, we can now predict the probability of earning less than  $j$  for those with and without concussions. From equation 2, we predict  $p_j^1 = E(p_j | post_t \times exposure_{git} = 1, exposure_g = 1, t \geq 0)$  and the counterfactual  $p_j^0 = E(p_j | post_t \times exposure_{git} = 0, exposure_g = 1, t \geq 0)$ . Plotting  $p_j^1$  and  $p_j^0$  over each value of salary  $j$ , and assuming rank stability, gives the cumulative density function of salary for the treated ( $p_j^1$ ) and the counterfactual observation of the treated had they not suffered concussions ( $p_j^0$ ). The difference between  $p_j^1$  and  $p_j^0$  is simply  $\delta_j$ . If the value of  $\delta_j$  monotonically moves towards zero as  $j$  increases until  $p_j^1 \approx p_j^0 \approx 1$  it indicates that exit from employment fully drives the effect of concussion on salary. If instead the value of  $\delta_j$  is constant or increasing across parts of the distribution, it instead indicates that a decrease in salary among those still receiving salary drives at least part of the effect.

Eq. 1 and eq. 2 are based on the parallel trends assumption. The assumption states that exposure and control groups follow parallel salary trajectories until individuals in the exposure group experiences a concussion, and that the parallel trends would have continued had the concussion not occurred. Whereas we cannot verify the counterfactual situation of parallel trends after exposure, we can use a dynamic model to test for systematic differences in salary trends between exposure and control group in the years leading up to the exposure group's concussion event. To do so, we estimate the following dynamic model:

$$Salary_{git} = \beta_0 + \sum_{t \neq -1, t = -4}^{\Delta-1} \delta_t \times I(t_g) \times exposure_g + \sum_{t = -4}^{\Delta-1} I(t_g) \eta_t + \gamma exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age} + \sum_{year=1999}^{2012} I(year) \eta_{year} + \epsilon_{git} \quad (S3)$$

Where we interact exposure group status ( $exposure_g$ ) with indicators  $I(t_g)$  capturing time from concussion. If the parallel trends assumption holds, then it must be the case  $\{\delta_{-4}, \delta_{-3}, \delta_{-2}\} = 0$ , whereas the size and sign of  $\{\delta_0, \dots, \delta_{\Delta-1}\}$  captures the dynamic effect of a concussion from the year of incidence and  $\Delta-1$  years onward. By estimating the effect of concussion on salary among different years of the study period, we are also able to capture how the impact of concussion on salary evolves year to year after the concussion has occurred. We further estimate eq. 3 for a series of related labor market outcomes (annual total income, annual amount of sickness benefits received, annual probability of being employed), to generate a more thorough understanding on how concussions affect labor market outcomes—i.e., if people experience a decrease in salary due to a concussion, are they then compensated through different types of welfare state services.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Supplemental results**

For peer review only

**Table S1. Test of parallel trends assumption pre-exposure incident against each control group separately using eq. S3 in supplementary methods. Separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period).**

Time to exposure group's concussion (exposure)	$\Delta=1$ Est (S.E.) p-value	$\Delta=2$ Est (S.E.) p-value	$\Delta=3$ Est (S.E.) p-value	$\Delta=4$ Est (S.E.) p-value	$\Delta=5$ Est (S.E.) p-value
Exposure-4y	-0.368 (0.226) p=.104	0.046 (0.363) p=.900	0.120 (0.362) p=.741	0.159 (0.313) p=.612	0.042 (0.329) p=.899
Exposure-3y	-0.094 (0.317) p=.768	0.227 (0.510) p=.656	0.167 (0.354) p=.637	0.537 (0.372) p=.148	0.113 (0.393) p=.774
Exposure-2y	-0.548 (0.312) p=.079	-0.082 (0.347) p=.812	-0.163 (0.236) p=.491	-0.124 (0.247) p=.617	0.082 (0.250) p=.744
Exposure-1y	Ref.	Ref.	Ref.	Ref.	Ref.
<b>N*T</b>	284115	273725	266120	260647	256337

Note: The table shows test for differences in pre-exposure trends between exposure and control group model using interactions between pre-exposure time dummies and the exposure indicator. There is no indication of substantial or significant pre-exposure differences in salary trajectories between exposure group and any of the control groups.

**Table S2. Effect of concussion on different labor market outcome parameters using separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period): In this exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group. Outcomes include annual salaried income (annual salary), total annual income (total income), annual sick leave benefits received (sick leave benefits) as well as a binary indicator of employment (probability of employment). Monetary outcomes were measured at 2015-level in € 1,000.**

Time to exposure group's concussion (exposure)	Annual Salary Est. (S.E.) p-value	Total income Est. (S.E.) p-value	Sick leave benefits Est. (S.E.) p-value	Probability of employment Est. (S.E.) p-value
Exposure-4y	0.012 (0.212) p=.954	0.164 (0.173) p=.343	0.035 (0.036) p=.320	0.001 (0.004) p=.803
Exposure-3y	0.059 (0.252) p=.814	0.305 (0.233) p=.190	0.022 (0.034) p=.529	-0.001 (0.003) p=.739
Exposure-2y	0.043 (0.160) p=.788	0.122 (0.147) p=.405	0.002 (0.029) p=.946	0.001 (0.003) p=.739
Exposure-1y				
Exposure	-0.611 (0.168) p<.001	-0.338 (0.140) 0.016	0.166 (0.030) p<.001	-0.003 (0.003) p=.317
Exposure+1y	-1.389 (0.209) p<.001	-0.608 (0.162) p<.001	0.288 (0.039) p<.001	-0.020 (0.003) p<.001
Exposure+2y	-1.568 (0.261) p<.001	-0.847 (0.231) p<.001	0.132 (0.039) p=.001	-0.023 (0.004) p<.001
Exposure+3y	-1.393 (0.246) p<.001	-0.497 (0.219) p=.023	0.031 (0.040) p=.432	-0.022 (0.004) p<.001
Exposure+4y	-1.319 (0.253) p<.001	-0.499 (0.218) p=.022	-0.076 (0.042) p=.075	-0.018 (0.004) p<.001
<b>N*T</b>	<b>577762</b>	<b>577758</b>	<b>577872</b>	<b>577872</b>

**Note:** Annual salary include all income from salary and employee fringe benefits, employee stock options, employer paid sick leave, net gains (including interests and capital gains) from own companies. Total income includes all income absent wealth. Sick leave includes only public health benefits (sick leave and paternity leave). Employment is a binary indicator measured last week of November for each year. Results obtained from estimations following Eq. (1). Models include controls for high school diploma, gender, age, and observation year. Results obtained using reghdfe in Stata. Total number of observations (N\*T) differ slightly between outcomes because all income information is not available for all observation all years.

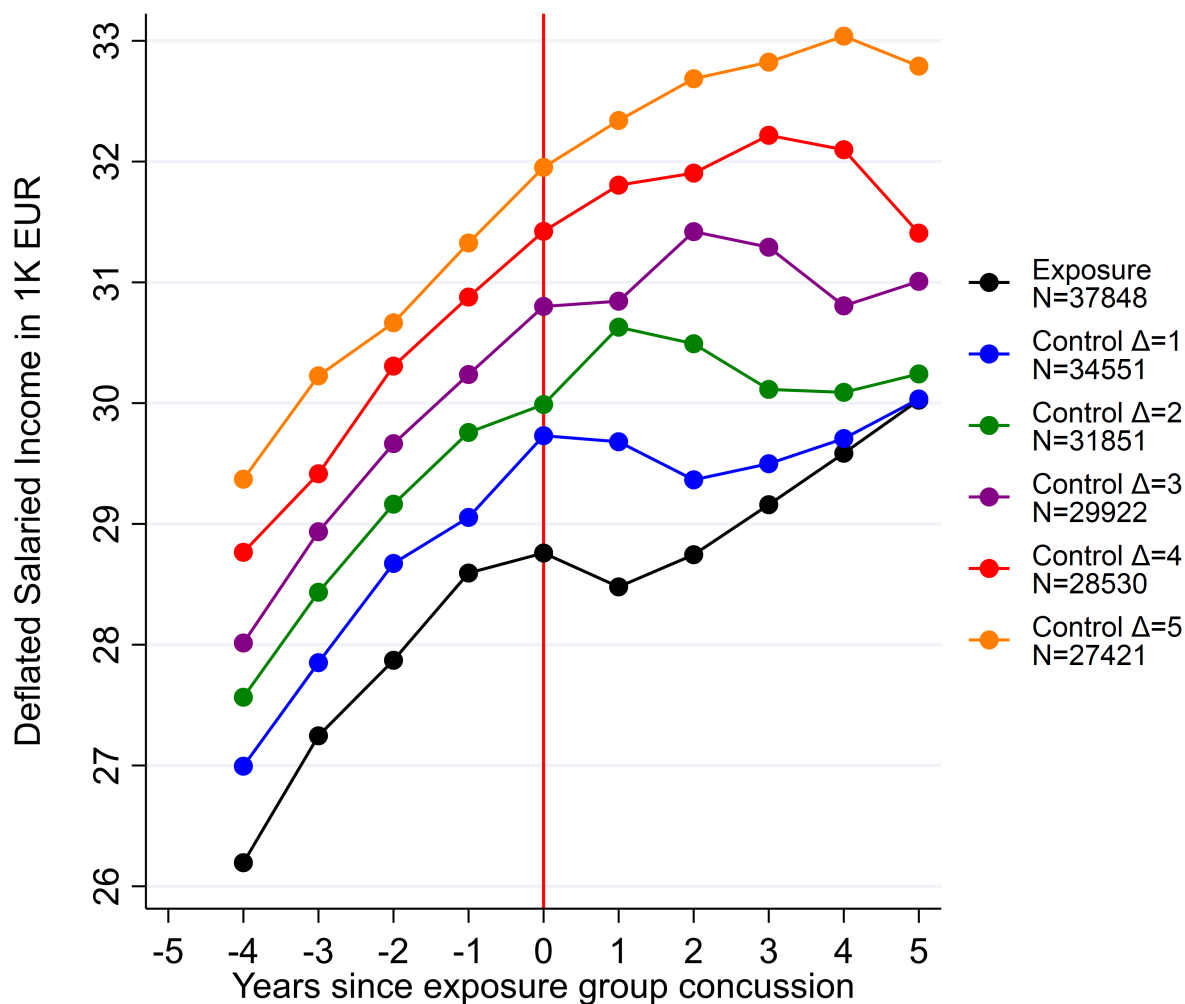
Source: Own calculations on data from Statistics Denmark.

**Table S3. Demographic factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years leading up to the concussion event in each of the groups. Factors include patient age (in years), share of sample female (1=100% female), and share of individuals with at least a high school degree (1=100%).**

		Exposure	Control, $\Delta=1$	Control, $\Delta=2$	Control, $\Delta=3$	Control, $\Delta=4$	Control, $\Delta=5$
<b>Pr(Female=1)</b>	Mean	.430	.438	.447	.458	.464	.473
	S.D.	(.495)	(.496)	(.497)	(.498)	(.499)	(.499)
	p-value		.030	<.001	<.001	<.001	<.001
<b>Age</b>	Mean	36.899	37.354	37.754	38.065	38.343	38.592
	S.D.	(11.856)	(11.857)	(11.718)	(11.630)	(11.584)	(11.491)
	p-value		<.001	<.001	<.001	<.001	<.001
<b>Pr(High school=1)</b>	Mean	.624	.632	.640	.646	.653	.660
	S.D.	(.484)	(.482)	(.480)	(.478)	(.476)	(.474)
	p-value		.026	<.001	<.001	<.001	<.001
<b>Total individuals</b>		37848	34551	31851	29922	28580	27484

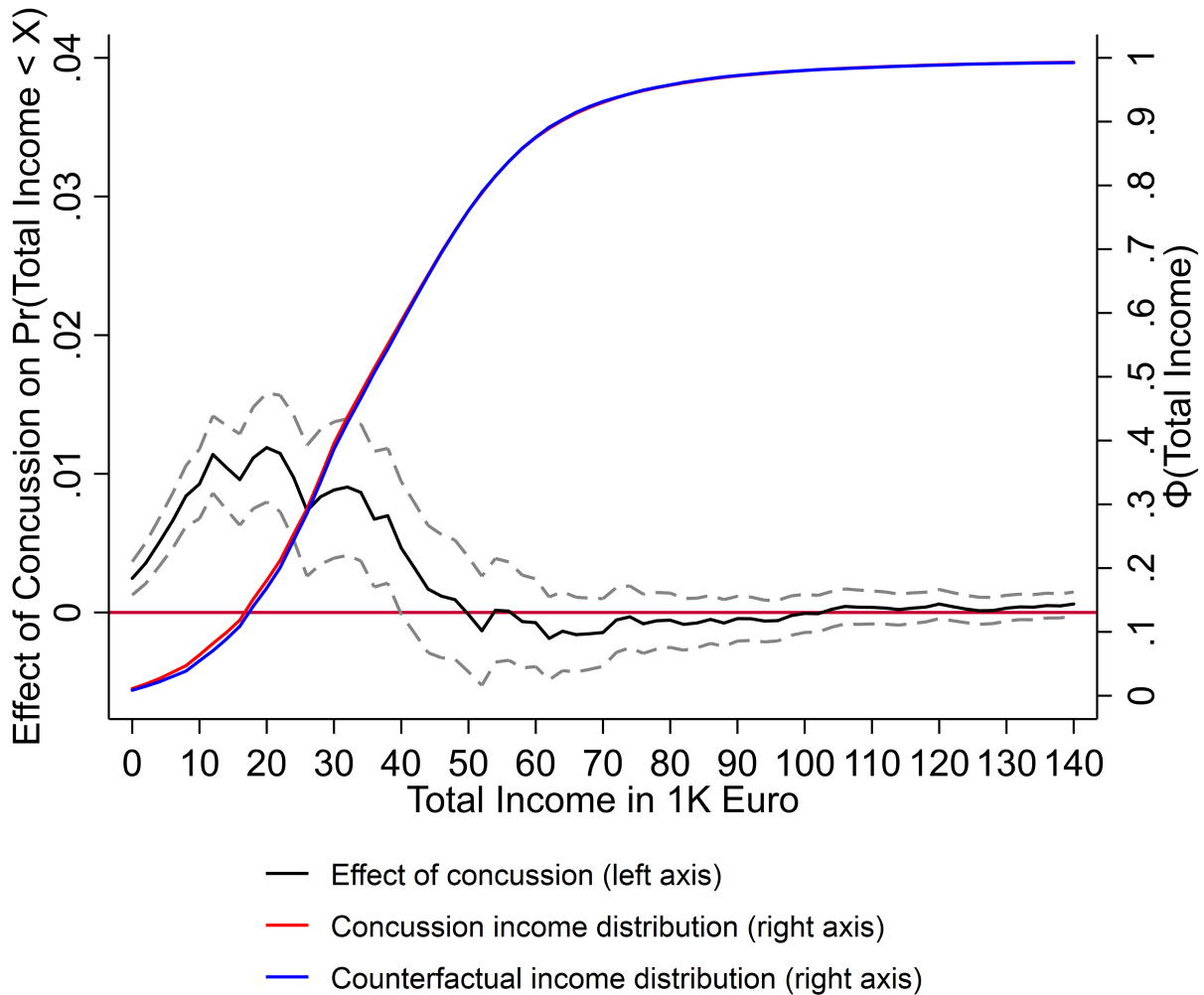
**Note:** S.D.: Standard deviation. P-values calculated using two-sided t-tests. All test performed between exposure group and each control group separately.

Figure S1. Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels



**Note:** Salary of the exposure group compared to salary of the 5 control groups, who experienced their concussions  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  years later than the exposure group. Salary progression is shown for the 5 years before and the 5 years after the exposure group suffered a concussion event. Table S1 demonstrates that the trends for salary progression pre-exposure incident are parallel between exposure group and each control group.

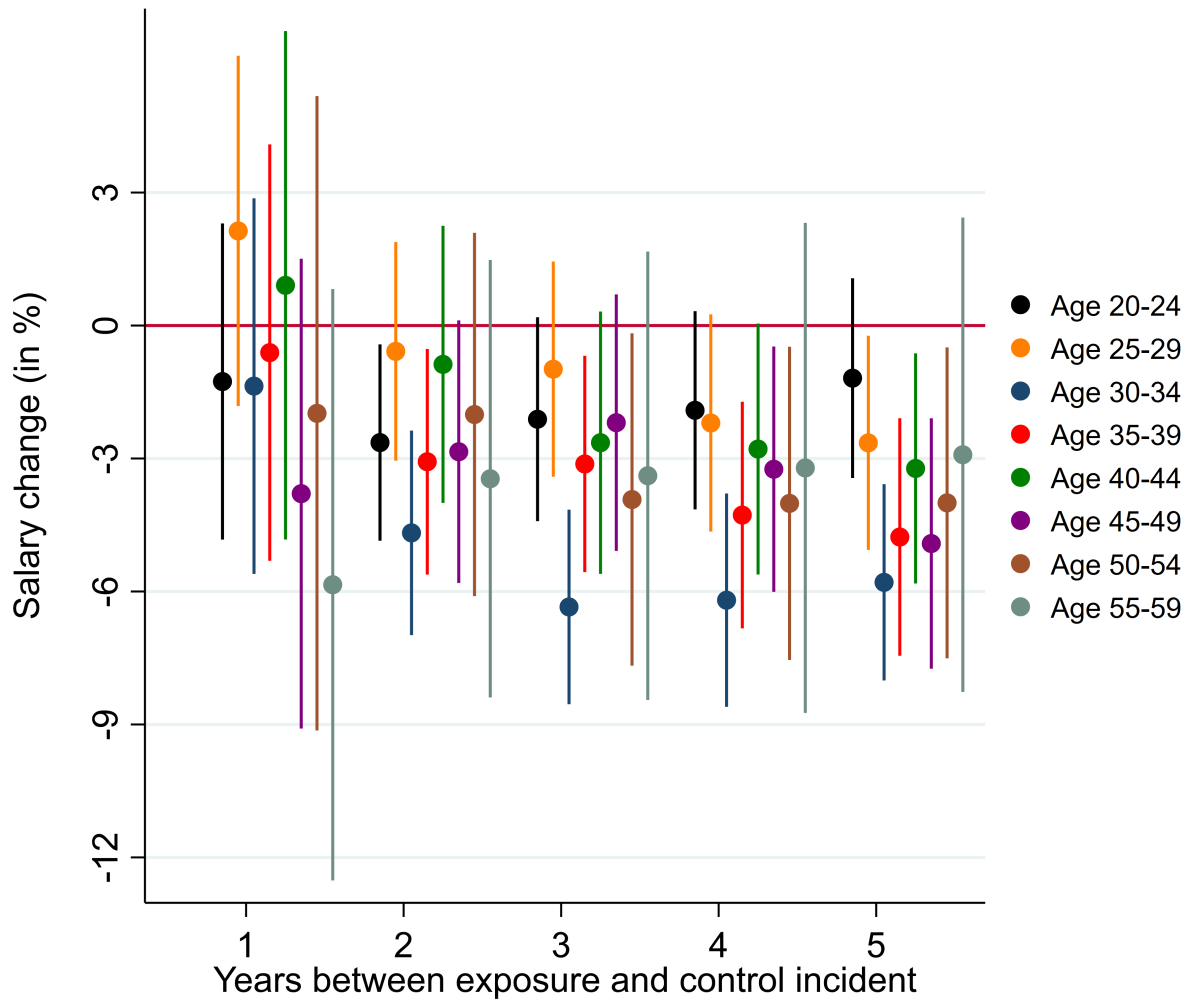
**Figure S2.** The Cumulative Density Function (CDF) for Total Income Post-Treatment among the Treatment Group and Their Counterfactual, and the Difference between the Two CDFs Expressed as the Effect of Concussion on the Probability of Total Income Below that Income-Level following Exposure Event.



**Note:** The figure shows the observed cumulative salary distribution following a concussion (red) and the expected counterfactual salary distribution absent the concussion (blue). The black line shows the difference between the observed and the counterfactual distribution, and the grey dash lines show the 95 % confidence interval. The bell-shape of the difference between the two distributions as the total income increase from 0 to 40,000 € indicates that the main part of the effect of concussions on total incomes is driven by low-income people shifting total income downwards following concussion, but not going to total income equal to zero.

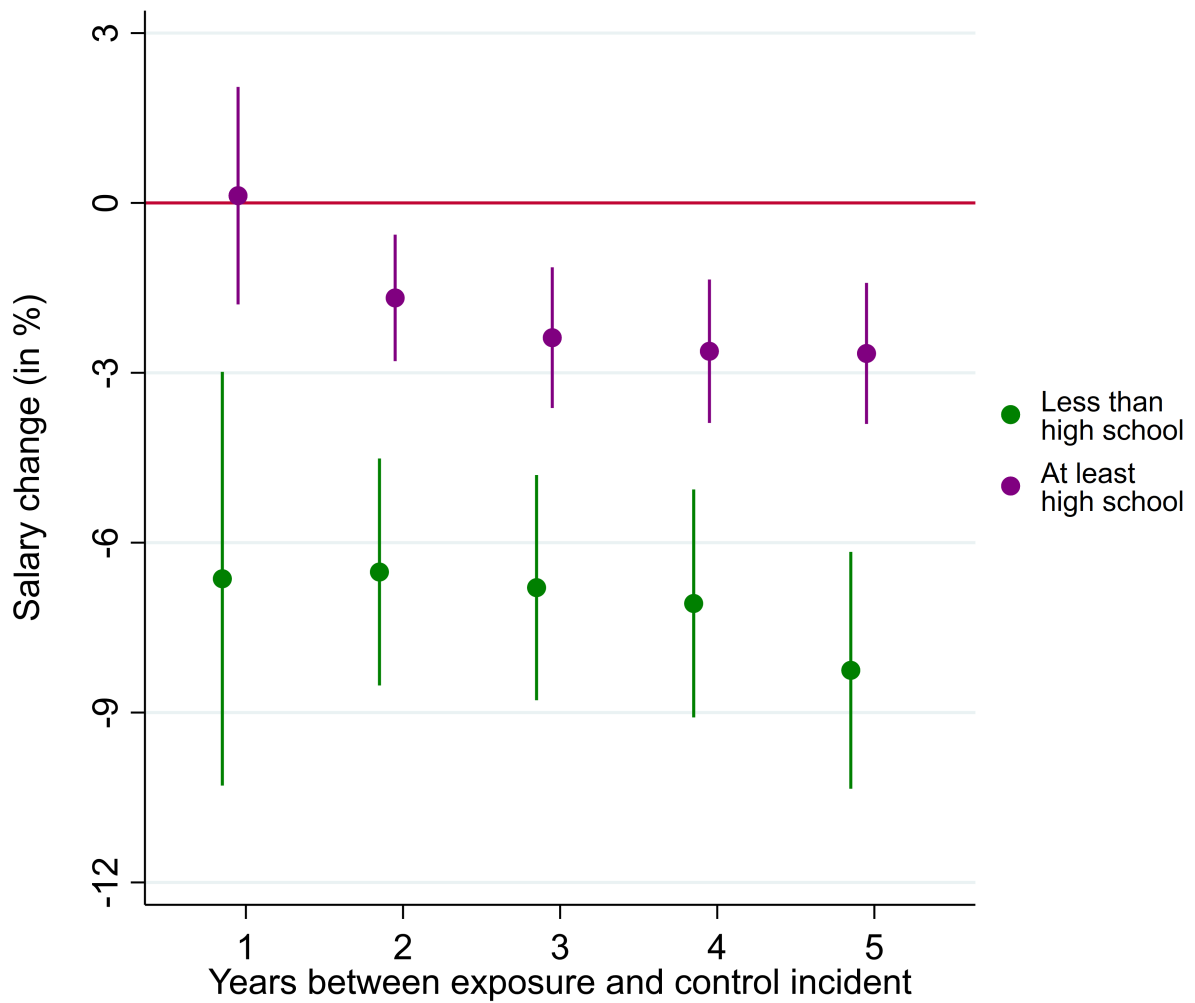


Figure S3. Percentage Effect of Concussion on Relative Salary Across Age Groups.

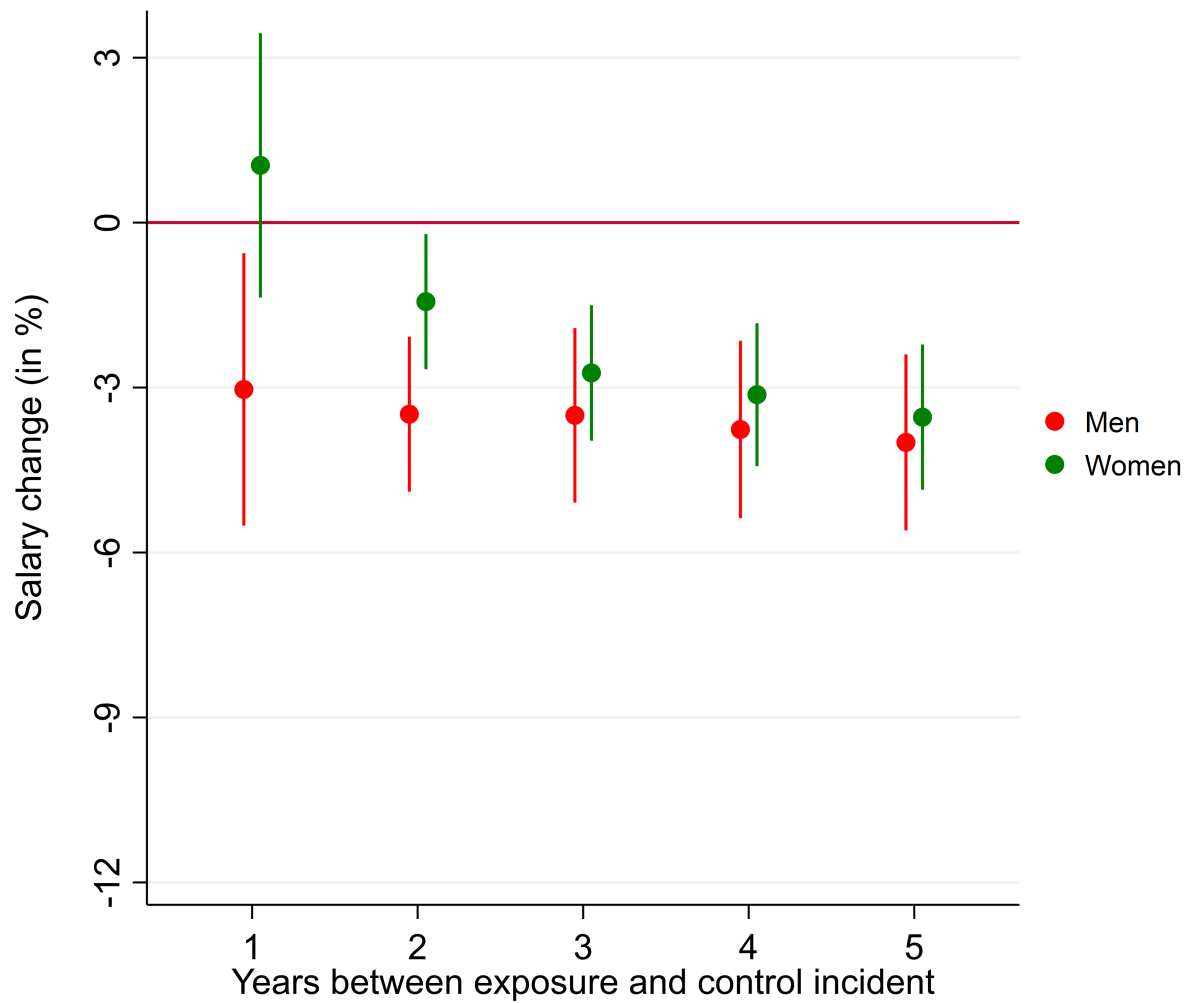


**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

**Figure S4. Percentage Effect of Concussion on Relative Salary Across High School Completion.**

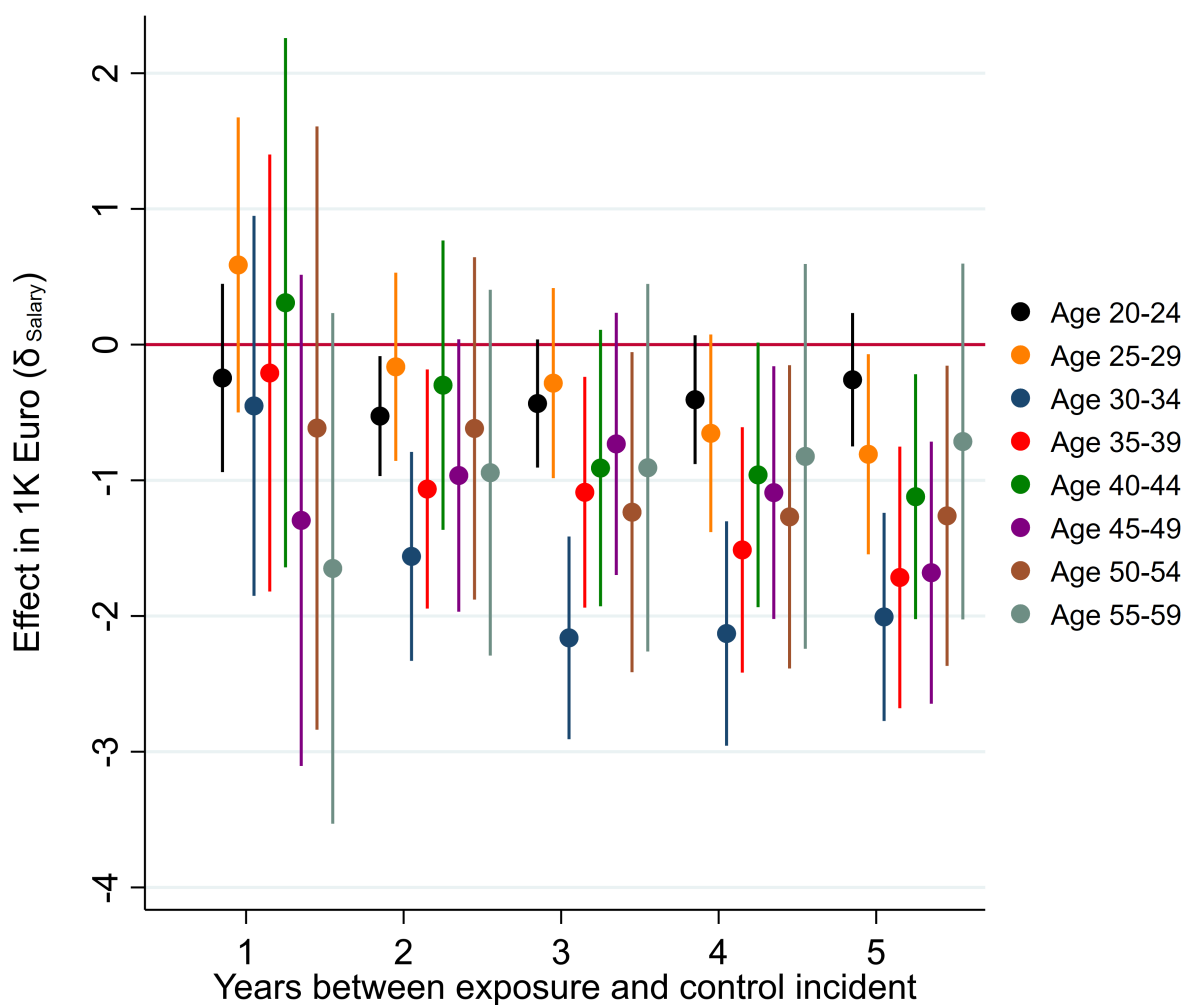


**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

**Figure S5. Percentage Effect of Concussion on Relative Salary Across Gender.**

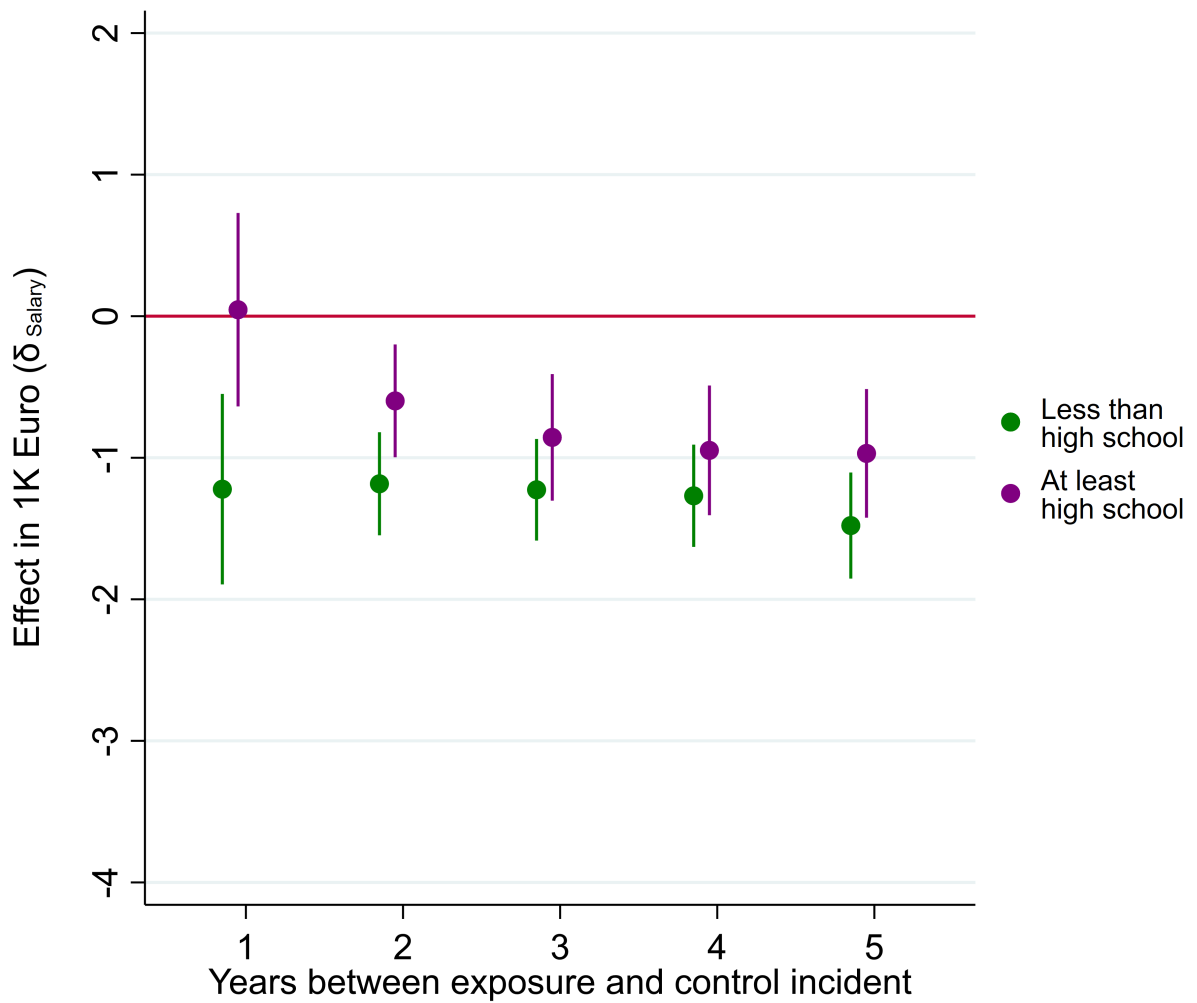
**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across gender. Graph shows parameter estimates and 95% CI.

Figure S6. Effect of Concussion on Absolute Salary in 1K Euro Across Age groups.



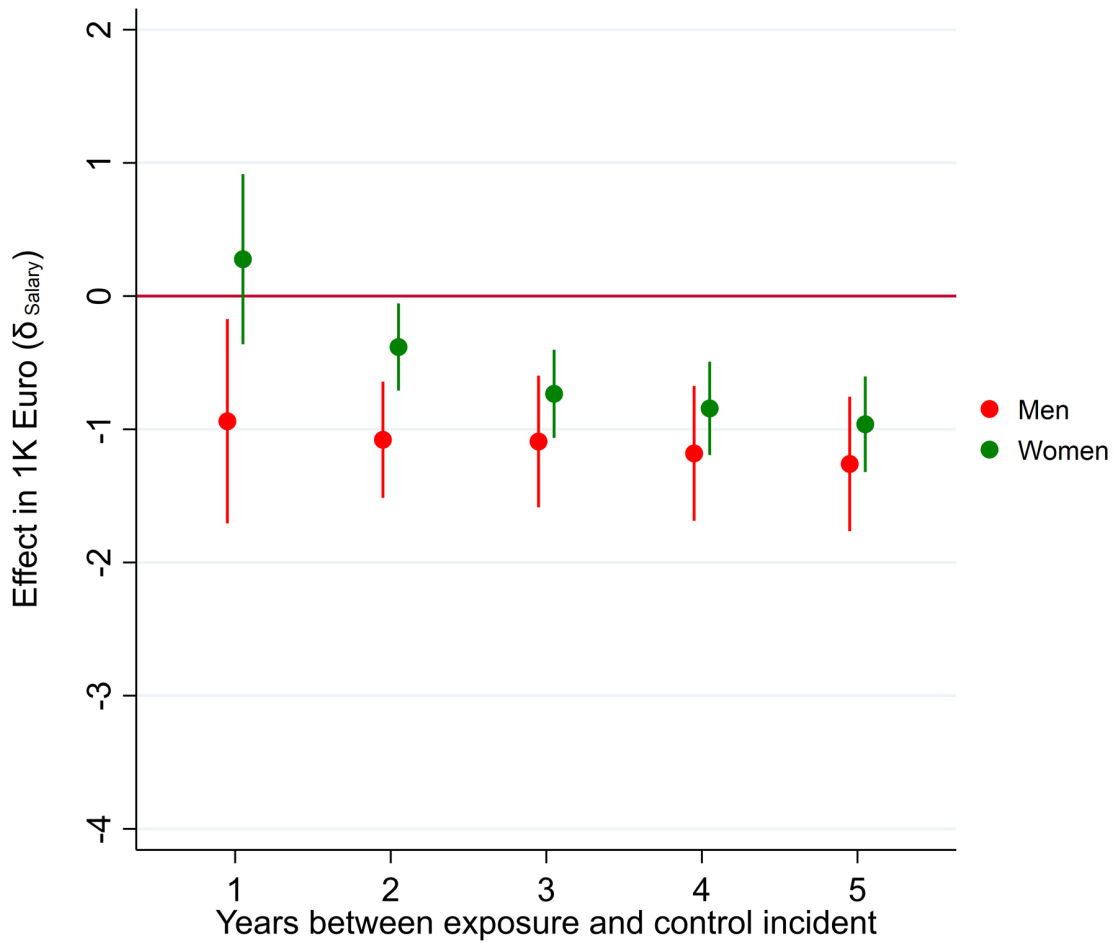
**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

Figure S7. Effect of Concussion on Absolute Salary in 1K Euro Across Education.



**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

**Figure S8. Effect of Concussion on Absolute Salary in 1K Euro Across Gender.**



**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across gender. Graph shows parameter estimates and 95% CI.

**The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.**

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
<b>Title and abstract</b>					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	title abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.  RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.  RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	title abstract  title abstract  abstract
<b>Introduction</b>					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	abstract introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	introduction		
<b>Methods</b>					
Study Design	4	Present key elements of study design early in the paper	introduction		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	materials and methods		

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27	Participants	6	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p><i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p> <p><i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>	materials and methods	<p>RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.</p> <p>RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.</p> <p>RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.</p>	<p>materials and methods</p> <p>materials and methods</p> <p>not included</p>
28 29 30 31 32 33 34	Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	materials and methods main text	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	materials and methods
35 36 37 38 39 40 41 42	Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	materials and methods		

43  
44  
45  
46  
47



1 2 3	Bias	9	Describe any efforts to address potential sources of bias	materials and methods and results	
4 5	Study size	10	Explain how the study size was arrived at	materials and methods	
6 7 8 9 10 11	Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	materials and methods	
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35	Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	a) materials and methods b) materials and methods c) materials and methods d-e) NA	
36 37 38 39 40 41 42 43 44	Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.  materials and methods  materials and methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	materials and methods
<b>Results</b>					
Participants	13	(a) Report the numbers of individuals at each stage of the study ( <i>e.g.</i> , numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a-c) materials and methods	RECORD 13.1: Describe in detail the selection of the persons included in the study ( <i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	materials and methods
Descriptive data	14	(a) Give characteristics of study participants ( <i>e.g.</i> , demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time ( <i>e.g.</i> , average and total amount)	a) materials and methods and Supplementary Table S3 b) materials and methods, Table 1 c) materials and methods		
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure	materials and methods and Supplementary Table S3, results		

		category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	a) results b) results c) results		
Other analyses	17	Report other analyses done— e.g., analyses of subgroups and interactions, and sensitivity analyses	results		
<b>Discussion</b>					
Key results	18	Summarise key results with reference to study objectives	results and discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	results and discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	NA
Interpretation	20	Give a cautious overall interpretation of results considering objectives,	discussion		

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results	discussion		
<b>Other Information</b>					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Title page		
Accessibility of protocol, raw data, and programming code		..	Reference to supplementary data throughout the text	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	materials and methods

\*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

\*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

# BMJ Open

## The Effect of Concussion on Salary and Employment-A Population-Based Event Time Study using a Quasi-Experimental Design

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038161.R1
Article Type:	Original research
Date Submitted by the Author:	21-Jul-2020
Complete List of Authors:	Fallesen, Peter; Stockholm University, Swedish Institute for Social Research Campos, Benito; Heidelberg University Hospital Department of Neurosurgery; Technical University of Denmark, Department of Health Tech
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Neurology
Keywords:	HEALTH ECONOMICS, Neurosurgery < SURGERY, Trauma management < ORTHOPAEDIC & TRAUMA SURGERY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 1 **The Effect of Concussion on Salary and Employment-A Population-Based Event Time**  
5 2 **Study using a Quasi-Experimental Design**  
6  
7  
8 3

9  
10 4 <sup>1,2</sup>Peter Fallesen, PhD, <sup>3,4</sup>Benito Campos, M.D.  
11  
12 5

13  
14  
15 6 <sup>1</sup>Swedish Institute for Social Research, Stockholm University, SE-106 91 Stockholm, Sweden

16 7 <sup>2</sup>ROCKWOOL Foundation, Ny Kongensgade 6, DK-1472 Copenhagen C, Denmark

17 8 <sup>3</sup>Department of Neurosurgery, University of Heidelberg, INF400, 69120 Heidelberg, Germany

18 9 <sup>4</sup>Department of Health Tech, Technical University of Denmark, DK-2800 Kongens Lyngby,  
19 10 Denmark  
20  
21  
22 11

23  
24 12 **Address of Correspondence:** Peter Fallesen, Swedish Institute for Social Research,  
25 13 Stockholm University, SE-106 91 Stockholm, Sweden, [peter.fallesen@sofi.su.se](mailto:peter.fallesen@sofi.su.se),  
26 14 +4561333184  
27  
28 15

29 16 **Keywords:** administrative data; concussion; income; productivity  
30  
31

32 17 **Abstract word count: 245**  
33

34 18 **Text word count: 3913 (excluding references and figure/table legends)**  
35

36 19 **Number of references: 21**  
37  
38

39 20 **Number of tables and/or figures: 2 tables and 4 figures**  
40

41 21 **Number of videos: 0**  
42  
43 22  
44  
45 23  
46  
47 24  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 25 **Abstract**

26 **Objective:** Concussions are the most frequent traumatic brain injuries. Yet, the socio-economic  
27 impact of concussions remains unclear. Socio-economic effects of concussions on working age  
28 adults were studied on a population scale.

29 **Design:** This population-based, event time study uses administrative data as well as hospital  
30 and emergency room records for the population of Denmark.

31 **Setting:** We study all Danish patients, aged 20-59 y, who were treated at a public hospital or  
32 emergency room between 2003-2017 after suffering a concussion without other intracranial or  
33 extracranial injuries (n=55,424 unique individuals) with no prior diagnosis of intra- or  
34 extracranial injury within the past ten years leading up to the incident.

35 **Primary and Secondary Outcome Measures:** As primary endpoint, we investigate the mean  
36 effect of concussion on annual salaried income within a five-year period after trauma. In an  
37 exploratory analysis, we study whether the potential impact of concussion on annual salaried  
38 income is driven by patient age, education, or economic cycle.

39 **Results:** Concussion was associated with an average change in annual salary income of -  
40 1,223€ (95% CI, -1,540€; -905€, p<.001) corresponding to a salary change of -4.2 % (95% CI,  
41 -5.2 %; -3.1 %). People between 30-39 y and those without high school degrees suffered the  
42 largest salary decreases. Affected individuals leaving the workforce drove the main part of the  
43 decrease. Absolute annual effect sizes were countercyclical to the unemployment rate.

44 **Conclusions:** Concussions have a large and long-lasting impact on salary and employment of  
45 working-age adults on a nationwide scale.



1  
2  
3  
4 47 **Strengths and limitations of this study**

- 5  
6 48 - Natural experiments used to obtain plausible causal effects between concussion and  
7  
8 49 salary/employment.  
9  
10  
11 50 - Large, population-based sample with multiple data layers.  
12  
13 51 - Analysis includes how economic cycles affect outcome measures.  
14  
15 52 - Data only captures concussions registered in ERs and hospitals.  
16  
17  
18 53 - Because concussions do not occur at random, causal estimate relies on stronger  
19  
20 54 assumptions than for a randomized control trial.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

## 56 Introduction

57 Concussions are by far the most frequently occurring intracranial injuries affecting  
58 approximately 450 to 600 per 100,000 people every year[1] . Immediate symptoms may last  
59 for days or weeks. Further, Danish cohort data[2] indicates that 10-15% of patients diagnosed  
60 with concussion suffer from long-term symptoms such as headache, fatigue, and intolerance to  
61 stress, whereas other studies place the upper bound as high as 30%[3,4]. Clinical practice has  
62 encouraged patients to restrict social, mental, and physical activity in the weeks following a  
63 concussion (see[5] for review), although prolonged inactivity may prolong symptoms. Thus,  
64 symptoms, comorbidities, and suggested treatment are associated with short- to long-term  
65 absence from work and lower productivity.

66 Yet, the causal effect of concussion on economic burdens for individuals and society through  
67 decreased labour market activity has not been identified. First, concussion is a sudden incident  
68 and thus not amenable to prospective study nor randomization. Cohort and case-control  
69 studies[6–10] provide some valuable evidence on employment and labour market outcomes  
70 among those who suffered concussions but are prone to selection bias. Individuals at high risk  
71 of concussions may differ on unobserved characteristics (e.g., risk aversion, routine activities)  
72 from those at low risk. People who are more likely to suffer concussions may also, on average,  
73 have more precarious or unstable employment trajectories prior to the incident, which may  
74 further bias prospective studies. Given the high incidence rate of concussion, even small losses  
75 of productivity and discrete drops in employment would have a significant socioeconomic  
76 impact and thus, it would require large patient cohorts with suitable controls to grasp the full  
77 socioeconomic impact of concussions. Thus, absent the possibility of randomization, using a

1  
2  
3  
4 78 natural or quasi-experimental design is the only likely option to parse out the *causal* effect of  
5  
6 79 concussions on labour market activity.[11]  
7  
8

9 80 We examine how concussions affect salary and employment of working age individuals in  
10  
11 81 Denmark, a representative north-European industrial nation with a strong welfare state and a  
12  
13 82 flexible labour market. We use administrative longitudinal data linked to hospital and emergency  
14  
15 83 room diagnostic data on all Danes, who received a primary diagnosis of concussion between  
16  
17 84 2003 and 2017. To address the problem of unmeasured bias between those that do and do not  
18  
19 85 experience a concussion, we use a quasi-experimental event-study approach[12,13] where we  
20  
21 86 compare similar individuals, who experienced their concussions at different time points. Under  
22  
23 87 mild assumptions of parallel trends in wage progression prior to concussion and random timing  
24  
25 88 of concussion event within a five-year time frame, the approach recovers a robust estimation  
26  
27 89 of the effect of concussion on annual salary and employment status.  
28  
29  
30  
31

32 90  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

## 91 **Material & Methods**

### 92 *Data Sources and Sample Construction*

93 Concussion data originates from the Danish National Patient Registry (DNPR) (see[14] for  
94 description). DNPR is published annually and holds information on all hospitalizations at public  
95 hospitals in Denmark since 1977, on all emergency room visits and outpatient treatments at  
96 public hospitals since 1994, and almost all private hospital treatments since 2003. With one  
97 single, short-lived exception, private hospitals do not operate emergency rooms in Denmark.  
98 Since 2003, the data cover 95 % of all treatments at private hospitals[14], yet only 13  
99 concussion were diagnosed in private hospital settings throughout the period covered by the  
100 data.

101 The combined exposure and control cohort includes all Danes aged 20-59 y, who received a  
102 primary diagnosis of concussion (ICD-10 code S06.0, ICD-8 code N850) between 2003 and  
103 2017 and did not sustain any kind of additional intracranial or extracranial injury. Individuals  
104 who regularly engage in activities associated with a high risk of sustaining multiple concussions  
105 may differ from the average concussion patient and would likely be over-represented in the  
106 exposure sample. To avoid such potential bias, all individuals who were diagnosed with any  
107 kind of brain trauma during a ten-year period prior to the concussion event were excluded.  
108 Altogether, the study included a cohort of 55,424 individuals. Only attrition is through mortality  
109 and out-migration, and out-migrated or deceased patients with missing spells in the follow up  
110 period is excluded in those periods.

111 As a measure of productivity, a price-index deflated annual salaried income was used. Salary  
112 information comes from Statistics Denmark's Income Statistics database. The database

1  
2  
3  
4 113 includes all declared annual incomes including income from self-employment. The Danish Tax  
5  
6 114 Authorities supply the data to Statistics Denmark. Overall accuracy is considered very good.[15]  
7  
8  
9 115 Table 1 reports number of observations for the samples and number of observations with  
10  
11 116 missing salary information. As evident, only between 0.01 to 0.02 percent of observations  
12  
13 117 across exposure and control groups have missing salary information. These observations were  
14  
15  
16 118 disregarded in the main analysis. Through social security numbers, information on salaried  
17  
18 119 income were linked to records on diagnosed concussions. Further, information on high school  
19  
20 120 or equivalent level degree at time of concussion was obtained using the Danish Education  
21  
22 121 Database. The Danish Population Database provided demographic information on age and  
23  
24  
25 122 gender for all respondents. Since the data used in the study come from de-identified  
26  
27 123 administrative registers that Statistics Denmark makes available for research purposes for  
28  
29  
30 124 approved institutions, no approval from an ethics committee was needed to carry out the study.  
31  
32 125 The research was carried out as part of project no. 706630 approved by Statistics Denmark.  
33  
34 126 Statistical analysis was carried out using Stata MP 15.1.  
35  
36  
37 127  
38

### 39 128 *Quasi-experimental design*

40  
41  
42 129 The study used a quasi-experimental, difference-in-differences event time approach previously  
43  
44 130 described in a health setting by Dobkin et al.[12] The approach compare two groups of  
45  
46 131 individuals from the same cohort, where both groups experience concussions, but at two  
47  
48 132 different time points ( $t_c, t_c + \Delta$ ). Specifically, the sample of 55,496 individuals was divided into six  
49  
50  
51 133 different subgroups: i) The *exposure group*, which includes all patients, who suffered their  
52  
53 134 concussion during the period 2003-2012 ( $n=37,848$ ) and ii) five *control groups*, which comprise

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

135 patients who experienced their concussions  $\Delta=\{1$  (n=34,551), 2 (n=31,851), 3 (n=29,922), 4  
 136 (n=28,530), and 5 (n=27,421)} years later than the exposure group and did not experience any  
 137 kind of brain injury in the  $10+\Delta$  years before the concussion event (note that the design allows  
 138 individuals to both be part of the exposure and control group). The model is built on the  
 139 assumption that the exact timing of a concussion is random for small enough values of  $\Delta$ , and  
 140 on the additional assumption that the exposure and the control groups would have displayed  
 141 parallel trends in salary if the control group had not suffered a concussion at  $t_c$  (i.e., assuming  
 142 that control and exposure group(s) would have continued to show similar trends in salaried  
 143 earnings had the exposure group not experienced concussions). Table 1 show the number of  
 144 patients in the exposure group and the five control groups for each year relative to exposure  
 145 group's concussion incident. Using multiple comparison groups makes it possible to gage the  
 146 validity of the assumption that the exact timing of a concussion is random for small enough  
 147 sizes of  $\Delta$

148 To estimate the impact of concussion on labour market outcomes, the analysis focuses on the  
 149 change in annual salary as the primary outcome, and, in further exploratory analyses, studies  
 150 additional outcomes such as income from health-related benefits, income from welfare benefits,  
 151 and employment rates. The data are nested within a three-level structure: Exposure or control  
 152 group  $g$ , which includes individuals  $l$ , at times to exposure-groups concussion incident  $t$ . First,  
 153 a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$   
 154 is estimated using ordinary least squares:

$$\begin{aligned}
 \text{Salary}_{git} = & \beta_0 + \gamma \text{Exposure}_g + \theta \text{Post}_t + \delta \text{Post} \times \text{Exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{\text{Age}=20}^{59+\Delta} I(\text{Age}) \eta_{\text{Age}} \\
 & + \sum_{\text{Year}=1999}^{2012} I(\text{Year}) \eta_{\text{Year}} + \epsilon_{git}
 \end{aligned}
 \tag{1}$$

8

1  
2  
3  
4 157 where  $Salary_{git}$  measures annual salaried income deflated to 2015-level;  $Exposure_g$  indicates  
5  
6  
7 158 whether the observation belongs to the exposure or control group;  $Post_t$  captures the period  
8  
9 159 after the exposure group's concussion occurred;  $Post_t \times Exposure_{git}$  captures the effect  
10  
11 160 concussion, measured as share of year  $t \geq 0$  affected by concussion (i.e., for year of incident  
12  
13  
14 161 exposure is expressed as share of year spent with post-exposure, for following years it is equal  
15  
16 162 to 1);  $X_i$  is a set of covariates that includes a high school indicator and a gender dummy;  $\epsilon_{git}$  is  
17  
18 163 the error-term; and the two last sets of indicator variables  $I(Age)$  and  $I(Year)$  capture age and  
19  
20  
21 164 incident year (for control group, the year indexed against). Under an assumption of parallel  
22  
23 165 trends in salaried earnings (i.e., assuming that control and exposure group(s) would have  
24  
25 166 continued to show similar trends in salaried earnings had the exposure group not experienced  
26  
27  
28 167 concussions),  $\delta$  then captures the annual causal effect of concussion on salary for people  
29  
30 168 exposed to concussions (see Supplemental Methods for further details). For additional  
31  
32 169 exploratory analyses, separate models across gender, educational level, and age, as well as  
33  
34  
35 170 across the salary distribution are also estimated (see Supplemental Methods, Supplemental  
36  
37 171 Digital Content 1, for further details). The authors document and make available all code  
38  
39 172 needed to reproduce the findings in the study (Supplemental Digital Content 2).

#### 44 174 *Standard Protocol Approvals, Registrations, and Patient Consents*

45  
46 175 Since the data used in the study come from de-identified administrative registers that Statistics  
47  
48  
49 176 Denmark makes available for research purposes for approved institutions, no approval from an  
50  
51 177 ethics committee was needed to carry out the study. The research was carried out as part of  
52  
53 178 project no. 706630 approved by Statistics Denmark.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

179

*Patient and Public Involvement*

There was no involvement from patients or members of the public in the design, or conduct, or reporting, or dissemination plans of the research.

For peer review only



## Results

### *Concussion leads to long-term loss in salaried income*

Individuals who suffered a concussion (exposure group) had a lower salaried income compared to individuals who experienced their concussion 1, 2, 3, 4 and 5 years after the exposure group (control groups). Compared to patients who experienced a concussion one year after the exposure group, salaried income was €423/£380 (95% CI: -€9129/-£8208; €73/£66,  $p=.095$ ; Table 2) lower, corresponding to a salary decrease of 1.5 % (95% CI: -0.3 %; 3.2 %; Figure 1). Compared to patients who experienced a concussion 5 years after the exposure group, however, salaried income in the exposure group was €1,243 (95% CI: -€1,564/-£1,406; -€922/-£829,  $p<.001$ ) lower, corresponding to a salary decrease of 4.2 % (95% CI: 3.1 %; 5.3 %; Figure 1). Normalized wage progression for the control groups, who suffered a concussion 1, 2, 3, 4, and 5 years after the exposure group, showed similar trends and similar levels pre-exposure, indicating that the parallel wage trends assumption was met (Figure 2 and table S1, Figures S1 in Supplemental Digital Content 3).

We hypothesized that the salary decreases resulted from a combination of lower salary and exit from the labour market, either through short- or long-term absence/unemployment. In an exploratory analysis, we tested whether labour force exit drove the full effect of concussion on salary (Figure 3). By comparing the cumulative distribution of salary density for the exposure group with the cumulative distribution of salary density for the  $\Delta=5$  control group (Figure 3, left panel), we found that the impact of concussion on salary was significant for individuals in the lower quartile of the salary distribution (at a 95 % significance level). Specifically, below a threshold salaried income of 40,000€ (£36,000) the presumed impact of concussion on salary

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

increased towards the bottom of the earnings pyramid; Figure 3, right panel).

Comparing the exposure group to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group, concussion was associated with a 2.6% (95% CI: 3.0 %; 2.2 %,  $p <.001$ ) increase in the risk of receiving € 0 in annual salary.

#### *Long-term loss in salaried income stems from exit from the labour market*

To further examine whether exit from the labour market was caused either through short- or long-term absence/unemployment, we estimated a dynamic model using the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group. Sick leave benefits payments were higher in the exposure group compared to the control groups for the first two years following concussion. Sick leave benefits were no longer different from year 3 while the difference in annual salary between exposure and control groups persisted. Further, employment in the exposure group remained lower than in the control group  $\Delta=5$  and remained so for the entire post-exposure period (see table S2, Supplemental Digital Content 3 for further details). To assess whether some form of public benefits covered part of the salary loss, total income decline was compared to salary decline following concussion. Indeed, total income decline was lower than the salary decline through a five year period (see Figure S2, Supplemental Digital Content 3 for further details).

#### *Younger patients without high school degree drove the effect of concussion on income loss*

The exposure group and all control groups differed slightly in terms of average patient age, male to female ratio, and for control group  $\Delta=5$ , in the frequency of individuals with at least a

1  
2  
3  
4 227 high school degree (see Table S3, Supplemental Digital Content 3 for further details). To ensure  
5  
6  
7 228 that differences in gender, education, or age did not influence our results, we subdivided our  
8  
9 229 exposure group into subgroups based on gender, education status, and age at time of  
10  
11 230 concussion. We then estimated the impact of concussion on salary and employment across all  
12  
13 231 values of  $\Delta$  and for all subgroups (see, Figures S3-S8, Supplemental Digital Content 3 for  
14  
15  
16 232 further details). Patients between age 30-39 and those without a high school degree  
17  
18 233 experienced the largest absolute and relative declines in salary.

19  
20  
21 234 Finally, we addressed the role of timing of concussion across different years. Given that per  
22  
23 235 design our exposure group always suffered their concussion earlier than the control groups do,  
24  
25 236 changing labour market conditions could moderate effects. Part of our sample suffered their  
26  
27 237 concussion during or just prior to the Great Recession in 2009-2010, which arguably presented  
28  
29  
30 238 the largest shock to both the global and local economy since the Great Depression in the 1930s.  
31  
32 239 In Denmark, the great recession was preceded by a series of years of economic growth, low  
33  
34 240 unemployment, and increasing salaries (see Figure S8, Supplemental Digital Content 3 for  
35  
36  
37 241 salary development from 1994 to 2017). We estimated the impact of concussion on salary  
38  
39 242 separately for each year from 2003-2012 and plotted the estimate against the percent of full-  
40  
41 243 time unemployment in the Danish labour force (Figure 4). Suffering a concussion during an  
42  
43 244 economic boom had a substantially higher impact on salary than doing so during a recession  
44  
45  
46 245 when comparing to control groups who suffered concussions two to five years later than  
47  
48 246 exposure group.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Discussion

The impact of concussion on employment and salary remains understudied. In a systematic review of four studies on the association between mild TBI and return to work, Cancelliere and colleagues[9] found that most workers return to work within 3-6 months of suffering a mild TBI, but that the long-term impact (more than one year after concussion) was not studied. In addition, studies included small to medium sample sizes, varied measures of return to work, and employed both case-control and cohort designs. Using an inception cohort study design, Theadom and colleagues[7] collected follow up data four years after mild TBI incidents on 245 New Zealanders who were employed prior to incident. They found a 3.6 percent productivity decline among those who suffered a concussion, compared to a 2.3 population average decline. The group suffering mild TBI also reported more difficulties carrying out work-related tasks. In a related study,[7] Theadom and colleagues further found that the mild TBI group had persisting cognitive symptoms four years after suffering their concussion compared to an age-sex matched control group. Also using a case-control design and data from Taiwan, Chu and colleagues[6] found that one month after incident, 26 percent of patients had still not managed to return to work, and a large share of those who did return scored below full-time employment on a work quality index. Only one other study by Graff and colleagues[10] include a large patient cohort (n=19,732). Using case-control they found an association between concussion and failing to return to work of 1.54 odds ratio, but also found that exposed individuals had lower labour market attachment and was more likely to receive health related benefits pre-incident compared to the control group.

In the present study, we overcame some of the obstacles faced by previous work on the impact

1  
2  
3  
4 270 of concussions on labour market outcomes by including a vast cohort of patients and exploiting  
5  
6  
7 271 a quasi-experimental design that allow us to plausibly account for unobserved difference  
8  
9 272 between exposure and control group. In such a quasi-experimental setup exposure and control  
10  
11 273 groups only differ in the timing of concussion. Since everyone in the control group experiences  
12  
13  
14 274 a concussion within five years after individuals in the exposure group, the groups are likely to  
15  
16 275 be balanced on unobservable characteristics. This is particularly important given the number of  
17  
18 276 potential factors that can influence employment after concussion[16,17]. Data from Donker-  
19  
20  
21 277 Cools et al.[17], for instance, suggests larger employers are more able to keep those who have  
22  
23 278 sustained brain injuries in work compared to smaller employers. However, studying the effect  
24  
25 279 of concussion on salaried income for individuals with employer of different size lies beyond the  
26  
27 280 granularity of our data.

28  
29  
30 281 In addition, salary and employment data reported here were compiled routinely through third-  
31  
32 282 party reporting and were mandatory for all subjects, thus giving a complete and comprehensive  
33  
34 283 picture of the economic impact of concussion on a nationwide scale. It should be mentioned  
35  
36  
37 284 that our study also included data from individuals diagnosed in private hospitals. However,  
38  
39 285 given the setup of the Danish health care system, i.e. private hospitals predominantly do  
40  
41 286 selective and overflow surgery and have no ERs[18], only 13 patients were diagnosed at a  
42  
43  
44 287 private hospital from 2003 onwards.

45  
46 288 Altogether, we showed that Danes between 20-59 year of age, who suffered a concussion  
47  
48 289 during the period 2003-2012 experienced average salary losses of 4.2%. The impact of  
49  
50  
51 290 concussions on salary already materialized one year after the incident and remained sizeable  
52  
53 291 for at least five years. This result is in line with a “burn-in” period in which the impact of

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

292 concussion on wages fully develops. First, concussions occur at some point during the year,  
293 thereby not affecting already earned salary that year. Second, in Denmark, most employees  
294 are entitled to receive their salary during sickness absence for an open ended, but not indefinite  
295 period. The results further showed that both in absolute and relative terms, people with an  
296 educational level at less than a high school degree saw substantially larger negative impact to  
297 salaried earnings than did those with at least a high school degree. Also, the group with less  
298 than a high school degree also saw an immediate impact on salary from their concussion (cf.  
299 Figure S4), indicating that the burn-in period present for workers with at least high school  
300 education likely expressed differences in types of employment and job protection.

301 In addition, total income decline was lower than the salary decline through a five year period  
302 (see Figure S2, Supplemental Digital Content 3 for further details), suggesting that the impact  
303 of concussions on salary largely stems from affected individuals leaving the labour force  
304 completely, likely sustaining themselves through early retirement, disability pensions, self-  
305 sufficiency, or other income sources instead.

306 It is important to mention that our study was restricted to individuals diagnosed in ER and  
307 hospital settings. Rowson et al., however, show that in concussed individuals, severity of the  
308 cranial injury is not strongly correlated with strength or length of subsequent symptoms[19].  
309 Thus, individuals diagnosed by a GP might suffer concussion effects as much as individuals  
310 who initially sustained a more severe cranial injury and sought medical attention in an ER or  
311 hospital setting. If this holds true, our results may have validity beyond individuals diagnosed in  
312 an ER or hospital setting.

1  
2  
3  
4 314 If we assume that people return to their expected salary levels after a five-year recovery period  
5  
6  
7 315 (a very conservative assumption that is not supported by our data), the mere net annual salary  
8  
9 316 loss in our sample would be approximately €23,000,000 (£21,000,000) measured in 2015-  
10  
11 317 value. That would neither include hospital charges, medical costs for the treatment of  
12  
13 318 concussion, the foregone tax from income, and the increased need for welfare spending, nor  
14  
15  
16 319 would it account for the large group of individuals who never seeks treatment[20] or receive  
17  
18 320 their diagnosis from their general practitioner rather than in a hospital or emergency room, and  
19  
20 321 thus escape our study. Thus, total public costs are likely substantially higher.

22  
23 322 In addition, both in absolute and relative terms, the early peak-working aged individuals (30-39  
24  
25 323 y) and the less-educated individuals in our cohort seemed to be most affected after suffering a  
26  
27 324 concussion. These findings might have an additional and yet unmeasured social impact,  
28  
29  
30 325 especially if our results are transferrable to other nations with a less established welfare state  
31  
32 326 and a less flexible labour market. In such countries, the impact on the young and less-educated  
33  
34 327 individuals suffering a concussion and thus on society might be accentuated.

36  
37 328 Comparing our hospital incidence rates to more complete canvases of incidences carried out  
38  
39 329 by Cassidy et al.[21], it seems likely that the actual cost in the population is more than twice as  
40  
41 330 large as what we estimate, assuming that individuals not diagnosed in a hospital setting on  
42  
43  
44 331 average suffer the same extent of concussion symptoms. If we were to consider the average  
45  
46 332 concussion incidence rates for six other advanced European countries that are somewhat  
47  
48 333 comparable to Denmark (Norway, Finland, Germany, Netherlands, England and France) and  
49  
50  
51 334 under the assumption that concussion have a similar impact on earnings in these countries, the  
52  
53 335 net annual salary loss would be approximately €1,099,400,000 (£988,4780,000) measured in

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

336 2015-value. While our study likely underestimates the total socioeconomic impact of  
337 concussion, it suggests that concussions has a large economic impact on a nationwide scale  
338 and on productivity and income at the patient level.

339

## 340 CONCLUSION

341 Using timing of concussion as a natural experiment, we provide first plausible causal estimates  
342 of the effect of concussion on salary and employment among patients treated for concussion in  
343 an emergency room or hospital setting in Denmark, 2003-2017. Our results show that among  
344 this patient group concussion has a large and long-term negative causal impact on salary and  
345 employment. People between 30-39 y and those without high school degrees suffered the  
346 largest salary decreases.



1  
2  
3  
4 348 **Disclosures:**

5  
6  
7 349 The authors report no conflict of interest.  
8

9 350  
10  
11 351 **Acknowledgements:**

12  
13  
14 352 NA  
15

16 353  
17  
18  
19 354 **Funding:**

20  
21 355 This work was supported by the Swedish Research Council for Health (PF), Working Life and  
22  
23  
24 356 Welfare (Grant no. 2016-07099) with additional funding from the ROCKWOOL Foundation (PF).

25  
26 357 The research was carried out independently of the funders.  
27

28 358  
29  
30  
31 359 **Authors contributions:** P.F. and B.C. conceived of the presented idea, P.F. performed the  
32  
33 360 computations. P.F. and B.C. verified the statistical methods. P.F. and B.C. discussed the results  
34  
35 361 and wrote the manuscript. The corresponding author confirms that he had full access to all the  
36  
37  
38 362 data in the study and had final responsibility for the decision to submit for publication.  
39

40 363  
41  
42 364 **Data Availability Statement**  
43

44  
45 365 The data used in this study has been made available through a trusted third party, Statistics  
46  
47 366 Denmark. Due to privacy concerns the data cannot be made available outside the hosted  
48  
49 367 research servers at Statistics Denmark. University-based and private Danish scientific  
50  
51  
52 368 organizations can be authorized to work with data within Statistics Denmark. Such organization  
53  
54 369 can provide access to individual scientists inside and outside of Denmark. Requests for data

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

370 may be sent to Statistics Denmark:  
371 <http://www.dst.dk/en/OmDS/organisation/TelefonbogOrg.aspx?kontor=13&tlfbogsort=sektion>  
372 or the Danish Data Protection Agency: <https://www.datatilsynet.dk/english/the-danish-data-protection-agency/contact/>. The authors document and make available all code needed to  
373 reproduce the findings in the study.  
374

For peer review only

## References

- 1 Cassidy JD, Carroll LJ, Peloso PM, *et al*. Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* 2004;:28–60. doi:10.1080/16501960410023732
- 2 Pinner M, Jensen R, Birket-Smith M, *et al*. Konsensusrapport om commotio cerebri (hjernerystelse). Videnscenter for Hjerneskode 2003.
- 3 Hou R, Moss-Morris R, Peveler R, *et al*. When a minor head injury results in enduring symptoms: a prospective investigation of risk factors for postconcussional syndrome after mild traumatic brain injury. *J Neurol Neurosurg Psychiatry* 2012;**83**:217–23. doi:10.1136/jnnp-2011-300767
- 4 Ruff R. Two decades of advances in understanding of mild traumatic brain injury. *J Head Trauma Rehabil* 2005;**20**:5–18. doi:10.1097/00001199-200501000-00003
- 5 Silverberg ND, Iverson GL. Is rest after concussion “the best medicine”? recommendations for activity resumption following concussion in athletes, civilians, and military service members. *The Journal of head trauma rehabilitation* 2013;**28**:250–9. doi:10.1097/HTR.0b013e31825ad658
- 6 Chu SY, Tsai YH, Xiao SH, *et al*. Quality of return to work in patients with mild traumatic brain injury: a prospective investigation of associations among post-concussion symptoms, neuropsychological functions, working status and stability. *Brain Injury* 2017;**31**:1674–82. doi:10.1080/02699052.2017.1332783
- 7 Theadom A, Barker-Collo S, Jones K, *et al*. Work Limitations 4 Years After Mild Traumatic Brain Injury: A Cohort Study. *Archives of Physical Medicine and Rehabilitation*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

398 2017;**98**:1560–6. doi:10.1016/j.apmr.2017.01.010

399 8 Theadom A, Starkey N, Barker-Collo S, *et al*. Population-based cohort study of the  
impacts of mild traumatic brain injury in adults four years post-injury. *PLOS ONE*  
400 2018;**13**:e0191655. doi:10.1371/journal.pone.0191655

402 9 Cancelliere C, Kristman VL, Cassidy JD, *et al*. Systematic Review of Return to  
Work After Mild Traumatic Brain Injury: Results of the International Collaboration on Mild  
403 Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation*  
404 2014;**95**:S201–9. doi:10.1016/j.apmr.2013.10.010

406 10 Graff HJ, Siersma V, Møller A, *et al*. Labour market attachment after mild traumatic  
407 brain injury: nationwide cohort study with 5-year register follow-up in Denmark. *BMJ open*  
408 2019;**9**:e026104. doi:10.1136/bmjopen-2018-026104

409 11 Ogilvie D, Adams J, Bauman A, *et al*. Using natural experimental studies to guide  
410 public health action: turning the evidence-based medicine paradigm on its head. *J Epidemiol*  
411 *Community Health* Published Online First: 19 November 2019. doi:10.1136/jech-2019-213085

412 12 Dobkin C, Finkelstein A, Kluender R, *et al*. The Economic Consequences of  
413 Hospital Admissions. *American Economic Review* 2018;**108**:308–52.  
414 doi:10.1257/aer.20161038

415 13 Fadlon I, Nielsen TH. Household Responses to Severe Health Shocks and the  
416 Design of Social Insurance. Cambridge, MA: 2015. doi:10.3386/w21352

417 14 Schmidt M, Schmidt SAJ, Sandegaard JL, *et al*. The Danish National Patient  
418 Registry: a review of content, data quality, and research potential. *Clinical epidemiology*  
419 2015;**7**:449–90. doi:10.2147/CLEP.S91125

22

- 1  
2  
3  
4 420 15 Statistics Denmark. Documentation of statistics: Income Statistics - Statistics  
5  
6  
7 421 Denmark.  
8  
9 422 2015.<https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/income->  
10  
11 423 statistics (accessed 20 Aug 2018).  
12
- 13  
14 424 16 Buck PW, Laster RG, Sagrati JS, *et al.* Working with mild traumatic brain injury:  
15  
16 425 voices from the field. *Rehabil Res Pract* 2012;**2012**:625621. doi:10.1155/2012/625621  
17
- 18  
19 426 17 Donker-Cools BHPM, Wind H, Frings-Dresen MHW. Prognostic factors of return to  
20  
21 427 work after traumatic or non-traumatic acquired brain injury. *Disabil Rehabil* 2016;**38**:733–41.  
22  
23 428 doi:10.3109/09638288.2015.1061608  
24
- 25  
26 429 18 Vrangbaek K, Østergren K, Birk HO, *et al.* Patient reactions to hospital choice in  
27  
28 430 Norway, Denmark, and Sweden. *Health Econ Policy Law* 2007;**2**:125–52.  
29  
30 431 doi:10.1017/S174413310700401X  
31
- 32  
33 432 19 Rowson S, Duma SM, Stemper BD, *et al.* Correlation of Concussion Symptom  
34  
35 433 Profile with Head Impact Biomechanics: A Case for Individual-Specific Injury Tolerance. *J*  
36  
37 434 *Neurotrauma* 2018;**35**:681–90. doi:10.1089/neu.2017.5169  
38
- 39  
40 435 20 Setnik L, Bazarian JJ. The characteristics of patients who do not seek medical  
41  
42 436 treatment for traumatic brain injury. *Brain Injury* 2007;**21**:1–9.  
43  
44 437 doi:10.1080/02699050601111419  
45
- 46  
47 438 21 Cassidy JD, Carroll LJ, Peloso PM, *et al.* Incidence, risk factors and prevention of  
48  
49 439 mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on mild  
50  
51 440 traumatic brain injury. *Journal of Rehabilitation Medicine* 2004;**36**:28–60.  
52  
53 441 doi:10.1080/16501960410023732  
54

1  
2  
3  
4 442  
5  
6  
7 443  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## Figure Legends

### Figure 1. Estimated effect of concussions in percentage on salary for the exposure group measured against each control group

Note: Figure shows the percentage change in salary experienced by the exposure group following their concussions compared to the expected trajectory absent the concussion (calculated from the control groups) with 95 % confidence intervals. See table 1 for separate p-values for each estimate.

### Figure 2. Salary development for exposure and control groups across time of exposure

Note: Figure shows the salary trajectories for the exposure group (black) who suffers concussion at year zero against normalized wage trajectories for the control groups who suffer their concussions one to five years later.  $\Delta$  indicates the number of years between exposure and control incident. Table shows that there are no significant differences in the normalized salary levels for exposure and control group prior to exposure incident (see Figure S1, Supplemental Digital Content 3 for unnormalized salary trajectories).

**Figure 3. (Left panel) The cumulative density function (cdf) for salary post-treatment among the treatment group and their counterfactual outcome had they not experienced their concussions, and (Right panel) the change in salary density for the exposure group compared to their counterfactual baseline expressed as the effect of concussion on the probability of earning below the salary-level expressed on the x-axis following exposure**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

465 **event.**

466 Note: The figure shows the observed cumulative salary distribution following concussion for the  
467 exposure group (red) and the expected counterfactual salary distribution absent suffering  
468 concussion in the exposure group (blue), when using the  $\Delta=5$  control group. The black line  
469 shows the difference between the observed and the counterfactual distribution, and the grey  
470 dash lines show the 95 % confidence interval. The close to constant decline of the difference  
471 between the two distributions as the salary increase indicates that the main part of the effect of  
472 concussions on salary are driven by people having a salary equal to zero.

473 **Figure 4. Effect of concussion on salary across incident years and control groups**  
474 **together with the percentage fulltime unemployed of the labor force.**

475 Note: Figure shows annual estimates of concussion against each control group separately  
476 mapped against the share of the labor force that is full time unemployed. 95 % confidence  
477 intervals. The estimates for the effect of concussion on salary almost uniformly increase in  
478 absolute magnitude when unemployment decreases, and decrease when unemployment  
479 increase, indicating that the effect of concussion on salary is countercyclical to the economic  
480 cycle.

481

482 **Supplemental Digital Content titles & legends**

483 **Supplemental Digital Content [#1].** Text file. Supplemental materials and methods. This file  
484 contains further details on our quasi-experimental, difference-in-differences event time  
485 approach.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

486 **Supplemental Digital Content [#2].** File. Code used for the analyses.

487

488 **Supplemental Digital Content [#3].** Table. Supplemental results Table S1: Test of parallel  
489 trends assumption pre-exposure incident against each control group separately using eq. S3 in  
490 supplementary methods. Separate exposure dummies for all time periods (except the year prior  
491 to exposure, which serves as reference period).

492

493 **Supplemental Digital Content [#3].** Table. Supplemental results Table S2: Effect of  
494 concussion on different labor market outcome parameters using separate exposure dummies  
495 for all time periods (except the year prior to exposure, which serves as reference period): In this  
496 exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers  
497 a concussion five years after the exposure group. Outcomes include annual salaried income  
498 (annual salary), total annual income (total income), annual sick leave benefits received (sick  
499 leave benefits) as well as a binary indicator of employment (probability of employment).  
500 Monetary outcomes were measured at 2015-level in € 1,000.

501

502 **Supplemental Digital Content [#3].** Table. Supplemental results Table S3: Demographic  
503 factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years  
504 leading up to the concussion event in each of the groups. Factors include patient age (in years),  
505 share of sample female (1=100% female), and share of individuals with at least a high school  
506 degree (1=100%).

507

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S1: Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S2: The Cumulative Density Function (CDF) for Total Income Post-Treatment among the Treatment Group and Their Counterfactual, and the Difference between the Two CDFs Expressed as the Effect of Concussion on the Probability of Total Income Below that Income-Level following Exposure Event.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S3: Percentage Effect of Concussion on Relative Salary Across Age Groups.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S4: Percentage Effect of Concussion on Relative Salary Across High School Completion.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S5: Percentage Effect of Concussion on Relative Salary Across Gender.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S6: Effect of Concussion on Absolute Salary in 1K Euro Across Age groups.

1  
2  
3  
4 529 **Supplemental Digital Content [#3].** Figure. Supplemental results Figure S7: Effect of  
5  
6  
7 530 Concussion on Absolute Salary in 1K Euro Across Education.

8  
9 531  
10  
11 532 **Supplemental Digital Content [#3].** Figure. Supplemental results Figure S8: Effect of  
12  
13  
14 533 Concussion on Absolute Salary in 1K Euro Across Gender.

15  
16 534  
17  
18  
19 535 **Table 1. Number of observations for exposure and control groups across time since**  
20  
21 536 **exposure and number of observations with missing salary information**

22 537

Years until exposure	Exposure					
	group	Control $\Delta=1$	Control $\Delta=2$	Control $\Delta=3$	Control $\Delta=4$	Control $\Delta=5$
-4	36,804	33,681	31,112	29,190	27,859	26,794
-3	36,978	33,834	31,245	29,366	27,973	26,907
-2	37,195	34,003	31,407	29,501	28,146	27,031
-1	37,449	34,224	31,582	29,687	28,288	27,220
0	37,848	34,551	31,851	29,922	28,530	27,421
1	37,467		31,755	29,832	28,433	27,337
2	36,940			29,807	28,421	27,295
3	36,484				28,421	27,304
4	36,084					27,314
Total observations	333,249	170,293	188,952	207,305	226,071	244,623
Observations with missing salary	81	32	31	44	35	29

23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41 538  
42 539 Control groups have not suffered a concussion in 10+ $\Delta$  years before incident, exposure group  
43  
44  
45 540 has not suffered concussion the 10 years before exposure incident.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 2. Estimated effect of concussion on salary of exposure group compared to control groups that suffered their concussion  $\Delta = 1, 2, 3, 4, 5$  y after the exposure group's concussion event, measured at 2015-level.  $N_{\text{Exposure}}$ : 37,848**

	<b>Estimated salary effect (<math>\delta</math>)</b>	<b>95 % CI</b>	<b>p</b>	<b><math>N_{\text{Control}}</math></b>
$\Delta = 1$ y	-423€	(-919€;73€)	.095	34,551
$\Delta = 2$ y	-825€	(-1,108€; -543€)	<.001	31,851
$\Delta = 3$ y	-1,019€	(-1,331€; -707€)	<.001	29,922
$\Delta = 4$ y	-1,126€	(-1,446€; -805€)	<.001	28,530
$\Delta = 5$ y	-1,243€	(-1,564€; -922€)	<.001	27,421

Results obtained from estimations following Eq. (1). Models include controls for high school diploma, gender, age, and observation year. Results obtained using reghdfe in Stata.

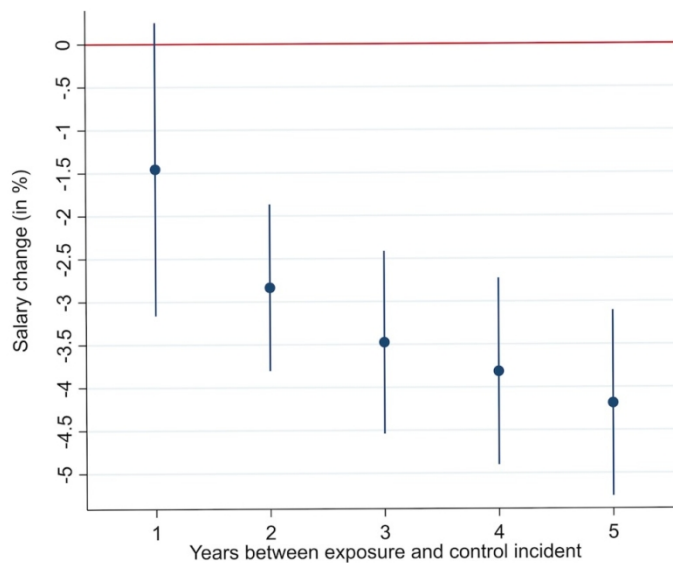


Figure 1

338x190mm (108 x 108 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

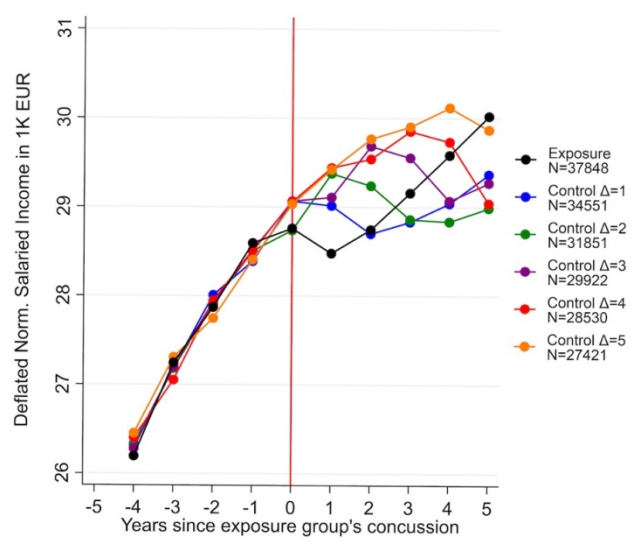


Figure 2

338x190mm (108 x 108 DPI)

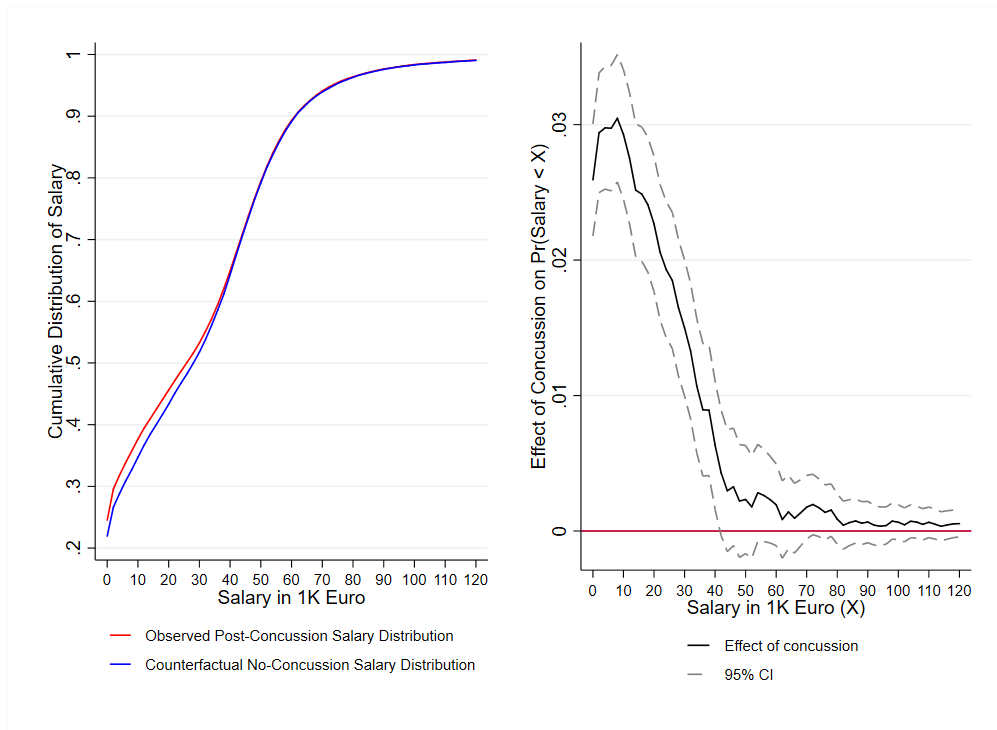


Figure 3

352x256mm (72 x 72 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

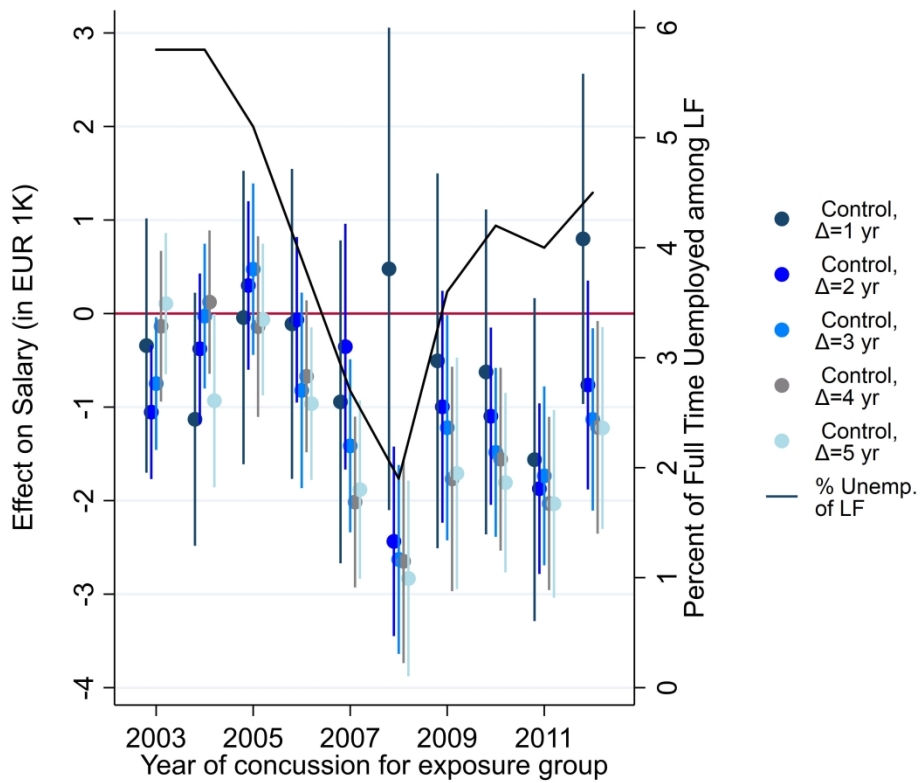


Figure 4



## SUPPLEMENTAL MATERIALS AND METHODS

Our quasi-experimental, difference-in-differences event time approach compares two groups of individuals from the same cohort, where both groups experience concussions, but at two different time points ( $t_c, t_c + \Delta$ ). For the simple situation where we have three periods ( $t=0, 1, 2$ ) and the exposure group (T) experiences their concussion at the start of period 1 ( $t_c=1$ ), and the control group (C) at the start of period 2 ( $t_c + \Delta=2$ ), the effect of concussion on salary (Y) is:

$$\Delta = (Y_1^T - Y_1^C) - (Y_0^T - Y_0^C)$$

The effect of concussion on salary in  $t=1$  is estimated by comparing the average difference in salary between exposure and control groups for the post-concussion period  $t=1$  ( $Y_1^T - Y_1^C$ ) to the average difference in salary for the pre-concussion, or baseline, interval  $t=0$  ( $Y_0^T - Y_0^C$ ). Assuming the exact timing of a concussion is random for small enough sizes of  $\Delta$ , and under the additional assumption that the exposure group would have had parallel trends in salary as the control group absent suffering concussion at  $t_c$ ,  $\delta$  captures the causal effect of concussion among those who suffer concussions – also known as the average effect on the treated (AT). The AT does not capture how concussions would affect a random person. The AT captures how concussions causally affect those who suffer concussions.

For our study, the parallel trends assumption states that exposure and control groups have parallel developments in salary leading up to the exposure group's concussion and the exposure and control groups would have further exhibited parallel salary trajectories if the concussion had not occurred. To test the parallel trends assumption, we estimate a dynamic version of the model specification (shown in supplementary table S1), which explicitly allows us to test whether the parallel trend assumption for our sample is probable.

To validate that the timing of concussion is random with our study period, we present estimates for effect of exposure across different periods between exposure and control incident ( $\Delta$ ). Most recorded concussions outside contact sports and military engagements stem from unforeseen events, such as falls or striking/being struck by an object<sup>25,26</sup>, so assuming random timing is likely valid. People who regular engage in activities that result in high risk of multiple concussions may be different than the average concussion patient and would be more likely to end up in the exposure sample than in the control sample, which could induce bias. To avoid such potential bias, we restrict our sample to individuals without prior diagnoses for intracranial injuries ten years prior to exposure.

At  $t=-1$ , i.e. one year before the exposure group suffered a concussion, the control groups were slightly smaller than the exposure group, and two control groups ( $\Delta=4$  and 5) differed slightly but significantly in terms of average patient age ( $p < .001$ ; supplementary table S2), male to female ratio ( $p < .001$ ), and for control group  $\Delta=5$ , in the frequency of individuals with at least a high school degree ( $p < .001$ ). However, the differences are numerically small. To test that composition differences between exposure and control do not drive our results, we provide separate results for individuals with and without high school degree, for males and females, and for different age groups across all different values of  $\Delta$ .

Further, our design inherently leads to the possibility of timing issues—our exposure group always suffers their concussion earlier (in terms of calendar time and age) than the control groups do. If the labor market is constantly improving or worsening during the period we consider, this could substantially influence our results. Therefore, we also estimate separate models across exposure incident year and control group. Estimating separate models allow us the added benefit of being able to examine whether the business cycle influences the effect of concussions on salary.

### Statistical model

To estimate the impact of concussion on salary, we define the following variables: Exposure or control group  $g$ , which includes individuals  $i$ , at times  $t$  to exposure-groups concussion incident  $t$ . First, we estimate a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  using ordinary least squares:

$$\text{Salary}_{git} = \beta_0 + \gamma \text{exposure}_g + \theta \text{post}_t + \delta \text{post}_t \times \text{exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{\text{Age}=26}^{48+\Delta} I(\text{Age}) \eta_{\text{age}} + \sum_{\text{year}=1999}^{2012} I(\text{year}) \eta_{\text{year}} + \epsilon_{git} \quad (\text{S1})$$

where  $\text{Salary}_{git}$  measures annual salaried income deflated to 2015-level,  $\text{exposure}_g$  indicates whether the observation belongs to the exposure or control group,  $\text{post}_t$  captures the period after the exposure group's concussion occurred, and  $\text{post}_t \times \text{exposure}_{git}$  captures the effect concussion, measured as share of year  $t \geq 0$  affected by concussion. In this way, someone who suffers a concussion July 1 has  $\text{post}_t \times \text{exposure}_{git} = 0.5$  for  $t = 0$  and  $\text{post}_t \times \text{exposure}_{git} = 1$  for  $t > 0$ .  $\mathbf{X}_i$  is a set of covariates that includes a high school indicator and a gender dummy,  $\epsilon_{git}$  is the error-term, and the two last sets of indicator variables  $I(\text{Age})$  and  $I(\text{Year})$  capture age and incident year levels (control group indexed against incident year). Under the parallel trends assumption,  $\delta$  then captures the annual effect of concussion on salary. In eq. 1,  $\text{exposure}_g$  normalizes any pre-exposure differences between the exposure and control group, thereby creating a joint baseline pre-exposure.

We estimate robust individual-level clustered standard errors to account for the possibility that individuals enter the data twice both as control (0) and exposure (1) individuals ( $g = \{0, 1\}$ ), and that they are observed for multiple periods ( $t = \{-4, \dots, \Delta - 1\}$ ). To calculate the relative salary decrease after concussion, we exploit the parallel trends assumption to generate the expected counterfactual salary level, i.e. had the concussion not occurred, and calculate the decline expressed in percentage as: % change =  $\delta / E(\widehat{Salary}_{git} | g = 1, post_t = 1, post_t \times exposure_{git} = 0)$ . In this way, we provide both absolute estimates measured in 1K Euro, as well as percentage change.

We expect  $\delta$  from eq. (1) to likely be negative. Yet, a decrease in annual salary can arrive through two different channels. Concussions may affect salary through either decreasing income among those employed or by reducing the number of individuals who are employed and earning any salary at. To parse out which of the two channels is driving the results, we examine how concussion affects the salary distribution among the exposure group following. Following Chernozhukov et al.<sup>27</sup> we estimate a series of regressions across the whole salary distribution, where, for a finite set of points, we predict how concussion affects the likelihood of having earnings on the left side of each finite point, as follows:

$$\sum_{j=0}^{\max(Salary)} p_j = \beta_{0j} + \delta_j post_t \times exposure_{git} + \theta post_t + \gamma_j exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age,j} + \sum_{year=1999}^{2012} I(year) \eta_{year,j} + \epsilon_{git,j} \quad (S2)$$

where  $p_j = \Pr(Salary_{git} \leq j)$  and  $j$  is the interval from 0 to  $\max(Salary)$ . Across the salary distribution, we can now predict the probability of earning less than  $j$  for those with and without concussions. From equation 2, we predict  $p_j^1 = E(p_j | post_t \times exposure_{git} = 1, exposure_g = 1, t \geq 0)$  and the counterfactual  $p_j^0 = E(p_j | post_t \times exposure_{git} = 0, exposure_g = 1, t \geq 0)$ . Plotting  $p_j^1$  and  $p_j^0$  over each value of salary  $j$ , and assuming rank stability, gives the cumulative density function of salary for the treated ( $p_j^1$ ) and the counterfactual observation of the treated had they not suffered concussions ( $p_j^0$ ). The difference between  $p_j^1$  and  $p_j^0$  is simply  $\delta_j$ . If the value of  $\delta_j$  monotonically moves towards zero as  $j$  increases until  $p_j^1 \approx p_j^0 \approx 1$  it indicates that exit from employment fully drives the effect of concussion on salary. If instead the value of  $\delta_j$  is constant or increasing across parts of the distribution, it instead indicates that a decrease in salary among those still receiving salary drives at least part of the effect.

Eq. 1 and eq. 2 are based on the parallel trends assumption. The assumption states that exposure and control groups follow parallel salary trajectories until individuals in the exposure group experiences a concussion, and that the parallel trends would have continued had the concussion not occurred. Whereas we cannot verify the counterfactual situation of parallel trends after exposure, we can use a dynamic model to test for systematic differences in salary trends between exposure and control group in the years leading up to the exposure group's concussion event. To do so, we estimate the following dynamic model:

$$Salary_{git} = \beta_0 + \sum_{t \neq -1, t = -4}^{\Delta-1} \delta_t \times I(t_g) \times exposure_g + \sum_{t = -4}^{\Delta-1} I(t_g) \eta_t + \gamma exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age} + \sum_{year=1999}^{2012} I(year) \eta_{year} + \epsilon_{git} \quad (S3)$$

Where we interact exposure group status ( $exposure_g$ ) with indicators  $I(t_g)$  capturing time from concussion. If the parallel trends assumption holds, then it must be the case  $\{\delta_{-4}, \delta_{-3}, \delta_{-2}\} = 0$ , whereas the size and sign of  $\{\delta_0, \dots, \delta_{\Delta-1}\}$  captures the dynamic effect of a concussion from the year of incidence and  $\Delta-1$  years onward. By estimating the effect of concussion on salary among different years of the study period, we are also able to capture how the impact of concussion on salary evolves year to year after the concussion has occurred. We further estimate eq. 3 for a series of related labor market outcomes (annual total income, annual amount of sickness benefits received, annual probability of being employed), to generate a more thorough understanding on how concussions affect labor market outcomes—i.e., if people experience a decrease in salary due to a concussion, are they then compensated through different types of welfare state services.

25 Faul M, Coronado V. Epidemiology of traumatic brain injury. *Handbook of clinical neurology* **127**:3-13, 2015

26 Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic Brain Injury–Related Emergency Department Visits, Hospitalizations, and Deaths — United States, 2007 and 2013. *MMWR Surveillance Summaries* **66**:1-16, 2017

27 Chernozhukov V, Fernández-Val I, Melly B. Inference on Counterfactual Distributions. *Econometrica* **81**:2205-2268, 2013

1  
2  
3 \*\*\*\*\*  
4 \*\*\*\*\*  
5 \*\*\*\*\*  
6 \*\*\*\*\*  
7 \*\*\*\*\*  
8 \*\*\*\*\*  
9 \*\*\*\*\*

10 \*\*\*\*\*

11 \*\*\*\*\* This study relies on restricted individual level  
12 administrative data obtained  
13 \*\*\*\*\* from Statistics Denmark.

14 \*\*\*\*\*

15 \*\*\*\*\* Only Danish research environments are granted authorization to  
16 access data from Statistics Denmark. \*\*\*\*\* Foreign  
17 \*\*\*\*\* researchers can, however, get access to micro data through an  
18 affiliation to a Danish authorized  
19 \*\*\*\*\* environment.

20 \*\*\*\*\* Access is given to anonymized micro data, i.e. data at an  
21 individual personal or corporate level. \*\*\*\*\* Access takes  
22 \*\*\*\*\* place through researcherís own pc over the Internet.

23 \*\*\*\*\*

24 \*\*\*\*\* See <https://www.dst.dk/en/TilSalg/Forskningservice> for detail  
25 \*\*\*\*\*

26 \*\*\*\*\* For the replication of present study, contact the ROCKWOOL  
27 Foundation for access.

28 \*\*\*\*\* <http://www.rockwoolfonden.dk/en/>

29 \*\*\*\*\*

30 \*\*\*\*\*  
31 \*\*\*\*\*  
32 \*\*\*\*\*  
33 \*\*\*\*\*  
34 \*\*\*\*\*  
35 \*\*\*\*\*  
36 \*\*\*\*\*  
37 \*\*\*\*\*

38 \*\*\*\*\*  
39 \*\*\*\*\*  
40 \*\*\*\*\*  
41 \*\*\*\*\*  
42 \*\*\*\*\*  
43 \*\*\*\*\*  
44 \*\*\*\*\*  
45 \*\*\*\*\*  
46 \*\*\*\*\*

47 \*\*\*\*\*

48 \*\*\*\*\* This is the master data do-file for the full package of do-  
49 files that generates the results for  
50 \*\*\*\*\* Fallesen and Campos (2020). Execute files in the order they  
51 are listed. For each file, specify a \*\*\*\*\* home registry in place of  
52 [home]. Further, generate subfolders [home]/data, [home]/tables, and  
53 \*\*\*\*\* [home]/highdef to capture auxiliary data sets, tables, and  
54 figures.

55 \*\*\*\*\*

56 \*\*\*\*\*  
57 \*\*\*\*\*  
58 \*\*\*\*\*  
59 \*\*\*\*\*  
60 \*\*\*\*\*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

```
*****  
*****
```

```
***Install user-written reghdfe command for faster computation of  
regressions  
ssc install reghdfe
```

```
***Generate concussion samples and merges on covariates and outcome  
do 01generate_sample.do
```

```
***Generates auxiliary data sets the includes social benefit  
reciprocity indicator and levels of benefits received  
do 02generate_benefits.do
```

```
***Generates auxiliary data sets the includes social benefit  
reciprocity indicator and levels of benefits received  
do 02generate_benefits.do
```

```
***Generate results  
do 03generate_figures_and_results
```

```
exit, clear
```

For peer review only

```
1
2
3 clear all
4
```

```
5 *****
6 *****
7 **
8 **          This program builds data for Fallesen & Campos
9 (2020)
10 **          study of concussion's impact on productivity
11 measured
12 **          through annual salary
13 **
14 **
15 *****
16 *****
```

```
17
18
19 **Global for path to registry data
20 global dorg "E:/data/rawdata/706630"
```

```
21
22 *Global for processed data
23 global data "[home]/data"
```

```
24
25 /*globals for price index to calculate income at 2015-level across
26 years*/
27 /*Price index obtained from www.dst.dk/en/statistik/emner/priser-og-
28 forbrug/forbrugeriser/nettoprisindeks */
29
```

```
30 {
31
32 global price1980 = .358
33 global price1981 = .398
34 global price1982 = .439
35 global price1983 = .466
36 global price1984 = .494
37 global price1985 = .517
38 global price1986 = .521
39 global price1987 = .537
40 global price1988 = .564
41 global price1989 = .594
42 global price1990 = .612
43 global price1991 = .628
44 global price1992 = .642
45 global price1993 = .651
46 global price1994 = .662
47 global price1995 = .674
48 global price1996 = .688
49 global price1997 = .703
50 global price1998 = .713
51 global price1999 = .728
52 global price2000 = .751
53 global price2001 = .769
54 global price2002 = .788
55 global price2003 = .806
56 global price2004 = .817
57 global price2005 = .833
58 global price2006 = .850
59
60
```

```

1
2
3     global price2007 = .867
4     global price2008 = .899
5     global price2009 = .917
6     global price2010 = .936
7     global price2011 = .960
8     global price2012 = .978
9     global price2013 = .986
10    global price2014 = .994
11    global price2015 =1.00
12    global price2016 =1.005
13    global price2017 =1.017
14
15 }
16
17
18 /*Locate concussions and other TBIs from the Danish National Patient
19 Registry */
20 /**/
21 forvalue t = 1977/2017{
22     if `t' < 1994 use $dorg/lpr_diag`t'.dta /// **uses ICD-8
23 codes until Dec. 31, 1993
24         if substr(c_diag,1,2)=="85"
25     if `t' > 1993 use $dorg/lpr_diag`t'.dta /// **uses ICD-10
26 codes from Jan. 1, 1994
27         if substr(c_diag,1,4)=="DS06"
28     **recovers encrypted social security number and admittance
29 date
30     merge m:m recnum using $dorg/lpr_adm`t', keepus(pnr d_ind*)
31 keep(3)
32     drop _merge recnum
33
34
35     **Keeps diagnosis, diagnosis type, and individual id (pnr)
36     keep pnr c_diag c_diagtype pnr d_ind
37
38
39     **generate year variable
40     gen year = year(d_ind)
41
42     **geenerate share of year with concussion
43     gen time_from_incident = 1-((d_ind-mdy(1,1,year(d_ind))))/
44 365)
45     drop d_ind
46
47
48     *save as one dataset
49     if `t' > 1977 append using $data/concussion.dta
50     if `t' == 2017 sort pnr year
51     save $data/concussion.dta, replace
52
53
54 }
55 */
56 /*****
57 Sets up datasets for treatment group (x = 0)
58 and the control groups who suffer concussion
59 1, 2, 3, 4, 5 years later (x = 1 2 3 4 5).
60

```

1  
2  
3  
4 Treatment group are not allowed to have suffered  
5 any type of TBI the last 10 years before concussion  
6 , control group are not allowed to have suffered any TBI  
7 concussion the last 10 + x years.  
8

```
9 *****/
10 /**/
11 forvalue x = 0/5{
12     use $data/concussion.dta, clear
13     sort pnr year
14
15     **generate measure of length between registered TBIs
16     by pnr: gen help = year-year[_n-1]
17
18     drop if help < 10+`x'
19
20     **Keep only concussion, and only when it was primary
21     diagnosis
22     keep if c_diagtype== "A" & /// Primary daignosis
23         (c_diag == "85099" | /// ICD-8 code for concussion
24         substr(c_diag,1,5) == "DS060") // ICD-10 code for
25     concussion
26
27     *generate treatment and control datasets
28     save $data/concussion_`x'.dta, replace
29
30 }
31
32
33
34
35
```

```
36 */
37 /*****
38 Generate datasets for analysis. First incident year is allowed
39 to be 1992, because it is the first year where we have full record
40 for the five year plus control group (1977+10+5 = 1992).
41
```

```
42 We generate seperate datasets for each incident year for treatment
43 group and different control groups.
44
```

```
45 *****/
46
47 forvalue time = 1/5{           //for different control groups
48
49     local post_period = 5     // local for number of years
50     observed post
51
52                                     //
53     concussion for treatment group
54
55     local endtime = 2017-`post_period'    /*last year where
56     we allow for treatment event
57
58                                     to
59     occur, in order to have long enough control
60                                     period.
```

```

1
2
3   Defined by latest year available data*/
4
5       forvalue count = 1992(1)`endtime'{
6           local t = `count'                // for
7   ease of coding
8           local n = `t'-4                  // first
9   pre-treatment event period
10          local c = `t'+`post_period'      //last post-event
11   period
12          local w = `t'+`time'            //time of
13   concussion for control
14
15          use pnr alder using $dorg/bef`t'  ///
16   Bring in all 30-49 yr olds
17          if inrange(alder`t',20,59), clear // from
18   the population register
19
20
21          **year variable
22          gen year = `t'
23
24          **limit sample to those who suffer a concussion in
25   `t'
26          merge 1:1 pnr year using $data/
27   concussion_0.dta, ///
28          keep(3) nogen
29
30          forvalue x=`n'/'c'{              //add longitudinal data
31              merge 1:1 pnr using $dorg/bef`x', ///
32              keep(1 3) keepus(efalle alder
33   koen) //add information on spouse,
34
35              //age, and gender
36              rename _merge merge`x' //indicator for
37   whether in DK that year
38
39
40              **Add salary information and ses
41   information
42              if `x' < 2017{
43                  merge 1:m pnr using $dorg/
44   ind`x', ///      m:1 to account for duplicates
45                  nogen keep(1 3)
46   keepus(erhvervsindk_13 pre_socio personindk dispon_13
47   aekvivadisp_13) // in data on non-important variables
48
49
50                  bysort _all: keep if _n ==1
51                  //drop perfect duplicates
52              **Align variable names and account
53   for inflation
54
55                  rename erhvervsindk_13 loenmv
56                  rename pre_socio pre_socio`x'
57
58                  foreach kk in personindk dispon_13
59   aekvivadisp_13 loenmv{
60

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

                                rename `kk'
`kk'\`x'
                                }
                                foreach kk in personindk dispon_13
aekvivadis_13 loenmv{
                                replace `kk'\`x' =
                                `kk'\`x'/{price`x'}
                                }
                                **Bring in educational information
                                merge 1:1 pnr using $dorg/
uddany`x', ///
                                nogen keep(1 3) keepus(hffsp)
                                }
                                if `x' == 2017{
                                merge 1:m pnr using $dorg/
ind`x', ///      m:1 to account for duplicates
                                nogen keep(1 3)
keepus(erhvervsindk_13 pre_socio personindk)      // in data on
non-important variables
                                bysort _all: keep if _n ==1
                                //drop perfect duplicates
**Align variable names and account
for inflation
                                rename erhvervsindk_13 loenmv
                                rename pre_socio pre_socio`x'
                                foreach kk in personindk loenmv{
                                rename `kk'
`kk'\`x'
                                }
                                foreach kk in personindk loenmv{
                                replace `kk'\`x' =
                                `kk'\`x'/{price`x'}
                                }
                                **Bring in educational information
                                merge 1:1 pnr using $dorg/
udda`x', ///
                                nogen keep(1 3)
keepus(hfaudd)
                                rename hfaudd hfaudd`x'
                                }
                                }
                                **Reshape data to panel structure
                                if `count'>= 2012 reshape long efalle alder koen
loenmv pre_socio personindk dispon_13 aekvivadis_13 hffsp merge
hfaudd , i(pnr) j(t)

```

```

1
2
3           if `count' < 2012 reshape long efalle alder koen
4 loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge ,
5 i(pnr) j(t)
6
7           gen count = t-year           //variable for time to
8 concussion
9           gen treatment =1 //treatment group indicator
10
11           save $data/sample_temp.dta, replace //temporary
12 dataset
13
14           /
15 *****
16           Now build control sample for time `time' and year
17 `count'
18 *****/
19
20
21
22
23
24           use pnr alder using $dorg/bef`t'           ///
25 Bring in all 30-49 yr olds
26           if inrange(alder`t',20,59), clear // from
27 the population register
28
29           **year variable
30           gen year = `w' //time for concussion for control
31 group `time'
32
33           **limit sample to those who suffer a concussion in
34 `t'
35           merge 1:1 pnr year using $data/
36 concussion_`time'.dta, ///
37           keep(3) nogen
38
39           forvalue x=`n'/'c'{           //add
40 longitudinal data
41           merge 1:1 pnr using $dorg/
42 bef`x', ///
43           keep(1 3) keepus(efalle
44 alder koen) //add information on spouse,
45
46           //age, and gender
47           rename _merge merge`x' //indicator
48 for whether in DK that year
49           if `x' < 2017{
50           **Add salary and SES
51 information
52           merge 1:m pnr using
53 $dorg/ind`x', ///           1:m to account for duplicates
54           nogen keep(1 3)
55 keepus(erhvervsindk_13 pre_socio personindk dispon_13
56 aekvivadisp_13) // in data on non-important variables
57
58
59
60

```

```

1
2
3
4 ==1 bysort _all: keep if _n
5 duplicates //drop perfect
6
7
8 **Align variable names
9 and account for inflation
10
11 rename erhvervsindk_13
12 loenmv
13 rename pre_socio
14 pre_socio`x'
15
16 foreach kk in personindk
17 dispon_13 aekvivadisp_13 loenmv{
18 rename `kk'
19 `kk'`x'
20 }
21
22 foreach kk in personindk
23 dispon_13 aekvivadisp_13 loenmv{
24 replace `kk'`x' =
25 `kk'`x'/{price`x'}
26 }
27
28
29 **Bring in educational
30 information
31 merge 1:1 pnr using
32 $dorg/uddany`x', ///
33 nogen keep(1 3)
34 keepus(hffsp)
35 }
36 if `x' == 2017{
37 merge 1:m pnr using $dorg/
38 ind`x', /// m:1 to account for duplicates
39 nogen keep(1 3)
40 keepus(erhvervsindk_13 pre_socio personindk) // in data on
41 non-important variables
42
43 bysort _all: keep if _n ==1
44 //drop perfect duplicates
45 **Align variable names and account
46 for inflation
47
48 rename erhvervsindk_13 loenmv
49 rename pre_socio pre_socio`x'
50
51 foreach kk in personindk loenmv{
52 rename `kk'
53 `kk'`x'
54 }
55
56 foreach kk in personindk loenmv{
57
58
59 replace `kk'`x' =
60

```

```

1
2
3      `kk`x'/{price`x'}
4
5      }
6
7      **Bring in educational information
8      merge 1:1 pnr using $dorg/
9
10     udda`x', ///
11
12     keepus(hfaudd)
13
14     }
15
16     **Reshape data to panel structure
17     if `count' >= 2012 reshape long efalle alder koen
18     loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge
19     hfaudd , i(pnr) j(t)
20     if `count' < 2012 reshape long efalle alder koen
21     loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge ,
22     i(pnr) j(t)
23     gen count = t-`t' //variable for
24     time to concussion for treatment
25     gen control`time' =1 //control indicator
26     save $data/control_temp, replace
27
28     **Build sample with treatment and control `time'
29     for year `count'
30
31     use $data/sample_temp
32     append using $data/control_temp
33
34     **fixes control and treatment indicators
35     replace control`time' = 0 if control`time'==.
36     replace treatment = 0 if treatment==.
37
38     **Picks up changes to education variable
39     if `count' >=2012{
40         tostring hfaudd, replace
41         rename hfaudd start
42         merge m:1 start using "\
43     \srvfsenas1\data\Formater\SAS formater i Danmarks
44     Statistik\STATA_datsaet\Disced\c_udd_niveau_l1l2_k.dta" , nogen
45     keep(1 3)
46
47         destring UDD, replace force
48         **Replace all with high school degree or
49     higher in HFAUDD to have HFFSP = 40000001
50         replace hffsp = 40000001 if t == 2017 &
51     inrange(UDD,30,80)
52         replace hffsp = 0 if t == 2017 & !
53     inrange(UDD,30,80)
54         drop UDD start
55     }
56
57     sort pnr t
58
59
60

```

```
1
2
3
4         save $data/sample_control_`count'_`time'.dta,
5 replace
6
7     }
8 }
9
10
11 forvalue time = 1/5{
12     forvalue count =2003/2012{
13         if `time' ==1 & `count' ==2003 use $data/
14 sample_control_`count'_`time'.dta, clear
15         else append using $data/
16 sample_control_`count'_`time'.dta
17
18         if `time' ==5 & `count' ==2012 bysort pnr: keep if
19 _n ==1
20         if `time' ==5 & `count' ==2012 count
21     }
22 }
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

Peer review only

```

1
2
3 clear
4
5 *****
6 *****
7 *****
8 *****
9 **
10 **          Calculate share of year on public benefits and size
11 of benefit
12 **          payments for Fallesen and Campos (2020)
13 **
14 **
15 **
16 *****
17 *****
18 *****
19 *****
20 /*globals for price index to calculate income at 2015-level across
21 years*/
22 /*Price index obtained from www.dst.dk/en/statistik/emner/priser-og-
23 forbrug/forbrugeriser/nettoprisindeks */
24
25 global price1980 = .358
26 global price1981 = .398
27 global price1982 = .439
28 global price1983 = .466
29 global price1984 = .494
30 global price1985 = .517
31 global price1986 = .521
32 global price1987 = .537
33 global price1988 = .564
34 global price1989 = .594
35 global price1990 = .612
36 global price1991 = .628
37 global price1992 = .642
38 global price1993 = .651
39 global price1994 = .662
40 global price1995 = .674
41 global price1996 = .688
42 global price1997 = .703
43 global price1998 = .713
44 global price1999 = .728
45 global price2000 = .751
46 global price2001 = .769
47 global price2002 = .788
48 global price2003 = .806
49 global price2004 = .817
50 global price2005 = .833
51 global price2006 = .850
52 global price2007 = .867
53 global price2008 = .899
54 global price2009 = .917
55 global price2010 = .936
56 global price2011 = .960
57 global price2012 = .978
58
59
60

```

```

1
2
3 global price2013 = .986
4 global price2014 = .994
5 global price2015 =1.00
6 global price2016 =1.005
7 global price2017 =1.017
8
9
10
11 **Global for path to registry data
12 global dorg "e:/data/rawdata/706630"
13 *Global for processed data
14 global data "E:/data/workdata/706630/pf/FallesenCampos/data"
15
16 forvalue t=1996/2017{
17
18     ** Read in data on social benefits reciprocity share of
19     weeks
20     ** from the DREAM database
21     use $dorg/dream`t'
22     gen share =0
23     forvalue y = 1/52{
24         if `y' < 10 replace share = share+1 if
25         y_0`y' !=.
26         if `y' > 9 replace share = share+1 if
27         y_`y' !=.
28         if `y' < 10 drop y_0`y'
29         if `y' > 9 drop y_`y'
30
31     }
32     **Generate annual measure of share of year receiving social
33     benefits
34     replace share = share/52
35     keep pnr share
36     gen t = `t'
37     if `t' > 1996 append using $data/temp.dta
38     save $data/temp.dta, replace
39
40 }
41
42
43
44 forvalue t=1998/2017{
45
46     **Read in information on size of different types of social
47     benefits
48
49     if `t' < 2002{
50         use pnr syg_barsel_13 konthj arblhum pre_socio
51         using $dorg/ind`t'.dta, clear
52         replace syg_barsel_13 = syg_barsel_13/{price`t'}
53         replace konthj = konthj /{price`t'}
54         replace arblhum = arblhum/{price`t'}
55         gen kont_dag = konthj+arblhum
56         drop konthj arblhum
57     }
58     if `t' >= 2002 & `t' < 2013{
59
60

```

```
1
2
3       use pnr syg_barsel_13 adagpagn konthj arblhum
4 pre_socio using $dorg/ind`t'.dta, clear
5       replace syg_barsel_13 = syg_barsel_13/{price`t'}
6       replace adagpagn = adagpagn/{price`t'}
7       replace konthj = konthj /{price`t'}
8       replace arblhum = arblhum/{price`t'}
9       gen kont_dag = konthj+arblhum
10      drop konthj arblhum
11    }
12  if `t' >= 2013{
13      use pnr syg_barsel_13 adagpagn dagpenge_kontant_13
14 pre_socio using $dorg/ind`t'.dta, clear
15      replace syg_barsel_13 = syg_barsel_13/{price`t'}
16      replace adagpagn = adagpagn/{price`t'}
17      replace dagpenge_kontant_13 =
18 dagpenge_kontant_13 /{price`t'}
19      gen kont_dag = dagpenge_kontant_13-syg_barsel_13
20      drop dagpenge_kontant_13
21    }
22  gen t = `t'
23  compress
24  bysort pnr: keep if _n ==1
25  if `t' > 1998 append using $data/temp2.dta
26  if `t' == 2017{
27      sort pnr t
28    }
29  save $data/temp2.dta, replace
30 }
31 }
32 }
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```



```

1
2
3 clear all
4
5 *****
6 *****
7 **
8 **
9 **          This program geenrates figerus and results for
10 Fallesen & Campos (2020)
11 **          study of concussion's impact on productivity
12 measured through
13 **          annual salary
14 **
15 *****
16 *****
17
18 **Global for path to registry data
19 global dorg "E:/data/rawdata/706630"
20
21 *Global for processed data
22 global data "[home]/data"
23
24 *Global on figures
25 global highdef "[home]\highdef"
26
27
28
29
30 forvalue time = 1/5{          //for different control groups
31
32     local post_period = 5    // local for number of years
33     observed post
34
35     concussion for exposure group          //
36
37     local endtime = 2017-`post_period'    /*last year where
38 we allow for exposure event
39
40 occur, in order to have long enough control          to
41
42 Defined by latest year available data*/          period.
43
44     **Matrixes to capture estimates
45     matrix results = J(15,6,.)    // For salary estimates
46     matrix results_p = J(15,6,.)  // For Pr(salary=0)
47 estimates
48     matrix t = J(15,6,.)    // For time indicators
49
50     forvalue count = 2003(1)`endtime'{
51         use $data/sample_control_`count'_'time', clear
52
53         gen female = koen==2
54
55         qui{
56             gen edu =0
57             replace edu = 1 if inrange(hffsp,
58 20000000,39000000) | ///
59
60

```

```

1
2
3
4     (hffsp >40000000 & hffsp!=.)
5         }
6
7         **exclude individuals in years where they do not
8 appear in data,
9         **due to either death or migration, as well as
10 periods from when
11         **the control group sufer their concussion
12 drop if merge ==1 | count > `time'-1
13
14         **generate concussion variable
15 gen treat = inrange(count,0,`time'-1) & treatment
16 ==1
17
18         replace treat = time_from_incident if count ==0 &
19 treatment ==1
20
21         **Generate pre-concussion income difference
22 for
23         **use in calculating marginal effects
24 sum loenmv if count <0 & treatment ==0
25 local control =r(mean)
26 sum loenmv if count <0 & treatment ==1
27 local treat =r(mean)
28 sum loenmv if count>=0 & treatment ==0
29 local control_post =r(mean)
30 gen post = count >=0
31
32         **estimate DiD model on salary
33 reghdfe loenmv treat, abs(alder female post
34 treatment edu year) cl(pnr)
35 matrix b =e(b) //regression coefficient
36 matrix V = e(V) // standard error^2
37 local n = `count'-2002 //time
38
39         matrix results[`n',1] = b[1,1]
40         matrix results[`n',2] = V[1,1]^0.5
41         matrix results[`n',3] = b[1,1]/
42 (`control_post'-(`control'-`treat'))
43         matrix results[`n',4] = `n'
44
45     }
46     svmat results
47
48     rename results1 est
49     rename results2 se
50     rename results3 marg
51     rename results4 time
52
53     replace time = time+(`time'-3)*.1 //jitter estimates for
54 graph
55
56
57
58
59
60

```

```

1
2
3     keep est* se* marg* time
4     keep if est !=.
5
6     replace est = est/7446    //estimate measured as 1000 Euro
7     replace se = se/7446    //S.E. measured as 1000 Euro
8
9     gen upper = est+se*1.96 // Upper CI
10    gen lower = est-se*1.96 // Lower CI
11
12
13
14    gen control = `time'           //indicate control group
15
16    if `time' >1 append using $data/results.dta
17    save $data/results.dta, replace
18 }
19
20 use $data/results.dta, clear
21
22 replace time = 2002+time
23
24
25 *reads in unemployment statistcis obtained from statistikbanken.dk/
26 en/
27
28 gen unemp = 5.8 if time ==2003
29 replace unemp = 5.8 if time ==2004
30 replace unemp = 5.1 if time ==2005
31 replace unemp = 3.9 if time ==2006
32 replace unemp = 2.7 if time ==2007
33 replace unemp = 1.9 if time ==2008
34 replace unemp = 3.6 if time ==2009
35 replace unemp = 4.2 if time ==2010
36 replace unemp = 4.0 if time ==2011
37 replace unemp = 4.5 if time ==2012
38
39
40 scatter est time if control ==1, mcolor(navy) yaxis(1) ysc(range(-4
41 3) axis(1)) ylab(-4(1)3) || ///
42 scatter est time if control ==2, mcolor(blue) || ///
43 scatter est time if control ==3, mcolor(midblue) || ///
44 scatter est time if control ==4, mcolor(gray) || ///
45 scatter est time if control ==5, mcolor(ltblue) || ///
46 rspike upper lower time if control ==1, lcolor(navy) || ///
47 rspike upper lower time if control ==2, lcolor(blue) || ///
48 rspike upper lower time if control ==3, lcolor(midblue) || ///
49 rspike upper lower time if control ==4, lcolor(gray) || ///
50 rspike upper lower time if control ==5, lcolor(ltblue) || ///
51 line unemp time , lcolor(black) yaxis(2) ysc(range(0 6) axis(2))
52 ylab(0(1)6, axis(2)) ///
53 xsc(range(2002.5 2012.5)) xlab(2003(2)2012) ///
54 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
55 xti("Year of concussion for exposure group") scale(.95) ///
56 legend(label(1 " Control," "{&Delta}=1 yr") ///
57 label(2 " Control," "{&Delta}=2 yr") ///
58 label(3 " Control," "{&Delta}=3 yr") ///
59
60

```

```

1
2
3 label(4 " Control," "{&Delta}=4 yr") ///
4 label(5 " Control," "{&Delta}=5 yr") ///
5 label(6 "% Unemp." "of LF") ///
6 c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
7 c(1) symx(4) region(lc(white))) ///
8 yti("Effect on Salary (in EUR 1K)", height(7) axis(1)) ///
9 yti("Percent of Full Time Uemployed among LF", height(7) axis(2))
10
11
12
13 graph export $highdef/marg_est.png, replace width(3900)
14
15
16 forvalue control_time=1/5{
17     local end = 2012 // last incident year in data
18     if `control_time' ==1     eststo clear
19
20     **build dataset for joint estimate across years
21     forvalue count=2003/`end'{
22         if `count'==2003{
23             use $data/
24 sample_control_`count'_`control_time'.dta, clear
25             gen time = `count' //incident year
26 indicator
27             }
28             else append using $data/
29 sample_control_`count'_`control_time'.dta
30             replace time = `count' if time ==.
31
32
33             **exclude individuals in years where they do not
34 appear in data,
35             **due to either death or migration, as well as
36 periods from when
37             **the control group suffer their concussion
38 drop if merge ==1 | count > `control_time'-1
39         }
40
41         gen female = koen==2
42
43         //build ident, so we can multivariate cluster for
44 individuals
45         //who occur both as control and exposure during the period
46 (id)
47
48         bysort pnr time: gen helpx = _n ==1
49         gen id= sum(helpx)
50         drop helpx
51
52         **Generate educational groups
53         qui{
54             gen edu =0
55             replace edu = 1 if inrange(hffsp,20000000,39000000)
56
57 | ///
58                                     (hffsp
59 >40000000 & hffsp!=.)
60

```

```

1
2
3     }
4
5     **Calculate number of observations for exposure and control
6     count if count==0 & treatment ==1
7     local Ntreated = r(N)
8     count if count==0 & treatment ==0
9     local Ncontrol = r(N)
10
11    **generate concussion variable
12    gen treat = inrange(count,0,`control_time'-1) & treatment
13
14 ==1
15    replace treat = time_from_incident if count ==0 & treatment
16 ==1
17
18    **Generate pre-concussion income difference          for
19    **use in calculating marginal effects
20    sum loenmv if count <0 & treatment ==0
21    local control =r(mean)
22    sum loenmv if count <0 & treatment ==1
23    local treat =r(mean)
24    sum loenmv if count>=0 & treatment ==0
25    local control_post =r(mean)
26
27    forvalue t=-4/4{
28        local n = `t'*-1
29        if `t' < -1 gen T_`n' = treatment ==1 & count ==`t'
30        if `t' > -1 gen T`'t' = treatment ==1 & count ==`t'
31
32    }
33
34
35    **estimate DiD model on salary
36    reghdfe loenmv T*, abs(alder female count time treatment
37 edu) cl(pnr id)
38    eststo est1_`control_time'
39    if `control_time'==1 matrix results = J(5,5,.) // matrix to
40 capture results
41    if `control_time'==1 matrix results_p = J(5,5,.) // matrix
42 to capture results
43    if `control_time'==1 matrix results_pre = J(5,5,.) //
44 matrix to capture results
45
46    matrix b = e(b)
47    matrix V = e(V)
48    local n = `control_time'
49
50    matrix results[`n',1] = b[1,1] / 7466 // capture beta
51 results as 1K Euro
52    matrix results[`n',2] = (V[1,1]^.5)/7466 //
53 capture standard error as 1K Euro
54    matrix results[`n',3] = b[1,1]/(`control_post'-
55 (`control'-`treat'))
56    matrix results[`n',4] = `n'
57
58
59    gen no_lon = loenmv<1 //dummy for no salary
60

```

```

1
2
3
4
5
6     **estimate DiD LP-model for pre-trends
7     xi: reghdfe loenmv T*, abs(alder female count time
8 treatment edu) cl(pnr id), if count <0
9     eststo est3_`control_time'
10    matrix b = e(b) //regression coefficient
11    matrix V = e(V) // standard error^2
12
13
14    matrix results_pre[`n',1] =      b[1,1]
15    matrix results_pre[`n',2] =      V[1,1]^5
16    matrix results_pre[`n',3] =      b[1,1]/(`control_post'-
17 (`control'-`treat'))
18    matrix results_pre[`n',4] =      `n'
19
20 }
21
22 esttab est1_* using [home]/tables/dynamic1.rtf, ///
23     replace se(1) b(1) compress nogap star(+ .1 * .05 ** .01
24 *** .001) ///
25     keep(T*)
26
27 esttab est2_* using [home]/tables/dynamic2.rtf, ///
28     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
29 *** .001) ///
30     keep(T*)
31
32 esttab est3_* using [home]/tables/pre_trends.rtf, ///
33     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
34 *** .001) ///
35     keep(T*)
36
37
38 forvalue control_time=1/5{
39     local end = 2012 // last incident year in data
40     if `control_time' ==1     eststo clear
41     qui{
42         **build dataset for joint estimate across years
43         forvalue count=2003/`end'{
44             if `count'==2003{
45                 use $data/
46 sample_control_`count'_`control_time'.dta, clear
47                 gen time = `count' //incident year
48 indicator
49             }
50             else append using $data/
51 sample_control_`count'_`control_time'.dta
52             replace time = `count' if time ==.
53
54
55             **exclude individuals in years where they
56 do not appear in data,
57             **due to either death or migration, as
58 well as periods from when
59             **the control group suffer their concussion
60

```

```

1
2
3           drop if merge ==1 | count >
4 `control_time'-1
5     }
6
7     gen female = koen==2
8
9           **Generate educational groups
10    gen edu =0
11    replace edu = 1 if inrange(hffsp,20000000,39000000)
12 | ///
13
14 (hffsp >40000000 & hffsp!=.)
15
16
17
18 //build ident, so we can multivariate cluster for
19 individuals //who occur both as control and exposure during the
20 period (id)
21
22
23    bysort pnr time: gen helpx = _n ==1
24    gen id= sum(helpx)
25    drop helpx
26
27
28 **Calculate number of observations for exposure and
29 control
30    count if count==0 & treatment ==1
31    local Ntreated = r(N)
32    count if count==0 & treatment ==0
33    local Ncontrol = r(N)
34
35 **generate concussion variable
36    gen treat = inrange(count,0,`control_time'-1) &
37 treatment ==1
38    replace treat = time_from_incident if count ==0 &
39 treatment ==1
40
41
42 **Generate pre-concussion income difference
43 for
44
45 **use in calculating marginal effects
46    sum loenmv if count <0 & treatment ==0
47    local control =r(mean)
48    sum loenmv if count <0 & treatment ==1
49    local treat =r(mean)
50    sum loenmv if count>=0 & treatment ==0
51    local control_post =r(mean)
52
53    forvalue t=-4/4{
54        local n = `t'*-1
55        if `t' < -1 gen T_`n' = treatment ==1 &
56 count ==`t'
57        if `t' > -1 gen T`'t' = treatment ==1 &
58 count ==`t'
59    }
60

```

```

1
2
3
4         gen post = count > -1
5
6         **estimate DiD model on salary
7         reghdfe loenmv treat, abs(alder female post time
8 treatment edu) cl(pnr id)
9         eststo est1_`control_time'
10        if `control_time'==1 matrix results = J(5,5,.) //
11 matrix to capture results
12        if `control_time'==1 matrix results_p = J(5,5,.) //
13 matrix to capture results
14
15
16        matrix b = e(b)
17        matrix V = e(V)
18        local n = `control_time'
19
20
21        matrix results[`n',1] = b[1,1] / 7466 //
22 capture beta results as 1K Euro
23        matrix results[`n',2] = (V[1,1]^5)/
24 7466 // capture standard error as 1K Euro
25        matrix results[`n',3] = b[1,1]/
26 (`control_post'-(`control'-`treat'))
27        matrix results[`n',4] = `n'
28
29
30    }
31    *examining balance of samples
32    di in ye `control_time'
33    bysort treatment: sum female alder edu if count ==0
34
35 }
36
37
38
39
40 svmat results
41
42 gen upper = results1+results2*1.96
43 gen lower = results1-results2*1.96
44
45 scatter results1 results4 if results4 ==1, mcolor(navy) || ///
46 scatter results1 results4 if results4 ==2, mcolor(navy) || ///
47 scatter results1 results4 if results4 ==3, mcolor(navy) || ///
48 scatter results1 results4 if results4 ==4, mcolor(navy) || ///
49 scatter results1 results4 if results4 ==5, mcolor(navy) || ///
50 rspike upper lower results4 if results4 ==1, lcolor(navy) || ///
51 rspike upper lower results4 if results4 ==2, lcolor(navy) || ///
52 rspike upper lower results4 if results4 ==3, lcolor(navy) || ///
53 rspike upper lower results4 if results4 ==4, lcolor(navy) || ///
54 rspike upper lower results4 if results4 ==5, lcolor(navy) ///
55 ysc(range(-2 1)) ylab(-2(.5)1) ///
56 xsc(range(.5 5.5)) xlab(1(1)5) ///
57 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
58 xti("Years between exposure and control incident") scale(.95) ///
59
60

```



```

1
2
3 legend(label(1 " Control," "{&Delta}=1 yr") ///
4 label(2 " Control," "{&Delta}=2 yr") ///
5 label(3 " Control," "{&Delta}=3 yr") ///
6 label(4 " Control," "{&Delta}=4 yr") ///
7 label(5 " Control," "{&Delta}=5 yr") ///
8 c(1) order(1 2 3 4 5) pos(3) size(small) ///
9 c(1) symx(4) region(lc(white))) ///
10 yti("Effect in 1K Euro ({&delta}{subscript: Salary})",
11 height(7)) ///
12 legend(off)
13
14
15 graph export $highdef/est2003_2011.png, replace width(3900)
16
17
18 **Reports marginal effects for period 2003–2011 in percent
19
20 gen upper2 = (results3+results2/(results1/results3)*1.96)*100
21 gen lower2 = (results3–results2/(results1/results3)*1.96)*100
22 replace results3 = results3*100
23
24
25 scatter results3 results4 , mcolor(navy) || ///
26 rspike upper2 lower2 results4, lcolor(navy) ///
27 ysc(range(-5 0)) ylab(-5(.5)0) ///
28 xsc(range(.5 5.5)) xlab(1(1)5) ///
29 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
30 /*title("Percentage change in salary, 2003–10")*/ ///
31 yti("Salary change (in %)", height(7)) ///
32 xti("Years between exposure and control incident") scale(.95) ///
33 legend(label(1 " Control," "{&Delta}=1 yr") ///
34 label(2 " Control," "{&Delta}=2 yr") ///
35 label(3 " Control," "{&Delta}=3 yr") ///
36 label(4 " Control," "{&Delta}=4 yr") ///
37 label(5 " Control," "{&Delta}=5 yr") ///
38 c(1) order(1 2 3 4 5) pos(3) size(small) ///
39 c(1) symx(4) region(lc(white))) legend(off)
40 /*
41 note("Marginal effects for exposure dummy across spacing of control
42 groups. Decrease " ///
43 "calculated by dividing {&delta} with the normalized control groups'
44 average salary " ///
45 "post-concussion. Control groups suffer concussions 1, 2, 3, 4, and
46 5 years (&Delta) after" ///
47 "the exposure group. Both control and exposure group are 30–49 years
48 of age when" ///
49 "exposure group suffers concussion. 95% confidence intervals.")*/
50
51
52 graph export $highdef/marginal2003_2011.png, replace width(3900)
53
54
55
56
57 *****
58 *****
59 **
60

```

```

1
2
3      **          Results for individuals with at least highschool
4      **
5      **
6      ****
7      ****
8
9
10     forvalue control_time=1/5{
11         local end = 2012 // last incident year in data
12
13         **build dataset for joint estimate across years
14         forvalue count=2003/`end'{
15             if `count'==2003{
16                 use $data/
17 sample_control_`count'_`control_time'.dta, clear
18                 gen time = `count' //incident year
19 indicator
20                 }
21                 else append using $data/
22 sample_control_`count'_`control_time'.dta
23                 replace time = `count' if time ==.
24
25                 **exclude individuals in years where they do not
26 appear in data,
27                 **due to either death or migration, as well as
28 periods from when
29                 **the control group suffer their concussion
30 drop if merge ==1 | count > `control_time'-1
31
32             }
33
34         gen female = koen==2
35
36         **Generate educational groups
37 qui{
38             gen edu =0
39             replace edu = 1 if inrange(hffsp,20000000,39000000)
40 | ///
41                                     (hffsp
42 >40000000 & hffsp!=.)
43             }
44
45         keep if edu==1
46
47         //build ident, so we can multivariate cluster for
48 individuals
49 //who occur both as control and exposure during the period
50 (id)
51
52         bysort pnr time: gen helpx = _n ==1
53         gen id= sum(helpx)
54         drop helpx help
55
56
57
58
59
60

```

```

1
2
3      **Calculate number of observations for exposure and control
4      count if count==0 & treatment ==1
5      local Ntreated = r(N)
6      count if count==0 & treatment ==0
7      local Ncontrol = r(N)
8
9      **generate concussion variable
10     gen treat = inrange(count,0,`control_time'-1) & treatment
11
12 ==1
13     replace treat = time_from_incident if count ==0 & treatment
14 ==1
15
16     **Generate pre-concussion income difference          for
17     **use in calculating marginal effects
18     sum loenmv if count <0 & treatment ==0
19     local control =r(mean)
20     sum loenmv if count <0 & treatment ==1
21     local treat =r(mean)
22     sum loenmv if count>=0 & treatment ==0
23     local control_post =r(mean)
24
25     gen post = count >=0
26
27     **estimate DiD model on salary
28     xi: reghdfe loenmv treat, abs(alder female post time
29     treatment) cl(id pnr)
30
31     if `control_time'==1 matrix results_edu = J(5,5,.) //
32     matrix to capture results
33     if `control_time'==1 matrix results_p_edu = J(5,5,.) //
34     matrix to capture results
35
36
37
38     matrix b = e(b)
39     matrix V = e(V)
40     local n = `control_time'
41
42     matrix results_edu[`n',1]          = b[1,1] / 7466 //
43     capture beta results as 1K Euro
44     matrix results_edu[`n',2]          = (V[1,1]^0.5)/
45     7466 // capture standard error as 1K Euro
46     matrix results_edu[`n',3] =      b[1,1]/(`control_post'-
47     (`control'-`treat'))
48     matrix results_edu[`n',4] =      `n'
49
50
51 }
52
53
54 *****
55 *****
56 **
57 **           Results for individuals with no high school+
58 **
59 **
60

```

```

1
2
3 *****
4 *****
5
6
7 forvalue control_time=1/5{
8     local end = 2012 // last incident year in data
9
10    **build dataset for joint estimate across years
11    forvalue count=2003/`end'{
12        if `count'==2003{
13            use $data/
14            sample_control_`count'_`control_time'.dta, clear
15            gen time = `count' //incident year
16            indicator
17                }
18            else append using $data/
19            sample_control_`count'_`control_time'.dta
20            replace time = `count' if time ==.
21
22
23            **exclude individuals in years where they do not
24            appear in data,
25            **due to either death or migration, as well as
26            periods from when
27            **the control group sufer their concussion
28            drop if merge ==1 | count > `control_time'-1
29
30
31        }
32
33        gen female = koen==2
34
35        //build ident, so we can multivariate cluster for
36        individuals
37        //who occur both as control and exposure during the period
38        (id)
39
40
41        bysort pnr time: gen helpx = _n ==1
42        gen id= sum(helpx)
43        drop helpx help
44
45        **Generate educational groups
46        qui{
47            gen edu =0
48            replace edu = 1 if inrange(hffsp,20000000,39000000)
49            | ///
50
51
52            >40000000 & hffsp!=.)
53            }
54
55        keep if edu==0
56
57
58
59        **Calculate number of observations for exposure and control
60

```

```

1
2
3     count if count==0 & treatment ==1
4     local Ntreated = r(N)
5     count if count==0 & treatment ==0
6     local Ncontrol = r(N)
7
8     **generate concussion variable
9     gen treat = inrange(count,0,`control_time'-1) & treatment
10
11 ==1
12     replace treat = time_from_incident if count ==0 & treatment
13 ==1
14
15     **Generate pre-concussion income difference          for
16     **use in calculating marginal effects
17     sum loenmv if count <0 & treatment ==0
18     local control =r(mean)
19     sum loenmv if count <0 & treatment ==1
20     local treat =r(mean)
21     sum loenmv if count>=0 & treatment ==0
22     local control_post =r(mean)
23     gen post = count >=0
24
25     **estimate DiD model on salary
26     xi: reghdfe loenmv treat, abs(alder female post time
27 treatment) cl(id pnr)
28
29     if `control_time'==1 matrix results_noedu = J(5,5,.) //
30 matrix to capture results
31     if `control_time'==1 matrix results_p_noedu = J(5,5,.) //
32 matrix to capture results
33
34
35
36     matrix b = e(b)
37     matrix V = e(V)
38     local n = `control_time'
39
40     matrix results_noedu[`n',1]      = b[1,1] / 7466 //
41 capture beta results as 1K Euro
42     matrix results_noedu[`n',2]      = (V[1,1]^0.5)/
43 7466 // capture standard error as 1K Euro
44     matrix results_noedu[`n',3]      =      b[1,1]/
45 (`control_post'-(`control'-`treat'))
46     matrix results_noedu[`n',4]      =      `n'
47
48
49
50
51 }
52 *****
53 *****
54 **
55 **     Draw figure for subgroups
56 **
57 **
58 *****
59 *****
60

```

```

1
2
3
4 local t= -.15
5
6 foreach x in noedu edu{
7     svmat results_`x'
8     replace results_`x'4= results_`x'4+`t'
9     svmat results_p_`x'
10    replace results_p_`x'4= results_p_`x'4+`t'
11
12
13    gen upper_`x' = results_`x'1+results_`x'2*1.96
14    gen lower_`x' = results_`x'1-results_`x'2*1.96
15
16    gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
17 results_`x'3)*1.96)*100
18    gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
19 results_`x'3)*1.96)*100
20    replace results_`x'3 = results_`x'3*100
21
22    gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
23    gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
24
25    local t = `t'+.1
26
27 }
28
29
30 keep results* upper* lower*
31 keep if _n <=5
32
33
34 **generate locals for figure
35
36 foreach x in noedu edu{
37     if "`x'" == "nopay" local color = "navy"
38     if "`x'" == "pay" local color = "red"
39     if "`x'" == "noedu" local color = "green"
40     if "`x'" == "edu" local color = "purple"
41
42     local figure_`x' "scatter results_`x'1 results_`x'4,
43 mcolor(`color') || rspike upper_`x' lower_`x' results_`x'4,
44 lcolor(`color') vertical"
45     if "`x'" == "nopay" local figure2_`x' "scatter results_`x'3
46 results_`x'4, mcolor(`color') || rspike upper2_`x' lower2_`x'
47 results_`x'4, lcolor(`color') vertical "
48     else local figure2_`x' "scatter results_`x'3 results_`x'4,
49 mcolor(`color') || rspike upper2_`x' lower2_`x' results_`x'4,
50 lcolor(`color') vertical "
51     local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
52 mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
53 lcolor(`color') vertical "
54 }
55
56
57
58 `figure_noedu' || `figure_edu' ///
59 legend( ///
60

```

```

1
2
3 label(1 "Less than" "high school") ///
4 label(3 "At least" "high school") ///
5 c(1) order(1 3 ) pos(3) size(small) ///
6 c(1) symx(4) region(lc(white))) ///
7 ysc(range(-4 2)) ylab(-4(1)2) ///
8 xsc(range(.5 5.5)) xlab(1(1)5) ///
9 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
10 ///title("Parameter estimates across control group, 2003-10") ///
11 xti("Years between exposure and control incident") scale(.95) ///
12 yti("Effect in 1K Euro ({&delta;}{subscript: Salary})",
13 height(7)) ///
14 /*note("Parameter estimates for exposure dummy across spacing of
15 control groups." ///
16 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
17 after the exposure group." ///
18 "Both control and exposure group are 30-49 years of age when
19 exposure group suffers " ///
20 "concussion. 95% confidence intervals.")
21 */
22
23
24 graph export $highdef/grouped_est2003_2011.png, replace width(3900)
25
26 preserve
27
28
29
30
31
32 `figure2_noedu' || `figure2_edu' ///
33 legend( ///
34 label(1 "Less than" "high school") ///
35 label(3 "At least" "high school") ///
36 c(1) order(1 3 5) pos(3) size(small) ///
37 c(1) symx(4) region(lc(white))) ///
38 ysc(range(-12 3)) ylab(-12(3)3) ///
39 xsc(range(.5 5.5)) xlab(1(1)5) ///
40 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
41 /// title("Percentage change in salary, 2003-10") ///
42 yti("Salary change (in %)", height(7)) ///
43 xti("Years between exposure and control incident") scale(.95) ///
44 /*note("Marginal effects for exposure dummy across spacing of
45 control groups." ///
46 "change calculated by {&delta;} with the normalized control groups'
47 average" ///
48 "salary post-concussion. Control groups suffer concussions 1, 2, 3,
49 4, and 5 years (&Delta) after" ///
50 "the exposure group. Both control and exposure group are 30-49 years
51 of age when" ///
52 "exposure group suffers concussion. 95% confidence intervals.")*/
53
54
55 graph export $highdef/grouped_marginal2003_2011.png, replace
56 width(3900)
57
58 restore
59
60

```

```

1
2
3
4 *****
5 *****
6 **
7 **           Results different age-groups
8 **
9 **
10 *****
11 *****
12
13
14 forvalue y=20(5)55{
15     forvalue control_time=1/5{
16         local end = 2012 // last incident year in data
17
18         **build dataset for joint estimate across years
19         forvalue count=2003/\`end'{
20             if `count'==2003{
21                 use $data/
22 sample_control_`count'_`control_time'.dta, clear
23                 gen time = `count' //incident year
24 indicator
25                 }
26                 else append using $data/
27 sample_control_`count'_`control_time'.dta
28                 replace time = `count' if time ==.
29
30                 **exclude individuals in years where they
31 do not appear in data,
32                 **due to either death or migration, as
33 well as periods from when
34                 **the control group suffer their concussion
35 drop if merge ==1 | count >
36 `control_time'-1
37
38
39
40     }
41
42     gen female = koen==2
43
44     //build ident, so we can multivariate cluster for
45 individuals
46     //who occur both as control and exposure during the
47 period (id)
48
49
50     bysort pnr time: gen helpx = _n ==1
51     gen id= sum(helpx)
52     drop helpx help
53
54     gen nopay = loenmv <1
55
56     **Generate age group
57     local z = `y'+4
58     gen help = count == 0 & inrange(alder,`y',`z')
59     bysort id: egen helpx =max(help)
60

```



```

1
2
3         keep if helpx == 1
4         drop helpx help
5
6
7
8
9
10        **Calculate number of observations for exposure and
11 control
12        count if count==0 & treatment ==1
13        local Ntreated = r(N)
14        count if count==0 & treatment ==0
15        local Ncontrol = r(N)
16
17        **generate concussion variable
18        gen treat = inrange(count,0,`control_time'-1) &
19 treatment ==1
20        replace treat = time_from_incident if count ==0 &
21 treatment ==1
22
23        **Generate pre-concussion income difference
24 for
25
26        **use in calculating marginal effects
27        sum loenmv if count <0 & treatment ==0
28        local control =r(mean)
29        sum loenmv if count <0 & treatment ==1
30        local treat =r(mean)
31        sum loenmv if count>=0 & treatment ==0
32        local control_post =r(mean)
33        gen post = count >=0
34
35        **estimate DiD model on salary
36        xi: reghdfe loenmv treat, abs(alder female post
37 time treatment) cl(id pnr)
38
39        if `control_time'==1 matrix results_`y' =
40 J(5,5,.) // matrix to capture results
41        if `control_time'==1 matrix results_p_`y' =
42 J(5,5,.) // matrix to capture results
43
44
45        matrix b = e(b)
46        matrix V = e(V)
47        local n = `control_time'
48
49        matrix results_`y'[`n',1]           = b[1,1] /
50 7466 // capture beta results as 1K Euro
51        matrix results_`y'[`n',2]           = (V[1,1]^0.5)/
52 7466 // capture standard error as 1K Euro
53        matrix results_`y'[`n',3] = b[1,1]/
54 (`control_post'-(`control'-`treat'))
55        matrix results_`y'[`n',4] = `n'
56
57
58
59     }
60

```

```

1
2
3     }
4
5
6
7     local t = -.15    //Jitter estimates along x-axis
8     forvalue x =20(5)55{
9         svmat results_`x'
10        replace results_`x'4= results_`x'4+`t'
11        svmat results_p_`x'
12        replace results_p_`x'4= results_p_`x'4+`t'
13
14
15        gen upper_`x' = results_`x'1+results_`x'2*1.96
16        gen lower_`x' = results_`x'1-results_`x'2*1.96
17
18        gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
19 results_`x'3)*1.96)*100
20        gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
21 results_`x'3)*1.96)*100
22        replace results_`x'3 = results_`x'3*100
23
24        gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
25        gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
26
27        local t = `t'+.1
28
29    }
30
31    keep results* upper* lower*
32    keep if _n <=5
33
34
35    **generate locals for figure
36
37    forvalue x =20(5)55{
38        if `x' == 20 local color = "black"
39        if `x' == 25 local color = "orange"
40        if `x' == 30 local color = "navy"
41        if `x' == 35 local color = "red"
42        if `x' == 40 local color = "green"
43        if `x' == 45 local color = "purple"
44        if `x' == 50 local color = "sienna"
45        if `x' == 55 local color = "teal"
46
47        local figure_`x' "scatter results_`x'1 results_`x'4,
48 mcolor(`color') || rspike upper_`x' lower_`x' results_`x'4,
49 lcolor(`color') vertical"
50        local figure2_`x' "scatter results_`x'3 results_`x'4,
51 mcolor(`color') || rspike upper2_`x' lower2_`x' results_`x'4,
52 lcolor(`color') vertical "
53        local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
54 mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
55 lcolor(`color') vertical "
56    }
57
58
59    /**/
60

```

```

1
2
3 `figure_20' || `figure_25' || `figure_30' || `figure_35' ///
4     || `figure_40' || `figure_45' || `figure_50' ||
5 `figure_55' ///
6 legend( ///
7 label(1 "Age 20-24") ///
8 label(3 "Age 25-29") ///
9 label(5 "Age 30-34") ///
10 label(7 "Age 35-39") ///
11 label(9 "Age 40-44") ///
12 label(11 "Age 45-49") ///
13 label(13 "Age 50-54") ///
14 label(15 "Age 55-59") ///
15 c(1) order(1 3 5 7 9 11 13 15) pos(3) size(small) ///
16 c(1) symx(4) region(lc(white))) ///
17 ysc(range(-4 2)) ylab(-4(1)2) ///
18 xsc(range(.5 5.5)) xlab(1(1)5) ///
19 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
20 /// title("Parameter estimates across control group, 2003-10") ///
21 xti("Years between exposure and control incident") scale(.95) ///
22 yti("Effect in 1K Euro ({&delta;}{subscript: Salary})",
23 height(7)) ///
24 /*note("Parameter estimates for exposure dummy across spacing of
25 control groups." ///
26 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
27 after the exposure group." ///
28 "Age group described age at time of exposure incident. 95%
29 confidence intervals.")*/
30
31
32 graph export $highdef/age_est2003_2011.png, replace width(3900)
33 */
34
35
36 `figure2_20' || `figure2_25' || `figure2_30' || `figure2_35' ///
37     || `figure2_40' || `figure2_45' || `figure2_50' ||
38 `figure2_55' ///
39 legend( ///
40 label(1 "Age 20-24") ///
41 label(3 "Age 25-29") ///
42 label(5 "Age 30-34") ///
43 label(7 "Age 35-39") ///
44 label(9 "Age 40-44") ///
45 label(11 "Age 45-49") ///
46 label(13 "Age 50-54") ///
47 label(15 "Age 55-59") ///
48 c(1) order(1 3 5 7 9 11 13 15) pos(3) size(small) ///
49 c(1) symx(4) region(lc(white))) ///
50 ysc(range(-12 3)) ylab(-12(3)3) ///
51 xsc(range(.5 5.5)) xlab(1(1)5) ///
52 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
53 yti("Salary change (in %)", height(7)) ///
54 xti("Years between exposure and control incident") scale(.95)
55
56
57
58 graph export $highdef/age_marginal2003_2011.png, replace width(3900)
59
60

```

```

1
2
3 `figure_p_30' || `figure_p_35' || `figure_p_40' ||
4 `figure_p_45' /// ///
5 legend(label(1 "Age 30-34") ///
6 label(3 "Age 35-39") ///
7 label(5 "Age 40-44") ///
8 label(7 "Age 45-49") ///
9 c(1) order(1 3 5 7) pos(3) size(small) ///
10 c(1) symx(4) region(lc(white))) ///
11 ysc(range(-.02 .050)) ylab(-0.02(.01)0.05) ///
12 xsc(range(.5 5.5)) xlab(1(1)5) ///
13 yline(0, lcolor(black)) ysize(10) xsize(12) graphr(c(white)) ///
14 title("Parameter estimates across control group, 2003-10") ///
15 yti("Effect on Pr(Salary=0)", height(7)) ///
16 xti("Years between exposure and control incident ({&Delta})")
17 scale(.95) ///
18 note("Parameter estimates for exposure dummy across spacing of
19 control groups." ///
20 "Control groups suffer concussions 1, 2, 3, 4, and 5 (&Delta) years
21 after the exposure group." ///
22 "Age group described age at time of exposure incident. 95%
23 confidence intervals.")
24
25
26 graph export $highdef/age_nopay2003_2011.png, replace width(3900)
27
28
29
30 *****
31 *****
32 **
33 ** Results across gender
34 **
35 **
36 *****
37 *****
38
39 forvalue y=0/1{
40     forvalue control_time=1/5{
41         local end = 2012 // last incident year in data
42
43         **build dataset for joint estimate across years
44         forvalue count=2003/`end'{
45             if `count'==2003{
46                 use $data/
47 sample_control_`count'_`control_time'.dta, clear
48                 gen time = `count' //incident year
49 indicator
50             }
51             else append using $data/
52 sample_control_`count'_`control_time'.dta
53             replace time = `count' if time ==.
54
55             **exclude individuals in years where they
56 do not appear in data,
57             **due to either death or migration, as
58 well as periods from when
59
60

```

```

1
2
3             **the control group suffer their concussion
4             drop if merge ==1 | count >
5 `control_time'-1
6
7
8     }
9
10    gen female = koen==2
11
12    //build ident, so we can multivariate cluster for
13 individuals    //who occur both as control and exposure during the
14 period (id)
15
16    bysort pnr time: gen helpx = _n ==1
17    gen id= sum(helpx)
18    drop helpx help
19
20    gen nopay = loenmv <1
21
22    **Generate age group
23    local z = `y'+4
24    gen help = count == 0 & female==`y'
25    bysort id: egen helpx =max(help)
26    keep if helpx == 1
27    drop helpx help
28
29
30
31
32
33
34
35    **Calculate number of observations for exposure and
36 control
37    count if count==0 & treatment ==1
38    local Ntreated = r(N)
39    count if count==0 & treatment ==0
40    local Ncontrol = r(N)
41
42    **generate concussion variable
43    gen treat = inrange(count,0,`control_time'-1) &
44 treatment ==1
45    replace treat = time_from_incident if count ==0 &
46 treatment ==1
47
48
49    **Generate pre-concussion income difference
50 for
51    **use in calculating marginal effects
52    sum loenmv if count <0 & treatment ==0
53    local control =r(mean)
54    sum loenmv if count <0 & treatment ==1
55    local treat =r(mean)
56    sum loenmv if count>=0 & treatment ==0
57    local control_post =r(mean)
58    gen post = count >=0
59
60

```

```

1
2
3         **estimate DiD model on salary
4         xi: reghdfe loenmv treat, abs(alder post time
5 treatment) cl(id pnr)
6
7         if `control_time'==1 matrix results_`y' =
8 J(5,5,.) // matrix to capture results
9         if `control_time'==1 matrix results_p_`y' =
10 J(5,5,.) // matrix to capture results
11
12
13         matrix b = e(b)
14         matrix V = e(V)
15         local n = `control_time'
16
17         matrix results_`y'[`n',1]           = b[1,1] /
18 7466 // capture beta results as 1K Euro
19         matrix results_`y'[`n',2]           = (V[1,1]^0.5)/
20 7466 // capture standard error as 1K Euro
21         matrix results_`y'[`n',3] = b[1,1]/
22 (`control_post'-(`control'-`treat'))
23         matrix results_`y'[`n',4] = `n'
24
25
26
27     }
28 }
29
30
31
32
33 local t = -.05 //Jitter estimates along x-axis
34 foreach x in 0 1{
35     svmat results_`x'
36     replace results_`x'4= results_`x'4+`t'
37     svmat results_p_`x'
38     replace results_p_`x'4= results_p_`x'4+`t'
39
40     gen upper_`x' = results_`x'1+results_`x'2*1.96
41     gen lower_`x' = results_`x'1-results_`x'2*1.96
42
43     gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
44 results_`x'3)*1.96)*100
45     gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
46 results_`x'3)*1.96)*100
47     replace results_`x'3 = results_`x'3*100
48
49     gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
50     gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
51
52     local t = `t'+.1
53
54 }
55
56
57 keep results* upper* lower*
58 keep if _n <=5
59
60

```

```

1
2
3 **generate locals for figure
4
5 foreach x in 0 1{
6     if `x' == 0 local color = "red"
7     if `x' == 1 local color = "green"
8
9         local figure_`x' "scatter results_`x'1 results_`x'4,
10 mcolor(`color') || rspike upper`x' lower`x' results_`x'4,
11 lcolor(`color') vertical"
12         local figure2_`x' "scatter results_`x'3 results_`x'4,
13 mcolor(`color') || rspike upper2`x' lower2`x' results_`x'4,
14 lcolor(`color') vertical "
15         local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
16 mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
17 lcolor(`color') vertical "
18     }
19
20
21
22 `figure_0' || `figure_1' ///
23 legend(label(1 "Men") ///
24 label(3 "Women") ///
25 c(1) order(1 3) pos(3) size(small) ///
26 c(1) symx(4) region(lc(white))) ///
27 ysc(range(-4 2)) ylab(-4(1)2) ///
28 xsc(range(.5 5.5)) xlab(1(1)5) ///
29 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
30 /// title("Parameter estimates across control group, 2003-10") ///
31 xti("Years between exposure and control incident") scale(.95) ///
32 yti("Effect in 1K Euro ({&delta;}{subscript: Salary})",
33 height(7)) ///
34 /*note("Parameter estimates for exposure dummy across spacing of
35 control groups." ///
36 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
37 after the exposure group." ///
38 "95% confidence intervals.")*/
39
40
41 graph export $highdef/gender_est2003_2011.png, replace width(3900)
42
43 `figure2_0' || `figure2_1' ///
44 legend( ///
45 label(1 "Men") ///
46 label(3 "Women") ///
47 c(1) order(1 3 5 7) pos(3) size(small) ///
48 c(1) symx(4) region(lc(white))) ///
49 ysc(range(-12 3)) ylab(-12(3)3) ///
50 xsc(range(.5 5.5)) xlab(1(1)5) ///
51 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
52 /// title("Percentage change in salary, 2003-10") ///
53 yti("Salary change (in %)", height(7)) ///
54 xti("Years between exposure and control incident") scale(.95) ///
55 /* note("Parameter estimates for exposure dummy across spacing of
56 control groups." ///
57 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
58 after the exposure group." ///
59

```

```

1
2
3 "95% confidence intervals.)*"/
4
5 graph export $highdef/gender_marginal2003_2011.png, replace
6 width(3900)
7
8 `figure_p_0' || `figure_p_1' ///
9 legend( ///
10 label(1 "Men") ///
11 label(3 "Women") ///
12 c(1) order(1 3 5 7) pos(3) size(small) ///
13 c(1) symx(4) region(lc(white))) ///
14 ysc(range(-.02 .050)) ylab(-0.02(.01)0.05) ///
15 xsc(range(.5 5.5)) xlab(1(1)5) ///
16 yline(0, lcolor(black)) ysize(10) xsize(12) graphr(c(white)) ///
17 /// title("Parameter estimates across control group, 2003-10") ///
18 yti("Effect on Pr(Salary=0)", height(7)) ///
19 xti("Years between exposure and control incident ({&Delta})")
20 scale(.95) ///
21 /*note("Parameter estimates for exposure dummy across spacing of
22 control groups." ///
23 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
24 after the exposure group." ///
25 "95% confidence intervals.)*"/
26
27
28 graph export $highdef/gender_nopay2003_2011.png, replace width(3900)
29
30
31
32 *****
33 *****
34 **
35 ** Generate descriptive figures of wage
36 development for exposure
37 ** and control group
38 **
39 **
40 *****
41 *****
42
43
44 forvalue control_time=1/5{
45     local end = 2012 // last incident year in data
46
47     **build dataset for joint estimate across years
48     forvalue count=2003/`end'{
49         if `control_time' ==1 & `count'==2003{
50             use $data/
51 sample_control_`count'_`control_time'.dta, clear
52             gen time = `count' //incident year
53 indicator
54             }
55             else append using $data/
56 sample_control_`count'_`control_time'.dta
57
58
59     **Drops exposure-group already in the data
60

```



```

1
2
3         if `control_time' > 1 drop if time ==. & treatment
4 ==1
5         replace time = `count' if time ==.
6         replace control1 = `control_time' if
7 control`control_time'==1
8         if `control_time' > 1 drop control`control_time'
9
10        **exclude individuals in years where they do not
11 appear in data,
12        **due to either death or migration, as well as
13 periods from when
14        **the control group suffer their concussion
15
16        }
17
18
19 }
20
21 **Generate mean salary for each period relative to exposure groups
22 concussion
23 **separately for exposure and each control group
24
25 bysort control1 count: egen mean_loen = mean(loenmv)
26
27 **Generate mean Pr(sal= for each period relative to exposure groups
28 concussion
29 **separately for exposure and each control group
30
31 gen no_lon = loenmv < 1
32
33 bysort control1 count: egen mean_no_lon = mean(no_lon)
34
35
36 **generate group size
37 bysort control1 count: gen Ncount=_N if count ==0
38
39 **generate pre-exposure mean levels for normalization
40 gen pre = count <0
41 bysort control1 pre: egen pre_mean_loen = mean(loenmv) if pre==1
42 bysort control1 pre: egen pre_mean_no_lon = mean(no_lon) if pre==1
43
44
45 **reduce data set size
46 bysort control1 count: keep if _n ==1
47
48 keep count mean* pre_* control1 Ncount
49
50
51 **standardize to 1k euro
52 replace mean_loen = mean_loen/7466
53 replace pre_mean_loen = pre_mean_loen/7466
54
55 sort control1 count
56
57
58 **Pre-treatment normalization of salary
59
60

```

```

1
2
3   gen norm_mean_lon =mean_loen
4   gen norm_mean_no_lon =mean_no_lon
5
6   forvalue t = 1/5{
7       qui sum pre_mean_loen if control == 0
8       local treat = r(mean)
9       qui sum pre_mean_loen if control == `t'
10      local control = r(mean)
11
12      **normalize with pre-concussion difference
13      qui replace norm_mean_lon = mean_loen - (`control'-`treat')
14  if control == `t'
15
16      qui sum pre_mean_no_lon if control == 0
17      local treat = r(mean)
18      qui sum pre_mean_no_lon if control == `t'
19      local control = r(mean)
20
21      **normalize with pre-concussion difference
22      qui replace norm_mean_no_lon = mean_no_lon - (`control'-
23  `treat') if control == `t'
24
25  }
26
27
28  **local indicators of group sizes
29  forvalue t=0/5{
30      qui sum Ncount if control == `t'
31      local C`t' = r(mean)
32  }
33
34
35
36  graph twoway ///
37      connect mean_l count if control1== 0, ///
38      lcolor(black) mcolor(black) || ///
39      connect mean_l count if control1== 1, ///
40      lcolor(blue) mcolor(blue) || ///
41      connect mean_l count if control1== 2, ///
42      lcolor(green) mcolor(green) || ///
43      connect mean_l count if control1== 3, ///
44      lcolor(purple) mcolor(purple) || ///
45      connect mean_l count if control1== 4, ///
46      lcolor(red) mcolor(red) || ///
47      connect mean_l count if control1== 5, ///
48      lcolor(orange) mcolor(orange) ///
49      legend( ///
50          label(1 "Exposure" ///
51              "N=`C0'") ///
52          label(2 "Control {&Delta}=1" ///
53              "N=`C1'") ///
54          label(3 "Control {&Delta}=2" ///
55              "N=`C2'") ///
56          label(4 "Control {&Delta}=3" ///
57              "N=`C3'") ///
58          label(5 "Control {&Delta}=4" ///
59
60

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
        "N=`C4'"') ///
        label(6 "Control {&Delta}=5" ///
            "N=`C5'"') ///
        c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
        c(1) symx(4) region(lc(white))) ///
        ysc(range(26 33)) ylab(26(1)33) ///
        xsc(range(-5 5)) xlab(-5(1)5) ///
        xline(0, lcolor(red)) ysize(10) xsize(12)
graphr(c(white)) ///
/// title("Impact of Concussion on Salary") ///
yti("Deflated Salaried Income in 1K EUR", height(7)) ///
xti("Years since exposure group concussion") scale(.95) ///
/* note("Control groups suffer concussion 1, 2, 3, 4, and 5
years (&Delta) after exposure group." ///
"Both control and exposure group at 30-49 years of
age at beginning of year =0." ///
"Vertical line indicates time of exposure group
concussion.")*/

graph export $highdef/FigureS2.png, replace width(3900)

graph twoway ///
connect norm_mean_l count if control1== 1, ///
lcolor(blue) mcolor(blue) || ///
connect norm_mean_l count if control1== 2, ///
lcolor(green) mcolor(green) || ///
connect norm_mean_l count if control1== 3, ///
lcolor(purple) mcolor(purple) || ///
connect norm_mean_l count if control1== 4, ///
lcolor(red) mcolor(red) || ///
connect norm_mean_l count if control1== 5, ///
lcolor(orange) mcolor(orange) || ///
connect mean_l count if control1== 0, ///
lcolor(black) mcolor(black) ///
legend( ///
label(1 "Control {&Delta}=1" ///
"N=`C1'"') ///
label(2 "Control {&Delta}=2" ///
"N=`C2'"') ///
label(3 "Control {&Delta}=3" ///
"N=`C3'"') ///
label(4 "Control {&Delta}=4" ///
"N=`C4'"') ///
label(5 "Control {&Delta}=5" ///
"N=`C5'"') ///
label(6 "Exposure" ///
"N=`C0'"') ///
c(1) order(6 1 2 3 4 5) pos(3) size(small) ///
c(1) symx(4) region(lc(white))) ///
ysc(range(26 31)) ylab(26(1)31) ///
xsc(range(-5 5)) xlab(-5(1)5) ///
xline(0, lcolor(red)) ysize(10) xsize(12)
graphr(c(white)) ///

```

```

1
2
3    ///    title("Impact of Concussion on Salary") ///
4    yti("Deflated Norm. Salaried Income in 1K EUR",
5    height(7)) ///
6    xti("Years since exposure group's concussion") scale(.95)
7    /*///
8    note("Control groups suffer concussion 1, 2, 3, 4, and 5
9    years (&Delta) after exposure group." ///
10    "Both control and exposure group at 30-49 years of
11    age at beginning of year =0." ///
12    "Vertical line indicates time of exposure group
13    concussion." ///
14    "Salary levels normalized with pre-concussion level
15    difference between" ///
16    "each control group and the exposure group")*/
17
18
19    graph export $highdef/Figure2.png, replace width(3900)
20
21
22    graph twoway ///
23    connect mean_no_l count if control1== 0, ///
24    lcolor(black) mcolor(black) || ///
25    connect mean_no_l count if control1== 1, ///
26    lcolor(blue) mcolor(blue) || ///
27    connect mean_no_l count if control1== 2, ///
28    lcolor(green) mcolor(green) || ///
29    connect mean_no_l count if control1== 3, ///
30    lcolor(purple) mcolor(purple) || ///
31    connect mean_no_l count if control1== 4, ///
32    lcolor(red) mcolor(red) || ///
33    connect mean_no_l count if control1== 5, ///
34    lcolor(orange) mcolor(orange) ///
35    legend( ///
36    label(1 "Exposure" ///
37    "N=`C0`") ///
38    label(2 "Control {&Delta}=1" ///
39    "N=`C1`") ///
40    label(3 "Control {&Delta}=2" ///
41    "N=`C2`") ///
42    label(4 "Control {&Delta}=3" ///
43    "N=`C3`") ///
44    label(5 "Control {&Delta}=4" ///
45    "N=`C4`") ///
46    label(6 "Control {&Delta}=5" ///
47    "N=`C5`") ///
48    c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
49    c(1) symx(4) region(lc(white))) ///
50    ysc(range(.2 .325)) ylab(.2(.025).325) ///
51    xsc(range(-5 5)) xlab(-5(1)5) ///
52    xline(0, lcolor(red)) ysize(10) xsize(12)
53    graphr(c(white)) ///
54    ///    title("Impact of Concussion on Prob(Salary=0)") ///
55    yti("Prob(Salary=0)", height(7)) ///
56    xti("Years since exposure group concussion") scale(.95) ///
57    /*
58    note("Control groups suffer concussion 1, 2, 3, 4, and 5
59
60

```

```

1
2
3 years (&Delta) after exposure group." ///
4         "Both control and exposure group at 30–49 years of
5 age at beginning of year =0." ///
6         "Vertical line indicates time of exposure group
7 concussion.)*"/
8
9 graph export $highdef/FigureS3.png, replace width(3900)
10
11 graph twoway ///
12     connect norm_mean_no_l count if control1== 1, ///
13     lcolor(blue) mcolor(blue) || ///
14     connect norm_mean_no_l count if control1== 2, ///
15     lcolor(green) mcolor(green) || ///
16     connect norm_mean_no_l count if control1== 3, ///
17     lcolor(purple) mcolor(purple) || ///
18     connect norm_mean_no_l count if control1== 4, ///
19     lcolor(red) mcolor(red) || ///
20     connect norm_mean_no_l count if control1== 5, ///
21     lcolor(orange) mcolor(orange) || ///
22     connect mean_no_l count if control1== 0, ///
23     lcolor(black) mcolor(black) ///
24     legend( ///
25         label(1 "Control {&Delta}=1" ///
26             "N=`C1'" ) ///
27         label(2 "Control {&Delta}=2" ///
28             "N=`C2'" ) ///
29         label(3 "Control {&Delta}=3" ///
30             "N=`C3'" ) ///
31         label(4 "Control {&Delta}=4" ///
32             "N=`C4'" ) ///
33         label(5 "Control {&Delta}=5" ///
34             "N=`C5'" ) ///
35         label(6 "Exposure" ///
36             "N=`C0'" ) ///
37         c(1) order(6 1 2 3 4 5) pos(3) size(small) ///
38         c(1) symx(4) region(lc(white))) ///
39     ysc(range(.2 .325)) ylab(.2(.025).325) ///
40     xsc(range(-5 5)) xlab(-5(1)5) ///
41     xline(0, lcolor(red)) ysize(10) xsize(12)
42 graphr(c(white)) ///
43 ///     title("Impact of Concussion on Prob(Salary=0)") ///
44     yti("Norm. Prob(Salary=0)", height(7)) ///
45     xti("Years since exposure group concussion") scale(.95) ///
46 /*     note("Control groups suffer concussion 1, 2, 3, 4, and 5
47 years (&Delta) after exposure group." ///
48         "Both control and exposure group at 30–49 years of
49 age at beginning of year =0." ///
50         "Vertical line indicates time of exposure group
51 concussion." ///
52         "Probability levels normalized with pre-concussion
53 level difference between" ///
54         "each control group and the exposure group")*/
55
56 graph export $highdef/Figure2A.png, replace width(3900)
57
58
59
60

```

```

1
2
3
4
5
6
7
8
9
10 *****
11 *****
12 **
13 **      Effects across the salary distribution
14 **
15 **
16 **
17 *****
18 *****
19
20
21 forvalue control_time=5/5{
22     local end = 2012 // last incident year in data
23
24     **build dataset for joint estimate across years
25     forvalue count=2003/`end'{
26         if `count'==2003{
27             use $data/
28 sample_control_`count'_`control_time'.dta, clear
29             gen time = `count' //incident year
30 indicator
31             }
32             else append using $data/
33 sample_control_`count'_`control_time'.dta
34             replace time = `count' if time ==.
35
36
37             **exclude individuals in years where they do not
38 appear in data,
39             **due to either death or migration, as well as
40 periods from when
41             **the control group sufer their concussion
42 drop if merge ==1 | count > `control_time'-1
43         }
44
45         gen female = koen==2
46         qui{
47             gen edu = 0
48             replace edu = 1 if inrange(hffsp,20000000,39000000)
49 | ///
50
51                                     (hffsp
52 >40000000 & hffsp!=.)
53         }
54
55         //build ident, so we can multivariate cluster for
56 individuals
57         //who occur both as control and exposure during the period
58 (id)
59
60

```

```

1
2
3     bysort pnr time: gen helpx = _n ==1
4     gen id= sum(helpx)
5     drop helpx
6
7
8     **generate concussion variable
9     gen treat = inrange(count,0,`control_time'-1) & treatment
10
11 ==1
12     replace treat = time_from_incident if count ==0 & treatment
13 ==1
14
15     **effect across salary distribution
16     if `control_time'==5 matrix results = J(81,5,.) // matrix
17 to capture results
18
19     local n = `control_time'
20     gen post = count >=0
21     replace count = count+6
22     local v =0
23
24     qui{
25         forvalue t= 0(14932)1045240{
26             local top = 42
27             local v =1+`v'
28             if `n' ==5 matrix results[`v',1]= `t'
29             gen D = personindk <=`t'
30             reg D treat i.treatment i.post i.edu
31 i.alder i.female i.time, cl(pnr)
32             matrix V = e(V)
33             matrix b= e(b)
34             matrix results[`v',2] = b[1,1]
35             margins, at(treat= (0 1) treatment=1
36 post=1)
37             matrix results[`v',5] = V[1,1]^5
38             matrix M = r(b)
39             matrix results[`v',4] = M[1,2] //
40             matrix results[`v',3] = M[1,1] //
41             drop D
42         }
43     }
44 }
45
46
47 svmat results
48
49 gen l5 = results2-1.96*results5
50 gen u5 = results2+1.96*results5
51
52
53
54 replace results1 = results1/7466
55
56 gen dif6 = results4-results3
57 gen udif6 = u5
58 gen ldif6 = l5
59
60

```

```

1
2
3 gr two rline udif6 ldif6 results1, ///
4     yaxis(2) color(gray) lp(dash) ylab(0(.01).04, axis(2))
5 ysc(range(-.006 .04) axis(2)) || ///
6     line dif6 results1, yaxis(2) yline(0, axis(2)) ///
7     lcolor(black) ylab(0(.01).04) ysc(range(-.006 .04)
8 axis(2)) || ///
9     line results4 results1, yaxis(1) lcolor(red) || ///
10    line results3 results1, yaxis(1) lcolor(blue) ///
11    ysc(range(0 1)) ylab(0(.1)1, nogrid) xlab(0(10)140)
12 xsc(range(0 140)) ///
13    ysize(10) xsize(12) graphr(c(white)) ///
14    xti("Total Income in 1K Euro") scale(1) ///
15    yti("Effect of Concussion on Pr(Total Income < X)",
16 axis(2)) ///
17    yti("{&Phi}(Total Income)", axis(1)) ///
18    legend( ///
19    label(2 "Effect of concussion (left axis)") ///
20    label(3 "Concussion income distribution (right axis)") ///
21    label(4 "Counterfactual income distribution (right
22 axis)") ///
23    c(1) order(2 3 4) pos(6) size(small) ///
24    symx(4) region(lc(white)))
25
26
27
28
29

```

```

30 graph export [home]\highdef\l5_income.png, replace width(3900)
31
32

```

```

33 cap graph drop g1 g2
34

```

```

35 gr two ///
36     line results4 results1, yaxis(1) lcolor(red) || ///
37     line results3 results1, yaxis(1) lcolor(blue) ///
38     ysc(range(.2 1)) ylab(0(.1)1) xlab(0(20)140,
39 labs(small)) ///
40     xsc(range(0 140)) ///
41     ysize(10) xsize(10) graphr(c(white)) ///
42     xti("Total Income in 1K Euro") scale(1) ///
43     yti("Cumulative Distribution of Total Income",
44 axis(1)) ///
45     legend( ///
46     label(1 "Observed Post-Concussion Total Income
47 Distribution") ///
48     label(2 "Counterfactual No-Concussion Total Income
49 Distribution") ///
50     c(1) order(1 2 4) pos(6) size(small) ///
51     symx(4) region(lc(white))) , name(g1)
52

```

```

53 gr two rline udif6 ldif6 results1, ///
54     color(gray) lp(dash) ysc(range(-.002 .035)) || ///
55     line dif6 results1, yline(0) ///
56     lcolor(black) ylab(0(.01).0325) ysc(range(-.002 .
57 035)) ///
58     xlab(0(20)140, labs(small)) xsc(range(0 140)) ///
59     ysize(10) xsize(10) graphr(c(white)) ///
60

```



```

1
2
3     xti("Total Income in 1K Euro (X)") scale(1) ///
4     yti("Effect of Concussion on Pr(Total Income < X)") ///
5         legend( label(1 "95% CI") ///
6             label(2 "Effect of concussion") ///
7             c(1) order(2 1) pos(6) size(small) ///
8                 symx(4) region(lc(white))) , name(g2)
9
10    graph combine g1 g2      ,graphr(c(white))
11
12    graph export "[home]\highdef\figure s2.tif", replace width(1000)
13
14
15
16
17    *****
18    *****
19    **
20    **      Effects across the salary distribution
21    **
22    **
23    **
24    *****
25    *****
26
27
28    forvalue control_time=5/5{
29        local end = 2012 // last incident year in data
30
31        **build dataset for joint estimate across years
32        forvalue count=2003/`end'{
33            if `count'==2003{
34                use $data/
35            sample_control_`count'_`control_time'.dta, clear
36                gen time = `count' //incident year
37            indicator
38                }
39                else append using $data/
40            sample_control_`count'_`control_time'.dta
41                replace time = `count' if time ==.
42
43                **exclude individuals in years where they do not
44            appear in data,
45                **due to either death or migration, as well as
46            periods from when
47                **the control group suffer their concussion
48            drop if merge ==1 | count > `control_time'-1
49        }
50
51        gen female = koen==2
52        qui{
53            gen edu = 0
54            replace edu = 1 if inrange(hffsp,20000000,39000000)
55            | ///
56
57
58
59            >40000000 & hffsp!=.)
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

1
2
3     }
4
5     //build ident, so we can multivariate cluster for
6 individuals
7     //who occur both as control and exposure during the period
8 (id)
9
10    bysort pnr time: gen helpx = _n ==1
11    gen id= sum(helpx)
12    drop helpx
13
14
15
16    **generate concussion variable
17    gen treat = inrange(count,0,`control_time'-1) & treatment
18 ==1
19    replace treat = time_from_incident if count ==0 & treatment
20 ==1
21
22    **effect across salary distribution
23    if `control_time'==5 matrix results = J(81,5,.) // matrix
24 to capture results
25
26    local n = `control_time'
27    gen post = count >=0
28    replace count = count+6
29    local v =0
30
31
32    **Estimate Pr(salary < X) across income distribution
33    qui{
34        forvalue t= 0(14932)895950{
35            local top = 42
36            local v =1+`v'
37            if `n' ==5 matrix results[`v',1]= `t'
38            gen D = loenmv <=`t'
39            reg D treat i.treatment i.post i.edu
40 i.alder i.female i.time, cl(pnr)
41            matrix V = e(V)
42            matrix b= e(b)
43            matrix results[`v',2] = b[1,1]
44            margins, at(treat= (0 1) treatment=1
45 post=1)
46            matrix results[`v',5] = V[1,1]^5
47            matrix M = r(b)
48            matrix results[`v',4] = M[1,2] //
49            matrix results[`v',3] = M[1,1] //
50            drop D
51        }
52    }
53 }
54
55
56 svmat results
57
58 *Generate 95% confidence intervals
59 gen l5 = results2-1.96*results5
60

```

```

1
2
3 gen u5 = results2+1.96*results5
4
5 *Correct to €
6 replace results1 = results1/7466
7
8 *Obtain difference between observed and counterfactual wage
9 distribution
10 gen dif6 = results4-results3
11 gen udif6 = u5
12 gen ldif6 = l5
13
14
15 gr two rline udif6 ldif6 results1, ///
16     yaxis(2) color(gray) lp(dash) ysc(range(-.006 .04) axis(2))
17 || ///
18     line dif6 results1, yaxis(2) yline(0, axis(2)) ///
19     lcolor(black) ylab(0(.01).0325) ysc(range(-.006 .04)
20 axis(2)) || ///
21     line results4 results1, yaxis(1) lcolor(red) || ///
22     line results3 results1, yaxis(1) lcolor(blue) ///
23     ysc(range(0 1)) ylab(0(.1)1, nogrid) xlab(0(10)120)
24 xsc(range(0 120)) ///
25     ysize(10) xsize(12) graphr(c(white)) ///
26     xti("Salary in 1K Euro") scale(1) ///
27     yti("Effect of Concussion on Pr(Salary < X)", axis(2)) ///
28     yti("{&Phi}(Salary)", axis(1)) ///
29     legend( ///
30     label(2 "Effect of concussion (left axis)") ///
31     label(3 "Concussion salary distribution (right axis)") ///
32     label(4 "Counterfactual salary distribution (right
33 axis)") ///
34     c(1) order(2 3 4) pos(6) size(small) ///
35     symx(4) region(lc(white)))
36
37
38 cap graph drop g1 g2
39
40 gr two ///
41     line results4 results1, yaxis(1) lcolor(red) || ///
42     line results3 results1, yaxis(1) lcolor(blue) ///
43     ysc(range(.2 1)) ylab(0.2(.1)1) xlab(0(10)120,
44 labs(small)) ///
45     xsc(range(0 120)) ///
46     ysize(10) xsize(10) graphr(c(white)) ///
47     xti("Salary in 1K Euro") scale(1) ///
48     yti("Cumulative Distribution of Salary", axis(1)) ///
49     legend( ///
50     label(1 "Observed Post-Concussion Salary Distribution") ///
51     label(2 "Counterfactual No-Concussion Salary
52 Distribution") ///
53     c(1) order(1 2 4) pos(6) size(small) ///
54     symx(4) region(lc(white))) , name(g1)
55
56
57 gr two rline udif6 ldif6 results1, ///
58     color(gray) lp(dash) ysc(range(-.002 .035)) || ///
59     line dif6 results1, yline(0) ///
60

```

```

1
2
3         lcolor(black) ylab(0(.01).0325) ysc(range(-.002 .
4 035)) ///
5         xlabel(0(10)120, labs(small)) xsc(range(0 120))      ///
6         ysize(10) xsize(10) graphr(c(white)) ///
7         xti("Salary in 1K Euro (X)") scale(1) ///
8         yti("Effect of Concussion on Pr(Salary < X)") ///
9             legend( label(1 "95% CI") ///
10                label(2 "Effect of concussion") ///
11                c(1) order(2 1) pos(6) size(small) ///
12                    symx(4) region(lc(white))) , name(g2)
13
14 graph combine g1 g2          ,graphr(c(white))
15
16 graph export "[home]\highdef\figure 3.tif",  replace width(1000)
17
18
19
20
21
22
23 **ESTIMATE SICK LEAVE
24
25 forvalue control_time=1/5{
26     local end = 2012 // last incident year in data
27     if `control_time' ==1     eststo clear
28
29     **build dataset for joint estimate across years
30     forvalue count=2003/`end'{
31         if `count'==2003{
32             use $data/
33             sample_control_`count'_`control_time'.dta, clear
34             gen time = `count' //incident year
35             indicator
36                 }
37             else append using $data/
38             sample_control_`count'_`control_time'.dta
39             replace time = `count' if time ==.
40
41
42             **exclude individuals in years where they do not
43 appear in data,
44             **due to either death or migration, as well as
45 periods from when
46             **the control group sufer their concussion
47             drop if merge ==1 | count > `control_time'-1
48         }
49
50
51     merge m:1 pnr t using $data/temp.dta, keep(1 3) nogen
52     replace share = 0 if share==.
53
54     gen female = koen==2
55
56     //build ident, so we can multivariate cluster for
57 individuals
58     //who occur both as control and exposure during the period
59 (id)
60

```

```

1
2
3
4     bysort pnr time: gen helpx = _n ==1
5     gen id= sum(helpx)
6     drop helpx
7
8
9
10    **Generate educational groups
11    qui{
12        gen edu =0
13        replace edu = 1 if inrange(hffsp,20000000,39000000)
14    | ///
15                                          (hffsp
16 >40000000 & hffsp!=.)
17    }
18
19
20    **Calculate number of observations for exposure and control
21    count if count==0 & treatment ==1
22    local Ntreated = r(N)
23    count if count==0 & treatment ==0
24    local Ncontrol = r(N)
25
26    **generate concussion variable
27    gen treat = inrange(count,0,`control_time'-1) & treatment
28
29 ==1
30
31 ==1
32
33    **Generate pre-concussion income difference          for
34    **use in calculating marginal effects
35    sum loenmv if count <0 & treatment ==0
36    local control =r(mean)
37    sum loenmv if count <0 & treatment ==1
38    local treat =r(mean)
39    sum loenmv if count>=0 & treatment ==0
40    local control_post =r(mean)
41
42    forvalue t=-4/4{
43        local n = `t'*-1
44        if `t' <-1 gen T_`n' = treatment ==1 & count ==`t'
45        if `t' > -1 gen T`t' = treatment ==1 & count ==`t'
46
47    }
48
49
50    **estimate DiD model on salary
51    reghdfe share T*, abs(alder female count time treatment
52 edu) cl(pnr id)
53    eststo est1_`control_time'
54    if `control_time'==1 matrix results = J(5,5,.) // matrix to
55 capture results
56    if `control_time'==1 matrix results_p = J(5,5,.) // matrix
57 to capture results
58
59
60

```

```

1
2
3     matrix b = e(b)
4     matrix V = e(V)
5     local n = `control_time'
6
7     matrix results[`n',1] = b[1,1] / 7466 // capture beta
8 results as 1K Euro
9     matrix results[`n',2] = (V[1,1]^5)/7466 //
10 capture standard error as 1K Euro
11     matrix results[`n',3] = b[1,1]/(`control_post'-
12 (`control'-`treat'))
13     matrix results[`n',4] = `n'
14
15
16     gen no_share = share >0 //dummy for no salary
17
18
19     **Generate pre-concussion probability difference for
20     **use in calculating marginal effects
21     sum no_share if count <0 & treatment ==0
22     local control =r(mean)
23     sum no_share if count <0 & treatment ==1
24     local treat =r(mean)
25     sum no_share if count>=0 & treatment ==0
26     local control_post =r(mean)
27
28
29     **estimate DiD LP-model on P(salary=0)
30     xi: reghdfe no_share T*, abs(alder female count time
31 treatment edu) cl(pnr id)
32     eststo est2_`control_time'
33     matrix b =e(b) //regression coefficient
34     matrix V = e(V) // standard error^2
35
36     matrix results_p[`n',1] = b[1,1]
37     matrix results_p[`n',2] = V[1,1]^5
38     matrix results_p[`n',3] = b[1,1]/(`control_post'-
39 (`control'-`treat'))
40     matrix results_p[`n',4] = `n'
41
42 }
43
44     esttab est1_* using [home]/tables/dynamic_share1.rtf, ///
45     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
46 *** .001) ///
47     keep(T*)
48
49     esttab est2_* using [home]/tables/dynamic_share2.rtf, ///
50     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
51 *** .001) ///
52     keep(T*)
53
54
55 **ESTIMATE SICK LEAVE
56
57 forvalue control_time=1/5{
58     local end = 2012 // last incident year in data
59     if `control_time' ==1 eststo clear
60

```

```

1
2
3
4     **build dataset for joint estimate across years
5     forvalue count=2003/`end'{
6         if `count'==2003{
7             use $data/
8 sample_control_`count'_`control_time'.dta, clear
9             gen time = `count' //incident year
10 indicator
11         }
12         else append using $data/
13 sample_control_`count'_`control_time'.dta
14         replace time = `count' if time ==.
15
16         **exclude individuals in years where they do not
17 appear in data,
18         **due to either death or migration, as well as
19 periods from when
20         **the control group suffer their concussion
21         drop if merge ==1 | count > `control_time'-1
22     }
23
24
25     merge m:1 pnr t using $data/temp2.dta, keep(1 3) nogen
26     replace syg_barsel_13= 0 if syg_barsel_13==.
27
28     gen female = koen==2
29
30     //build ident, so we can multivariate cluster for
31 individuals
32     //who occur both as control and exposure during the period
33 (id)
34
35
36     bysort pnr time: gen helpx = _n ==1
37     gen id= sum(helpx)
38     drop helpx
39
40
41     **Generate educational groups
42     qui{
43         gen edu =0
44         replace edu = 1 if inrange(hffsp,20000000,39000000)
45 | ///
46                                     (hffsp
47 >40000000 & hffsp!=.)
48     }
49
50
51
52     **Calculate number of observations for exposure and control
53     count if count==0 & treatment ==1
54     local Ntreated = r(N)
55     count if count==0 & treatment ==0
56     local Ncontrol = r(N)
57
58     **generate concussion variable
59     gen treat = inrange(count,0,`control_time'-1) & treatment
60

```

```

1
2
3 ==1
4 replace treat = time_from_incident if count ==0 & treatment
5 ==1
6
7 **Generate pre-concussion income difference for
8 **use in calculating marginal effects
9
10 sum loenmv if count <0 & treatment ==0
11 local control =r(mean)
12 sum loenmv if count <0 & treatment ==1
13 local treat =r(mean)
14 sum loenmv if count >=0 & treatment ==0
15 local control_post =r(mean)
16
17 forvalue t=-4/4{
18     local n = `t'*-1
19     if `t' < -1 gen T_`n' = treatment ==1 & count ==`t'
20     if `t' > -1 gen T`t' = treatment ==1 & count ==`t'
21
22 }
23
24 **estimate DiD model on salary
25 reghdfe syg_barsel_13 T*, abs(alder female count time
26 treatment edu) cl(pnr id)
27 eststo est1_`control_time'
28 if `control_time'==1 matrix results = J(5,5,.) // matrix to
29 capture results
30 if `control_time'==1 matrix results_p = J(5,5,.) // matrix
31 to capture results
32
33
34
35 matrix b = e(b)
36 matrix V = e(V)
37 local n = `control_time'
38
39 matrix results[`n',1] = b[1,1] / 7466 // capture beta
40 results as 1K Euro
41 matrix results[`n',2] = (V[1,1]^5)/7466 //
42 capture standard error as 1K Euro
43 matrix results[`n',3] = b[1,1]/(`control_post'-
44 (`control'-`treat'))
45 matrix results[`n',4] = `n'
46
47
48
49
50
51 **Generate pre-concussion probability difference for
52 **use in calculating marginal effects
53
54 **estimate DiD LP-model on P(salary=0)
55 xi: reghdfe kont_dag T*, abs(alder female count time
56 treatment edu) cl(pnr id)
57 eststo est2_`control_time'
58 matrix b =e(b) //regression coefficient
59 matrix V = e(V) // standard error^2
60

```



```
1
2
3
4         matrix results_p[`n',1] =      b[1,1]
5         matrix results_p[`n',2] =      V[1,1]^5
6         *matrix results_p[`n',3] =      b[1,1]/(`control_post'-
7 (`control'-`treat'))
8         matrix results_p[`n',4] =      `n'
9
10      }
11
12      esttab est1_* using [home]/tables/dynamic_sickpay1.rtf, ///
13          replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
14 *** .001) ///
15          keep(T*)
16
17
18      esttab est2_* using [home]/tables/dynamic_welfare2.rtf, ///
19          replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
20 *** .001) ///
21          keep(T*)
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Supplemental results**

For peer review only

**Table S1. Test of parallel trends assumption pre-exposure incident against each control group separately using eq. S3 in supplementary methods. Separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period).**

Time to exposure group's concussion (exposure)	$\Delta=1$ Est (S.E.) p-value	$\Delta=2$ Est (S.E.) p-value	$\Delta=3$ Est (S.E.) p-value	$\Delta=4$ Est (S.E.) p-value	$\Delta=5$ Est (S.E.) p-value
Exposure-4y	-0.368 (0.226) p=.104	0.046 (0.363) p=.900	0.120 (0.362) p=.741	0.159 (0.313) p=.612	0.042 (0.329) p=.899
Exposure-3y	-0.094 (0.317) p=.768	0.227 (0.510) p=.656	0.167 (0.354) p=.637	0.537 (0.372) p=.148	0.113 (0.393) p=.774
Exposure-2y	-0.548 (0.312) p=.079	-0.082 (0.347) p=.812	-0.163 (0.236) p=.491	-0.124 (0.247) p=.617	0.082 (0.250) p=.744
Exposure-1y	Ref.	Ref.	Ref.	Ref.	Ref.
<b>N*T</b>	284115	273725	266120	260647	256337

Note: The table shows test for differences in pre-exposure trends between exposure and control group model using interactions between pre-exposure time dummies and the exposure indicator. There is no indication of substantial or significant pre-exposure differences in salary trajectories between exposure group and any of the control groups.

**Table S2.** Effect of concussion on different labor market outcome parameters using separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period): In this exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group. Outcomes include annual salaried income (annual salary), total annual income (total income), annual sick leave benefits received (sick leave benefits) as well as a binary indicator of employment (probability of employment). Monetary outcomes were measured at 2015-level in € 1,000.

Time to exposure group's concussion (exposure)	Annual Salary Est. (S.E.) p-value	Total income Est. (S.E.) p-value	Sick leave benefits Est. (S.E.) p-value	Probability of employment Est. (S.E.) p-value
Exposure-4y	0.012 (0.212) p=.954	0.164 (0.173) p=.343	0.035 (0.036) p=.320	0.001 (0.004) p=.803
Exposure-3y	0.059 (0.252) p=.814	0.305 (0.233) p=.190	0.022 (0.034) p=.529	-0.001 (0.003) p=.739
Exposure-2y	0.043 (0.160) p=.788	0.122 (0.147) p=.405	0.002 (0.029) p=.946	0.001 (0.003) p=.739
Exposure-1y				
Exposure	-0.611 (0.168) p<.001	-0.338 (0.140) 0.016	0.166 (0.030) p<.001	-0.003 (0.003) p=.317
Exposure+1y	-1.389 (0.209) p<.001	-0.608 (0.162) p<.001	0.288 (0.039) p<.001	-0.020 (0.003) p<.001
Exposure+2y	-1.568 (0.261) p<.001	-0.847 (0.231) p<.001	0.132 (0.039) p=.001	-0.023 (0.004) p<.001
Exposure+3y	-1.393 (0.246) p<.001	-0.497 (0.219) p=.023	0.031 (0.040) p=.432	-0.022 (0.004) p<.001
Exposure+4y	-1.319 (0.253) p<.001	-0.499 (0.218) p=.022	-0.076 (0.042) p=.075	-0.018 (0.004) p<.001
<b>N*T</b>	<b>577762</b>	<b>577758</b>	<b>577872</b>	<b>577872</b>

**Note:** Annual salary include all income from salary and employee fringe benefits, employee stock options, employer paid sick leave, net gains (including interests and capital gains) from own companies. Total income includes all income absent wealth. Sick leave includes only public health benefits (sick leave and paternity leave). Employment is a binary indicator measured last week of November for each year. Results obtained from estimations following Eq. (1). Models include controls for high school diploma, gender, age, and observation year. Results obtained using reghdfe in Stata. Total number of observations (N\*T) differ slightly between outcomes because all income information is not available for all observation all years.

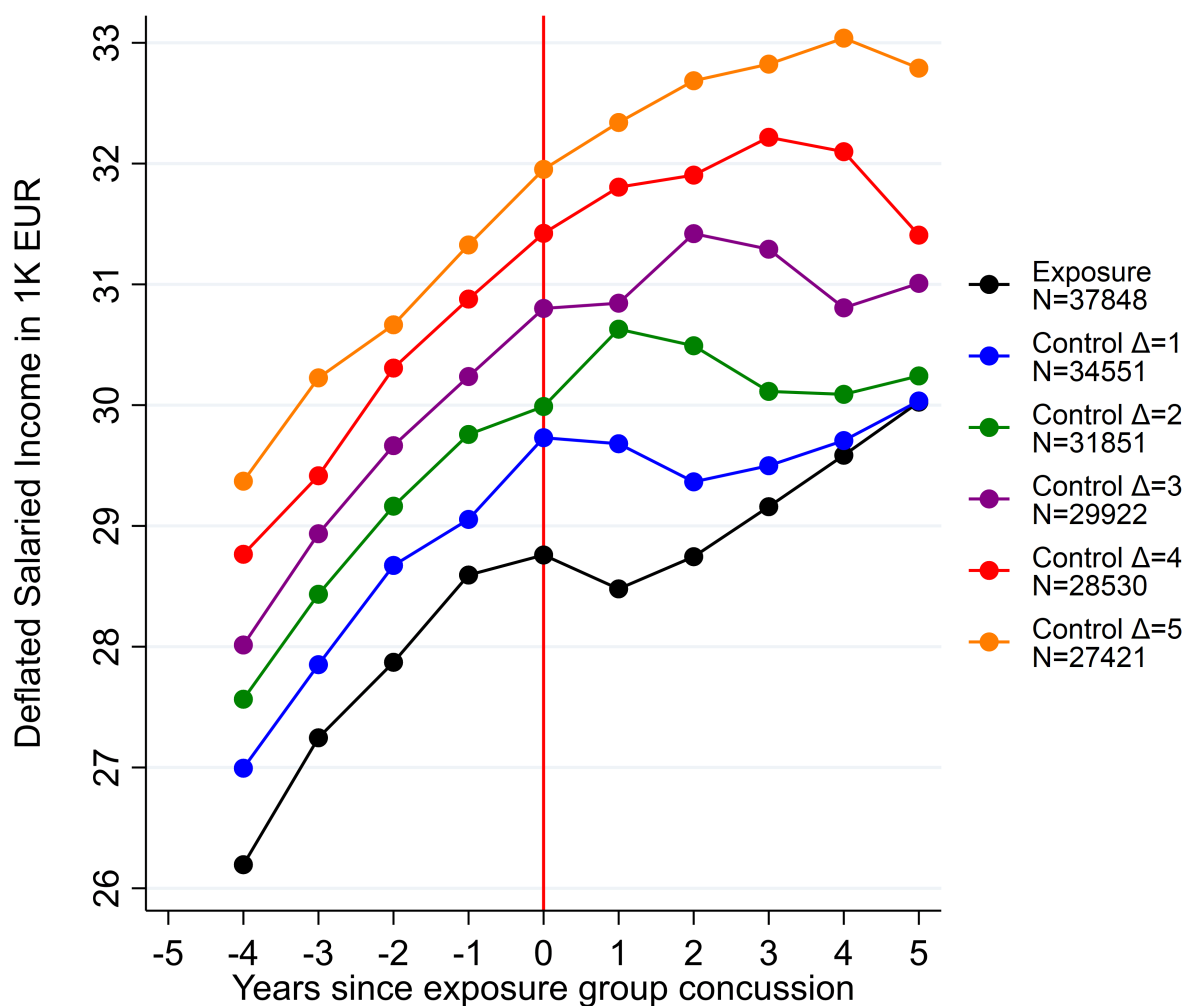
Source: Own calculations on data from Statistics Denmark.

**Table S3. Demographic factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years leading up to the concussion event in each of the groups. Factors include patient age (in years), share of sample female (1=100% female), and share of individuals with at least a high school degree (1=100%).**

		Exposure	Control, $\Delta=1$	Control, $\Delta=2$	Control, $\Delta=3$	Control, $\Delta=4$	Control, $\Delta=5$
<b>Pr(Female=1)</b>	Mean	.430	.438	.447	.458	.464	.473
	S.D.	(.495)	(.496)	(.497)	(.498)	(.499)	(.499)
	p-value		.030	<.001	<.001	<.001	<.001
<b>Age</b>	Mean	36.899	37.354	37.754	38.065	38.343	38.592
	S.D.	(11.856)	(11.857)	(11.718)	(11.630)	(11.584)	(11.491)
	p-value		<.001	<.001	<.001	<.001	<.001
<b>Pr(High school=1)</b>	Mean	.624	.632	.640	.646	.653	.660
	S.D.	(.484)	(.482)	(.480)	(.478)	(.476)	(.474)
	p-value		.026	<.001	<.001	<.001	<.001
<b>Total individuals</b>		37848	34551	31851	29922	28580	27484

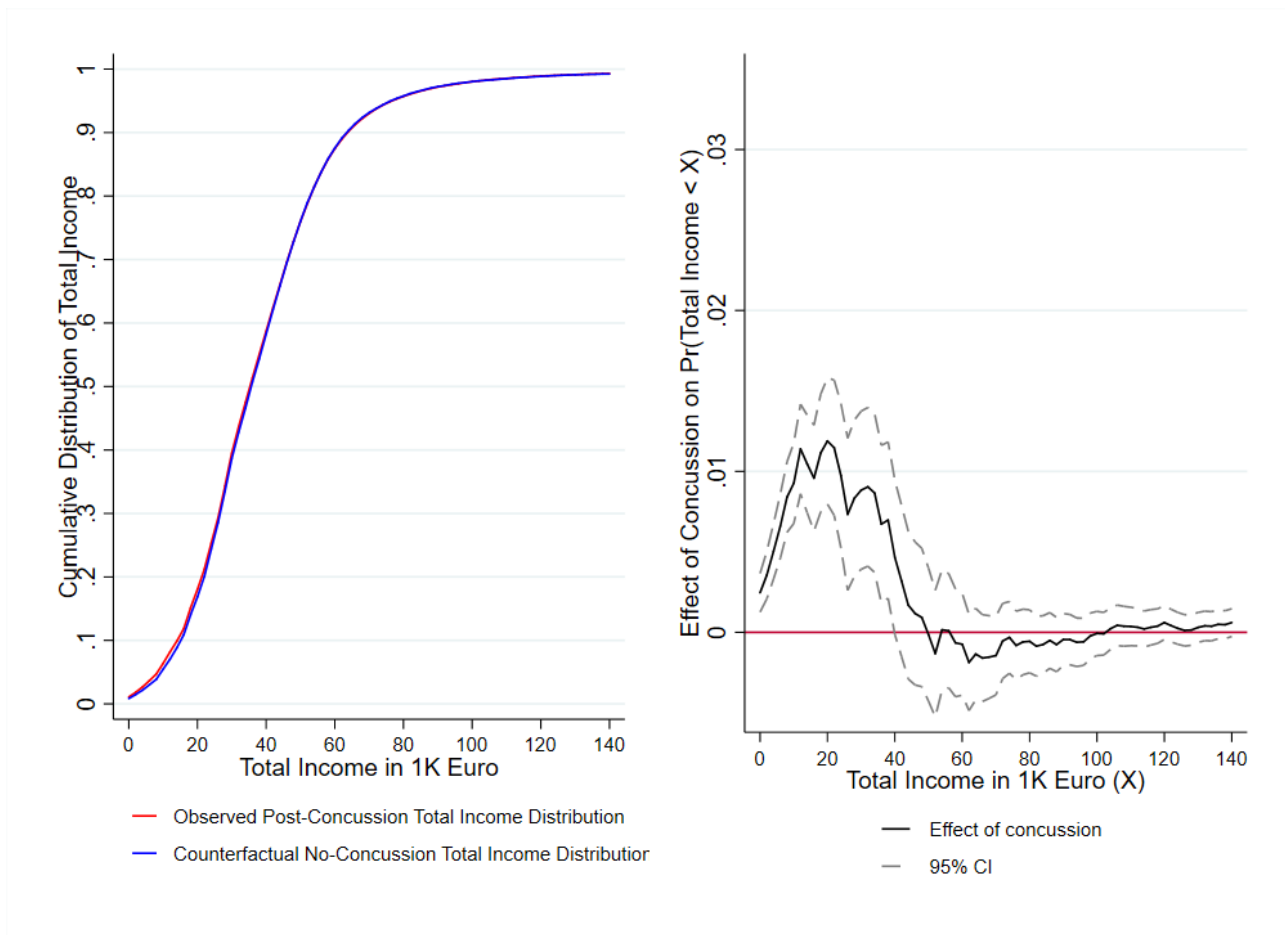
**Note:** S.D.: Standard deviation. P-values calculated using two-sided t-tests. All test performed between exposure group and each control group separately.

Figure S1. Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels



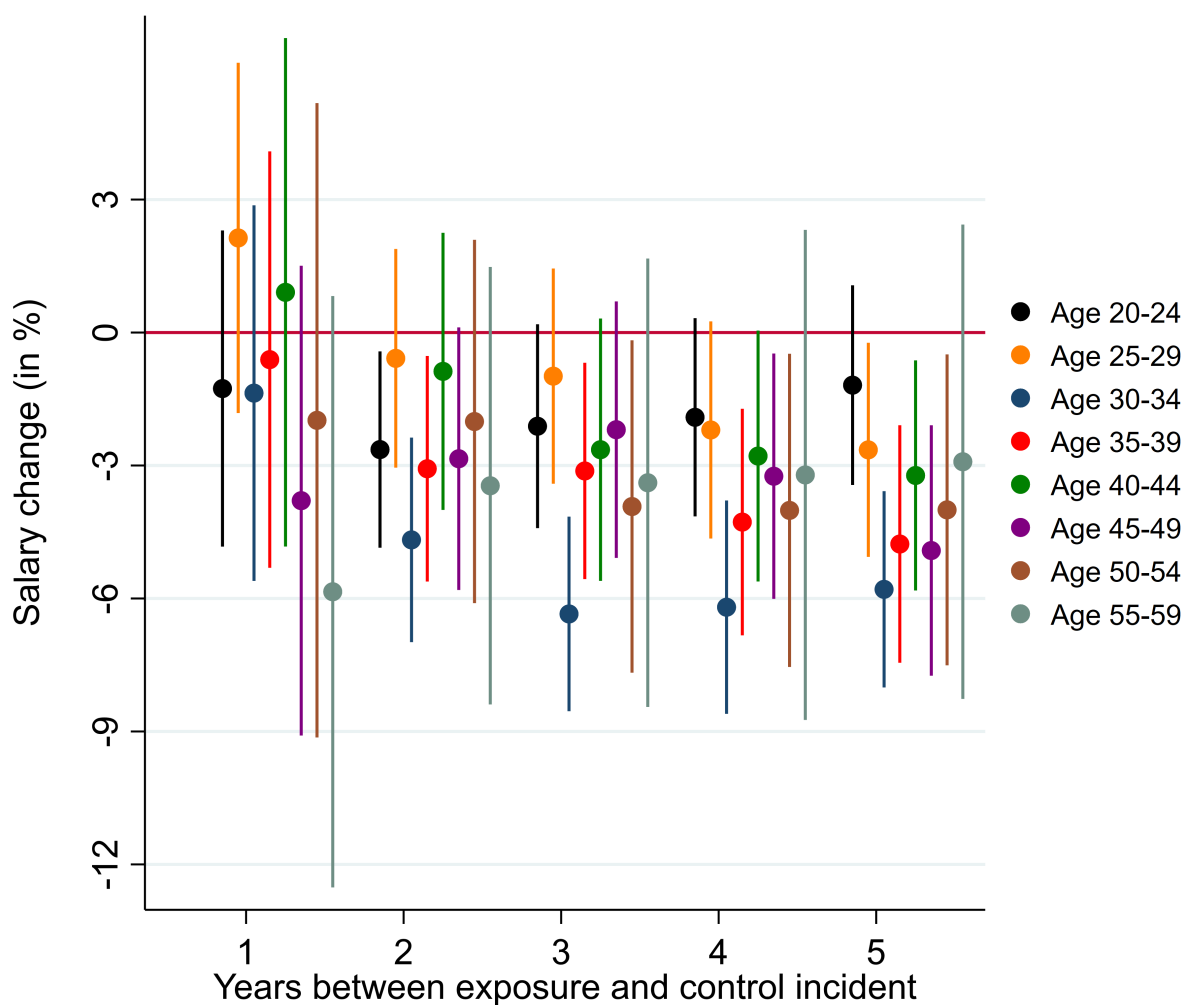
**Note:** Salary of the exposure group compared to salary of the 5 control groups, who experienced their concussions  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  years later than the exposure group. Salary progression is shown for the 5 years before and the 5 years after the exposure group suffered a concussion event. Table S1 demonstrates that the trends for salary progression pre-exposure incident are parallel between exposure group and each control group.

**Figure S2. (Left Panel) The Cumulative Distribution for Total Income Post-Treatment among the Treatment Group and Their Counterfactual, and (Right Panel) the Difference between the Two CDFs Expressed as the Effect of Concussion on the Probability of Total Income Below that Income-Level Expressed on the X-Axis following Exposure Event.**



**Note:** The figure shows the observed cumulative salary distribution following a concussion (red) and the expected counterfactual salary distribution absent the concussion (blue). The black line shows the difference between the observed and the counterfactual distribution, and the grey dash lines show the 95 % confidence interval. The bell-shape of the difference between the two distributions as the total income increase from 0 to 40,000 € indicates that the main part of the effect of concussions on total incomes is driven by low-income people shifting total income downwards following concussion, but not going to total income equal to zero.

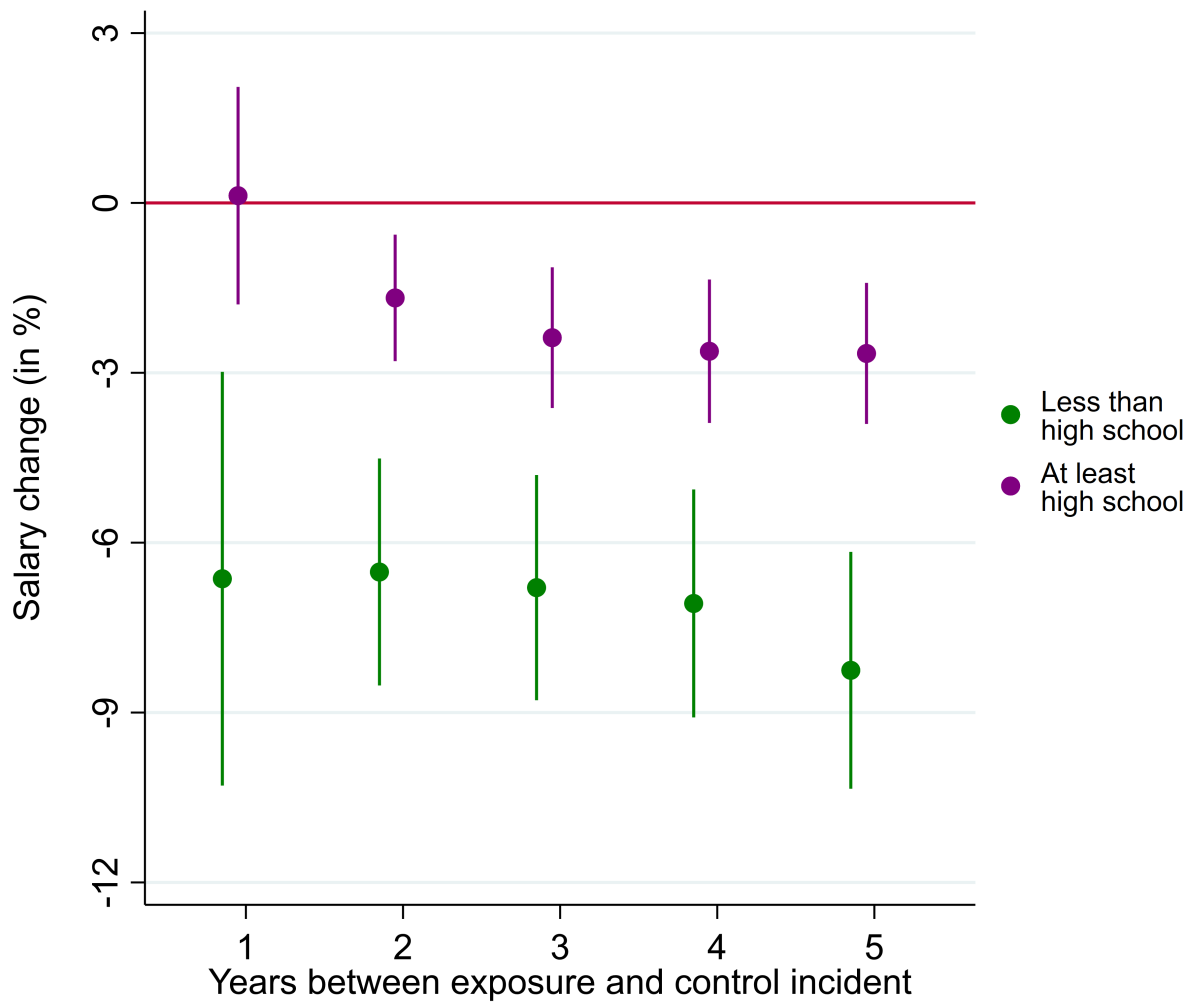
Figure S3. Percentage Effect of Concussion on Relative Salary Across Age Groups.



**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

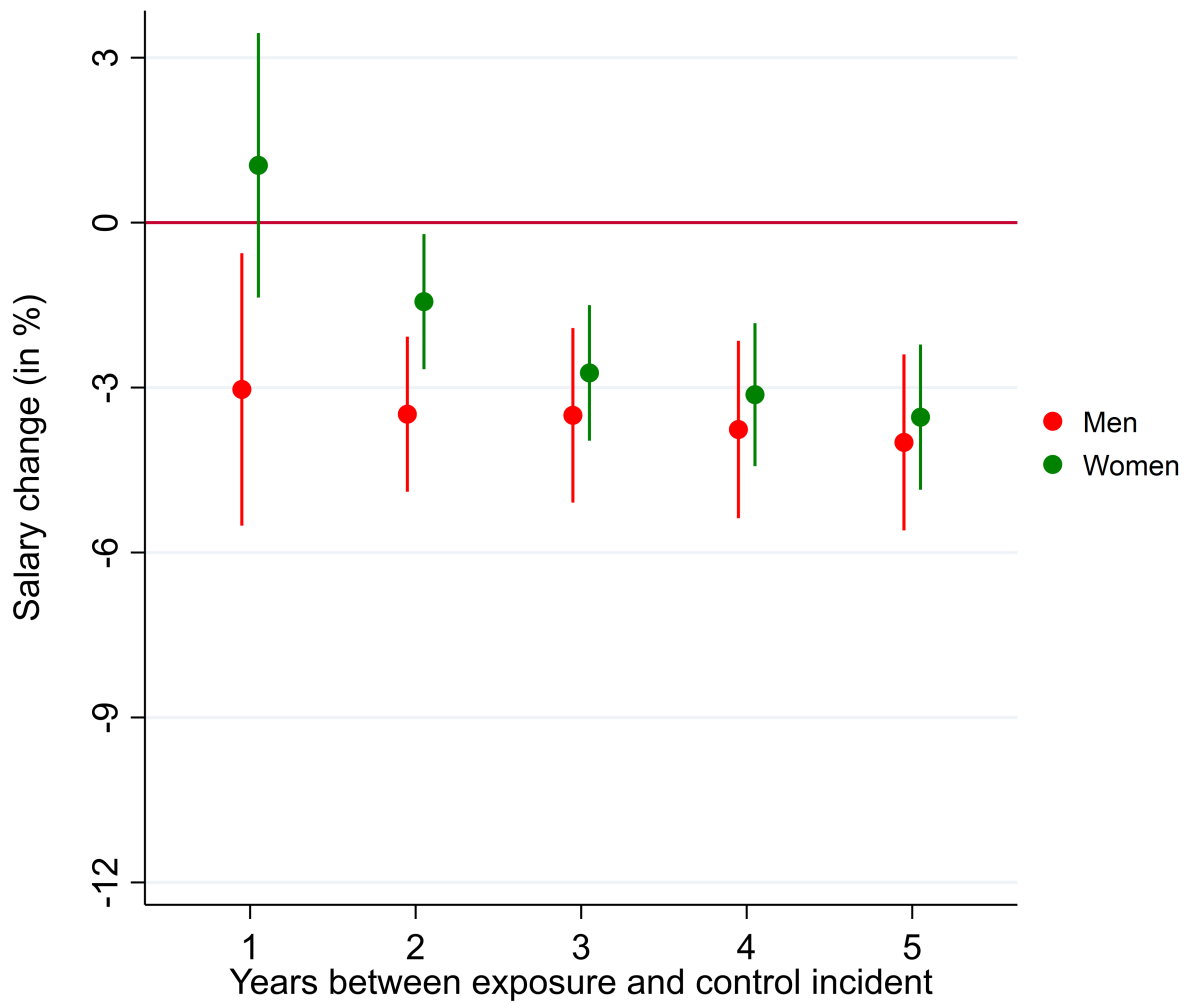


**Figure S4. Percentage Effect of Concussion on Relative Salary Across High School Completion.**



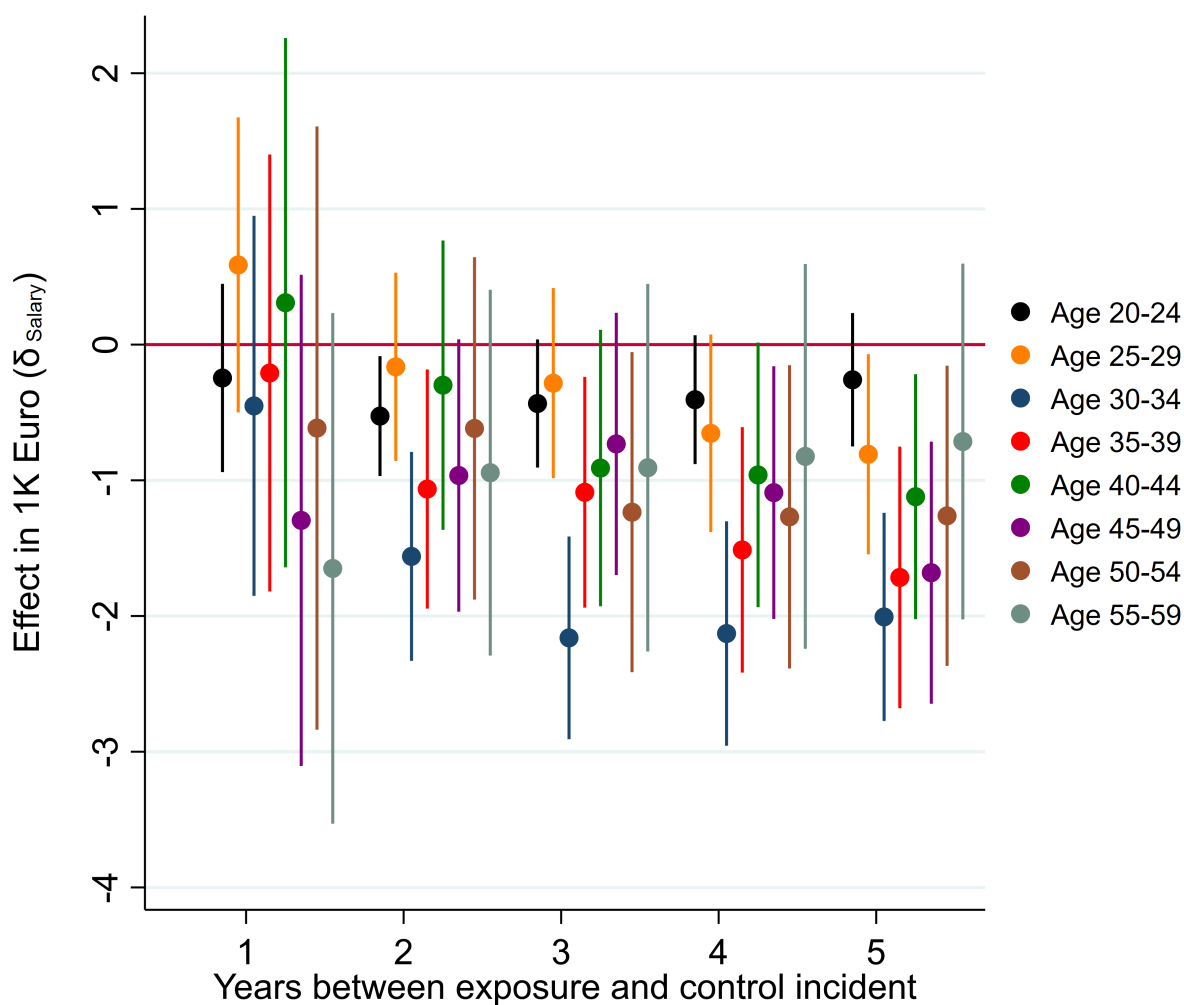
**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

Figure S5. Percentage Effect of Concussion on Relative Salary Across Gender.

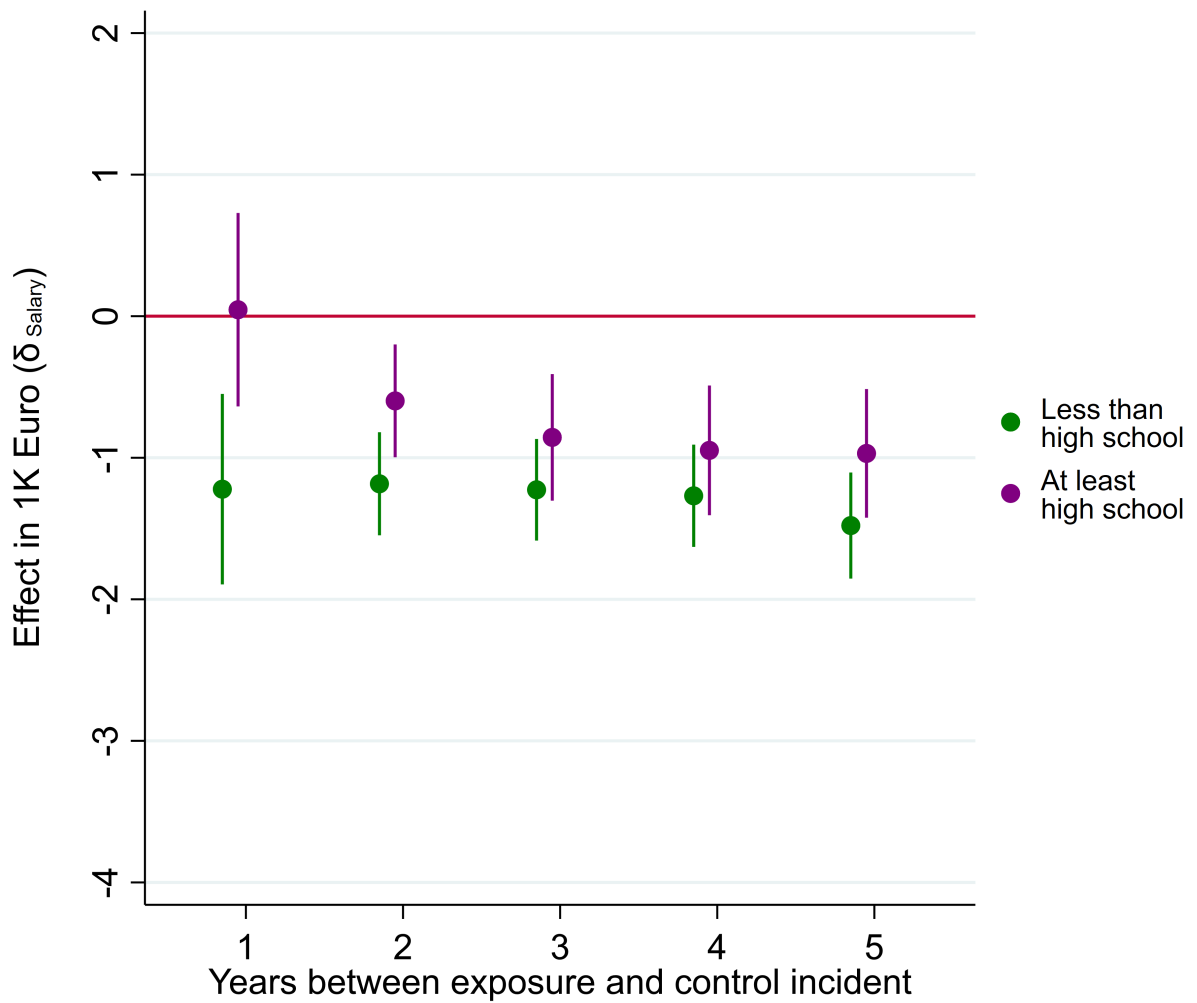


**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across gender. Graph shows parameter estimates and 95% CI.

Figure S6. Effect of Concussion on Absolute Salary in 1K Euro Across Age groups.

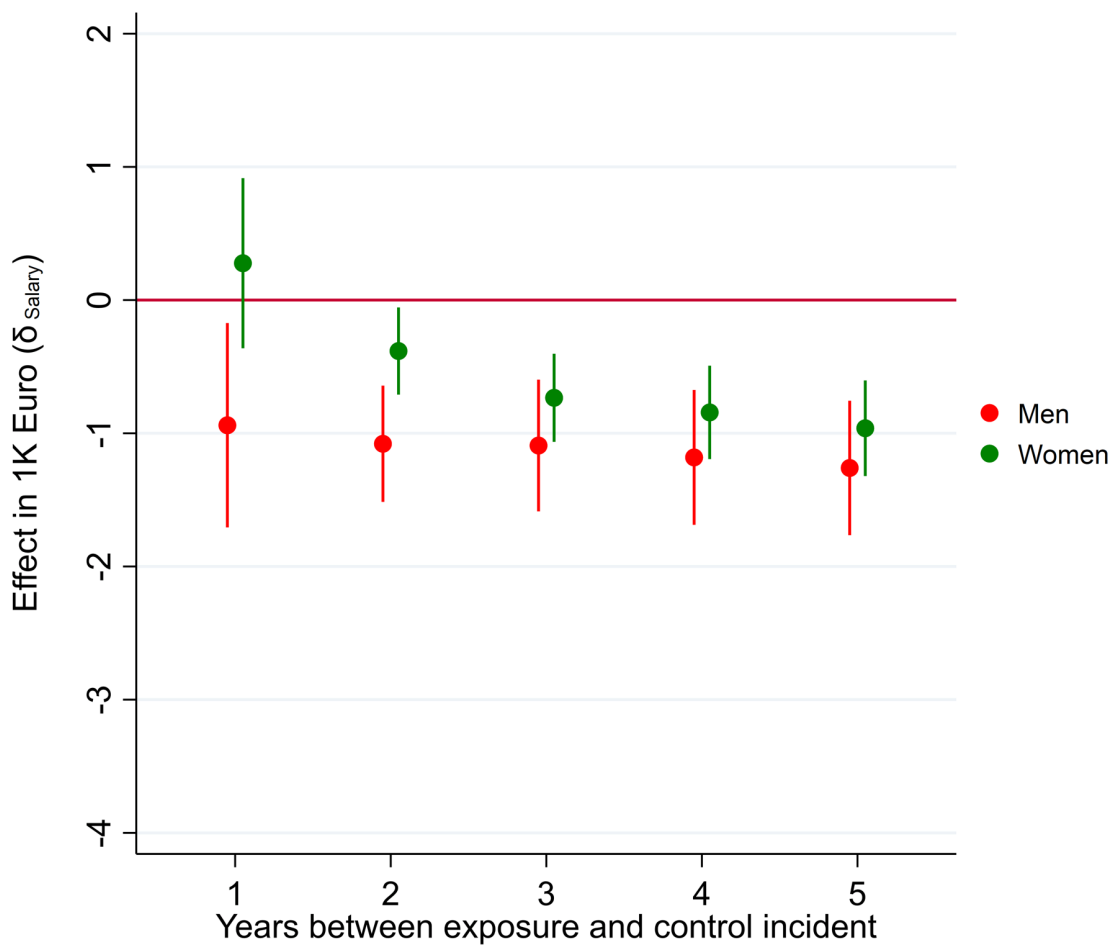


**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

**Figure S7. Effect of Concussion on Absolute Salary in 1K Euro Across Education.**

**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

**Figure S8. Effect of Concussion on Absolute Salary in 1K Euro Across Gender.**



**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across gender. Graph shows parameter estimates and 95% CI.

**The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.**

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
<b>Title and abstract</b>					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	title abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.  RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.  RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	title abstract  title abstract  abstract
<b>Introduction</b>					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	abstract introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	introduction		
<b>Methods</b>					
Study Design	4	Present key elements of study design early in the paper	introduction		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	materials and methods		

Participants	6	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p><i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p> <p><i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>	materials and methods	<p>RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.</p> <p>RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.</p> <p>RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.</p>	<p>materials and methods</p> <p>materials and methods</p> <p>not included</p>
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	materials and methods main text	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	materials and methods
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	materials and methods		

1 2 3	Bias	9	Describe any efforts to address potential sources of bias	materials and methods and results	
4 5	Study size	10	Explain how the study size was arrived at	materials and methods	
6 7 8 9 10 11	Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	materials and methods	
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35	Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	a) materials and methods b) materials and methods c) materials and methods d-e) NA	
36 37 38 39 40 41 42 43 44	Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.  materials and methods  materials and methods



				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	materials and methods
<b>Results</b>					
Participants	13	(a) Report the numbers of individuals at each stage of the study ( <i>e.g.</i> , numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a-c) materials and methods	RECORD 13.1: Describe in detail the selection of the persons included in the study ( <i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	materials and methods
Descriptive data	14	(a) Give characteristics of study participants ( <i>e.g.</i> , demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time ( <i>e.g.</i> , average and total amount)	a) materials and methods and Supplementary Table S3 b) materials and methods, Table 1 c) materials and methods		
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure	materials and methods and Supplementary Table S3, results		

		category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	a) results b) results c) results		
Other analyses	17	Report other analyses done— e.g., analyses of subgroups and interactions, and sensitivity analyses	results		
<b>Discussion</b>					
Key results	18	Summarise key results with reference to study objectives	results and discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	results and discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	NA
Interpretation	20	Give a cautious overall interpretation of results considering objectives,	discussion		

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results	discussion		
<b>Other Information</b>					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Title page		
Accessibility of protocol, raw data, and programming code		..	Reference to supplementary data throughout the text	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	materials and methods

\*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

\*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

# BMJ Open

## The Effect of Concussion on Salary and Employment-A Population-Based Event Time Study using a Quasi- Experimental Design

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038161.R2
Article Type:	Original research
Date Submitted by the Author:	10-Sep-2020
Complete List of Authors:	Fallesen, Peter; Stockholm University, Swedish Institute for Social Research Campos, Benito; Heidelberg University Hospital Department of Neurosurgery; Technical University of Denmark, Department of Health Tech
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Neurology
Keywords:	HEALTH ECONOMICS, Neurosurgery < SURGERY, Trauma management < ORTHOPAEDIC & TRAUMA SURGERY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 1 **The Effect of Concussion on Salary and Employment-A Population-Based Event Time**  
5 2 **Study using a Quasi-Experimental Design**  
6  
7  
8 3

9  
10 4 <sup>1,2</sup>Peter Fallesen, PhD, <sup>3,4</sup>Benito Campos, M.D.  
11  
12 5

13  
14  
15 6 <sup>1</sup>Swedish Institute for Social Research, Stockholm University, SE-106 91 Stockholm, Sweden

16 7 <sup>2</sup>ROCKWOOL Foundation, Ny Kongensgade 6, DK-1472 Copenhagen C, Denmark

17 8 <sup>3</sup>Department of Neurosurgery, University of Heidelberg, INF400, 69120 Heidelberg, Germany

18 9 <sup>4</sup>Department of Health Tech, Technical University of Denmark, DK-2800 Kongens Lyngby,  
19 10 Denmark  
20  
21 10  
22 11

23  
24 12 **Address of Correspondence:** Peter Fallesen, Swedish Institute for Social Research,  
25 13 Stockholm University, SE-106 91 Stockholm, Sweden, [peter.fallesen@sofi.su.se](mailto:peter.fallesen@sofi.su.se),  
26 14 +4561333184  
27 15

28  
29 16 **Keywords:** administrative data; concussion; income; productivity  
30  
31

32 17 **Abstract word count: 245**  
33

34 18 **Text word count: 3989 (excluding references and figure/table legends)**  
35

36 19 **Number of references: 21**  
37

38 20 **Number of tables and/or figures: 2 tables and 4 figures**  
39  
40

41 21 **Number of videos: 0**  
42  
43 22  
44 23  
45 24  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 25 **Abstract**

26 **Objective:** Concussions are the most frequent traumatic brain injuries. Yet, the socio-economic  
27 impact of concussions remains unclear. Socio-economic effects of concussions on working age  
28 adults were studied on a population scale.

29 **Design:** This population-based, event time study uses administrative data as well as hospital  
30 and emergency room records for the population of Denmark.

31 **Setting:** We study all Danish patients, aged 20-59 y, who were treated at a public hospital or  
32 at an emergency room between 2003-2017 after suffering a concussion without other  
33 intracranial or extracranial injuries (n=55,424 unique individuals). None of the patients had a  
34 prior diagnosis of intra- or extracranial injuries within the past ten years leading up to the  
35 incident.

36 **Primary and Secondary Outcome Measures:** As primary endpoint, we investigate the mean  
37 effect of concussion on annual salaried income within a five-year period after trauma. In an  
38 exploratory analysis, we study whether the potential impact of concussion on annual salaried  
39 income is driven by patient age, education, or economic cycle.

40 **Results:** Concussion was associated with an average change in annual salary income of -  
41 1,223€ (95% CI, -1,540€; -905€, p<.001) corresponding to a salary change of -4.2 % (95% CI,  
42 -5.2 %; -3.1 %). People between 30-39 y and those without high school degrees suffered the  
43 largest salary decreases. Affected individuals leaving the workforce drove the main part of the  
44 decrease. Absolute annual effect sizes were countercyclical to the unemployment rate.

45 **Conclusions:** Concussions have a large and long-lasting impact on salary and employment of  
46 working-age adults on a nationwide scale.

1  
2  
3  
4 47  
56 48 **Strengths and limitations of this study**  
7

- 8 49 - Natural experiments used to obtain plausible causal effects between concussion and  
9 salary/employment.  
10 50  
11  
12 51 - Large, population-based sample with multiple data layers.  
13  
14  
15 52 - Analysis includes how economic cycles affect outcome measures.  
16  
17 53 - Data only captures concussions registered in ERs and hospitals.  
18  
19  
20 54 - Because concussions do not occur at random, causal estimate relies on stronger  
21 assumptions than for a randomized control trial.  
22 55  
23

24 56  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## 57 Introduction

58 Concussions are by far the most frequently occurring intracranial injuries affecting  
59 approximately 450 to 600 per 100,000 people every year[1] . Immediate symptoms may last  
60 for days or weeks. Further, Danish cohort data[2] indicates that 10-15% of patients diagnosed  
61 with concussion suffer from long-term symptoms such as headache, fatigue, and intolerance to  
62 stress, whereas other studies place the upper bound as high as 30%[3,4]. Clinical practice has  
63 encouraged patients to restrict social, mental, and physical activity in the weeks following a  
64 concussion (see[5] for review), although prolonged inactivity may prolong symptoms. Thus,  
65 symptoms, comorbidities, and suggested treatment are associated with short- to long-term  
66 absence from work and lower productivity.

67 Yet, the causal effect of concussion on economic burdens for individuals and society through  
68 decreased labour market activity has not been identified. First, concussion is a sudden incident  
69 and thus not amenable to prospective study nor randomization. Cohort and case-control  
70 studies[6–10] provide some valuable evidence on employment and labour market outcomes  
71 among those who suffered concussions but are prone to selection bias. Individuals at high risk  
72 of concussions may differ on unobserved characteristics (e.g., risk aversion, routine activities)  
73 from those at low risk. People who are more likely to suffer concussions may also, on average,  
74 have more precarious or unstable employment trajectories prior to the incident, which may  
75 further bias prospective studies. Given the high incidence rate of concussion, even small losses  
76 of productivity and discrete drops in employment would have a significant socioeconomic  
77 impact and thus, it would require large patient cohorts with suitable controls to grasp the full  
78 socioeconomic impact of concussions. Thus, absent the possibility of randomization, using a

1  
2  
3  
4 79 natural or quasi-experimental design is the only likely option to parse out the *causal* effect of  
5  
6  
7 80 concussions on labour market activity.[11]  
8

9 81 We examine how concussions affect salary and employment of working age individuals in  
10  
11 82 Denmark, a representative north-European industrial nation with a strong welfare state and a  
12  
13 83 flexible labour market. We use administrative longitudinal data linked to hospital and emergency  
14  
15  
16 84 room diagnostic data on all Danes, who received a primary diagnosis of concussion between  
17  
18 85 2003 and 2017. To address the problem of unmeasured bias between those that do and do not  
19  
20  
21 86 experience a concussion, we use a quasi-experimental event-study approach[12,13] where we  
22  
23 87 compare similar individuals, who experienced their concussions at different time points. Under  
24  
25 88 mild assumptions of parallel trends in wage progression prior to concussion and random timing  
26  
27 89 of concussion event within a five-year time frame, the approach recovers a robust estimation  
28  
29  
30 90 of the effect of concussion on annual salary and employment status.  
31

32 91  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

## 92 **Material & Methods**

### 93 *Data Sources and Sample Construction*

94 Concussion data originates from the Danish National Patient Registry (DNPR) (see[14] for  
95 description). DNPR is published annually and holds information on all hospitalizations at public  
96 hospitals in Denmark since 1977, on all emergency room visits and outpatient treatments at  
97 public hospitals since 1994, and almost all private hospital treatments since 2003. With one  
98 single, short-lived exception, private hospitals do not operate emergency rooms in Denmark.  
99 Since 2003, the data cover 95 % of all treatments at private hospitals[14], yet only 13  
100 concussion were diagnosed in private hospital settings throughout the period covered by the  
101 data.

102 The combined exposure and control cohort includes all Danes aged 20-59 y, who received a  
103 primary diagnosis of concussion (ICD-10 code S06.0, ICD-8 code N850) between 2003 and  
104 2017 and did not sustain any kind of additional intracranial or extracranial injury. Individuals  
105 who regularly engage in activities associated with a high risk of sustaining multiple concussions  
106 may differ from the average concussion patient and would likely be over-represented in the  
107 exposure sample. To avoid such potential bias, all individuals who were diagnosed with any  
108 kind of brain trauma during a ten-year period prior to the concussion event were excluded.  
109 Altogether, the study included a cohort of 55,424 individuals. Only attrition is through mortality  
110 and out-migration, and out-migrated or deceased patients with missing spells in the follow up  
111 period is excluded in those periods.

112 As a measure of productivity, a price-index deflated annual salaried income was used. Salary  
113 information comes from Statistics Denmark's Income Statistics database. The database

1  
2  
3  
4 114 includes all declared annual incomes including income from self-employment. The Danish Tax  
5  
6  
7 115 Authorities supply the data to Statistics Denmark. Overall accuracy is considered very good.[15]  
8  
9 116 Table 1 reports number of observations for the samples and number of observations with  
10  
11 117 missing salary information. As evident, only between 0.01 to 0.02 percent of observations  
12  
13 118 across exposure and control groups have missing salary information. These observations were  
14  
15  
16 119 disregarded in the main analysis. Through social security numbers, information on salaried  
17  
18 120 income were linked to records on diagnosed concussions. Further, information on high school  
19  
20 121 or equivalent level degree at time of concussion was obtained using the Danish Education  
21  
22  
23 122 Database. The Danish Population Database provided demographic information on age and  
24  
25 123 gender for all respondents. Since the data used in the study come from de-identified  
26  
27 124 administrative registers that Statistics Denmark makes available for research purposes for  
28  
29  
30 125 approved institutions, no approval from an ethics committee was needed to carry out the study.  
31  
32 126 The research was carried out as part of project no. 706630 approved by Statistics Denmark.  
33  
34 127 Statistical analysis was carried out using Stata MP 15.1.  
35  
36  
37 128

### 39 129 *Quasi-experimental design*

40  
41  
42 130 The study used a quasi-experimental, difference-in-differences event time approach previously  
43  
44 131 described in a health setting by Dobkin et al.[12] The approach compare two groups of  
45  
46 132 individuals from the same cohort, where both groups experience concussions, but at two  
47  
48  
49 133 different time points ( $t_c, t_c + \Delta$ ). Specifically, the sample of 55,496 individuals was divided into six  
50  
51 134 different subgroups: i) The *exposure group*, which includes all patients, who suffered their  
52  
53 135 concussion during the period 2003-2012 ( $n=37,848$ ) and ii) five *control groups*, which comprise

1  
2  
3  
4 136 patients who experienced their concussions  $\Delta=\{1$  (n=34,551), 2 (n=31,851), 3 (n=29,922), 4  
5  
6  
7 137 (n=28,530), and 5 (n=27,421)} years later than the exposure group and did not experience any  
8  
9 138 kind of brain injury in the  $10+\Delta$  years before the concussion event (note that the design allows  
10  
11 139 individuals to both be part of the exposure and control group). The model is built on the  
12  
13 140 assumption that the exact timing of a concussion is random for small enough values of  $\Delta$ , and  
14  
15  
16 141 on the additional assumption that the exposure and the control groups would have displayed  
17  
18 142 parallel trends in salary if the control group had not suffered a concussion at  $t_c$  (i.e., assuming  
19  
20 143 that control and exposure group(s) would have continued to show similar trends in salaried  
21  
22 144 earnings had the exposure group not experienced concussions). Table 1 show the number of  
23  
24  
25 145 patients in the exposure group and the five control groups for each year relative to exposure  
26  
27 146 group's concussion incident. Using multiple comparison groups makes it possible to gage the  
28  
29  
30 147 validity of the assumption that the exact timing of a concussion is random for small enough  
31  
32 148 sizes of  $\Delta$

33  
34 149 To estimate the impact of concussion on labour market outcomes, the analysis focuses on the  
35  
36  
37 150 change in annual salary as the primary outcome, and, in further exploratory analyses, studies  
38  
39 151 additional outcomes such as income from health-related benefits, income from welfare benefits,  
40  
41 152 and employment rates. The data are nested within a three-level structure: Exposure or control  
42  
43  
44 153 group  $g$ , which includes individuals  $l$ , at times to exposure-groups concussion incident  $t$ . First,  
45  
46 154 a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$   
47  
48 155 is estimated using ordinary least squares:

$$\begin{aligned}
 & \text{Salary}_{git} = \beta_0 + \gamma \text{Exposure}_g + \theta \text{Post}_t + \delta \text{Post} \times \text{Exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{\text{Age}=20}^{59+\Delta} I(\text{Age}) \eta_{\text{Age}} \\
 & + \sum_{\text{Year}=1999}^{2012} I(\text{Year}) \eta_{\text{Year}} + \epsilon_{git} \quad (1)
 \end{aligned}$$

54  
55 8  
56  
57  
58  
59  
60

1  
2  
3  
4 158 where  $Salary_{git}$  measures annual salaried income deflated to 2015-level;  $Exposure_g$  indicates  
5  
6  
7 159 whether the observation belongs to the exposure or control group;  $Post_t$  captures the period  
8  
9 160 after the exposure group's concussion occurred;  $Post_t \times Exposure_{git}$  captures the effect  
10  
11 161 concussion, measured as share of year  $t \geq 0$  affected by concussion (i.e., for year of incident  
12  
13 162 exposure is expressed as share of year spent with post-exposure, for following years it is equal  
14  
15  
16 163 to 1);  $X_i$  is a set of covariates that includes a high school indicator and a gender dummy;  $\epsilon_{git}$  is  
17  
18 164 the error-term; and the two last sets of indicator variables  $I(Age)$  and  $I(Year)$  capture age and  
19  
20  
21 165 incident year (for control group, the year indexed against). Under an assumption of parallel  
22  
23 166 trends in salaried earnings (i.e., assuming that control and exposure group(s) would have  
24  
25 167 continued to show similar trends in salaried earnings had the exposure group not experienced  
26  
27  
28 168 concussions),  $\delta$  then captures the annual causal effect of concussion on salary for people  
29  
30 169 exposed to concussions (see Supplemental Methods for further details). For additional  
31  
32 170 exploratory analyses, separate models across gender, educational level, and age, as well as  
33  
34  
35 171 across the salary distribution are also estimated (see Supplemental Methods, Supplemental  
36  
37 172 Digital Content 1, for further details). The authors document and make available all code  
38  
39 173 needed to reproduce the findings in the study (Supplemental Digital Content 2).  
40  
41  
42 174

#### 44 175 *Standard Protocol Approvals, Registrations, and Patient Consents*

45

46 176 Since the data used in the study come from de-identified administrative registers that Statistics  
47  
48  
49 177 Denmark makes available for research purposes for approved institutions, no approval from an  
50  
51 178 ethics committee was needed to carry out the study. The research was carried out as part of  
52  
53 179 project no. 706630 approved by Statistics Denmark.  
54

1  
2  
3  
4 180  
5  
6  
7 181  
8  
9 182  
10  
11 183  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Patient and Public Involvement*

There was no involvement from patients or members of the public in the design, or conduct, or reporting, or dissemination plans of the research.

For peer review only

## Results

### *Concussion leads to long-term loss in salaried income*

Individuals who suffered a concussion (exposure group) had a lower salaried income compared to individuals who experienced their concussion 1, 2, 3, 4 and 5 years after the exposure group (control groups). Compared to patients who experienced a concussion one year after the exposure group, salaried income was €423/£380 (95% CI: -€9129/-£8208; €73/£66,  $p=.095$ ; Table 2) lower, corresponding to a salary decrease of 1.5 % (95% CI: -0.3 %; 3.2 %; Figure 1). Compared to patients who experienced a concussion 5 years after the exposure group, however, salaried income in the exposure group was €1,243 (95% CI: -€1,564/-£1,406; -€922/-£829,  $p<.001$ ) lower, corresponding to a salary decrease of 4.2 % (95% CI: 3.1 %; 5.3 %; Figure 1). Normalized wage progression for the control groups, who suffered a concussion 1, 2, 3, 4, and 5 years after the exposure group, showed similar trends and similar levels pre-exposure, indicating that the parallel wage trends assumption was met (Figure 2 and table S1, Figures S1 in Supplemental Digital Content 3).

We hypothesized that the salary decreases resulted from a combination of lower salary and exit from the labour market, either through short- or long-term absence/unemployment. In an exploratory analysis, we tested whether labour force exit drove the full effect of concussion on salary (Figure 3). By comparing the cumulative distribution of salary density for the exposure group with the cumulative distribution of salary density for the  $\Delta=5$  control group (Figure 3, left panel), we found that the impact of concussion on salary was significant for individuals in the lower quartile of the salary distribution (at a 95 % significance level). Specifically, below a threshold salaried income of 40,000€ (£36,000) the presumed impact of concussion on salary



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

206 increased towards the bottom of the earnings pyramid; Figure 3, right panel).

207 Comparing the exposure group to the control group  $\Delta=5$ , which suffers a concussion five years  
208 after the exposure group, concussion was associated with a 2.6% (95% CI: 3.0 %; 2.2 %, p  
209 <.001) increase in the risk of receiving € 0 in annual salary.

### 211 *Long-term loss in salaried income stems from exit from the labour market*

212 To further examine whether exit from the labour market was caused either through short- or  
213 long-term absence/unemployment, we estimated a dynamic model using the control group  $\Delta=5$ ,  
214 which suffers a concussion five years after the exposure group. Sick leave benefits payments  
215 were higher in the exposure group compared to the control groups for the first two years  
216 following concussion. Sick leave benefits were no longer different from year 3 while the  
217 difference in annual salary between exposure and control groups persisted. Further,  
218 employment in the exposure group remained lower than in the control group  $\Delta=5$  and remained  
219 so for the entire post-exposure period (see table S2, Supplemental Digital Content 3 for further  
220 details). To assess whether some form of public benefits covered part of the salary loss, total  
221 income decline was compared to salary decline following concussion. Indeed, total income  
222 decline was lower than the salary decline through a five year period (see Figure S2,  
223 Supplemental Digital Content 3 for further details).

### 225 *Younger patients without high school degree drove the effect of concussion on income loss*

226 The exposure group and all control groups differed slightly in terms of average patient age,  
227 male to female ratio, and for control group  $\Delta=5$ , in the frequency of individuals with at least a

1  
2  
3  
4 228 high school degree (see Table S3, Supplemental Digital Content 3 for further details). To ensure  
5  
6 229 that differences in gender, education, or age did not influence our results, we subdivided our  
7  
8  
9 230 exposure group into subgroups based on gender, education status, and age at time of  
10  
11 231 concussion. We then estimated the impact of concussion on salary and employment across all  
12  
13 232 values of  $\Delta$  and for all subgroups (see, Figures S3-S8, Supplemental Digital Content 3 for  
14  
15  
16 233 further details). Patients between age 30-39 and those without a high school degree  
17  
18 234 experienced the largest absolute and relative declines in salary.

19  
20 235 Finally, we addressed the role of timing of concussion across different years. Given that per  
21  
22  
23 236 design our exposure group always suffered their concussion earlier than the control groups do,  
24  
25 237 changing labour market conditions could moderate effects. Part of our sample suffered their  
26  
27 238 concussion during or just prior to the Great Recession in 2009-2010, which arguably presented  
28  
29  
30 239 the largest shock to both the global and local economy since the Great Depression in the 1930s.  
31  
32 240 In Denmark, the great recession was preceded by a series of years of economic growth, low  
33  
34 241 unemployment, and increasing salaries (see Figure S8, Supplemental Digital Content 3 for  
35  
36  
37 242 salary development from 1994 to 2017). We estimated the impact of concussion on salary  
38  
39 243 separately for each year from 2003-2012 and plotted the estimate against the percent of full-  
40  
41 244 time unemployment in the Danish labour force (Figure 4). Suffering a concussion during an  
42  
43 245 economic boom had a substantially higher impact on salary than doing so during a recession  
44  
45  
46 246 when comparing to control groups who suffered concussions two to five years later than  
47  
48 247 exposure group.  
49  
50

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Discussion

The impact of concussion on employment and salary remains understudied. In a systematic review of four studies on the association between mild TBI and return to work, Cancelliere and colleagues[9] found that most workers return to work within 3-6 months of suffering a mild TBI, but that the long-term impact (more than one year after concussion) was not studied. In addition, studies included small to medium sample sizes, varied measures of return to work, and employed both case-control and cohort designs. Using an inception cohort study design, Theadom and colleagues[7] collected follow up data four years after mild TBI incidents on 245 New Zealanders who were employed prior to incident. They found a 3.6 percent productivity decline among those who suffered a concussion, compared to a 2.3 population average decline. The group suffering mild TBI also reported more difficulties carrying out work-related tasks. In a related study,[7] Theadom and colleagues further found that the mild TBI group had persisting cognitive symptoms four years after suffering their concussion compared to an age-sex matched control group. Also using a case-control design and data from Taiwan, Chu and colleagues[6] found that one month after incident, 26 percent of patients had still not managed to return to work, and a large share of those who did return scored below full-time employment on a work quality index. Only one other study by Graff and colleagues[10] include a large patient cohort (n=19,732). Using case-control they found an association between concussion and failing to return to work of 1.54 odds ratio, but also found that exposed individuals had lower labour market attachment and was more likely to receive health related benefits pre-incident compared to the control group.

In the present study, we overcame some of the obstacles faced by previous work on the impact

1  
2  
3  
4 271 of concussions on labour market outcomes by including a vast cohort of patients and exploiting  
5  
6  
7 272 a quasi-experimental design that allow us to plausibly account for unobserved difference  
8  
9 273 between exposure and control group. In such a quasi-experimental setup exposure and control  
10  
11 274 groups only differ in the timing of concussion. Since everyone in the control group experiences  
12  
13 275 a concussion within five years after individuals in the exposure group, the groups are likely to  
14  
15  
16 276 be balanced on unobservable characteristics. This is particularly important given the number of  
17  
18 277 potential factors that can influence employment after concussion[16,17]. Data from Donker-  
19  
20 278 Cools et al.[17], for instance, suggests larger employers are more able to keep those who have  
21  
22  
23 279 sustained brain injuries in work compared to smaller employers. Furthermore, since our data  
24  
25 280 did not include health-related data such as past psychiatric history, we cannot exclude that  
26  
27 281 exposure and control groups differed in health-related aspects and that these differences  
28  
29  
30 282 biased our results, i.e. that an individual left the workforce for a concussion-unrelated cause  
31  
32 283 like a psychiatric disease triggered by the stress of a concussion event. Thus, even if we believe  
33  
34 284 that our quasi-experimental setup leaves us with exposure and control groups that only differ  
35  
36  
37 285 in the timing of concussion, especially given the reported sample sizes and the finding that  
38  
39 286 exposure and control groups show similar pre-exposure trends on both primary and secondary  
40  
41 287 outcomes (Figure 1 and Tables S1 and S2),, this aspect needs to be discussed as a potential  
42  
43 288 limitation of our study.

44  
45  
46 289 In addition, salary and employment data reported here were compiled routinely through third-  
47  
48 290 party reporting and were mandatory for all subjects, thus giving a complete and comprehensive  
49  
50  
51 291 picture of the economic impact of concussion on a nationwide scale. It should be mentioned  
52  
53 292 that our study also included data from individuals diagnosed in private hospitals. However,

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

293 given the setup of the Danish health care system, i.e. private hospitals predominantly do  
294 selective and overflow surgery and have no ERs[18], only 13 patients were diagnosed at a  
295 private hospital from 2003 onwards.

296 Altogether, we showed that Danes between 20-59 year of age, who suffered a concussion  
297 during the period 2003-2012 experienced average salary losses of 4.2%. The impact of  
298 concussions on salary already materialized one year after the incident and remained sizeable  
299 for at least five years. This result is in line with a “burn-in” period in which the impact of  
300 concussion on wages fully develops. First, concussions occur at some point during the year,  
301 thereby not affecting already earned salary that year. Second, in Denmark, most employees  
302 are entitled to receive their salary during sickness absence for an open ended, but not indefinite  
303 period. The results further showed that both in absolute and relative terms, people with an  
304 educational level at less than a high school degree saw substantially larger negative impact to  
305 salaried earnings than did those with at least a high school degree. Also, the group with less  
306 than a high school degree also saw an immediate impact on salary from their concussion (cf.  
307 Figure S4), indicating that the burn-in period present for workers with at least high school  
308 education likely expressed differences in types of employment and job protection.

309 In addition, total income decline was lower than the salary decline through a five year period  
310 (see Figure S2, Supplemental Digital Content 3 for further details), suggesting that the impact  
311 of concussions on salary largely stems from affected individuals leaving the labour force  
312 completely, likely sustaining themselves through early retirement, disability pensions, self-  
313 sufficiency, or other income sources instead.

314 It is important to mention that our study was restricted to individuals diagnosed in ER and

1  
2  
3  
4 315 hospital settings and individuals diagnosed by a GP might differ from the population studied  
5  
6  
7 316 here. Rowson et al., however, show that in concussed individuals, severity of the cranial injury  
8  
9 317 is not strongly correlated with strength or length of subsequent symptoms[19]. Thus, individuals  
10  
11 318 diagnosed by a GP might suffer concussion effects as much as individuals who initially  
12  
13 319 sustained a more severe cranial injury and sought medical attention in an ER or hospital setting.  
14  
15  
16 320 If this holds true, our results may have validity beyond individuals diagnosed in an ER or hospital  
17  
18 321 setting.

19  
20  
21 322  
22  
23 323 If we assume that people return to their expected salary levels after a five-year recovery period  
24  
25 324 (a very conservative assumption that is not supported by our data), the mere net annual salary  
26  
27  
28 325 loss in our sample would be approximately €23,000,000 (£21,000,000) measured in 2015-  
29  
30 326 value. That would neither include hospital charges, medical costs for the treatment of  
31  
32 327 concussion, the foregone tax from income, and the increased need for welfare spending, nor  
33  
34  
35 328 would it account for the large group of individuals who never seeks treatment[20] or receive  
36  
37 329 their diagnosis from their general practitioner rather than in a hospital or emergency room, and  
38  
39 330 thus escape our study. Thus, total public costs are likely substantially higher.

40  
41  
42 331 In addition, both in absolute and relative terms, the early peak-working aged individuals (30-39  
43  
44 332 y) and the less-educated individuals in our cohort seemed to be most affected after suffering a  
45  
46 333 concussion. These findings might have an additional and yet unmeasured social impact,  
47  
48  
49 334 especially if our results are transferrable to other nations with a less established welfare state  
50  
51 335 and a less flexible labour market. In such countries, the impact on the young and less-educated  
52  
53 336 individuals suffering a concussion and thus on society might be accentuated.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

337 Comparing our hospital incidence rates to more complete canvases of incidences carried out  
338 by Cassidy et al.[21], it seems likely that the actual cost in the population is more than twice as  
339 large as what we estimate, assuming that individuals not diagnosed in a hospital setting on  
340 average suffer the same extent of concussion symptoms. If we were to consider the average  
341 concussion incidence rates for six other advanced European countries that are somewhat  
342 comparable to Denmark (Norway, Finland, Germany, Netherlands, England and France) and  
343 under the assumption that concussion have a similar impact on earnings in these countries, the  
344 net annual salary loss would be approximately €1,099,400,000 (£988,4780,000) measured in  
345 2015-value. While our study likely underestimates the total socioeconomic impact of  
346 concussion, it suggests that concussions has a large economic impact on a nationwide scale  
347 and on productivity and income at the patient level.

## 349 CONCLUSION

350 Using timing of concussion as a natural experiment, we provide first plausible causal estimates  
351 of the effect of concussion on salary and employment among patients treated for concussion in  
352 an emergency room or hospital setting in Denmark, 2003-2017. Our results show that among  
353 this patient group concussion has a large and long-term negative causal impact on salary and  
354 employment. People between 30-39 y and those without high school degrees suffered the  
355 largest salary decreases.

1  
2  
3  
4 357 **Disclosures:**

5  
6  
7 358 The authors report no conflict of interest.  
8

9 359  
10  
11 360 **Acknowledgements:**

12  
13  
14 361 The authors thank Torben Heien Nielsen for helpful comments on an earlier draft of the  
15  
16 362 manuscript.  
17  
18

19 363  
20  
21 364 **Funding:**

22  
23  
24 365 This work was supported by the Swedish Research Council for Health (PF), Working Life and  
25  
26 366 Welfare (Grant no. 2016-07099) with additional funding from the ROCKWOOL Foundation (PF).  
27

28 367 The research was carried out independently of the funders.  
29  
30

31 368  
32  
33 369 **Authors contributions:** P.F. and B.C. conceived of the presented idea, P.F. performed the  
34  
35 370 computations. P.F. and B.C. verified the statistical methods. P.F. and B.C. discussed the results  
36  
37 371 and wrote the manuscript. The corresponding author confirms that he had full access to all the  
38  
39  
40 372 data in the study and had final responsibility for the decision to submit for publication.  
41

42 373  
43  
44  
45 374 **Data Availability Statement**

46  
47 375 The data used in this study has been made available through a trusted third party, Statistics  
48  
49 376 Denmark. Due to privacy concerns the data cannot be made available outside the hosted  
50  
51  
52 377 research servers at Statistics Denmark. University-based and private Danish scientific  
53  
54 378 organizations can be authorized to work with data within Statistics Denmark. Such organization



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

379 can provide access to individual scientists inside and outside of Denmark. Requests for data  
380 may be sent to Statistics Denmark:  
381 <http://www.dst.dk/en/OmDS/organisation/TelefonbogOrg.aspx?kontor=13&tlfbogsort=sektion>  
382 or the Danish Data Protection Agency: <https://www.datatilsynet.dk/english/the-danish-data-protection-agency/contact/>. The authors document and make available all code needed to  
383 reproduce the findings in the study.  
384

385

## References

- 1 Cassidy JD, Carroll LJ, Peloso PM, *et al*. Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med* 2004;:28–60. doi:10.1080/16501960410023732
- 2 Pinner M, Jensen R, Birket-Smith M, *et al*. Konsensusrapport om commotio cerebri (hjernerystelse). Videnscenter for Hjerneskode 2003.
- 3 Hou R, Moss-Morris R, Peveler R, *et al*. When a minor head injury results in enduring symptoms: a prospective investigation of risk factors for postconcussional syndrome after mild traumatic brain injury. *J Neurol Neurosurg Psychiatry* 2012;**83**:217–23. doi:10.1136/jnnp-2011-300767
- 4 Ruff R. Two decades of advances in understanding of mild traumatic brain injury. *J Head Trauma Rehabil* 2005;**20**:5–18. doi:10.1097/00001199-200501000-00003
- 5 Silverberg ND, Iverson GL. Is rest after concussion “the best medicine”? recommendations for activity resumption following concussion in athletes, civilians, and military service members. *The Journal of head trauma rehabilitation* 2013;**28**:250–9. doi:10.1097/HTR.0b013e31825ad658
- 6 Chu SY, Tsai YH, Xiao SH, *et al*. Quality of return to work in patients with mild traumatic brain injury: a prospective investigation of associations among post-concussion symptoms, neuropsychological functions, working status and stability. *Brain Injury* 2017;**31**:1674–82. doi:10.1080/02699052.2017.1332783
- 7 Theadom A, Barker-Collo S, Jones K, *et al*. Work Limitations 4 Years After Mild Traumatic Brain Injury: A Cohort Study. *Archives of Physical Medicine and Rehabilitation*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

408 2017;**98**:1560–6. doi:10.1016/j.apmr.2017.01.010

409 8 Theadom A, Starkey N, Barker-Collo S, *et al.* Population-based cohort study of the  
impacts of mild traumatic brain injury in adults four years post-injury. *PLOS ONE*  
410 2018;**13**:e0191655. doi:10.1371/journal.pone.0191655

412 9 Cancelliere C, Kristman VL, Cassidy JD, *et al.* Systematic Review of Return to  
Work After Mild Traumatic Brain Injury: Results of the International Collaboration on Mild  
413 Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation*  
414 2014;**95**:S201–9. doi:10.1016/j.apmr.2013.10.010

416 10 Graff HJ, Siersma V, Møller A, *et al.* Labour market attachment after mild traumatic  
417 brain injury: nationwide cohort study with 5-year register follow-up in Denmark. *BMJ open*  
418 2019;**9**:e026104. doi:10.1136/bmjopen-2018-026104

419 11 Ogilvie D, Adams J, Bauman A, *et al.* Using natural experimental studies to guide  
420 public health action: turning the evidence-based medicine paradigm on its head. *J Epidemiol*  
421 *Community Health* Published Online First: 19 November 2019. doi:10.1136/jech-2019-213085

422 12 Dobkin C, Finkelstein A, Kluender R, *et al.* The Economic Consequences of  
423 Hospital Admissions. *American Economic Review* 2018;**108**:308–52.  
424 doi:10.1257/aer.20161038

425 13 Fadlon I, Nielsen TH. Household Responses to Severe Health Shocks and the  
426 Design of Social Insurance. Cambridge, MA: 2015. doi:10.3386/w21352

427 14 Schmidt M, Schmidt SAJ, Sandegaard JL, *et al.* The Danish National Patient  
428 Registry: a review of content, data quality, and research potential. *Clinical epidemiology*  
429 2015;**7**:449–90. doi:10.2147/CLEP.S91125

22

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

430 15 Statistics Denmark. Documentation of statistics: Income Statistics - Statistics

431 Denmark.

432 2015.<https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/income->

433 statistics (accessed 20 Aug 2018).

434 16 Buck PW, Laster RG, Sagrati JS, *et al.* Working with mild traumatic brain injury:

435 voices from the field. *Rehabil Res Pract* 2012;**2012**:625621. doi:10.1155/2012/625621

436 17 Donker-Cools BHPM, Wind H, Frings-Dresen MHW. Prognostic factors of return to

437 work after traumatic or non-traumatic acquired brain injury. *Disabil Rehabil* 2016;**38**:733–41.

438 doi:10.3109/09638288.2015.1061608

439 18 Vrangbaek K, Østergren K, Birk HO, *et al.* Patient reactions to hospital choice in

440 Norway, Denmark, and Sweden. *Health Econ Policy Law* 2007;**2**:125–52.

441 doi:10.1017/S174413310700401X

442 19 Rowson S, Duma SM, Stemper BD, *et al.* Correlation of Concussion Symptom

443 Profile with Head Impact Biomechanics: A Case for Individual-Specific Injury Tolerance. *J*

444 *Neurotrauma* 2018;**35**:681–90. doi:10.1089/neu.2017.5169

445 20 Setnik L, Bazarian JJ. The characteristics of patients who do not seek medical

446 treatment for traumatic brain injury. *Brain Injury* 2007;**21**:1–9.

447 doi:10.1080/02699050601111419

448 21 Cassidy JD, Carroll LJ, Peloso PM, *et al.* Incidence, risk factors and prevention of

449 mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on mild

450 traumatic brain injury. *Journal of Rehabilitation Medicine* 2004;**36**:28–60.

451 doi:10.1080/16501960410023732

23

1  
2  
3  
4 452  
5  
6  
7 453  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 454 **Figure Legends**

### 455 **Figure 1. Estimated effect of concussions in percentage on salary for the exposure group** 456 **measured against each control group**

457 Note: Figure shows the percentage change in salary experienced by the exposure group  
458 following their concussions compared to the expected trajectory absent the concussion  
459 (calculated from the control groups) with 95 % confidence intervals. See table 1 for separate p-  
460 values for each estimate.

### 462 **Figure 2. Salary development for exposure and control groups across time of exposure**

463 Note: Figure shows the salary trajectories for the exposure group (black) who suffers  
464 concussion at year zero against normalized wage trajectories for the control groups who suffer  
465 their concussions one to five years later.  $\Delta$  indicates the number of years between exposure  
466 and control incident. Table shows that there are no significant differences in the normalized  
467 salary levels for exposure and control group prior to exposure incident (see Figure S1,  
468 Supplemental Digital Content 3 for unnormalized salary trajectories).

470 **Figure 3. (Left panel) The cumulative density function (cdf) for salary post-treatment**  
471 **among the treatment group and their counterfactual outcome had they not experienced**  
472 **their concussions, and (Right panel) the change in salary density for the exposure group**  
473 **compared to their counterfactual baseline expressed as the effect of concussion on the**  
474 **probability of earning below the salary-level expressed on the x-axis following exposure**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

475 **event.**

476 Note: The figure shows the observed cumulative salary distribution following concussion for the  
477 exposure group (red) and the expected counterfactual salary distribution absent suffering  
478 concussion in the exposure group (blue), when using the  $\Delta=5$  control group. The black line  
479 shows the difference between the observed and the counterfactual distribution, and the grey  
480 dash lines show the 95 % confidence interval. The close to constant decline of the difference  
481 between the two distributions as the salary increase indicates that the main part of the effect of  
482 concussions on salary are driven by people having a salary equal to zero.

483 **Figure 4. Effect of concussion on salary across incident years and control groups**  
484 **together with the percentage fulltime unemployed of the labor force.**

485 Note: Figure shows annual estimates of concussion against each control group separately  
486 mapped against the share of the labor force that is full time unemployed. 95 % confidence  
487 intervals. The estimates for the effect of concussion on salary almost uniformly increase in  
488 absolute magnitude when unemployment decreases, and decrease when unemployment  
489 increase, indicating that the effect of concussion on salary is countercyclical to the economic  
490 cycle.

492 **Supplemental Digital Content titles & legends**

493 **Supplemental Digital Content [#1].** Text file. Supplemental materials and methods. This file  
494 contains further details on our quasi-experimental, difference-in-differences event time  
495 approach.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

496 **Supplemental Digital Content [#2]**. File. Code used for the analyses.

497

498 **Supplemental Digital Content [#3]**. Table. Supplemental results Table S1: Test of parallel  
499 trends assumption pre-exposure incident against each control group separately using eq. S3 in  
500 supplementary methods. Separate exposure dummies for all time periods (except the year prior  
501 to exposure, which serves as reference period).

502

503 **Supplemental Digital Content [#3]**. Table. Supplemental results Table S2: Effect of  
504 concussion on different labor market outcome parameters using separate exposure dummies  
505 for all time periods (except the year prior to exposure, which serves as reference period): In this  
506 exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers  
507 a concussion five years after the exposure group. Outcomes include annual salaried income  
508 (annual salary), total annual income (total income), annual sick leave benefits received (sick  
509 leave benefits) as well as a binary indicator of employment (probability of employment).  
510 Monetary outcomes were measured at 2015-level in € 1,000.

511

512 **Supplemental Digital Content [#3]**. Table. Supplemental results Table S3: Demographic  
513 factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years  
514 leading up to the concussion event in each of the groups. Factors include patient age (in years),  
515 share of sample female (1=100% female), and share of individuals with at least a high school  
516 degree (1=100%).

517



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S1: Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S2: The Cumulative Density Function (CDF) for Total Income Post-Treatment among the Treatment Group and Their Counterfactual, and the Difference between the Two CDFs Expressed as the Effect of Concussion on the Probability of Total Income Below that Income-Level following Exposure Event.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S3: Percentage Effect of Concussion on Relative Salary Across Age Groups.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S4: Percentage Effect of Concussion on Relative Salary Across High School Completion.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S5: Percentage Effect of Concussion on Relative Salary Across Gender.

**Supplemental Digital Content [#3].** Figure. Supplemental results Figure S6: Effect of Concussion on Absolute Salary in 1K Euro Across Age groups.

Supplemental Digital Content [#3]. Figure. Supplemental results Figure S7: Effect of Concussion on Absolute Salary in 1K Euro Across Education.

Supplemental Digital Content [#3]. Figure. Supplemental results Figure S8: Effect of Concussion on Absolute Salary in 1K Euro Across Gender.

**Table 1. Number of observations for exposure and control groups across time since exposure and number of observations with missing salary information**

Years until exposure	Exposure					
	group	Control $\Delta=1$	Control $\Delta=2$	Control $\Delta=3$	Control $\Delta=4$	Control $\Delta=5$
-4	36,804	33,681	31,112	29,190	27,859	26,794
-3	36,978	33,834	31,245	29,366	27,973	26,907
-2	37,195	34,003	31,407	29,501	28,146	27,031
-1	37,449	34,224	31,582	29,687	28,288	27,220
0	37,848	34,551	31,851	29,922	28,530	27,421
1	37,467		31,755	29,832	28,433	27,337
2	36,940			29,807	28,421	27,295
3	36,484				28,421	27,304
4	36,084					27,314
Total observations	333,249	170,293	188,952	207,305	226,071	244,623
Observations with missing salary	81	32	31	44	35	29

Control groups have not suffered a concussion in 10+ $\Delta$  years before incident, exposure group has not suffered concussion the 10 years before exposure incident.

1

2

3

4 552 **Table 2. Estimated effect of concussion on salary of exposure group compared to**

5

6 553 **control groups that suffered their concussion  $\Delta = 1, 2, 3, 4, 5$  y after the exposure**

7

8 554 **group's concussion event, measured at 2015-level.  $N_{\text{Exposure}}$ : 37,848**

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

	<b>Estimated salary effect (<math>\delta</math>)</b>	<b>95 % CI</b>	<b><math>p</math></b>	<b><math>N_{\text{Control}}</math></b>
$\Delta = 1$ y	-423€	(-919€;73€)	.095	34,551
$\Delta = 2$ y	-825€	(-1,108€; -543€)	<.001	31,851
$\Delta = 3$ y	-1,019€	(-1,331€; -707€)	<.001	29,922
$\Delta = 4$ y	-1,126€	(-1,446€; -805€)	<.001	28,530
$\Delta = 5$ y	-1,243€	(-1,564€; -922€)	<.001	27,421

21 556

22 557 Results obtained from estimations following Eq. (1). Models include controls for high school  
23 diploma, gender, age, and observation year. Results obtained using reghdfe in Stata.

24 558

25 559

26 559

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

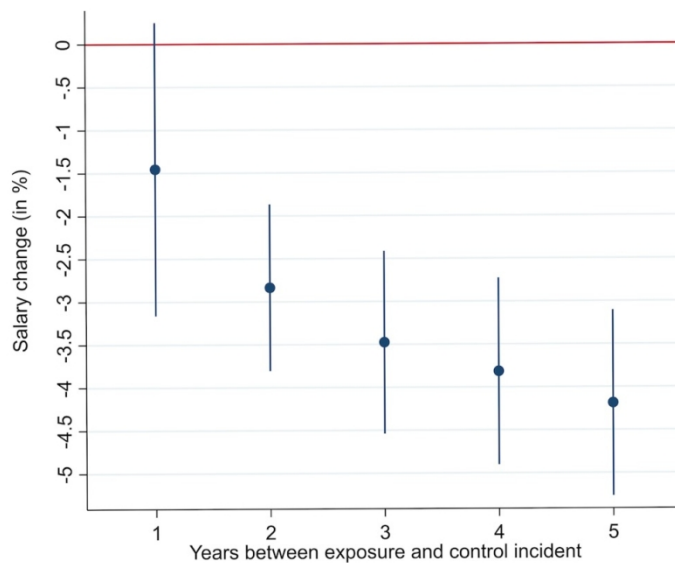


Figure 1

338x190mm (108 x 108 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

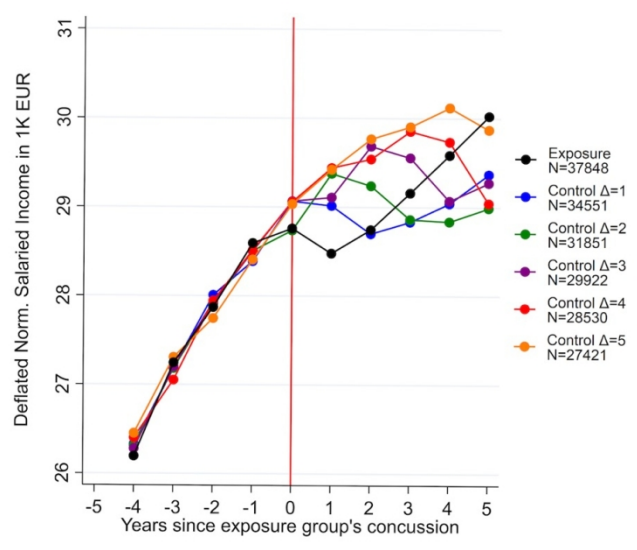


Figure 2

338x190mm (108 x 108 DPI)

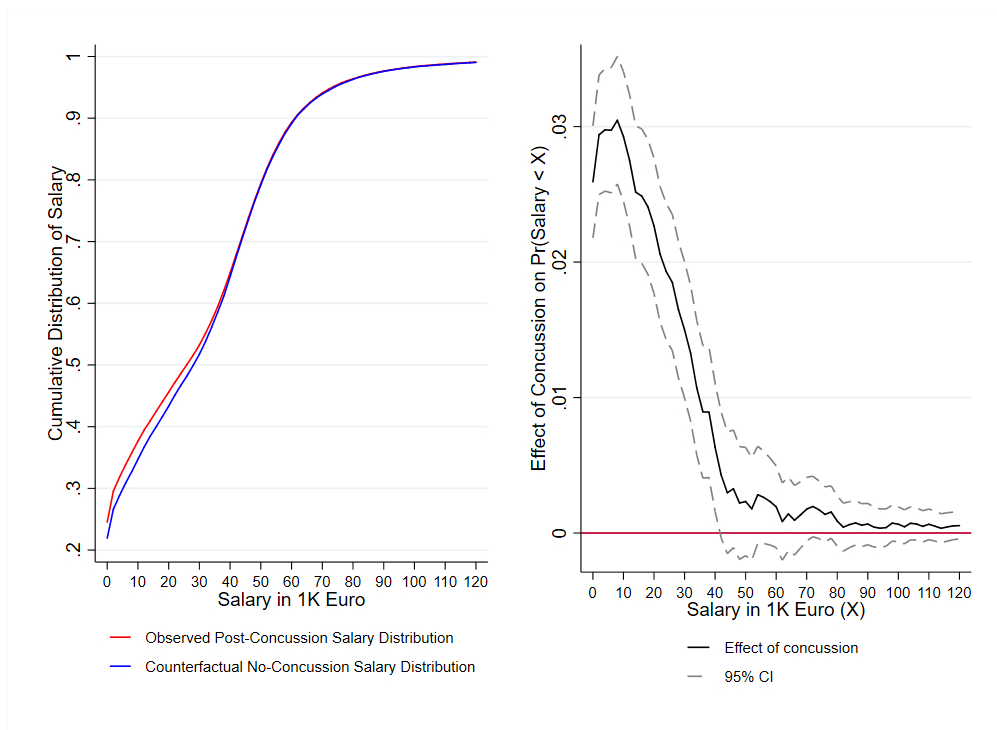


Figure 3

352x256mm (72 x 72 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

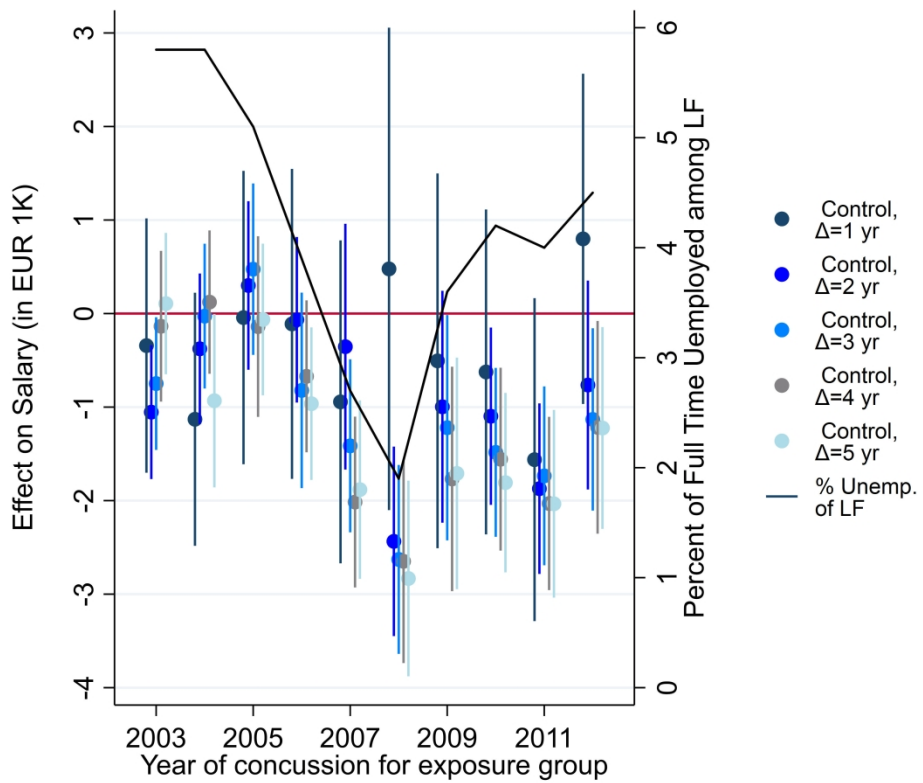


Figure 4

## SUPPLEMENTAL MATERIALS AND METHODS

Our quasi-experimental, difference-in-differences event time approach compares two groups of individuals from the same cohort, where both groups experience concussions, but at two different time points ( $t_c, t_c + \Delta$ ). For the simple situation where we have three periods ( $t=0, 1, 2$ ) and the exposure group (T) experiences their concussion at the start of period 1 ( $t_c=1$ ), and the control group (C) at the start of period 2 ( $t_c + \Delta=2$ ), the effect of concussion on salary (Y) is:

$$\Delta = (Y_1^T - Y_1^C) - (Y_0^T - Y_0^C)$$

The effect of concussion on salary in  $t=1$  is estimated by comparing the average difference in salary between exposure and control groups for the post-concussion period  $t=1$  ( $Y_1^T - Y_1^C$ ) to the average difference in salary for the pre-concussion, or baseline, interval  $t=0$  ( $Y_0^T - Y_0^C$ ). Assuming the exact timing of a concussion is random for small enough sizes of  $\Delta$ , and under the additional assumption that the exposure group would have had parallel trends in salary as the control group absent suffering concussion at  $t_c$ ,  $\delta$  captures the causal effect of concussion among those who suffer concussions – also known as the average effect on the treated (AT). The AT does not capture how concussions would affect a random person. The AT captures how concussions causally affect those who suffer concussions.

For our study, the parallel trends assumption states that exposure and control groups have parallel developments in salary leading up to the exposure group's concussion and the exposure and control groups would have further exhibited parallel salary trajectories if the concussion had not occurred. To test the parallel trends assumption, we estimate a dynamic version of the model specification (shown in supplementary table S1), which explicitly allows us to test whether the parallel trend assumption for our sample is probable.

To validate that the timing of concussion is random with our study period, we present estimates for effect of exposure across different periods between exposure and control incident ( $\Delta$ ). Most recorded concussions outside contact sports and military engagements stem from unforeseen events, such as falls or striking/being struck by an object<sup>25,26</sup>, so assuming random timing is likely valid. People who regular engage in activities that result in high risk of multiple concussions may be different than the average concussion patient and would be more likely to end up in the exposure sample than in the control sample, which could induce bias. To avoid such potential bias, we restrict our sample to individuals without prior diagnoses for intracranial injuries ten years prior to exposure.

At  $t=-1$ , i.e. one year before the exposure group suffered a concussion, the control groups were slightly smaller than the exposure group, and two control groups ( $\Delta=4$  and 5) differed slightly but significantly in terms of average patient age ( $p < .001$ ; supplementary table S2), male to female ratio ( $p < .001$ ), and for control group  $\Delta=5$ , in the frequency of individuals with at least a high school degree ( $p < .001$ ). However, the differences are numerically small. To test that composition differences between exposure and control do not drive our results, we provide separate results for individuals with and without high school degree, for males and females, and for different age groups across all different values of  $\Delta$ .

Further, our design inherently leads to the possibility of timing issues—our exposure group always suffers their concussion earlier (in terms of calendar time and age) than the control groups do. If the labor market is constantly improving or worsening during the period we consider, this could substantially influence our results. Therefore, we also estimate separate models across exposure incident year and control group. Estimating separate models allow us the added benefit of being able to examine whether the business cycle influences the effect of concussions on salary.

### Statistical model

To estimate the impact of concussion on salary, we define the following variables: Exposure or control group  $g$ , which includes individuals  $i$ , at times  $t$  to exposure-groups concussion incident  $t$ . First, we estimate a standard difference in differences model for each separate control group  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  using ordinary least squares:

$$\text{Salary}_{git} = \beta_0 + \gamma \text{exposure}_g + \theta \text{post}_t + \delta \text{post}_t \times \text{exposure}_{git} + \mathbf{X}_i \boldsymbol{\beta} + \sum_{\text{Age}=26}^{48+\Delta} I(\text{Age}) \eta_{\text{age}} + \sum_{\text{year}=1999}^{2012} I(\text{year}) \eta_{\text{year}} + \epsilon_{git} \quad (\text{S1})$$

where  $\text{Salary}_{git}$  measures annual salaried income deflated to 2015-level,  $\text{exposure}_g$  indicates whether the observation belongs to the exposure or control group,  $\text{post}_t$  captures the period after the exposure group's concussion occurred, and  $\text{post}_t \times \text{exposure}_{git}$  captures the effect concussion, measured as share of year  $t \geq 0$  affected by concussion. In this way, someone who suffers a concussion July 1 has  $\text{post}_t \times \text{exposure}_{git} = 0.5$  for  $t = 0$  and  $\text{post}_t \times \text{exposure}_{git} = 1$  for  $t > 0$ .  $\mathbf{X}_i$  is a set of covariates that includes a high school indicator and a gender dummy,  $\epsilon_{git}$  is the error-term, and the two last sets of indicator variables  $I(\text{Age})$  and  $I(\text{Year})$  capture age and incident year levels (control group indexed against incident year). Under the parallel trends assumption,  $\delta$  then captures the annual effect of concussion on salary. In eq. 1,  $\text{exposure}_g$  normalizes any pre-exposure differences between the exposure and control group, thereby creating a joint baseline pre-exposure.



We estimate robust individual-level clustered standard errors to account for the possibility that individuals enter the data twice both as control (0) and exposure (1) individuals ( $g=\{0,1\}$ ), and that they are observed for multiple periods ( $t=\{-4, \dots, \Delta-1\}$ ). To calculate the relative salary decrease after concussion, we exploit the parallel trends assumption to generate the expected counterfactual salary level, i.e. had the concussion not occurred, and calculate the decline expressed in percentage as: % change =  $\delta / E(\widehat{Salary}_{git} | g = 1, post_t = 1, post_t \times exposure_{git} = 0)$ . In this way, we provide both absolute estimates measured in 1K Euro, as well as percentage change.

We expect  $\delta$  from eq. (1) to likely be negative. Yet, a decrease in annual salary can arrive through two different channels. Concussions may affect salary through either decreasing income among those employed or by reducing the number of individuals who are employed and earning any salary at. To parse out which of the two channels is driving the results, we examine how concussion affects the salary distribution among the exposure group following. Following Chernozhukov et al.<sup>27</sup> we estimate a series of regressions across the whole salary distribution, where, for a finite set of points, we predict how concussion affects the likelihood of having earnings on the left side of each finite point, as follows:

$$\sum_{j=0}^{\max(Salary)} p_j = \beta_{0j} + \delta_j post_t \times exposure_{git} + \theta post_t + \gamma_j exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age,j} + \sum_{year=1999}^{2012} I(year) \eta_{year,j} + \epsilon_{git,j} \quad (S2)$$

where  $p_j = \Pr(Salary_{git} \leq j)$  and  $j$  is the interval from 0 to  $\max(Salary)$ . Across the salary distribution, we can now predict the probability of earning less than  $j$  for those with and without concussions. From equation 2, we predict  $p_j^1 = E(p_j | post_t \times exposure_{git} = 1, exposure_g = 1, t \geq 0)$  and the counterfactual  $p_j^0 = E(p_j | post_t \times exposure_{git} = 0, exposure_g = 1, t \geq 0)$ . Plotting  $p_j^1$  and  $p_j^0$  over each value of salary  $j$ , and assuming rank stability, gives the cumulative density function of salary for the treated ( $p_j^1$ ) and the counterfactual observation of the treated had they not suffered concussions ( $p_j^0$ ). The difference between  $p_j^1$  and  $p_j^0$  is simply  $\delta_j$ . If the value of  $\delta_j$  monotonically moves towards zero as  $j$  increases until  $p_j^1 \approx p_j^0 \approx 1$  it indicates that exit from employment fully drives the effect of concussion on salary. If instead the value of  $\delta_j$  is constant or increasing across parts of the distribution, it instead indicates that a decrease in salary among those still receiving salary drives at least part of the effect.

Eq. 1 and eq. 2 are based on the parallel trends assumption. The assumption states that exposure and control groups follow parallel salary trajectories until individuals in the exposure group experiences a concussion, and that the parallel trends would have continued had the concussion not occurred. Whereas we cannot verify the counterfactual situation of parallel trends after exposure, we can use a dynamic model to test for systematic differences in salary trends between exposure and control group in the years leading up to the exposure group's concussion event. To do so, we estimate the following dynamic model:

$$Salary_{git} = \beta_0 + \sum_{t \neq -1, t = -4}^{\Delta-1} \delta_t \times I(t_g) \times exposure_g + \sum_{t=-4}^{\Delta-1} I(t_g) \eta_t + \gamma exposure_g + \mathbf{X}_i \boldsymbol{\beta} + \sum_{Age=26}^{48+\Delta} I(Age) \eta_{age} + \sum_{year=1999}^{2012} I(year) \eta_{year} + \epsilon_{git} \quad (S3)$$

Where we interact exposure group status ( $exposure_g$ ) with indicators  $I(t_g)$  capturing time from concussion. If the parallel trends assumption holds, then it must be the case  $\{\delta_{-4}, \delta_{-3}, \delta_{-2}\} = 0$ , whereas the size and sign of  $\{\delta_0, \dots, \delta_{\Delta-1}\}$  captures the dynamic effect of a concussion from the year of incidence and  $\Delta-1$  years onward. By estimating the effect of concussion on salary among different years of the study period, we are also able to capture how the impact of concussion on salary evolves year to year after the concussion has occurred. We further estimate eq. 3 for a series of related labor market outcomes (annual total income, annual amount of sickness benefits received, annual probability of being employed), to generate a more thorough understanding on how concussions affect labor market outcomes—i.e., if people experience a decrease in salary due to a concussion, are they then compensated through different types of welfare state services.

25 Faul M, Coronado V. Epidemiology of traumatic brain injury. *Handbook of clinical neurology* **127**:3-13, 2015

26 Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic Brain Injury–Related Emergency Department Visits, Hospitalizations, and Deaths — United States, 2007 and 2013. *MMWR Surveillance Summaries* **66**:1-16, 2017

27 Chernozhukov V, Fernández-Val I, Melly B. Inference on Counterfactual Distributions. *Econometrica* **81**:2205-2268, 2013

1  
2  
3 \*\*\*\*\*  
4 \*\*\*\*\*  
5 \*\*\*\*\*  
6 \*\*\*\*\*  
7 \*\*\*\*\*  
8 \*\*\*\*\*  
9 \*\*\*\*\*

10 \*\*\*\*\*

11 \*\*\*\*\* This study relies on restricted individual level  
12 administrative data obtained  
13 \*\*\*\*\* from Statistics Denmark.

14 \*\*\*\*\*

15 \*\*\*\*\* Only Danish research environments are granted authorization to  
16 access data from Statistics Denmark. \*\*\*\*\* Foreign  
17 \*\*\*\*\* researchers can, however, get access to micro data through an  
18 affiliation to a Danish authorized  
19 \*\*\*\*\* environment.

20 \*\*\*\*\* Access is given to anonymized micro data, i.e. data at an  
21 individual personal or corporate level. \*\*\*\*\* Access takes  
22 \*\*\*\*\* place through researcherís own pc over the Internet.

23 \*\*\*\*\*

24 \*\*\*\*\* See <https://www.dst.dk/en/TilSalg/Forskningservice> for detail  
25 \*\*\*\*\*

26 \*\*\*\*\* For the replication of present study, contact the ROCKWOOL  
27 Foundation for access.

28 \*\*\*\*\* <http://www.rockwoolfonden.dk/en/>

29 \*\*\*\*\*

30 \*\*\*\*\*  
31 \*\*\*\*\*  
32 \*\*\*\*\*  
33 \*\*\*\*\*  
34 \*\*\*\*\*  
35 \*\*\*\*\*  
36 \*\*\*\*\*

37  
38  
39  
40 \*\*\*\*\*  
41 \*\*\*\*\*  
42 \*\*\*\*\*  
43 \*\*\*\*\*  
44 \*\*\*\*\*  
45 \*\*\*\*\*  
46 \*\*\*\*\*

47 \*\*\*\*\*

48 \*\*\*\*\* This is the master data do-file for the full package of do-  
49 files that generates the results for  
50 \*\*\*\*\* Fallesen and Campos (2020). Execute files in the order they  
51 are listed. For each file, specify a \*\*\*\*\* home registry in place of  
52 [home]. Further, generate subfolders [home]/data, [home]/tables, and  
53 \*\*\*\*\* [home]/highdef to capture auxiliary data sets, tables, and  
54 figures.

55 \*\*\*\*\*

56 \*\*\*\*\*  
57 \*\*\*\*\*  
58 \*\*\*\*\*  
59 \*\*\*\*\*  
60 \*\*\*\*\*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

```
*****  
*****
```

```
***Install user-written reghdfe command for faster computation of  
regressions  
ssc install reghdfe
```

```
***Generate concussion samples and merges on covariates and outcome  
do 01generate_sample.do
```

```
***Generates auxiliary data sets the includes social benefit  
reciprocity indicator and levels of benefits received  
do 02generate_benefits.do
```

```
***Generates auxiliary data sets the includes social benefit  
reciprocity indicator and levels of benefits received  
do 02generate_benefits.do
```

```
***Generate results  
do 03generate_figures_and_results
```

```
exit, clear
```

For peer review only

```
1
2
3 clear all
4
```

```
5 *****
6 *****
7 **
8 **          This program builds data for Fallesen & Campos
9 (2020)
10 **          study of concussion's impact on productivity
11 measured
12 **          through annual salary
13 **
14 **
15 *****
16 *****
```

```
17
18
19 **Global for path to registry data
20 global dorg "E:/data/rawdata/706630"
```

```
21
22 *Global for processed data
23 global data "[home]/data"
```

```
24
25 /*globals for price index to calculate income at 2015-level across
26 years*/
27 /*Price index obtained from www.dst.dk/en/statistik/emner/priser-og-
28 forbrug/forbrugeriser/nettoprisindeks */
29
```

```
30 {
31
32 global price1980 = .358
33 global price1981 = .398
34 global price1982 = .439
35 global price1983 = .466
36 global price1984 = .494
37 global price1985 = .517
38 global price1986 = .521
39 global price1987 = .537
40 global price1988 = .564
41 global price1989 = .594
42 global price1990 = .612
43 global price1991 = .628
44 global price1992 = .642
45 global price1993 = .651
46 global price1994 = .662
47 global price1995 = .674
48 global price1996 = .688
49 global price1997 = .703
50 global price1998 = .713
51 global price1999 = .728
52 global price2000 = .751
53 global price2001 = .769
54 global price2002 = .788
55 global price2003 = .806
56 global price2004 = .817
57 global price2005 = .833
58 global price2006 = .850
59
60
```

```

1
2
3     global price2007 = .867
4     global price2008 = .899
5     global price2009 = .917
6     global price2010 = .936
7     global price2011 = .960
8     global price2012 = .978
9     global price2013 = .986
10    global price2014 = .994
11    global price2015 =1.00
12    global price2016 =1.005
13    global price2017 =1.017
14
15 }
16
17
18 /*Locate concussions and other TBIs from the Danish National Patient
19 Registry */
20 /**/
21 forvalue t = 1977/2017{
22     if `t' < 1994 use $dorg/lpr_diag`t'.dta /// **uses ICD-8
23 codes until Dec. 31, 1993
24         if substr(c_diag,1,2)=="85"
25     if `t' > 1993 use $dorg/lpr_diag`t'.dta /// **uses ICD-10
26 codes from Jan. 1, 1994
27         if substr(c_diag,1,4)=="DS06"
28     **recovers encrypted social security number and admittance
29 date
30     merge m:m recnum using $dorg/lpr_adm`t', keepus(pnr d_ind*)
31 keep(3)
32     drop _merge recnum
33
34
35     **Keeps diagnosis, diagnosis type, and individual id (pnr)
36     keep pnr c_diag c_diagtype pnr d_ind
37
38
39     **generate year variable
40     gen year = year(d_ind)
41
42     **geenerate share of year with concussion
43     gen time_from_incident = 1-((d_ind-mdy(1,1,year(d_ind))))/
44 365)
45     drop d_ind
46
47
48     *save as one dataset
49     if `t' > 1977 append using $data/concussion.dta
50     if `t' == 2017 sort pnr year
51     save $data/concussion.dta, replace
52
53
54 }
55 */
56 /*****
57 Sets up datasets for treatment group (x = 0)
58 and the control groups who suffer concussion
59 1, 2, 3, 4, 5 years later (x = 1 2 3 4 5).
60

```

1  
2  
3  
4 Treatment group are not allowed to have suffered  
5 any type of TBI the last 10 years before concussion  
6 , control group are not allowed to have suffered any TBI  
7 concussion the last 10 + x years.  
8

```
9 *****/
10 /**/
11 forvalue x = 0/5{
12     use $data/concussion.dta, clear
13     sort pnr year
14
15     **generate measure of length between registered TBIs
16     by pnr: gen help = year-year[_n-1]
17
18     drop if help < 10+`x'
19
20     **Keep only concussion, and only when it was primary
21     diagnosis
22     keep if c_diagtype== "A" & /// Primary daignosis
23         (c_diag == "85099" | /// ICD-8 code for concussion
24         substr(c_diag,1,5) == "DS060") // ICD-10 code for
25     concussion
26
27     *generate treatment and control datasets
28     save $data/concussion_`x'.dta, replace
29
30 }
31
32
33
34
35
```

```
36 */
37 /*****
38 Generate datasets for analysis. First incident year is allowed
39 to be 1992, because it is the first year where we have full record
40 for the five year plus control group (1977+10+5 = 1992).
41
```

42 We generate seperate datasets for each incident year for treatment  
43 group and different control groups.  
44

```
45 *****/
46
47 forvalue time = 1/5{           //for different control groups
48
49     local post_period = 5     // local for number of years
50     observed post
51
52                                     //
53     concussion for treatment group
54
55     local endtime = 2017-`post_period'    /*last year where
56     we allow for treatment event
57
58                                     to
59     occur, in order to have long enough control
60                                     period.
```

```

1
2
3   Defined by latest year available data*/
4
5       forvalue count = 1992(1)`endtime'{
6           local t = `count'                // for
7   ease of coding
8           local n = `t'-4                // first
9   pre-treatment event period
10          local c = `t'+`post_period'     //last post-event
11   period
12          local w = `t'+`time'           //time of
13   concussion for control
14
15          use pnr alder using $dorg/bef`t'  ///
16   Bring in all 30-49 yr olds
17          if inrange(alder`t',20,59), clear // from
18   the population register
19
20
21          **year variable
22          gen year = `t'
23
24          **limit sample to those who suffer a concussion in
25   `t'
26          merge 1:1 pnr year using $data/
27   concussion_0.dta, ///
28          keep(3) nogen
29
30          forvalue x=`n'/'c'{           //add longitudinal data
31              merge 1:1 pnr using $dorg/bef`x', ///
32              keep(1 3) keepus(efalle alder
33   koen) //add information on spouse,
34
35              //age, and gender
36              rename _merge merge`x' //indicator for
37   whether in DK that year
38
39
40              **Add salary information and ses
41   information
42              if `x' < 2017{
43                  merge 1:m pnr using $dorg/
44   ind`x', ///      m:1 to account for duplicates
45                  nogen keep(1 3)
46   keepus(erhvervsindk_13 pre_socio personindk dispon_13
47   aekvivadisp_13) // in data on non-important variables
48
49
50                  bysort _all: keep if _n ==1
51                  //drop perfect duplicates
52              **Align variable names and account
53   for inflation
54
55                  rename erhvervsindk_13 loenmv
56                  rename pre_socio pre_socio`x'
57
58                  foreach kk in personindk dispon_13
59   aekvivadisp_13 loenmv{
60

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

                                rename `kk'
`kk'\`x'
                                }
                                foreach kk in personindk dispon_13
aekvivadis_13 loenmv{
                                replace `kk'\`x' =
                                `kk'\`x'/{price`x'}
                                }
                                **Bring in educational information
                                merge 1:1 pnr using $dorg/
uddany`x', ///
                                nogen keep(1 3) keepus(hffsp)
                                }
                                if `x' == 2017{
                                merge 1:m pnr using $dorg/
ind`x', ///      m:1 to account for duplicates
                                nogen keep(1 3)
keepus(erhvervsindk_13 pre_socio personindk) // in data on
non-important variables
                                bysort _all: keep if _n ==1
                                //drop perfect duplicates
**Align variable names and account
for inflation
                                rename erhvervsindk_13 loenmv
                                rename pre_socio pre_socio`x'
                                foreach kk in personindk loenmv{
                                rename `kk'
`kk'\`x'
                                }
                                foreach kk in personindk loenmv{
                                replace `kk'\`x' =
                                `kk'\`x'/{price`x'}
                                }
                                **Bring in educational information
                                merge 1:1 pnr using $dorg/
udda`x', ///
                                nogen keep(1 3)
keepus(hfaudd)
                                rename hfaudd hfaudd`x'
                                }
                                }
                                **Reshape data to panel structure
                                if `count'>= 2012 reshape long efalle alder koen
loenmv pre_socio personindk dispon_13 aekvivadis_13 hffsp merge
hfaudd , i(pnr) j(t)

```



```

1
2
3           if `count' < 2012 reshape long efalle alder koen
4 loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge ,
5 i(pnr) j(t)
6
7           gen count = t-year           //variable for time to
8 concussion
9           gen treatment =1 //treatment group indicator
10
11          save $data/sample_temp.dta, replace //temporary
12 dataset
13
14          /
15 *****
16          Now build control sample for time `time' and year
17 `count'
18 *****/
19
20
21
22
23
24          use pnr alder using $dorg/bef`t'           ///
25 Bring in all 30-49 yr olds
26          if inrange(alder`t',20,59), clear // from
27 the population register
28
29          **year variable
30          gen year = `w' //time for concussion for control
31 group `time'
32
33          **limit sample to those who suffer a concussion in
34 `t'
35          merge 1:1 pnr year using $data/
36 concussion_`time'.dta, ///
37          keep(3) nogen
38
39          forvalue x=`n'/'c'{           //add
40 longitudinal data
41          merge 1:1 pnr using $dorg/
42 bef`x', ///
43          keep(1 3) keepus(efalle
44 alder koen) //add information on spouse,
45
46          //age, and gender
47          rename _merge merge`x' //indicator
48 for whether in DK that year
49          if `x' < 2017{
50          **Add salary and SES
51 information
52          merge 1:m pnr using
53 $dorg/ind`x', ///          1:m to account for duplicates
54          nogen keep(1 3)
55 keepus(erhvervsindk_13 pre_socio personindk dispon_13
56 aekvivadisp_13) // in data on non-important variables
57
58
59
60

```

```

1
2
3
4 ==1 bysort _all: keep if _n
5 duplicates //drop perfect
6
7 **Align variable names
8 and account for inflation
9
10 rename erhvervsindk_13
11 loenmv
12 rename pre_socio
13 pre_socio`x'
14
15
16 foreach kk in personindk
17 dispon_13 aekvivadisp_13 loenmv{
18 rename `kk'
19 `kk'`x'
20 }
21
22 foreach kk in personindk
23 dispon_13 aekvivadisp_13 loenmv{
24 replace `kk'`x' =
25 `kk'`x'/{price`x'}
26 }
27
28 **Bring in educational
29 information
30 merge 1:1 pnr using
31 $dorg/uddany`x', ///
32 noegen keep(1 3)
33 keepus(hffsp)
34 }
35 if `x' == 2017{
36 merge 1:m pnr using $dorg/
37 ind`x', /// m:1 to account for duplicates
38 noegen keep(1 3)
39 keepus(erhvervsindk_13 pre_socio personindk) // in data on
40 non-important variables
41
42
43 bysort _all: keep if _n ==1
44 //drop perfect duplicates
45 **Align variable names and account
46 for inflation
47
48 rename erhvervsindk_13 loenmv
49 rename pre_socio pre_socio`x'
50
51
52 foreach kk in personindk loenmv{
53 rename `kk'
54 `kk'`x'
55 }
56
57 foreach kk in personindk loenmv{
58
59 replace `kk'`x' =
60

```

```

1
2
3      `kk`x'/{price`x'}
4
5      }
6
7      **Bring in educational information
8      merge 1:1 pnr using $dorg/
9
10     udda`x', ///
11
12     keepus(hfaudd)
13
14     }
15
16     **Reshape data to panel structure
17     if `count' >= 2012 reshape long efalle alder koen
18     loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge
19     hfaudd , i(pnr) j(t)
20     if `count' < 2012 reshape long efalle alder koen
21     loenmv pre_socio personindk dispon_13 aekvivadisp_13 hffsp merge ,
22     i(pnr) j(t)
23     gen count = t-`t' //variable for
24     time to concussion for treatment
25     gen control`time' =1 //control indicator
26     save $data/control_temp, replace
27
28     **Build sample with treatment and control `time'
29     for year `count'
30
31     use $data/sample_temp
32     append using $data/control_temp
33
34     **fixes control and treatment indicators
35     replace control`time' = 0 if control`time'==.
36     replace treatment = 0 if treatment==.
37
38     **Picks up changes to education variable
39     if `count' >=2012{
40         tostring hfaudd, replace
41         rename hfaudd start
42         merge m:1 start using "\
43     \srvfsenas1\data\Formater\SAS formater i Danmarks
44     Statistik\STATA_datsaet\Disced\c_udd_niveau_l1l2_k.dta" , nogen
45     keep(1 3)
46
47         destring UDD, replace force
48         **Replace all with high school degree or
49         higher in HFAUDD to have HFFSP = 40000001
50         replace hffsp = 40000001 if t == 2017 &
51         inrange(UDD,30,80)
52         replace hffsp = 0 if t == 2017 & !
53         inrange(UDD,30,80)
54         drop UDD start
55     }
56
57     sort pnr t
58
59
60

```

```
1
2
3
4         save $data/sample_control_`count'_`time'.dta,
5 replace
6
7     }
8 }
9
10
11 forvalue time = 1/5{
12     forvalue count =2003/2012{
13         if `time' ==1 & `count' ==2003 use $data/
14 sample_control_`count'_`time'.dta, clear
15         else append using $data/
16 sample_control_`count'_`time'.dta
17
18         if `time' ==5 & `count' ==2012 bysort pnr: keep if
19 _n ==1
20         if `time' ==5 & `count' ==2012 count
21     }
22 }
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

Peer review only

```

1
2
3 clear
4
5 *****
6 *****
7 *****
8 *****
9 **
10 **          Calculate share of year on public benefits and size
11 of benefit
12 **          payments for Fallesen and Campos (2020)
13 **
14 **
15 **
16 *****
17 *****
18 *****
19 *****
20 /*globals for price index to calculate income at 2015-level across
21 years*/
22 /*Price index obtained from www.dst.dk/en/statistik/emner/priser-og-
23 forbrug/forbrugeriser/nettoprisindeks */
24
25 global price1980 = .358
26 global price1981 = .398
27 global price1982 = .439
28 global price1983 = .466
29 global price1984 = .494
30 global price1985 = .517
31 global price1986 = .521
32 global price1987 = .537
33 global price1988 = .564
34 global price1989 = .594
35 global price1990 = .612
36 global price1991 = .628
37 global price1992 = .642
38 global price1993 = .651
39 global price1994 = .662
40 global price1995 = .674
41 global price1996 = .688
42 global price1997 = .703
43 global price1998 = .713
44 global price1999 = .728
45 global price2000 = .751
46 global price2001 = .769
47 global price2002 = .788
48 global price2003 = .806
49 global price2004 = .817
50 global price2005 = .833
51 global price2006 = .850
52 global price2007 = .867
53 global price2008 = .899
54 global price2009 = .917
55 global price2010 = .936
56 global price2011 = .960
57 global price2012 = .978
58
59
60

```

```

1
2
3 global price2013 = .986
4 global price2014 = .994
5 global price2015 =1.00
6 global price2016 =1.005
7 global price2017 =1.017
8
9
10
11 **Global for path to registry data
12 global dorg "e:/data/rawdata/706630"
13 *Global for processed data
14 global data "E:/data/workdata/706630/pf/FallesenCampos/data"
15
16 forvalue t=1996/2017{
17
18     ** Read in data on social benefits reciprocity share of
19 weeks
20     ** from the DREAM database
21     use $dorg/dream`t'
22     gen share =0
23     forvalue y = 1/52{
24         if `y' < 10 replace share = share+1 if
25 y_0`y' !=.
26         if `y' > 9 replace share = share+1 if
27 y_`y' !=.
28         if `y' < 10 drop y_0`y'
29         if `y' > 9 drop y_`y'
30
31     }
32     **Generate annual measure of share of year receiving social
33 benefits
34     replace share = share/52
35     keep pnr share
36     gen t = `t'
37     if `t' > 1996 append using $data/temp.dta
38     save $data/temp.dta, replace
39
40 }
41
42
43
44 forvalue t=1998/2017{
45
46     **Read in information on size of different types of social
47 benefits
48
49     if `t' < 2002{
50         use pnr syg_barsel_13 konthj arblhum pre_socio
51 using $dorg/ind`t'.dta, clear
52         replace syg_barsel_13 = syg_barsel_13/{price`t'}
53         replace konthj = konthj /{price`t'}
54         replace arblhum = arblhum/{price`t'}
55         gen kont_dag = konthj+arblhum
56         drop konthj arblhum
57     }
58     if `t' >= 2002 & `t' < 2013{
59
60

```

```
1
2
3       use pnr syg_barsel_13 adagpagn konthj arblhum
4 pre_socio using $dorg/ind`t'.dta, clear
5       replace syg_barsel_13 = syg_barsel_13/{price`t'}
6       replace adagpagn = adagpagn/{price`t'}
7       replace konthj = konthj /{price`t'}
8       replace arblhum = arblhum/{price`t'}
9       gen kont_dag = konthj+arblhum
10      drop konthj arblhum
11    }
12  if `t' >= 2013{
13    use pnr syg_barsel_13 adagpagn dagpenge_kontant_13
14 pre_socio using $dorg/ind`t'.dta, clear
15    replace syg_barsel_13 = syg_barsel_13/{price`t'}
16    replace adagpagn = adagpagn/{price`t'}
17    replace dagpenge_kontant_13 =
18 dagpenge_kontant_13 /{price`t'}
19    gen kont_dag = dagpenge_kontant_13-syg_barsel_13
20    drop dagpenge_kontant_13
21  }
22  gen t = `t'
23  compress
24  bysort pnr: keep if _n ==1
25  if `t' > 1998 append using $data/temp2.dta
26  if `t' == 2017{
27    sort pnr t
28  }
29  save $data/temp2.dta, replace
30 }
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```

1
2
3 clear all
4
5 *****
6 *****
7 **
8 **
9 **          This program geenrates figerus and results for
10 Fallesen & Campos (2020)
11 **          study of concussion's impact on productivity
12 measured through
13 **          annual salary
14 **
15 *****
16 *****
17
18 **Global for path to registry data
19 global dorg "E:/data/rawdata/706630"
20
21 *Global for processed data
22 global data "[home]/data"
23
24 *Global on figures
25 global highdef "[home]\highdef"
26
27
28
29
30 forvalue time = 1/5{          //for different control groups
31
32     local post_period = 5    // local for number of years
33     observed post
34
35     concussion for exposure group          //
36
37     local endtime = 2017-`post_period'    /*last year where
38 we allow for exposure event
39
40 occur, in order to have long enough control          to
41
42 Defined by latest year available data*/          period.
43
44     **Matrixes to capture estimates
45     matrix results = J(15,6,.)          // For salary estimates
46     matrix results_p = J(15,6,.)      // For Pr(salary=0)
47 estimates
48     matrix t = J(15,6,.)          // For time indicators
49
50     forvalue count = 2003(1)`endtime'{
51         use $data/sample_control_`count'`time', clear
52
53         gen female = koen==2
54
55         qui{
56             gen edu =0
57             replace edu = 1 if inrange(hffsp,
58 20000000,39000000) | ///
59
60

```



```

1
2
3
4     (hffsp >40000000 & hffsp!=.)
5         }
6
7         **exclude individuals in years where they do not
8 appear in data,
9         **due to either death or migration, as well as
10 periods from when
11         **the control group sufer their concussion
12 drop if merge ==1 | count > `time'-1
13
14         **generate concussion variable
15 gen treat = inrange(count,0,`time'-1) & treatment
16 ==1
17
18         replace treat = time_from_incident if count ==0 &
19 treatment ==1
20
21         **Generate pre-concussion income difference
22 for
23         **use in calculating marginal effects
24 sum loenmv if count <0 & treatment ==0
25 local control =r(mean)
26 sum loenmv if count <0 & treatment ==1
27 local treat =r(mean)
28 sum loenmv if count>=0 & treatment ==0
29 local control_post =r(mean)
30 gen post = count >=0
31
32         **estimate DiD model on salary
33 reghdfe loenmv treat, abs(alder female post
34 treatment edu year) cl(pnr)
35 matrix b =e(b) //regression coefficient
36 matrix V = e(V) // standard error^2
37 local n = `count'-2002 //time
38
39         matrix results[`n',1] = b[1,1]
40         matrix results[`n',2] = V[1,1]^0.5
41         matrix results[`n',3] = b[1,1]/
42 (`control_post'-(`control'-`treat'))
43         matrix results[`n',4] = `n'
44
45     }
46     svmat results
47
48     rename results1 est
49     rename results2 se
50     rename results3 marg
51     rename results4 time
52
53     replace time = time+(`time'-3)*.1 //jitter estimates for
54 graph
55
56
57
58
59
60

```

```

1
2
3     keep est* se* marg* time
4     keep if est !=.
5
6     replace est = est/7446    //estimate measured as 1000 Euro
7     replace se = se/7446     //S.E. measured as 1000 Euro
8
9     gen upper = est+se*1.96 // Upper CI
10    gen lower = est-se*1.96 // Lower CI
11
12
13
14    gen control = `time'           //indicate control group
15
16    if `time' >1 append using $data/results.dta
17    save $data/results.dta, replace
18 }
19
20 use $data/results.dta, clear
21
22 replace time = 2002+time
23
24
25 *reads in unemployment statistcis obtained from statistikbanken.dk/
26 en/
27
28 gen unemp = 5.8 if time ==2003
29 replace unemp = 5.8 if time ==2004
30 replace unemp = 5.1 if time ==2005
31 replace unemp = 3.9 if time ==2006
32 replace unemp = 2.7 if time ==2007
33 replace unemp = 1.9 if time ==2008
34 replace unemp = 3.6 if time ==2009
35 replace unemp = 4.2 if time ==2010
36 replace unemp = 4.0 if time ==2011
37 replace unemp = 4.5 if time ==2012
38
39
40 scatter est time if control ==1, mcolor(navy) yaxis(1) ysc(range(-4
41 3) axis(1)) ylab(-4(1)3) || ///
42 scatter est time if control ==2, mcolor(blue) || ///
43 scatter est time if control ==3, mcolor(midblue) || ///
44 scatter est time if control ==4, mcolor(gray) || ///
45 scatter est time if control ==5, mcolor(ltblue) || ///
46 rspike upper lower time if control ==1, lcolor(navy) || ///
47 rspike upper lower time if control ==2, lcolor(blue) || ///
48 rspike upper lower time if control ==3, lcolor(midblue) || ///
49 rspike upper lower time if control ==4, lcolor(gray) || ///
50 rspike upper lower time if control ==5, lcolor(ltblue) || ///
51 line unemp time , lcolor(black) yaxis(2) ysc(range(0 6) axis(2))
52 ylab(0(1)6, axis(2)) ///
53 xsc(range(2002.5 2012.5)) xlab(2003(2)2012) ///
54 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
55 xti("Year of concussion for exposure group") scale(.95) ///
56 legend(label(1 " Control," "{&Delta}=1 yr") ///
57 label(2 " Control," "{&Delta}=2 yr") ///
58 label(3 " Control," "{&Delta}=3 yr") ///
59
60

```

```

1
2
3 label(4 " Control," "{&Delta}=4 yr") ///
4 label(5 " Control," "{&Delta}=5 yr") ///
5 label(6 "% Unemp." "of LF") ///
6 c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
7 c(1) symx(4) region(lc(white))) ///
8 yti("Effect on Salary (in EUR 1K)", height(7) axis(1)) ///
9 yti("Percent of Full Time Uemployed among LF", height(7) axis(2))
10
11
12
13 graph export $highdef/marg_est.png, replace width(3900)
14
15
16 forvalue control_time=1/5{
17     local end = 2012 // last incident year in data
18     if `control_time' ==1     eststo clear
19
20     **build dataset for joint estimate across years
21     forvalue count=2003/`end'{
22         if `count'==2003{
23             use $data/
24 sample_control_`count'_`control_time'.dta, clear
25             gen time = `count' //incident year
26 indicator
27             }
28             else append using $data/
29 sample_control_`count'_`control_time'.dta
30             replace time = `count' if time ==.
31
32
33             **exclude individuals in years where they do not
34 appear in data,
35             **due to either death or migration, as well as
36 periods from when
37             **the control group sufer their concussion
38 drop if merge ==1 | count > `control_time'-1
39         }
40
41         gen female = koen==2
42
43         //build ident, so we can multivariate cluster for
44 individuals
45         //who occur both as control and exposure during the period
46 (id)
47
48         bysort pnr time: gen helpx = _n ==1
49         gen id= sum(helpx)
50         drop helpx
51
52         **Generate educational groups
53         qui{
54             gen edu =0
55             replace edu = 1 if inrange(hffsp,20000000,39000000)
56
57 | ///
58                                     (hffsp
59 >40000000 & hffsp!=.)
60

```

```

1
2
3     }
4
5     **Calculate number of observations for exposure and control
6     count if count==0 & treatment ==1
7     local Ntreated = r(N)
8     count if count==0 & treatment ==0
9     local Ncontrol = r(N)
10
11    **generate concussion variable
12    gen treat = inrange(count,0,`control_time'-1) & treatment
13
14 ==1
15    replace treat = time_from_incident if count ==0 & treatment
16 ==1
17
18    **Generate pre-concussion income difference          for
19    **use in calculating marginal effects
20    sum loenmv if count <0 & treatment ==0
21    local control =r(mean)
22    sum loenmv if count <0 & treatment ==1
23    local treat =r(mean)
24    sum loenmv if count>=0 & treatment ==0
25    local control_post =r(mean)
26
27    forvalue t=-4/4{
28        local n = `t'*-1
29        if `t' < -1 gen T_`n' = treatment ==1 & count ==`t'
30        if `t' > -1 gen T`t' = treatment ==1 & count ==`t'
31
32    }
33
34
35    **estimate DiD model on salary
36    reghdfe loenmv T*, abs(alder female count time treatment
37 edu) cl(pnr id)
38    eststo est1_`control_time'
39    if `control_time'==1 matrix results = J(5,5,.) // matrix to
40 capture results
41    if `control_time'==1 matrix results_p = J(5,5,.) // matrix
42 to capture results
43    if `control_time'==1 matrix results_pre = J(5,5,.) //
44 matrix to capture results
45
46    matrix b = e(b)
47    matrix V = e(V)
48    local n = `control_time'
49
50    matrix results[`n',1]    = b[1,1] / 7466 // capture beta
51 results as 1K Euro
52    matrix results[`n',2]    = (V[1,1]^.5)/7466 //
53 capture standard error as 1K Euro
54    matrix results[`n',3]    =      b[1,1]/(`control_post'-
55 (`control'-`treat'))
56    matrix results[`n',4]    =      `n'
57
58
59    gen no_lon = loenmv<1 //dummy for no salary
60

```

```

1
2
3
4
5
6     **estimate DiD LP-model for pre-trends
7     xi: reghdfe loenmv T*, abs(alder female count time
8 treatment edu) cl(pnr id), if count <0
9     eststo est3_`control_time'
10    matrix b = e(b) //regression coefficient
11    matrix V = e(V) // standard error^2
12
13
14    matrix results_pre[`n',1] =      b[1,1]
15    matrix results_pre[`n',2] =      V[1,1]^.5
16    matrix results_pre[`n',3] =      b[1,1]/(`control_post'-
17 (`control'-`treat'))
18    matrix results_pre[`n',4] =      `n'
19
20 }
21
22 esttab est1_* using [home]/tables/dynamic1.rtf, ///
23     replace se(1) b(1) compress nogap star(+ .1 * .05 ** .01
24 *** .001) ///
25     keep(T*)
26
27 esttab est2_* using [home]/tables/dynamic2.rtf, ///
28     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
29 *** .001) ///
30     keep(T*)
31
32 esttab est3_* using [home]/tables/pre_trends.rtf, ///
33     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
34 *** .001) ///
35     keep(T*)
36
37
38 forvalue control_time=1/5{
39     local end = 2012 // last incident year in data
40     if `control_time' ==1     eststo clear
41     qui{
42         **build dataset for joint estimate across years
43         forvalue count=2003/`end'{
44             if `count'==2003{
45                 use $data/
46 sample_control_`count'_`control_time'.dta, clear
47                 gen time = `count' //incident year
48 indicator
49             }
50             else append using $data/
51 sample_control_`count'_`control_time'.dta
52                 replace time = `count' if time ==.
53
54
55         **exclude individuals in years where they
56 do not appear in data,
57         **due to either death or migration, as
58 well as periods from when
59         **the control group suffer their concussion
60

```

```

1
2
3           drop if merge ==1 | count >
4 `control_time'-1
5     }
6
7     gen female = koen==2
8
9           **Generate educational groups
10    gen edu =0
11    replace edu = 1 if inrange(hffsp,20000000,39000000)
12 | ///
13
14 (hffsp >40000000 & hffsp!=.)
15
16
17
18 //build ident, so we can multivariate cluster for
19 individuals //who occur both as control and exposure during the
20 period (id)
21
22
23 bysort pnr time: gen helpx = _n ==1
24 gen id= sum(helpx)
25 drop helpx
26
27
28 **Calculate number of observations for exposure and
29 control
30 count if count==0 & treatment ==1
31 local Ntreated = r(N)
32 count if count==0 & treatment ==0
33 local Ncontrol = r(N)
34
35 **generate concussion variable
36 gen treat = inrange(count,0,`control_time'-1) &
37 treatment ==1
38 replace treat = time_from_incident if count ==0 &
39 treatment ==1
40
41
42 **Generate pre-concussion income difference
43 for
44 **use in calculating marginal effects
45 sum loenmv if count <0 & treatment ==0
46 local control =r(mean)
47 sum loenmv if count <0 & treatment ==1
48 local treat =r(mean)
49 sum loenmv if count>=0 & treatment ==0
50 local control_post =r(mean)
51
52 forvalue t=-4/4{
53     local n = `t'*-1
54     if `t' < -1 gen T_`n' = treatment ==1 &
55 count ==`t'
56     if `t' > -1 gen T`t' = treatment ==1 &
57 count ==`t'
58
59 }
60

```

```

1
2
3
4         gen post = count > -1
5
6         **estimate DiD model on salary
7         reghdfe loenmv treat, abs(alder female post time
8 treatment edu) cl(pnr id)
9         eststo est1_`control_time'
10        if `control_time'==1 matrix results = J(5,5,.) //
11 matrix to capture results
12        if `control_time'==1 matrix results_p = J(5,5,.) //
13 matrix to capture results
14
15
16
17        matrix b = e(b)
18        matrix V = e(V)
19        local n = `control_time'
20
21        matrix results[`n',1]      = b[1,1] / 7466 //
22 capture beta results as 1K Euro
23        matrix results[`n',2]      = (V[1,1]^5)/
24 7466 // capture standard error as 1K Euro
25        matrix results[`n',3]      =      b[1,1]/
26 (`control_post'-(`control'-`treat'))
27        matrix results[`n',4]      =      `n'
28
29
30    }
31    *examining balance of samples
32    di in ye `control_time'
33    bysort treatment: sum female alder edu if count ==0
34
35 }
36
37
38
39
40 svmat results
41
42 gen upper = results1+results2*1.96
43 gen lower = results1-results2*1.96
44
45 scatter results1 results4 if results4 ==1, mcolor(navy) || ///
46 scatter results1 results4 if results4 ==2, mcolor(navy) || ///
47 scatter results1 results4 if results4 ==3, mcolor(navy) || ///
48 scatter results1 results4 if results4 ==4, mcolor(navy) || ///
49 scatter results1 results4 if results4 ==5, mcolor(navy) || ///
50 rspike upper lower results4 if results4 ==1, lcolor(navy) || ///
51 rspike upper lower results4 if results4 ==2, lcolor(navy) || ///
52 rspike upper lower results4 if results4 ==3, lcolor(navy) || ///
53 rspike upper lower results4 if results4 ==4, lcolor(navy) || ///
54 rspike upper lower results4 if results4 ==5, lcolor(navy) ///
55 ysc(range(-2 1)) ylab(-2(.5)1) ///
56 xsc(range(.5 5.5)) xlab(1(1)5) ///
57 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
58 xti("Years between exposure and control incident") scale(.95) ///
59
60

```

```

1
2
3 legend(label(1 " Control," "{&Delta}=1 yr") ///
4 label(2 " Control," "{&Delta}=2 yr") ///
5 label(3 " Control," "{&Delta}=3 yr") ///
6 label(4 " Control," "{&Delta}=4 yr") ///
7 label(5 " Control," "{&Delta}=5 yr") ///
8 c(1) order(1 2 3 4 5) pos(3) size(small) ///
9 c(1) symx(4) region(lc(white))) ///
10 yti("Effect in 1K Euro ({&delta}{subscript: Salary})",
11 height(7)) ///
12 legend(off)
13
14
15 graph export $highdef/est2003_2011.png, replace width(3900)
16
17
18 **Reports marginal effects for period 2003–2011 in percent
19
20 gen upper2 = (results3+results2/(results1/results3)*1.96)*100
21 gen lower2 = (results3–results2/(results1/results3)*1.96)*100
22 replace results3 = results3*100
23
24
25 scatter results3 results4 , mcolor(navy) || ///
26 rspike upper2 lower2 results4, lcolor(navy) ///
27 ysc(range(-5 0)) ylab(-5(.5)0) ///
28 xsc(range(.5 5.5)) xlab(1(1)5) ///
29 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
30 /*title("Percentage change in salary, 2003–10")*/ ///
31 yti("Salary change (in %)", height(7)) ///
32 xti("Years between exposure and control incident") scale(.95) ///
33 legend(label(1 " Control," "{&Delta}=1 yr") ///
34 label(2 " Control," "{&Delta}=2 yr") ///
35 label(3 " Control," "{&Delta}=3 yr") ///
36 label(4 " Control," "{&Delta}=4 yr") ///
37 label(5 " Control," "{&Delta}=5 yr") ///
38 c(1) order(1 2 3 4 5) pos(3) size(small) ///
39 c(1) symx(4) region(lc(white))) legend(off)
40 /*
41 note("Marginal effects for exposure dummy across spacing of control
42 groups. Decrease " ///
43 "calculated by dividing {&delta} with the normalized control groups'
44 average salary " ///
45 "post-concussion. Control groups suffer concussions 1, 2, 3, 4, and
46 5 years (&Delta) after" ///
47 "the exposure group. Both control and exposure group are 30–49 years
48 of age when" ///
49 "exposure group suffers concussion. 95% confidence intervals.")*/
50
51
52 graph export $highdef/marginal2003_2011.png, replace width(3900)
53
54
55
56
57 *****
58 *****
59 **
60

```



```

1
2
3      **          Results for individuals with at least highschool
4      **
5      **
6      ****
7      ****
8
9
10     forvalue control_time=1/5{
11         local end = 2012 // last incident year in data
12
13         **build dataset for joint estimate across years
14         forvalue count=2003/`end'{
15             if `count'==2003{
16                 use $data/
17 sample_control_`count'_`control_time'.dta, clear
18                 gen time = `count' //incident year
19 indicator
20                 }
21                 else append using $data/
22 sample_control_`count'_`control_time'.dta
23                 replace time = `count' if time ==.
24
25                 **exclude individuals in years where they do not
26 appear in data,
27                 **due to either death or migration, as well as
28 periods from when
29                 **the control group suffer their concussion
30 drop if merge ==1 | count > `control_time'-1
31
32             }
33
34
35         gen female = koen==2
36
37         **Generate educational groups
38         qui{
39             gen edu =0
40             replace edu = 1 if inrange(hffsp,20000000,39000000)
41 | ///
42                                     (hffsp
43 >40000000 & hffsp!=.)
44                                     }
45
46         keep if edu==1
47
48         //build ident, so we can multivariate cluster for
49 individuals
50 //who occur both as control and exposure during the period
51 (id)
52
53         bysort pnr time: gen helpx = _n ==1
54         gen id= sum(helpx)
55         drop helpx help
56
57
58
59
60

```

```

1
2
3      **Calculate number of observations for exposure and control
4      count if count==0 & treatment ==1
5      local Ntreated = r(N)
6      count if count==0 & treatment ==0
7      local Ncontrol = r(N)
8
9      **generate concussion variable
10     gen treat = inrange(count,0,`control_time'-1) & treatment
11
12 ==1
13     replace treat = time_from_incident if count ==0 & treatment
14 ==1
15
16     **Generate pre-concussion income difference          for
17     **use in calculating marginal effects
18     sum loenmv if count <0 & treatment ==0
19     local control =r(mean)
20     sum loenmv if count <0 & treatment ==1
21     local treat =r(mean)
22     sum loenmv if count>=0 & treatment ==0
23     local control_post =r(mean)
24
25     gen post = count >=0
26
27     **estimate DiD model on salary
28     xi: reghdfe loenmv treat, abs(alder female post time
29     treatment) cl(id pnr)
30
31     if `control_time'==1 matrix results_edu = J(5,5,.) //
32     matrix to capture results
33     if `control_time'==1 matrix results_p_edu = J(5,5,.) //
34     matrix to capture results
35
36
37
38     matrix b = e(b)
39     matrix V = e(V)
40     local n = `control_time'
41
42     matrix results_edu[`n',1]          = b[1,1] / 7466 //
43     capture beta results as 1K Euro
44     matrix results_edu[`n',2]          = (V[1,1]^0.5)/
45     7466 // capture standard error as 1K Euro
46     matrix results_edu[`n',3] =      b[1,1]/(`control_post'-
47     (`control'-`treat'))
48     matrix results_edu[`n',4] =      `n'
49
50
51 }
52
53
54 *****
55 *****
56 **
57 **           Results for individuals with no high school+
58 **
59 **
60

```

```

1
2
3 *****
4 *****
5
6
7 forvalue control_time=1/5{
8     local end = 2012 // last incident year in data
9
10    **build dataset for joint estimate across years
11    forvalue count=2003/`end'{
12        if `count'==2003{
13            use $data/
14            sample_control_`count'_`control_time'.dta, clear
15            gen time = `count' //incident year
16            indicator
17                }
18            else append using $data/
19            sample_control_`count'_`control_time'.dta
20            replace time = `count' if time ==.
21
22
23            **exclude individuals in years where they do not
24            appear in data,
25            **due to either death or migration, as well as
26            periods from when
27            **the control group sufer their concussion
28            drop if merge ==1 | count > `control_time'-1
29
30
31        }
32
33        gen female = koen==2
34
35        //build ident, so we can multivariate cluster for
36        individuals
37        //who occur both as control and exposure during the period
38        (id)
39
40
41        bysort pnr time: gen helpx = _n ==1
42        gen id= sum(helpx)
43        drop helpx help
44
45        **Generate educational groups
46        qui{
47            gen edu =0
48            replace edu = 1 if inrange(hffsp,20000000,39000000)
49            | ///
50
51
52            >40000000 & hffsp!=.)
53            }
54
55            keep if edu==0
56
57
58
59            **Calculate number of observations for exposure and control
60

```

```

1
2
3     count if count==0 & treatment ==1
4     local Ntreated = r(N)
5     count if count==0 & treatment ==0
6     local Ncontrol = r(N)
7
8     **generate concussion variable
9     gen treat = inrange(count,0,`control_time'-1) & treatment
10
11 ==1
12     replace treat = time_from_incident if count ==0 & treatment
13 ==1
14
15     **Generate pre-concussion income difference          for
16     **use in calculating marginal effects
17     sum loenmv if count <0 & treatment ==0
18     local control =r(mean)
19     sum loenmv if count <0 & treatment ==1
20     local treat =r(mean)
21     sum loenmv if count>=0 & treatment ==0
22     local control_post =r(mean)
23     gen post = count >=0
24
25     **estimate DiD model on salary
26     xi: reghdfe loenmv treat, abs(alder female post time
27 treatment) cl(id pnr)
28
29     if `control_time'==1 matrix results_noedu = J(5,5,.) //
30 matrix to capture results
31     if `control_time'==1 matrix results_p_noedu = J(5,5,.) //
32 matrix to capture results
33
34
35
36     matrix b = e(b)
37     matrix V = e(V)
38     local n = `control_time'
39
40     matrix results_noedu[`n',1]      = b[1,1] / 7466 //
41 capture beta results as 1K Euro
42     matrix results_noedu[`n',2]      = (V[1,1]^0.5)/
43 7466 // capture standard error as 1K Euro
44     matrix results_noedu[`n',3]      =      b[1,1]/
45 (`control_post'-(`control'-`treat'))
46     matrix results_noedu[`n',4]      =      `n'
47
48
49
50
51 }
52 *****
53 *****
54 **
55 **     Draw figure for subgroups
56 **
57 **
58 *****
59 *****
60

```

```

1
2
3
4 local t= -.15
5
6 foreach x in noedu edu{
7     svmat results_`x'
8     replace results_`x'4= results_`x'4+`t'
9     svmat results_p_`x'
10    replace results_p_`x'4= results_p_`x'4+`t'
11
12
13    gen upper_`x' = results_`x'1+results_`x'2*1.96
14    gen lower_`x' = results_`x'1-results_`x'2*1.96
15
16    gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
17 results_`x'3)*1.96)*100
18    gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
19 results_`x'3)*1.96)*100
20    replace results_`x'3 = results_`x'3*100
21
22    gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
23    gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
24
25    local t = `t'+.1
26
27 }
28
29
30 keep results* upper* lower*
31 keep if _n <=5
32
33
34 **generate locals for figure
35
36 foreach x in noedu edu{
37     if "`x'" == "nopay" local color = "navy"
38     if "`x'" == "pay"     local color = "red"
39     if "`x'" == "noedu" local color = "green"
40     if "`x'" == "edu"     local color = "purple"
41
42     local figure_`x' "scatter results_`x'1 results_`x'4,
43 mcolor(`color') || rspike upper_`x' lower_`x' results_`x'4,
44 lcolor(`color') vertical"
45     if "`x'" == "nopay" local figure2_`x' "scatter results_`x'3
46 results_`x'4, mcolor(`color') || rspike upper2_`x' lower2_`x'
47 results_`x'4, lcolor(`color') vertical "
48     else local figure2_`x' "scatter results_`x'3 results_`x'4,
49 mcolor(`color') || rspike upper2_`x' lower2_`x' results_`x'4,
50 lcolor(`color') vertical "
51     local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
52 mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
53 lcolor(`color') vertical "
54 }
55
56
57
58 `figure_noedu' || `figure_edu' ///
59 legend( ///
60

```

```

1
2
3 label(1 "Less than" "high school") ///
4 label(3 "At least" "high school") ///
5 c(1) order(1 3 ) pos(3) size(small) ///
6 c(1) symx(4) region(lc(white))) ///
7 ysc(range(-4 2)) ylab(-4(1)2) ///
8 xsc(range(.5 5.5)) xlab(1(1)5) ///
9 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
10 ///title("Parameter estimates across control group, 2003-10") ///
11 xti("Years between exposure and control incident") scale(.95) ///
12 yti("Effect in 1K Euro ({&delta;}{subscript: Salary})",
13 height(7)) ///
14 /*note("Parameter estimates for exposure dummy across spacing of
15 control groups." ///
16 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
17 after the exposure group." ///
18 "Both control and exposure group are 30-49 years of age when
19 exposure group suffers " ///
20 "concussion. 95% confidence intervals.")
21 */
22
23
24 graph export $highdef/grouped_est2003_2011.png, replace width(3900)
25
26 preserve
27
28
29
30
31
32 `figure2_noedu' || `figure2_edu' ///
33 legend( ///
34 label(1 "Less than" "high school") ///
35 label(3 "At least" "high school") ///
36 c(1) order(1 3 5) pos(3) size(small) ///
37 c(1) symx(4) region(lc(white))) ///
38 ysc(range(-12 3)) ylab(-12(3)3) ///
39 xsc(range(.5 5.5)) xlab(1(1)5) ///
40 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
41 /// title("Percentage change in salary, 2003-10") ///
42 yti("Salary change (in %)", height(7)) ///
43 xti("Years between exposure and control incident") scale(.95) ///
44 /*note("Marginal effects for exposure dummy across spacing of
45 control groups." ///
46 "change calculated by {&delta;} with the normalized control groups'
47 average" ///
48 "salary post-concussion. Control groups suffer concussions 1, 2, 3,
49 4, and 5 years (&Delta) after" ///
50 "the exposure group. Both control and exposure group are 30-49 years
51 of age when" ///
52 "exposure group suffers concussion. 95% confidence intervals.")*/
53
54
55 graph export $highdef/grouped_marginal2003_2011.png, replace
56 width(3900)
57
58 restore
59
60

```

```

1
2
3
4 *****
5 *****
6 **
7 **           Results different age-groups
8 **
9 **
10 *****
11 *****
12
13
14 forvalue y=20(5)55{
15     forvalue control_time=1/5{
16         local end = 2012 // last incident year in data
17
18         **build dataset for joint estimate across years
19         forvalue count=2003/\`end'{
20             if `count'==2003{
21                 use $data/
22 sample_control_`count'_`control_time'.dta, clear
23                 gen time = `count' //incident year
24 indicator
25                 }
26                 else append using $data/
27 sample_control_`count'_`control_time'.dta
28                 replace time = `count' if time ==.
29
30                 **exclude individuals in years where they
31 do not appear in data,
32                 **due to either death or migration, as
33 well as periods from when
34                 **the control group suffer their concussion
35 drop if merge ==1 | count >
36 `control_time'-1
37
38
39
40     }
41
42     gen female = koen==2
43
44     //build ident, so we can multivariate cluster for
45 individuals
46     //who occur both as control and exposure during the
47 period (id)
48
49
50     bysort pnr time: gen helpx = _n ==1
51     gen id= sum(helpx)
52     drop helpx help
53
54     gen nopay = loenmv <1
55
56     **Generate age group
57     local z = `y'+4
58     gen help = count == 0 & inrange(alder,`y',`z')
59     bysort id: egen helpx =max(help)
60

```

```

1
2
3         keep if helpx == 1
4         drop helpx help
5
6
7
8
9
10        **Calculate number of observations for exposure and
11 control
12        count if count==0 & treatment ==1
13        local Ntreated = r(N)
14        count if count==0 & treatment ==0
15        local Ncontrol = r(N)
16
17        **generate concussion variable
18        gen treat = inrange(count,0,`control_time'-1) &
19 treatment ==1
20        replace treat = time_from_incident if count ==0 &
21 treatment ==1
22
23        **Generate pre-concussion income difference
24 for
25
26        **use in calculating marginal effects
27        sum loenmv if count <0 & treatment ==0
28        local control =r(mean)
29        sum loenmv if count <0 & treatment ==1
30        local treat =r(mean)
31        sum loenmv if count>=0 & treatment ==0
32        local control_post =r(mean)
33        gen post = count >=0
34
35        **estimate DiD model on salary
36        xi: reghdfe loenmv treat, abs(alder female post
37 time treatment) cl(id pnr)
38
39        if `control_time'==1 matrix results_`y' =
40 J(5,5,.) // matrix to capture results
41        if `control_time'==1 matrix results_p_`y' =
42 J(5,5,.) // matrix to capture results
43
44
45        matrix b = e(b)
46        matrix V = e(V)
47        local n = `control_time'
48
49        matrix results_`y'[`n',1]           = b[1,1] /
50 7466 // capture beta results as 1K Euro
51        matrix results_`y'[`n',2]           = (V[1,1]^0.5)/
52 7466 // capture standard error as 1K Euro
53        matrix results_`y'[`n',3] = b[1,1]/
54 (`control_post'-(`control'-`treat'))
55        matrix results_`y'[`n',4] = `n'
56
57
58
59     }
60

```



```

1
2
3     }
4
5
6
7     local t = -.15    //Jitter estimates along x-axis
8     forvalue x =20(5)55{
9         svmat results_`x'
10        replace results_`x'4= results_`x'4+`t'
11        svmat results_p_`x'
12        replace results_p_`x'4= results_p_`x'4+`t'
13
14
15        gen upper_`x' = results_`x'1+results_`x'2*1.96
16        gen lower_`x' = results_`x'1-results_`x'2*1.96
17
18        gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
19        results_`x'3)*1.96)*100
20        gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
21        results_`x'3)*1.96)*100
22        replace results_`x'3 = results_`x'3*100
23
24        gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
25        gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
26
27        local t = `t'+.1
28
29    }
30
31    keep results* upper* lower*
32    keep if _n <=5
33
34
35    **generate locals for figure
36
37    forvalue x =20(5)55{
38        if `x' == 20 local color = "black"
39        if `x' == 25 local color = "orange"
40        if `x' == 30 local color = "navy"
41        if `x' == 35 local color = "red"
42        if `x' == 40 local color = "green"
43        if `x' == 45 local color = "purple"
44        if `x' == 50 local color = "sienna"
45        if `x' == 55 local color = "teal"
46
47        local figure_`x' "scatter results_`x'1 results_`x'4,
48        mcolor(`color') || rspike upper_`x' lower_`x' results_`x'4,
49        lcolor(`color') vertical"
50        local figure2_`x' "scatter results_`x'3 results_`x'4,
51        mcolor(`color') || rspike upper2_`x' lower2_`x' results_`x'4,
52        lcolor(`color') vertical "
53        local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
54        mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
55        lcolor(`color') vertical "
56    }
57
58
59    /**/
60

```

```

1
2
3 `figure_20' || `figure_25' || `figure_30' || `figure_35' ///
4     || `figure_40' || `figure_45' || `figure_50' ||
5 `figure_55' ///
6 legend( ///
7 label(1 "Age 20-24") ///
8 label(3 "Age 25-29") ///
9 label(5 "Age 30-34") ///
10 label(7 "Age 35-39") ///
11 label(9 "Age 40-44") ///
12 label(11 "Age 45-49") ///
13 label(13 "Age 50-54") ///
14 label(15 "Age 55-59") ///
15 c(1) order(1 3 5 7 9 11 13 15) pos(3) size(small) ///
16 c(1) symx(4) region(lc(white))) ///
17 ysc(range(-4 2)) ylab(-4(1)2) ///
18 xsc(range(.5 5.5)) xlab(1(1)5) ///
19 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
20 /// title("Parameter estimates across control group, 2003-10") ///
21 xti("Years between exposure and control incident") scale(.95) ///
22 yti("Effect in 1K Euro ({&delta;}{subscript: Salary})",
23 height(7)) ///
24 /*note("Parameter estimates for exposure dummy across spacing of
25 control groups." ///
26 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
27 after the exposure group." ///
28 "Age group described age at time of exposure incident. 95%
29 confidence intervals.")*/
30
31
32 graph export $highdef/age_est2003_2011.png, replace width(3900)
33 */
34
35
36 `figure2_20' || `figure2_25' || `figure2_30' || `figure2_35' ///
37     || `figure2_40' || `figure2_45' || `figure2_50' ||
38 `figure2_55' ///
39 legend( ///
40 label(1 "Age 20-24") ///
41 label(3 "Age 25-29") ///
42 label(5 "Age 30-34") ///
43 label(7 "Age 35-39") ///
44 label(9 "Age 40-44") ///
45 label(11 "Age 45-49") ///
46 label(13 "Age 50-54") ///
47 label(15 "Age 55-59") ///
48 c(1) order(1 3 5 7 9 11 13 15) pos(3) size(small) ///
49 c(1) symx(4) region(lc(white))) ///
50 ysc(range(-12 3)) ylab(-12(3)3) ///
51 xsc(range(.5 5.5)) xlab(1(1)5) ///
52 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
53 yti("Salary change (in %)", height(7)) ///
54 xti("Years between exposure and control incident") scale(.95)
55
56
57
58 graph export $highdef/age_marginal2003_2011.png, replace width(3900)
59
60

```

```

1
2
3 `figure_p_30' || `figure_p_35' || `figure_p_40' ||
4 `figure_p_45' /// ///
5 legend(label(1 "Age 30-34") ///
6 label(3 "Age 35-39") ///
7 label(5 "Age 40-44") ///
8 label(7 "Age 45-49") ///
9 c(1) order(1 3 5 7) pos(3) size(small) ///
10 c(1) symx(4) region(lc(white))) ///
11 ysc(range(-.02 .050)) ylab(-0.02(.01)0.05) ///
12 xsc(range(.5 5.5)) xlab(1(1)5) ///
13 yline(0, lcolor(black)) ysize(10) xsize(12) graphr(c(white)) ///
14 title("Parameter estimates across control group, 2003-10") ///
15 yti("Effect on Pr(Salary=0)", height(7)) ///
16 xti("Years between exposure and control incident ({&Delta})")
17 scale(.95) ///
18 note("Parameter estimates for exposure dummy across spacing of
19 control groups." ///
20 "Control groups suffer concussions 1, 2, 3, 4, and 5 (&Delta) years
21 after the exposure group." ///
22 "Age group described age at time of exposure incident. 95%
23 confidence intervals.")
24
25
26 graph export $highdef/age_nopay2003_2011.png, replace width(3900)
27
28
29
30 *****
31 *****
32 **
33 **           Results accross gender
34 **
35 **
36 *****
37 *****
38
39 forvalue y=0/1{
40     forvalue control_time=1/5{
41         local end = 2012 // last incident year in data
42
43         **build dataset for joint estimate across years
44         forvalue count=2003/`end'{
45             if `count'==2003{
46                 use $data/
47 sample_control_`count'_`control_time'.dta, clear
48                 gen time = `count' //incident year
49 indicator
50             }
51             else append using $data/
52 sample_control_`count'_`control_time'.dta
53             replace time = `count' if time ==.
54
55             **exclude individuals in years where they
56 do not appear in data,
57             **due to either death or migration, as
58 well as periods from when
59
60

```

```

1
2
3             **the control group suffer their concussion
4             drop if merge ==1 | count >
5 `control_time'-1
6
7
8         }
9
10        gen female = koen==2
11
12        //build ident, so we can multivariate cluster for
13 individuals
14        //who occur both as control and exposure during the
15 period (id)
16
17        bysort pnr time: gen helpx = _n ==1
18        gen id= sum(helpx)
19        drop helpx help
20
21        gen nopay = loenmv <1
22
23        **Generate age group
24        local z = `y'+4
25        gen help = count == 0 & female==`y'
26        bysort id: egen helpx =max(help)
27        keep if helpx == 1
28        drop helpx help
29
30
31
32
33
34
35        **Calculate number of observations for exposure and
36 control
37        count if count==0 & treatment ==1
38        local Ntreated = r(N)
39        count if count==0 & treatment ==0
40        local Ncontrol = r(N)
41
42        **generate concussion variable
43        gen treat = inrange(count,0,`control_time'-1) &
44 treatment ==1
45        replace treat = time_from_incident if count ==0 &
46 treatment ==1
47
48        **Generate pre-concussion income difference
49 for
50
51        **use in calculating marginal effects
52        sum loenmv if count <0 & treatment ==0
53        local control =r(mean)
54        sum loenmv if count <0 & treatment ==1
55        local treat =r(mean)
56        sum loenmv if count>=0 & treatment ==0
57        local control_post =r(mean)
58        gen post = count >=0
59
60

```

```

1
2
3           **estimate DiD model on salary
4           xi: reghdfe loenmv treat, abs(alder post time
5 treatment) cl(id pnr)
6
7           if `control_time'==1 matrix results_`y' =
8 J(5,5,.) // matrix to capture results
9           if `control_time'==1 matrix results_p_`y' =
10 J(5,5,.) // matrix to capture results
11
12
13           matrix b = e(b)
14           matrix V = e(V)
15           local n = `control_time'
16
17           matrix results_`y'[`n',1]           = b[1,1] /
18 7466 // capture beta results as 1K Euro           = (V[1,1]^0.5)/
19           matrix results_`y'[`n',2]           =
20 7466 // capture standard error as 1K Euro           b[1,1]/
21           matrix results_`y'[`n',3] =           (`control_post'-(`control'-`treat'))
22           matrix results_`y'[`n',4] =           `n'
23
24
25
26
27     }
28 }
29
30
31
32
33 local t = -.05 //Jitter estimates along x-axis
34 foreach x in 0 1{
35     svmat results_`x'
36     replace results_`x'4= results_`x'4+`t'
37     svmat results_p_`x'
38     replace results_p_`x'4= results_p_`x'4+`t'
39
40     gen upper_`x' = results_`x'1+results_`x'2*1.96
41     gen lower_`x' = results_`x'1-results_`x'2*1.96
42
43     gen upper2_`x' = (results_`x'3+results_`x'2/(results_`x'1/
44 results_`x'3)*1.96)*100
45     gen lower2_`x' = (results_`x'3-results_`x'2/(results_`x'1/
46 results_`x'3)*1.96)*100
47     replace results_`x'3 = results_`x'3*100
48
49     gen upper_p_`x' = results_p_`x'1+results_p_`x'2*1.96
50     gen lower_p_`x' = results_p_`x'1-results_p_`x'2*1.96
51
52     local t = `t'+.1
53
54 }
55
56
57 keep results* upper* lower*
58 keep if _n <=5
59
60

```

```

1
2
3 **generate locals for figure
4
5 foreach x in 0 1{
6     if `x' == 0 local color = "red"
7     if `x' == 1 local color = "green"
8
9         local figure_`x' "scatter results_`x'1 results_`x'4,
10 mcolor(`color') || rspike upper`x' lower`x' results_`x'4,
11 lcolor(`color') vertical"
12         local figure2_`x' "scatter results_`x'3 results_`x'4,
13 mcolor(`color') || rspike upper2`x' lower2`x' results_`x'4,
14 lcolor(`color') vertical "
15         local figure_p_`x' "scatter results_p_`x'1 results_p_`x'4,
16 mcolor(`color') || rspike upper_p_`x' lower_p_`x' results_p_`x'4,
17 lcolor(`color') vertical "
18     }
19 }
20
21
22 `figure_0' || `figure_1' ///
23 legend(label(1 "Men") ///
24 label(3 "Women") ///
25 c(1) order(1 3) pos(3) size(small) ///
26 c(1) symx(4) region(lc(white))) ///
27 ysc(range(-4 2)) ylab(-4(1)2) ///
28 xsc(range(.5 5.5)) xlab(1(1)5) ///
29 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
30 /// title("Parameter estimates across control group, 2003-10") ///
31 xti("Years between exposure and control incident") scale(.95) ///
32 yti("Effect in 1K Euro ({&delta}{subscript: Salary})",
33 height(7)) ///
34 /*note("Parameter estimates for exposure dummy across spacing of
35 control groups." ///
36 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
37 after the exposure group." ///
38 "95% confidence intervals.")*/
39
40
41 graph export $highdef/gender_est2003_2011.png, replace width(3900)
42
43 `figure2_0' || `figure2_1' ///
44 legend( ///
45 label(1 "Men") ///
46 label(3 "Women") ///
47 c(1) order(1 3 5 7) pos(3) size(small) ///
48 c(1) symx(4) region(lc(white))) ///
49 ysc(range(-12 3)) ylab(-12(3)3) ///
50 xsc(range(.5 5.5)) xlab(1(1)5) ///
51 yline(0) ysize(10) xsize(12) graphr(c(white)) ///
52 /// title("Percentage change in salary, 2003-10") ///
53 yti("Salary change (in %)", height(7)) ///
54 xti("Years between exposure and control incident") scale(.95) ///
55 /* note("Parameter estimates for exposure dummy across spacing of
56 control groups." ///
57 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
58 after the exposure group." ///
59

```

```

1
2
3 "95% confidence intervals.)*"/
4
5 graph export $highdef/gender_marginal2003_2011.png, replace
6 width(3900)
7
8 `figure_p_0' || `figure_p_1' ///
9 legend( ///
10 label(1 "Men") ///
11 label(3 "Women") ///
12 c(1) order(1 3 5 7) pos(3) size(small) ///
13 c(1) symx(4) region(lc(white))) ///
14 ysc(range(-.02 .050)) ylab(-0.02(.01)0.05) ///
15 xsc(range(.5 5.5)) xlab(1(1)5) ///
16 yline(0, lcolor(black)) ysize(10) xsize(12) graphr(c(white)) ///
17 /// title("Parameter estimates across control group, 2003-10") ///
18 yti("Effect on Pr(Salary=0)", height(7)) ///
19 xti("Years between exposure and control incident ({&Delta})")
20 scale(.95) ///
21 /*note("Parameter estimates for exposure dummy across spacing of
22 control groups." ///
23 "Control groups suffer concussions 1, 2, 3, 4, and 5 years (&Delta)
24 after the exposure group." ///
25 "95% confidence intervals.)*"/
26
27
28 graph export $highdef/gender_nopay2003_2011.png, replace width(3900)
29
30
31
32 *****
33 *****
34 **
35 ** Generate descriptive figures of wage
36 development for exposure
37 ** and control group
38 **
39 **
40 *****
41 *****
42
43
44 forvalue control_time=1/5{
45     local end = 2012 // last incident year in data
46
47     **build dataset for joint estimate across years
48     forvalue count=2003/`end'{
49         if `control_time' ==1 & `count'==2003{
50             use $data/
51 sample_control_`count'_`control_time'.dta, clear
52             gen time = `count' //incident year
53 indicator
54             }
55             else append using $data/
56 sample_control_`count'_`control_time'.dta
57
58
59     **Drops exposure-group already in the data
60

```

```

1
2
3         if `control_time' > 1 drop if time ==. & treatment
4 ==1
5         replace time = `count' if time ==.
6         replace control1 = `control_time' if
7 control`control_time'==1
8         if `control_time' > 1 drop control`control_time'
9
10        **exclude individuals in years where they do not
11 appear in data,
12        **due to either death or migration, as well as
13 periods from when
14        **the control group suffer their concussion
15
16        }
17
18
19    }
20
21    **Generate mean salary for each period relative to exposure groups
22    concussion
23    **separately for exposure and each control group
24
25    bysort control1 count: egen mean_loen = mean(loenmv)
26
27    **Generate mean Pr(sal= for each period relative to exposure groups
28    concussion
29    **separately for exposure and each control group
30
31    gen no_lon = loenmv < 1
32
33    bysort control1 count: egen mean_no_lon = mean(no_lon)
34
35
36    **generate group size
37    bysort control1 count: gen Ncount=_N if count ==0
38
39    **generate pre-exposure mean levels for normalization
40    gen pre = count <0
41    bysort control1 pre: egen pre_mean_loen = mean(loenmv) if pre==1
42    bysort control1 pre: egen pre_mean_no_lon = mean(no_lon) if pre==1
43
44
45    **reduce data set size
46    bysort control1 count: keep if _n ==1
47
48
49    keep count mean* pre_* control1 Ncount
50
51    **standardize to 1k euro
52    replace mean_loen = mean_loen/7466
53    replace pre_mean_loen = pre_mean_loen/7466
54
55    sort control1 count
56
57
58    **Pre-treatment normalization of salary
59
60

```



```

1
2
3 gen norm_mean_lon =mean_loen
4 gen norm_mean_no_lon =mean_no_lon
5
6 forvalue t = 1/5{
7     qui sum pre_mean_loen if control == 0
8     local treat = r(mean)
9     qui sum pre_mean_loen if control == `t'
10    local control = r(mean)
11
12    **normalize with pre-concussion difference
13    qui replace norm_mean_lon = mean_loen - (`control'-`treat')
14 if control == `t'
15
16    qui sum pre_mean_no_lon if control == 0
17    local treat = r(mean)
18    qui sum pre_mean_no_lon if control == `t'
19    local control = r(mean)
20
21    **normalize with pre-concussion difference
22    qui replace norm_mean_no_lon = mean_no_lon - (`control'-
23 `treat') if control == `t'
24
25 }
26
27
28 **local indicators of group sizes
29 forvalue t=0/5{
30     qui sum Ncount if control == `t'
31     local C`t' = r(mean)
32 }
33
34
35
36 graph twoway ///
37     connect mean_l count if control1== 0, ///
38     lcolor(black) mcolor(black) || ///
39     connect mean_l count if control1== 1, ///
40     lcolor(blue) mcolor(blue) || ///
41     connect mean_l count if control1== 2, ///
42     lcolor(green) mcolor(green) || ///
43     connect mean_l count if control1== 3, ///
44     lcolor(purple) mcolor(purple) || ///
45     connect mean_l count if control1== 4, ///
46     lcolor(red) mcolor(red) || ///
47     connect mean_l count if control1== 5, ///
48     lcolor(orange) mcolor(orange) ///
49     legend( ///
50         label(1 "Exposure" ///
51             "N=`C0'") ///
52         label(2 "Control {&Delta}=1" ///
53             "N=`C1'") ///
54         label(3 "Control {&Delta}=2" ///
55             "N=`C2'") ///
56         label(4 "Control {&Delta}=3" ///
57             "N=`C3'") ///
58         label(5 "Control {&Delta}=4" ///
59

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
        "N=`C4'"') ///
        label(6 "Control {&Delta}=5" ///
            "N=`C5'"') ///
        c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
        c(1) symx(4) region(lc(white))) ///
        ysc(range(26 33)) ylab(26(1)33) ///
        xsc(range(-5 5)) xlab(-5(1)5) ///
        xline(0, lcolor(red)) ysize(10) xsize(12)
graphr(c(white)) ///
/// title("Impact of Concussion on Salary") ///
yti("Deflated Salaried Income in 1K EUR", height(7)) ///
xti("Years since exposure group concussion") scale(.95) ///
/* note("Control groups suffer concussion 1, 2, 3, 4, and 5
years (&Delta) after exposure group." ///
"Both control and exposure group at 30-49 years of
age at beginning of year =0." ///
"Vertical line indicates time of exposure group
concussion.")*/

graph export $highdef/FigureS2.png, replace width(3900)

graph twoway ///
connect norm_mean_l count if control1== 1, ///
lcolor(blue) mcolor(blue) || ///
connect norm_mean_l count if control1== 2, ///
lcolor(green) mcolor(green) || ///
connect norm_mean_l count if control1== 3, ///
lcolor(purple) mcolor(purple) || ///
connect norm_mean_l count if control1== 4, ///
lcolor(red) mcolor(red) || ///
connect norm_mean_l count if control1== 5, ///
lcolor(orange) mcolor(orange) || ///
connect mean_l count if control1== 0, ///
lcolor(black) mcolor(black) ///
legend( ///
label(1 "Control {&Delta}=1" ///
"N=`C1'"') ///
label(2 "Control {&Delta}=2" ///
"N=`C2'"') ///
label(3 "Control {&Delta}=3" ///
"N=`C3'"') ///
label(4 "Control {&Delta}=4" ///
"N=`C4'"') ///
label(5 "Control {&Delta}=5" ///
"N=`C5'"') ///
label(6 "Exposure" ///
"N=`C0'"') ///
c(1) order(6 1 2 3 4 5) pos(3) size(small) ///
c(1) symx(4) region(lc(white))) ///
ysc(range(26 31)) ylab(26(1)31) ///
xsc(range(-5 5)) xlab(-5(1)5) ///
xline(0, lcolor(red)) ysize(10) xsize(12)
graphr(c(white)) ///

```

```

1
2
3    ///    title("Impact of Concussion on Salary") ///
4    yti("Deflated Norm. Salaried Income in 1K EUR",
5    height(7)) ///
6    xti("Years since exposure group's concussion") scale(.95)
7    /*///
8    note("Control groups suffer concussion 1, 2, 3, 4, and 5
9    years (&Delta) after exposure group." ///
10    "Both control and exposure group at 30-49 years of
11    age at beginning of year =0." ///
12    "Vertical line indicates time of exposure group
13    concussion." ///
14    "Salary levels normalized with pre-concussion level
15    difference between" ///
16    "each control group and the exposure group")*/
17
18
19    graph export $highdef/Figure2.png, replace width(3900)
20
21
22    graph twoway ///
23    connect mean_no_l count if control1== 0, ///
24    lcolor(black) mcolor(black) || ///
25    connect mean_no_l count if control1== 1, ///
26    lcolor(blue) mcolor(blue) || ///
27    connect mean_no_l count if control1== 2, ///
28    lcolor(green) mcolor(green) || ///
29    connect mean_no_l count if control1== 3, ///
30    lcolor(purple) mcolor(purple) || ///
31    connect mean_no_l count if control1== 4, ///
32    lcolor(red) mcolor(red) || ///
33    connect mean_no_l count if control1== 5, ///
34    lcolor(orange) mcolor(orange) ///
35    legend( ///
36    label(1 "Exposure" ///
37    "N=`C0`") ///
38    label(2 "Control {&Delta}=1" ///
39    "N=`C1`") ///
40    label(3 "Control {&Delta}=2" ///
41    "N=`C2`") ///
42    label(4 "Control {&Delta}=3" ///
43    "N=`C3`") ///
44    label(5 "Control {&Delta}=4" ///
45    "N=`C4`") ///
46    label(6 "Control {&Delta}=5" ///
47    "N=`C5`") ///
48    c(1) order(1 2 3 4 5 6) pos(3) size(small) ///
49    c(1) symx(4) region(lc(white))) ///
50    ysc(range(.2 .325)) ylab(.2(.025).325) ///
51    xsc(range(-5 5)) xlab(-5(1)5) ///
52    xline(0, lcolor(red)) ysize(10) xsize(12)
53    graphr(c(white)) ///
54    ///    title("Impact of Concussion on Prob(Salary=0)") ///
55    yti("Prob(Salary=0)", height(7)) ///
56    xti("Years since exposure group concussion") scale(.95) ///
57    /*
58    note("Control groups suffer concussion 1, 2, 3, 4, and 5
59
60

```

```

1
2
3 years (&Delta) after exposure group." ///
4         "Both control and exposure group at 30–49 years of
5 age at beginning of year =0." ///
6         "Vertical line indicates time of exposure group
7 concussion.))*/
8
9 graph export $highdef/FigureS3.png, replace width(3900)
10
11 graph twoway ///
12     connect norm_mean_no_l count if control1== 1, ///
13     lcolor(blue) mcolor(blue) || ///
14     connect norm_mean_no_l count if control1== 2, ///
15     lcolor(green) mcolor(green) || ///
16     connect norm_mean_no_l count if control1== 3, ///
17     lcolor(purple) mcolor(purple) || ///
18     connect norm_mean_no_l count if control1== 4, ///
19     lcolor(red) mcolor(red) || ///
20     connect norm_mean_no_l count if control1== 5, ///
21     lcolor(orange) mcolor(orange) || ///
22     connect mean_no_l count if control1== 0, ///
23     lcolor(black) mcolor(black) ///
24     legend( ///
25         label(1 "Control {&Delta}=1" ///
26             "N=`C1'" ) ///
27         label(2 "Control {&Delta}=2" ///
28             "N=`C2'" ) ///
29         label(3 "Control {&Delta}=3" ///
30             "N=`C3'" ) ///
31         label(4 "Control {&Delta}=4" ///
32             "N=`C4'" ) ///
33         label(5 "Control {&Delta}=5" ///
34             "N=`C5'" ) ///
35         label(6 "Exposure" ///
36             "N=`C0'" ) ///
37         c(1) order(6 1 2 3 4 5) pos(3) size(small) ///
38         c(1) symx(4) region(lc(white))) ///
39     ysc(range(.2 .325)) ylab(.2(.025).325) ///
40     xsc(range(-5 5)) xlab(-5(1)5) ///
41     xline(0, lcolor(red)) ysize(10) xsize(12)
42 graphr(c(white)) ///
43 ///     title("Impact of Concussion on Prob(Salary=0)") ///
44     yti("Norm. Prob(Salary=0)", height(7)) ///
45     xti("Years since exposure group concussion") scale(.95) ///
46 /*     note("Control groups suffer concussion 1, 2, 3, 4, and 5
47 years (&Delta) after exposure group." ///
48         "Both control and exposure group at 30–49 years of
49 age at beginning of year =0." ///
50         "Vertical line indicates time of exposure group
51 concussion." ///
52         "Probability levels normalized with pre-concussion
53 level difference between" ///
54         "each control group and the exposure group"))*/
55
56 graph export $highdef/Figure2A.png, replace width(3900)
57
58
59
60

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

```

*****
*****
**
**       Effects across the salary distribution
**
**
**
*****
*****

forvalue control_time=5/5{
    local end = 2012 // last incident year in data

    **build dataset for joint estimate across years
    forvalue count=2003/`end'{
        if `count'==2003{
            use $data/
sample_control_`count'_`control_time'.dta, clear
            gen time = `count' //incident year
indicator
            }
            else append using $data/
sample_control_`count'_`control_time'.dta
            replace time = `count' if time ==.

            **exclude individuals in years where they do not
appear in data,
            **due to either death or migration, as well as
periods from when
            **the control group sufer their concussion
drop if merge ==1 | count > `control_time'-1
        }

        gen female = koen==2
        qui{
            gen edu = 0
            replace edu = 1 if inrange(hffsp,20000000,39000000)
| ///
            (hffsp
>40000000 & hffsp!=.)
        }

        //build ident, so we can multivariate cluster for
individuals
        //who occur both as control and exposure during the period
(id)

```

```

1
2
3     bysort pnr time: gen helpx = _n ==1
4     gen id= sum(helpx)
5     drop helpx
6
7
8     **generate concussion variable
9     gen treat = inrange(count,0,`control_time'-1) & treatment
10
11 ==1
12     replace treat = time_from_incident if count ==0 & treatment
13 ==1
14
15     **effect across salary distribution
16     if `control_time'==5 matrix results = J(81,5,.) // matrix
17 to capture results
18
19     local n = `control_time'
20     gen post = count >=0
21     replace count = count+6
22     local v =0
23
24     qui{
25         forvalue t= 0(14932)1045240{
26             local top = 42
27             local v =1+`v'
28             if `n' ==5 matrix results[`v',1]= `t'
29             gen D = personindk <=`t'
30             reg D treat i.treatment i.post i.edu
31 i.alder i.female i.time, cl(pnr)
32             matrix V = e(V)
33             matrix b= e(b)
34             matrix results[`v',2] = b[1,1]
35             margins, at(treat= (0 1) treatment=1
36 post=1)
37             matrix results[`v',5] = V[1,1]^5
38             matrix M = r(b)
39             matrix results[`v',4] = M[1,2] //
40             matrix results[`v',3] = M[1,1] //
41             drop D
42         }
43     }
44 }
45
46
47 svmat results
48
49 gen l5 = results2-1.96*results5
50 gen u5 = results2+1.96*results5
51
52
53
54 replace results1 = results1/7466
55
56 gen dif6 = results4-results3
57 gen udif6 = u5
58 gen ldif6 = l5
59
60

```

```

1
2
3 gr two rline udif6 ldif6 results1, ///
4     yaxis(2) color(gray) lp(dash) ylab(0(.01).04, axis(2))
5 ysc(range(-.006 .04) axis(2)) || ///
6     line dif6 results1, yaxis(2) yline(0, axis(2)) ///
7     lcolor(black) ylab(0(.01).04) ysc(range(-.006 .04)
8 axis(2)) || ///
9     line results4 results1, yaxis(1) lcolor(red) || ///
10    line results3 results1, yaxis(1) lcolor(blue) ///
11    ysc(range(0 1)) ylab(0(.1)1, nogrid) xlab(0(10)140)
12 xsc(range(0 140)) ///
13    ysize(10) xsize(12) graphr(c(white)) ///
14    xti("Total Income in 1K Euro") scale(1) ///
15    yti("Effect of Concussion on Pr(Total Income < X)",
16 axis(2)) ///
17    yti("{&Phi}(Total Income)", axis(1)) ///
18    legend( ///
19    label(2 "Effect of concussion (left axis)") ///
20    label(3 "Concussion income distribution (right axis)") ///
21    label(4 "Counterfactual income distribution (right
22 axis)") ///
23    c(1) order(2 3 4) pos(6) size(small) ///
24    symx(4) region(lc(white)))
25
26
27
28
29
30 graph export [home]\highdef\l5_income.png, replace width(3900)
31
32 cap graph drop g1 g2
33
34 gr two ///
35     line results4 results1, yaxis(1) lcolor(red) || ///
36     line results3 results1, yaxis(1) lcolor(blue) ///
37     ysc(range(.2 1)) ylab(0(.1)1) xlab(0(20)140,
38 labs(small)) ///
39     xsc(range(0 140)) ///
40     ysize(10) xsize(10) graphr(c(white)) ///
41     xti("Total Income in 1K Euro") scale(1) ///
42     yti("Cumulative Distribution of Total Income",
43 axis(1)) ///
44     legend( ///
45     label(1 "Observed Post-Concussion Total Income
46 Distribution") ///
47     label(2 "Counterfactual No-Concussion Total Income
48 Distribution") ///
49     c(1) order(1 2 4) pos(6) size(small) ///
50     symx(4) region(lc(white))) , name(g1)
51
52
53 gr two rline udif6 ldif6 results1, ///
54     color(gray) lp(dash) ysc(range(-.002 .035)) || ///
55     line dif6 results1, yline(0) ///
56     lcolor(black) ylab(0(.01).0325) ysc(range(-.002 .
57 035)) ///
58     xlab(0(20)140, labs(small)) xsc(range(0 140)) ///
59     ysize(10) xsize(10) graphr(c(white)) ///
60

```

```

1
2
3     xti("Total Income in 1K Euro (X)") scale(1) ///
4     yti("Effect of Concussion on Pr(Total Income < X)") ///
5         legend( label(1 "95% CI") ///
6             label(2 "Effect of concussion") ///
7             c(1) order(2 1) pos(6) size(small) ///
8                 symx(4) region(lc(white))) , name(g2)
9
10    graph combine g1 g2          ,graphr(c(white))
11
12    graph export "[home]\highdef\figure s2.tif", replace width(1000)
13
14
15
16
17    *****
18    *****
19    **
20    **     Effects across the salary distribution
21    **
22    **
23    **
24    *****
25    *****
26
27
28    forvalue control_time=5/5{
29        local end = 2012 // last incident year in data
30
31        **build dataset for joint estimate across years
32        forvalue count=2003/`end'{
33            if `count'==2003{
34                use $data/
35            sample_control_`count'_`control_time'.dta, clear
36                gen time = `count' //incident year
37            indicator
38                }
39                else append using $data/
40            sample_control_`count'_`control_time'.dta
41                replace time = `count' if time ==.
42
43                **exclude individuals in years where they do not
44            appear in data,
45                **due to either death or migration, as well as
46            periods from when
47                **the control group suffer their concussion
48            drop if merge ==1 | count > `control_time'-1
49        }
50
51        gen female = koen==2
52        qui{
53            gen edu = 0
54            replace edu = 1 if inrange(hffsp,20000000,39000000)
55
56        | ///
57
58
59        >40000000 & hffsp!=.)
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```



```

1
2
3     }
4
5     //build ident, so we can multivariate cluster for
6 individuals
7     //who occur both as control and exposure during the period
8 (id)
9
10    bysort pnr time: gen helpx = _n ==1
11    gen id= sum(helpx)
12    drop helpx
13
14
15
16    **generate concussion variable
17    gen treat = inrange(count,0,`control_time'-1) & treatment
18 ==1
19    replace treat = time_from_incident if count ==0 & treatment
20 ==1
21
22    **effect across salary distribution
23    if `control_time'==5 matrix results = J(81,5,.) // matrix
24 to capture results
25
26    local n = `control_time'
27    gen post = count >=0
28    replace count = count+6
29    local v =0
30
31
32    **Estimate Pr(salary < X) across income distribution
33    qui{
34        forvalue t= 0(14932)895950{
35            local top = 42
36            local v =1+`v'
37            if `n' ==5 matrix results[`v',1]= `t'
38            gen D = loenmv <=`t'
39            reg D treat i.treatment i.post i.edu
40 i.alder i.female i.time, cl(pnr)
41            matrix V = e(V)
42            matrix b= e(b)
43            matrix results[`v',2] = b[1,1]
44            margins, at(treat= (0 1) treatment=1
45 post=1)
46            matrix results[`v',5] = V[1,1]^5
47            matrix M = r(b)
48            matrix results[`v',4] = M[1,2] //
49            matrix results[`v',3] = M[1,1] //
50            drop D
51        }
52    }
53 }
54
55
56 svmat results
57
58 *Generate 95% confidence intervals
59 gen l5 = results2-1.96*results5
60

```

```

1
2
3 gen u5 = results2+1.96*results5
4
5 *Correct to €
6 replace results1 = results1/7466
7
8 *Obtain difference between observed and counterfactual wage
9 distribution
10 gen dif6 = results4-results3
11 gen udif6 = u5
12 gen ldif6 = l5
13
14
15 gr two rline udif6 ldif6 results1, ///
16     yaxis(2) color(gray) lp(dash) ysc(range(-.006 .04) axis(2))
17 || ///
18     line dif6 results1, yaxis(2) yline(0, axis(2)) ///
19     lcolor(black) ylab(0(.01).0325) ysc(range(-.006 .04)
20 axis(2)) || ///
21     line results4 results1, yaxis(1) lcolor(red) || ///
22     line results3 results1, yaxis(1) lcolor(blue) ///
23     ysc(range(0 1)) ylab(0(.1)1, nogrid) xlab(0(10)120)
24 xsc(range(0 120)) ///
25     ysize(10) xsize(12) graphr(c(white)) ///
26     xti("Salary in 1K Euro") scale(1) ///
27     yti("Effect of Concussion on Pr(Salary < X)", axis(2)) ///
28     yti("{&Phi}(Salary)", axis(1)) ///
29     legend( ///
30     label(2 "Effect of concussion (left axis)") ///
31     label(3 "Concussion salary distribution (right axis)") ///
32     label(4 "Counterfactual salary distribution (right
33 axis)") ///
34     c(1) order(2 3 4) pos(6) size(small) ///
35     symx(4) region(lc(white)))
36
37
38 cap graph drop g1 g2
39
40 gr two ///
41     line results4 results1, yaxis(1) lcolor(red) || ///
42     line results3 results1, yaxis(1) lcolor(blue) ///
43     ysc(range(.2 1)) ylab(0.2(.1)1) xlab(0(10)120,
44 labs(small)) ///
45     xsc(range(0 120)) ///
46     ysize(10) xsize(10) graphr(c(white)) ///
47     xti("Salary in 1K Euro") scale(1) ///
48     yti("Cumulative Distribution of Salary", axis(1)) ///
49     legend( ///
50     label(1 "Observed Post-Concussion Salary Distribution") ///
51     label(2 "Counterfactual No-Concussion Salary
52 Distribution") ///
53     c(1) order(1 2 4) pos(6) size(small) ///
54     symx(4) region(lc(white))) , name(g1)
55
56
57 gr two rline udif6 ldif6 results1, ///
58     color(gray) lp(dash) ysc(range(-.002 .035)) || ///
59     line dif6 results1, yline(0) ///
60

```

```

1
2
3         lcolor(black) ylab(0(.01).0325) ysc(range(-.002 .
4 035)) ///
5         xlabel(0(10)120, labs(small)) xsc(range(0 120))      ///
6         ysize(10) xsize(10) graphr(c(white)) ///
7         xti("Salary in 1K Euro (X)") scale(1) ///
8         yti("Effect of Concussion on Pr(Salary < X)") ///
9             legend( label(1 "95% CI") ///
10                label(2 "Effect of concussion") ///
11                c(1) order(2 1) pos(6) size(small) ///
12                    symx(4) region(lc(white))) , name(g2)
13
14 graph combine g1 g2          ,graphr(c(white))
15
16 graph export "[home]\highdef\figure 3.tif",  replace width(1000)
17
18
19
20
21
22
23 **ESTIMATE SICK LEAVE
24
25 forvalue control_time=1/5{
26     local end = 2012 // last incident year in data
27     if `control_time' ==1     eststo clear
28
29     **build dataset for joint estimate across years
30     forvalue count=2003/`end'{
31         if `count'==2003{
32             use $data/
33             sample_control_`count'_`control_time'.dta, clear
34             gen time = `count' //incident year
35             indicator
36                 }
37             else append using $data/
38             sample_control_`count'_`control_time'.dta
39             replace time = `count' if time ==.
40
41
42             **exclude individuals in years where they do not
43 appear in data,
44             **due to either death or migration, as well as
45 periods from when
46             **the control group sufer their concussion
47             drop if merge ==1 | count > `control_time'-1
48         }
49
50
51     merge m:1 pnr t using $data/temp.dta, keep(1 3) nogen
52     replace share = 0 if share==.
53
54     gen female = koen==2
55
56     //build ident, so we can multivariate cluster for
57 individuals
58     //who occur both as control and exposure during the period
59 (id)
60

```

```

1
2
3
4     bysort pnr time: gen helpx = _n ==1
5     gen id= sum(helpx)
6     drop helpx
7
8
9     **Generate educational groups
10    qui{
11        gen edu =0
12        replace edu = 1 if inrange(hffsp,20000000,39000000)
13    | ///
14
15
16    >40000000 & hffsp!=.)
17    }
18
19
20    **Calculate number of observations for exposure and control
21    count if count==0 & treatment ==1
22    local Ntreated = r(N)
23    count if count==0 & treatment ==0
24    local Ncontrol = r(N)
25
26    **generate concussion variable
27    gen treat = inrange(count,0,`control_time'-1) & treatment
28
29    ==1
30    replace treat = time_from_incident if count ==0 & treatment
31    ==1
32
33    **Generate pre-concussion income difference          for
34    **use in calculating marginal effects
35    sum loenmv if count <0 & treatment ==0
36    local control =r(mean)
37    sum loenmv if count <0 & treatment ==1
38    local treat =r(mean)
39    sum loenmv if count>=0 & treatment ==0
40    local control_post =r(mean)
41
42    forvalue t=-4/4{
43        local n = `t'*-1
44        if `t' <-1 gen T_`n' = treatment ==1 & count ==`t'
45        if `t' > -1 gen T`t' = treatment ==1 & count ==`t'
46
47    }
48
49
50    **estimate DiD model on salary
51    reghdfe share T*, abs(alder female count time treatment
52    edu) cl(pnr id)
53    eststo est1_`control_time'
54    if `control_time'==1 matrix results = J(5,5,.) // matrix to
55    capture results
56    if `control_time'==1 matrix results_p = J(5,5,.) // matrix
57    to capture results
58
59
60

```

```

1
2
3     matrix b = e(b)
4     matrix V = e(V)
5     local n = `control_time'
6
7     matrix results[`n',1] = b[1,1] / 7466 // capture beta
8 results as 1K Euro
9     matrix results[`n',2] = (V[1,1]^5)/7466 //
10 capture standard error as 1K Euro
11     matrix results[`n',3] = b[1,1]/(`control_post'-
12 (`control'-`treat'))
13     matrix results[`n',4] = `n'
14
15
16     gen no_share = share >0 //dummy for no salary
17
18
19     **Generate pre-concussion probability difference for
20     **use in calculating marginal effects
21     sum no_share if count <0 & treatment ==0
22     local control =r(mean)
23     sum no_share if count <0 & treatment ==1
24     local treat =r(mean)
25     sum no_share if count>=0 & treatment ==0
26     local control_post =r(mean)
27
28
29     **estimate DiD LP-model on P(salary=0)
30     xi: reghdfe no_share T*, abs(alder female count time
31 treatment edu) cl(pnr id)
32     eststo est2_`control_time'
33     matrix b =e(b) //regression coefficient
34     matrix V = e(V) // standard error^2
35
36     matrix results_p[`n',1] = b[1,1]
37     matrix results_p[`n',2] = V[1,1]^5
38     matrix results_p[`n',3] = b[1,1]/(`control_post'-
39 (`control'-`treat'))
40     matrix results_p[`n',4] = `n'
41
42 }
43
44     esttab est1_* using [home]/tables/dynamic_share1.rtf, ///
45     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
46 *** .001) ///
47     keep(T*)
48
49     esttab est2_* using [home]/tables/dynamic_share2.rtf, ///
50     replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
51 *** .001) ///
52     keep(T*)
53
54
55 **ESTIMATE SICK LEAVE
56
57 forvalue control_time=1/5{
58     local end = 2012 // last incident year in data
59     if `control_time' ==1 eststo clear
60

```

```

1
2
3
4     **build dataset for joint estimate across years
5     forvalue count=2003/`end'{
6         if `count'==2003{
7             use $data/
8 sample_control_`count'_`control_time'.dta, clear
9             gen time = `count' //incident year
10 indicator
11         }
12         else append using $data/
13 sample_control_`count'_`control_time'.dta
14         replace time = `count' if time ==.
15
16         **exclude individuals in years where they do not
17 appear in data,
18         **due to either death or migration, as well as
19 periods from when
20         **the control group suffer their concussion
21         drop if merge ==1 | count > `control_time'-1
22     }
23
24
25     merge m:1 pnr t using $data/temp2.dta, keep(1 3) nogen
26     replace syg_barsel_13= 0 if syg_barsel_13==.
27
28     gen female = koen==2
29
30     //build ident, so we can multivariate cluster for
31 individuals
32     //who occur both as control and exposure during the period
33 (id)
34
35
36     bysort pnr time: gen helpx = _n ==1
37     gen id= sum(helpx)
38     drop helpx
39
40
41     **Generate educational groups
42     qui{
43         gen edu =0
44         replace edu = 1 if inrange(hffsp,20000000,39000000)
45 | ///
46                                     (hffsp
47 >40000000 & hffsp!=.)
48     }
49
50
51
52     **Calculate number of observations for exposure and control
53     count if count==0 & treatment ==1
54     local Ntreated = r(N)
55     count if count==0 & treatment ==0
56     local Ncontrol = r(N)
57
58     **generate concussion variable
59     gen treat = inrange(count,0,`control_time'-1) & treatment
60

```

```

1
2
3 ==1
4 replace treat = time_from_incident if count ==0 & treatment
5 ==1
6
7 **Generate pre-concussion income difference for
8 **use in calculating marginal effects
9 sum loenmv if count <0 & treatment ==0
10 local control =r(mean)
11 sum loenmv if count <0 & treatment ==1
12 local treat =r(mean)
13 sum loenmv if count >=0 & treatment ==0
14 local control_post =r(mean)
15
16
17 forvalue t=-4/4{
18     local n = `t'*-1
19     if `t' < -1 gen T_`n' = treatment ==1 & count ==`t'
20     if `t' > -1 gen T`t' = treatment ==1 & count ==`t'
21
22 }
23
24 **estimate DiD model on salary
25 reghdfe syg_barsel_13 T*, abs(alder female count time
26 treatment edu) cl(pnr id)
27 eststo est1_`control_time'
28 if `control_time'==1 matrix results = J(5,5,.) // matrix to
29 capture results
30 if `control_time'==1 matrix results_p = J(5,5,.) // matrix
31 to capture results
32
33
34
35 matrix b = e(b)
36 matrix V = e(V)
37 local n = `control_time'
38
39 matrix results[`n',1] = b[1,1] / 7466 // capture beta
40 results as 1K Euro
41 matrix results[`n',2] = (V[1,1]^5)/7466 //
42 capture standard error as 1K Euro
43 matrix results[`n',3] = b[1,1]/(`control_post'-
44 (`control'-`treat'))
45 matrix results[`n',4] = `n'
46
47
48
49
50
51 **Generate pre-concussion probability difference for
52 **use in calculating marginal effects
53
54 **estimate DiD LP-model on P(salary=0)
55 xi: reghdfe kont_dag T*, abs(alder female count time
56 treatment edu) cl(pnr id)
57 eststo est2_`control_time'
58 matrix b =e(b) //regression coefficient
59 matrix V = e(V) // standard error^2
60

```

```
1
2
3
4         matrix results_p[`n',1] =      b[1,1]
5         matrix results_p[`n',2] =      V[1,1]^5
6         *matrix results_p[`n',3] =      b[1,1]/(`control_post'-
7 (`control'-`treat'))
8         matrix results_p[`n',4] =      `n'
9
10      }
11
12      esttab est1_* using [home]/tables/dynamic_sickpay1.rtf, ///
13          replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
14 *** .001) ///
15          keep(T*)
16
17      esttab est2_* using [home]/tables/dynamic_welfare2.rtf, ///
18          replace se(3) b(3) compress nogap star(+ .1 * .05 ** .01
19 *** .001) ///
20          keep(T*)
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Supplemental results**

For peer review only

**Table S1. Test of parallel trends assumption pre-exposure incident against each control group separately using eq. S3 in supplementary methods. Separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period).**

Time to exposure group's concussion (exposure)	$\Delta=1$ Est (S.E.) p-value	$\Delta=2$ Est (S.E.) p-value	$\Delta=3$ Est (S.E.) p-value	$\Delta=4$ Est (S.E.) p-value	$\Delta=5$ Est (S.E.) p-value
Exposure-4y	-0.368 (0.226) p=.104	0.046 (0.363) p=.900	0.120 (0.362) p=.741	0.159 (0.313) p=.612	0.042 (0.329) p=.899
Exposure-3y	-0.094 (0.317) p=.768	0.227 (0.510) p=.656	0.167 (0.354) p=.637	0.537 (0.372) p=.148	0.113 (0.393) p=.774
Exposure-2y	-0.548 (0.312) p=.079	-0.082 (0.347) p=.812	-0.163 (0.236) p=.491	-0.124 (0.247) p=.617	0.082 (0.250) p=.744
Exposure-1y	Ref.	Ref.	Ref.	Ref.	Ref.
<b>N*T</b>	284115	273725	266120	260647	256337

Note: The table shows test for differences in pre-exposure trends between exposure and control group model using interactions between pre-exposure time dummies and the exposure indicator. There is no indication of substantial or significant pre-exposure differences in salary trajectories between exposure group and any of the control groups.

**Table S2.** Effect of concussion on different labor market outcome parameters using separate exposure dummies for all time periods (except the year prior to exposure, which serves as reference period): In this exploratory analysis, the exposure group is compared to the control group  $\Delta=5$ , which suffers a concussion five years after the exposure group. Outcomes include annual salaried income (annual salary), total annual income (total income), annual sick leave benefits received (sick leave benefits) as well as a binary indicator of employment (probability of employment). Monetary outcomes were measured at 2015-level in € 1,000.

Time to exposure group's concussion (exposure)	Annual Salary Est. (S.E.) p-value	Total income Est. (S.E.) p-value	Sick leave benefits Est. (S.E.) p-value	Probability of employment Est. (S.E.) p-value
Exposure-4y	0.012 (0.212) p=.954	0.164 (0.173) p=.343	0.035 (0.036) p=.320	0.001 (0.004) p=.803
Exposure-3y	0.059 (0.252) p=.814	0.305 (0.233) p=.190	0.022 (0.034) p=.529	-0.001 (0.003) p=.739
Exposure-2y	0.043 (0.160) p=.788	0.122 (0.147) p=.405	0.002 (0.029) p=.946	0.001 (0.003) p=.739
Exposure-1y				
Exposure	-0.611 (0.168) p<.001	-0.338 (0.140) 0.016	0.166 (0.030) p<.001	-0.003 (0.003) p=.317
Exposure+1y	-1.389 (0.209) p<.001	-0.608 (0.162) p<.001	0.288 (0.039) p<.001	-0.020 (0.003) p<.001
Exposure+2y	-1.568 (0.261) p<.001	-0.847 (0.231) p<.001	0.132 (0.039) p=.001	-0.023 (0.004) p<.001
Exposure+3y	-1.393 (0.246) p<.001	-0.497 (0.219) p=.023	0.031 (0.040) p=.432	-0.022 (0.004) p<.001
Exposure+4y	-1.319 (0.253) p<.001	-0.499 (0.218) p=.022	-0.076 (0.042) p=.075	-0.018 (0.004) p<.001
N*T	577762	577758	577872	577872

**Note:** Annual salary include all income from salary and employee fringe benefits, employee stock options, employer paid sick leave, net gains (including interests and capital gains) from own companies. Total income includes all income absent wealth. Sick leave includes only public health benefits (sick leave and paternity leave). Employment is a binary indicator measured last week of November for each year. Results obtained from estimations following Eq. (1). Models include controls for high school diploma, gender, age, and observation year. Results obtained using reghdfe in Stata. Total number of observations (N\*T) differ slightly between outcomes because all income information is not available for all observation all years.

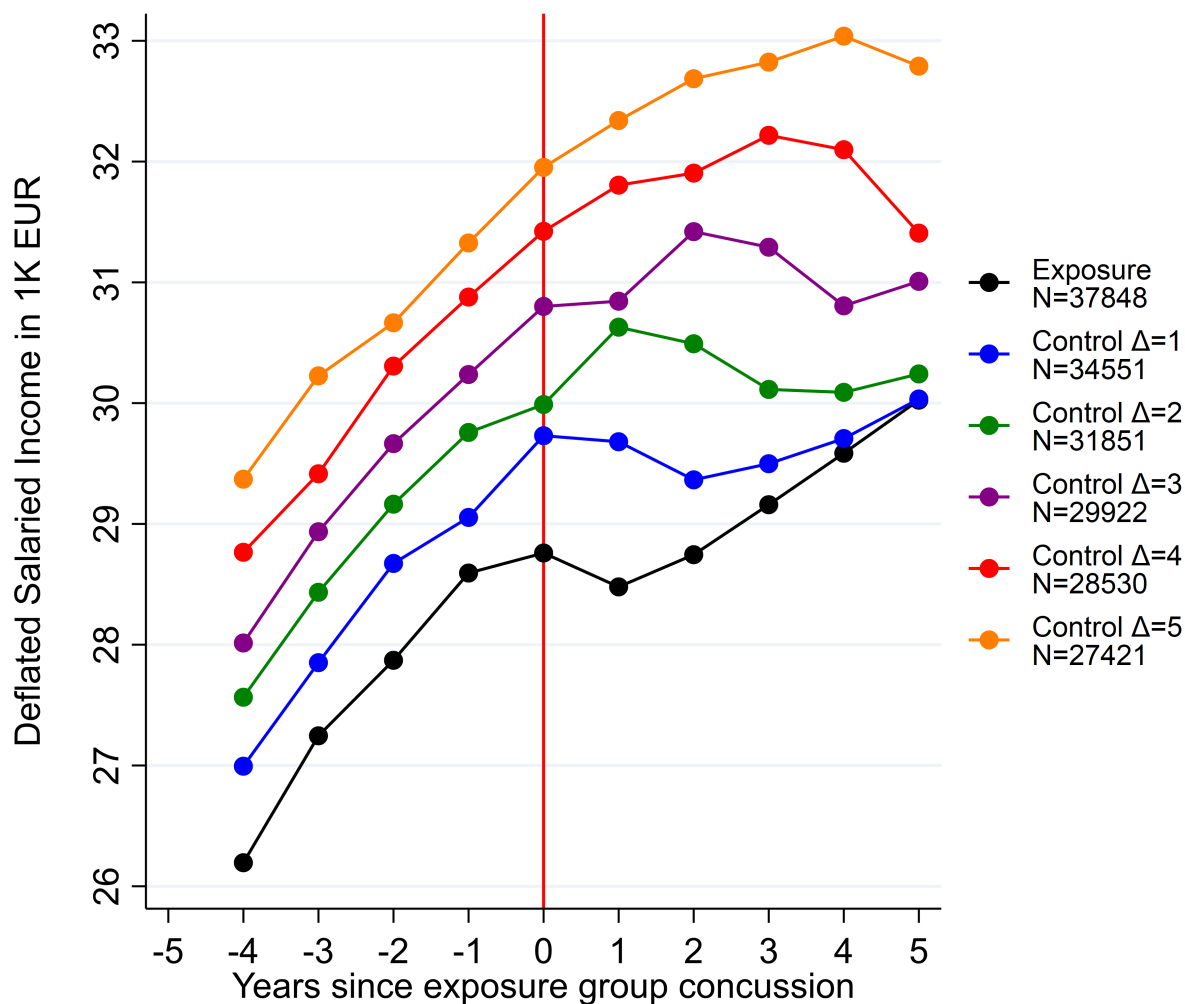
Source: Own calculations on data from Statistics Denmark.

**Table S3. Demographic factors for exposure group and control groups ( $\Delta=1, 2, 3, 4, 5$ ) averaged over the 5 years leading up to the concussion event in each of the groups. Factors include patient age (in years), share of sample female (1=100% female), and share of individuals with at least a high school degree (1=100%).**

		Exposure	Control, $\Delta=1$	Control, $\Delta=2$	Control, $\Delta=3$	Control, $\Delta=4$	Control, $\Delta=5$
<b>Pr(Female=1)</b>	Mean	.430	.438	.447	.458	.464	.473
	S.D.	(.495)	(.496)	(.497)	(.498)	(.499)	(.499)
	p-value		.030	<.001	<.001	<.001	<.001
<b>Age</b>	Mean	36.899	37.354	37.754	38.065	38.343	38.592
	S.D.	(11.856)	(11.857)	(11.718)	(11.630)	(11.584)	(11.491)
	p-value		<.001	<.001	<.001	<.001	<.001
<b>Pr(High school=1)</b>	Mean	.624	.632	.640	.646	.653	.660
	S.D.	(.484)	(.482)	(.480)	(.478)	(.476)	(.474)
	p-value		.026	<.001	<.001	<.001	<.001
<b>Total individuals</b>		37848	34551	31851	29922	28580	27484

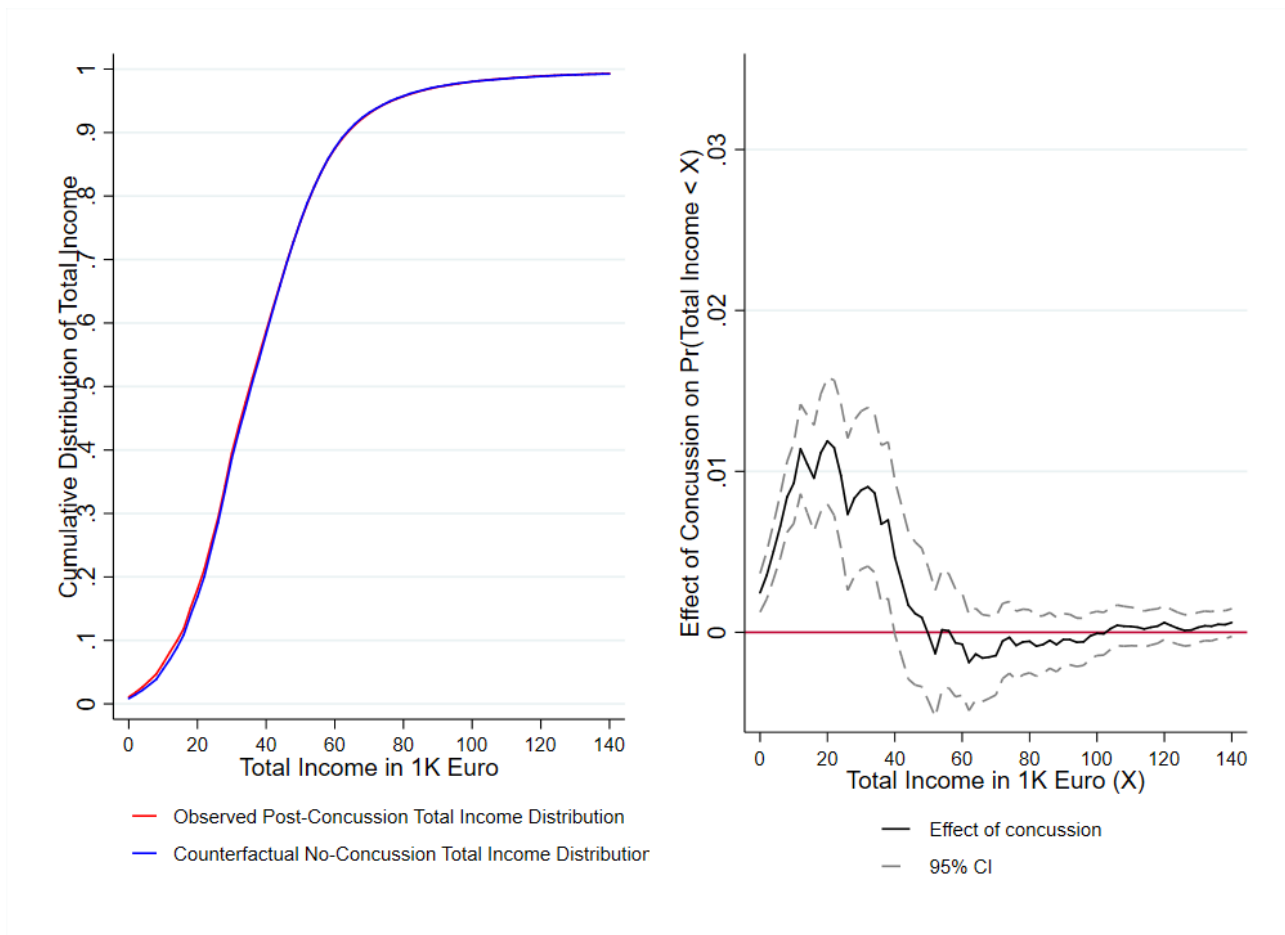
**Note:** S.D.: Standard deviation. P-values calculated using two-sided t-tests. All test performed between exposure group and each control group separately.

Figure S1. Unnormalized Average Salary for Treatment and Control Groups Measured in 1K € at 2015-levels



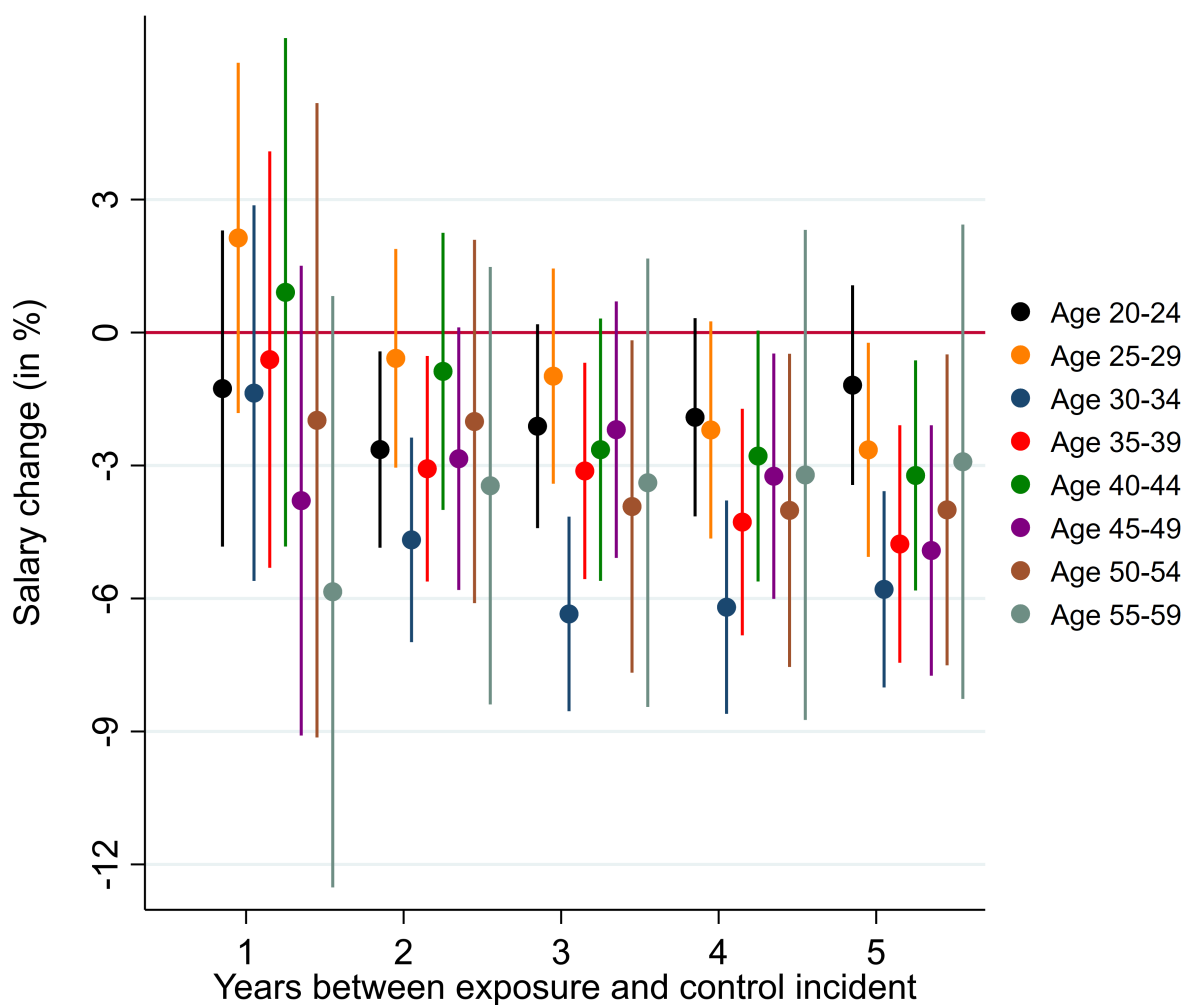
**Note:** Salary of the exposure group compared to salary of the 5 control groups, who experienced their concussions  $\Delta=\{1, 2, 3, 4, \text{ and } 5\}$  years later than the exposure group. Salary progression is shown for the 5 years before and the 5 years after the exposure group suffered a concussion event. Table S1 demonstrates that the trends for salary progression pre-exposure incident are parallel between exposure group and each control group.

**Figure S2. (Left Panel) The Cumulative Distribution for Total Income Post-Treatment among the Treatment Group and Their Counterfactual, and (Right Panel) the Difference between the Two CDFs Expressed as the Effect of Concussion on the Probability of Total Income Below that Income-Level Expressed on the X-Axis following Exposure Event.**



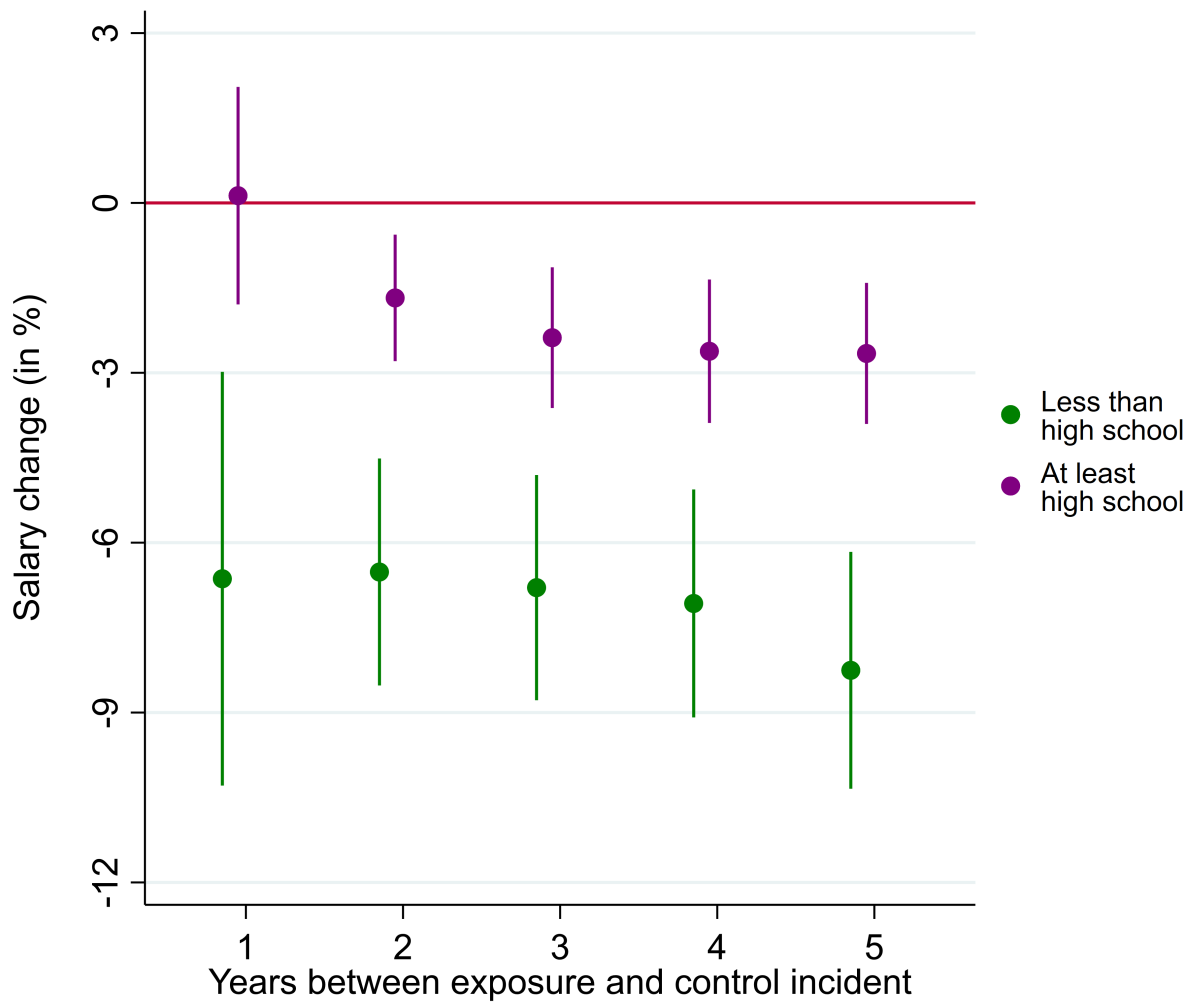
**Note:** The figure shows the observed cumulative salary distribution following a concussion (red) and the expected counterfactual salary distribution absent the concussion (blue). The black line shows the difference between the observed and the counterfactual distribution, and the grey dash lines show the 95 % confidence interval. The bell-shape of the difference between the two distributions as the total income increase from 0 to 40,000 € indicates that the main part of the effect of concussions on total incomes is driven by low-income people shifting total income downwards following concussion, but not going to total income equal to zero.

Figure S3. Percentage Effect of Concussion on Relative Salary Across Age Groups.



**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

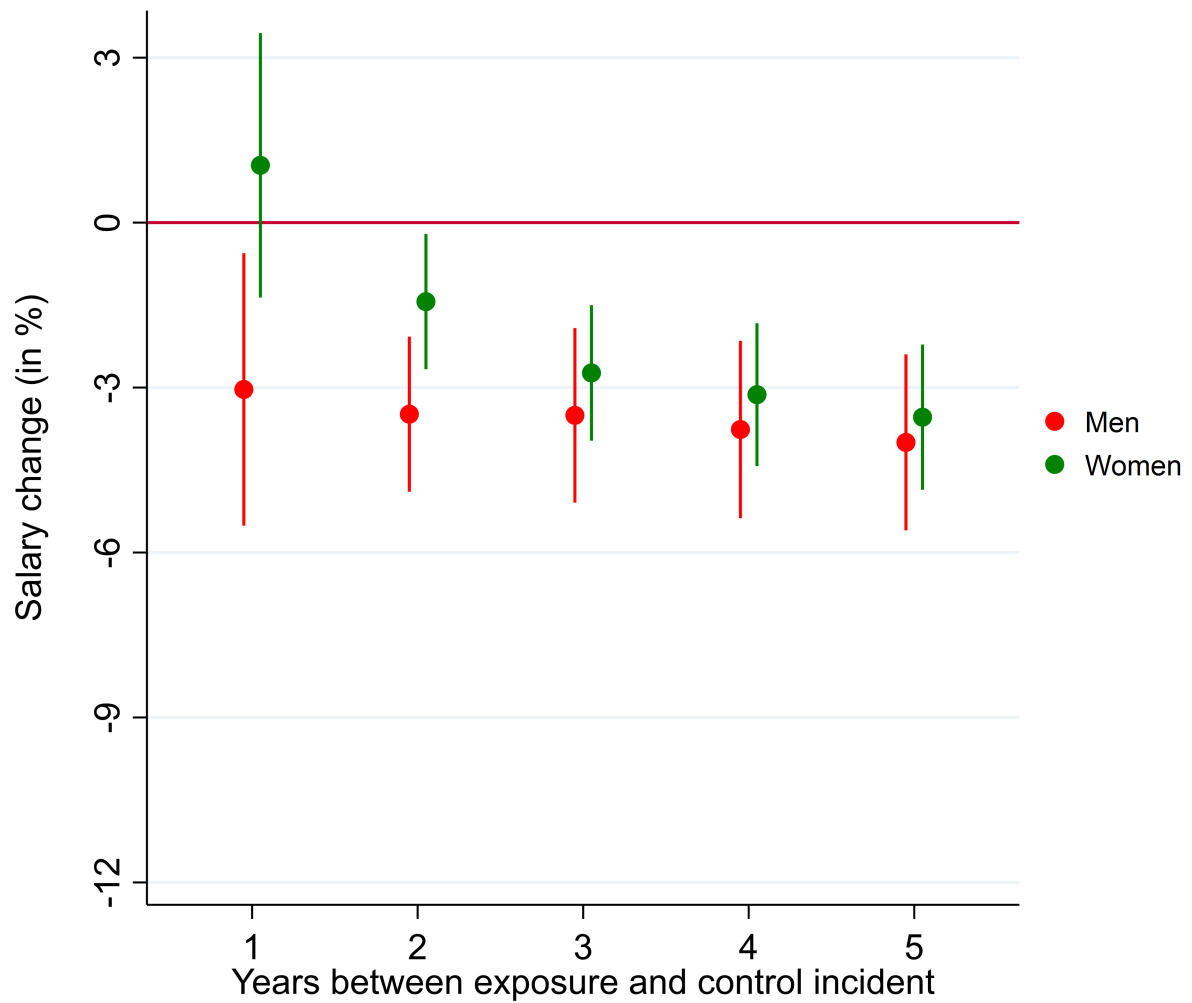
**Figure S4. Percentage Effect of Concussion on Relative Salary Across High School Completion.**



**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

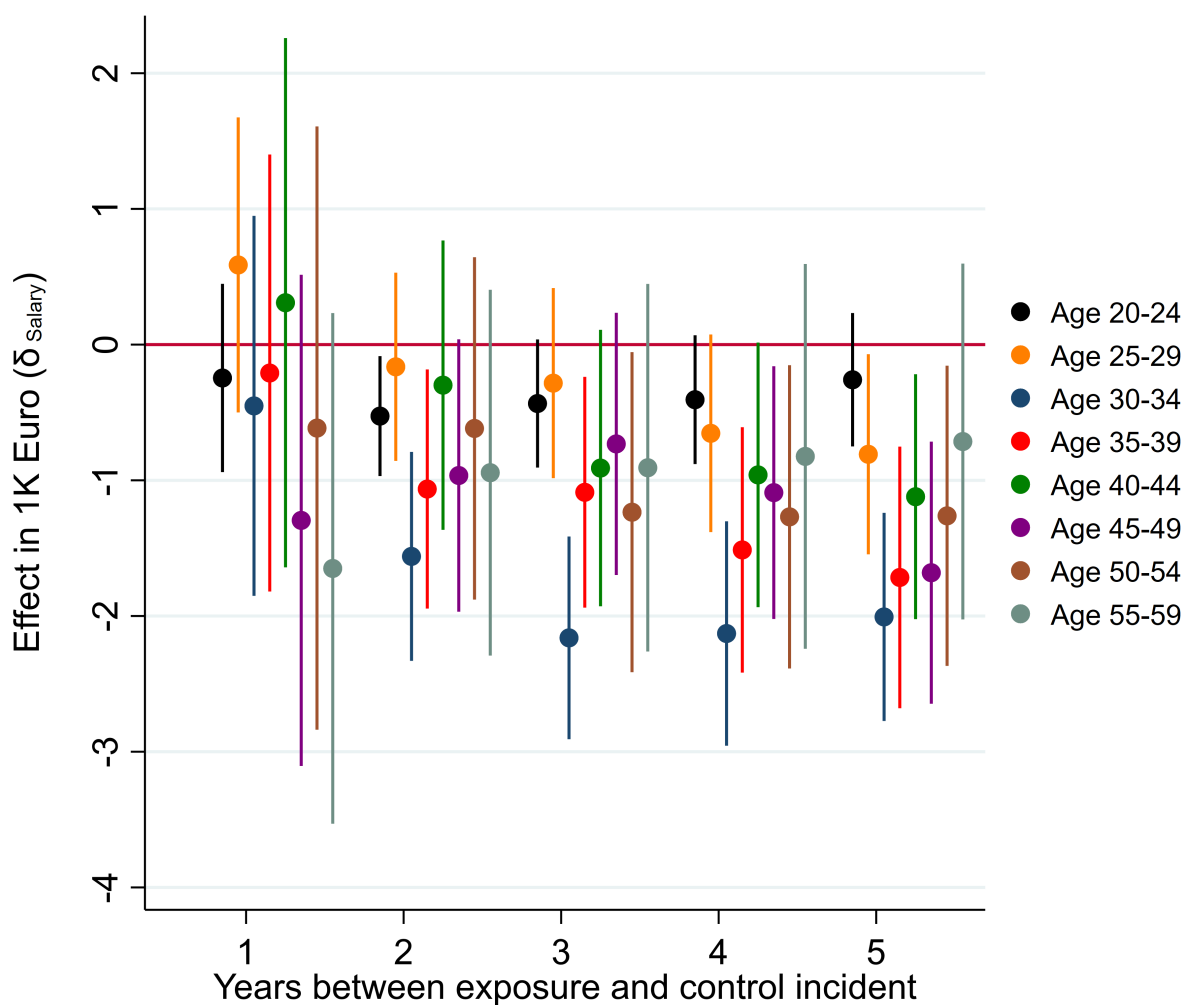


Figure S5. Percentage Effect of Concussion on Relative Salary Across Gender.

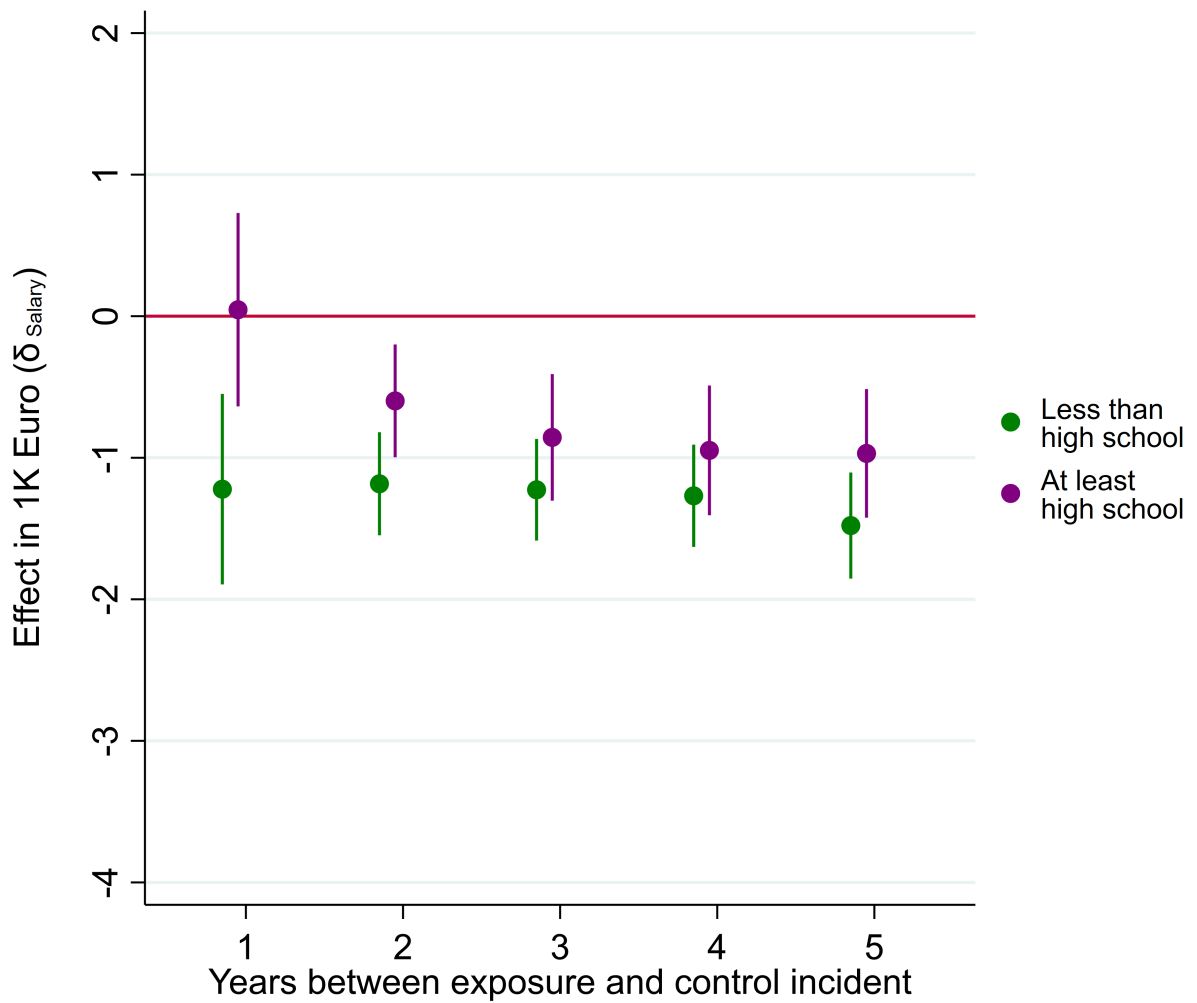


**Note:** Relative salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Relative salary changes are shown across gender. Graph shows parameter estimates and 95% CI.

Figure S6. Effect of Concussion on Absolute Salary in 1K Euro Across Age groups.

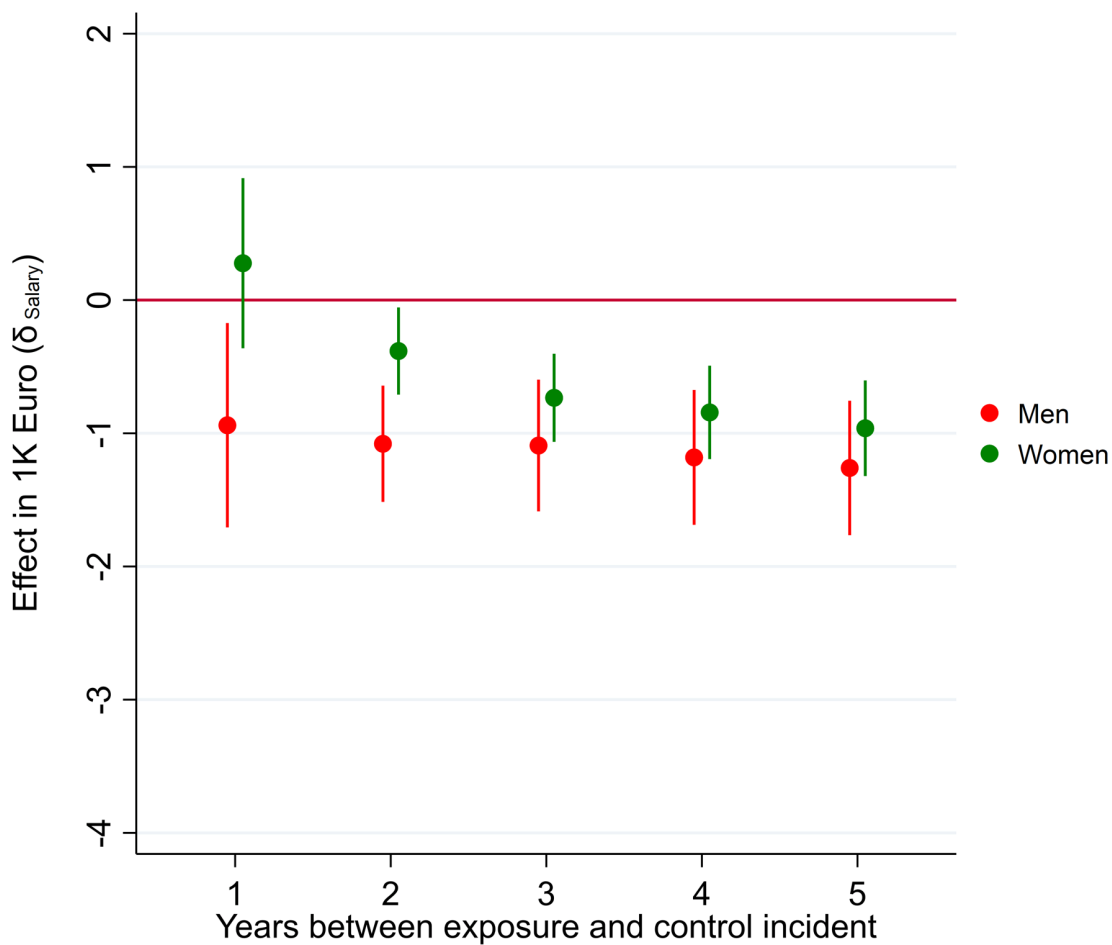


**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute salary changes are shown across different age groups. Graph shows parameter estimates and 95% CI.

**Figure S7. Effect of Concussion on Absolute Salary in 1K Euro Across Education.**

**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across whether individuals had obtained at least a high school diploma (ISCED > 2). Graph shows parameter estimates and 95% CI.

**Figure S8. Effect of Concussion on Absolute Salary in 1K Euro Across Gender.**



**Note:** Absolute salary drop of the exposure group compared to the control groups, who suffer a concussion 1, 2, 3, 4, and 5 years after the exposure group. Absolute changes in salary are shown across gender. Graph shows parameter estimates and 95% CI.

**The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.**

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
<b>Title and abstract</b>					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	title abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.  RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.  RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	title abstract  title abstract  abstract
<b>Introduction</b>					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	abstract introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	introduction		
<b>Methods</b>					
Study Design	4	Present key elements of study design early in the paper	introduction		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	materials and methods		

Participants	6	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p><i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p> <p><i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>	materials and methods	<p>RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.</p> <p>RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.</p> <p>RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.</p>	<p>materials and methods</p> <p>materials and methods</p> <p>not included</p>
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	materials and methods main text	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	materials and methods
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	materials and methods		

1 2 3	Bias	9	Describe any efforts to address potential sources of bias	materials and methods and results	
4 5	Study size	10	Explain how the study size was arrived at	materials and methods	
6 7 8 9 10 11	Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	materials and methods	
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35	Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	a) materials and methods b) materials and methods c) materials and methods d-e) NA	
36 37 38 39 40 41 42 43 44	Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.  materials and methods  materials and methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	materials and methods
<b>Results</b>					
Participants	13	(a) Report the numbers of individuals at each stage of the study ( <i>e.g.</i> , numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a-c) materials and methods	RECORD 13.1: Describe in detail the selection of the persons included in the study ( <i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	materials and methods
Descriptive data	14	(a) Give characteristics of study participants ( <i>e.g.</i> , demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time ( <i>e.g.</i> , average and total amount)	a) materials and methods and Supplementary Table S3 b) materials and methods, Table 1 c) materials and methods		
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure	materials and methods and Supplementary Table S3, results		



		category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	a) results b) results c) results		
Other analyses	17	Report other analyses done— e.g., analyses of subgroups and interactions, and sensitivity analyses	results		
<b>Discussion</b>					
Key results	18	Summarise key results with reference to study objectives	results and discussion		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	results and discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	NA
Interpretation	20	Give a cautious overall interpretation of results considering objectives,	discussion		

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results	discussion		
<b>Other Information</b>					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Title page		
Accessibility of protocol, raw data, and programming code		..	Reference to supplementary data throughout the text	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	materials and methods

\*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

\*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.