Unequal excess mortality during the Spanish Flu pandemic

Auke Rijpma[1], Ingrid K. van Dijk[2,3], Ruben Schalk[1], Richard L. Zijdeman[4], and Rick J. Mourits[2,4]

[1]Utrecht University
[2]Radboud University
[3]Lund University
[4]International Institute of Social History

November 3, 2021

Abstract

A century after the Spanish Flu the COVID-19 pandemic has re-ignited scholarly interest in socioeconomic and occupational differences in mortality in the earlier pandemic. The magnitude of these differences, whether they were context-specific, and the underlying pathways tying together increased mortality and occupations remain unclear. In this paper, we explore the relation between occupational characteristics and excess mortality during the pandemic in the Netherlands. Our aim is to disentangle social standing and conditions for viral transmission by creating a new occupational coding for exposure to disease at work. We use a data set based on death certificates to calculate excess mortality rates by age group, sex, and occupational group. Using OLS regression models, we estimate whether social position, regular interaction in the workplace, and working in an enclosed space affected excess mortality in the Netherlands in the autumn of 1918. We find that people with occupations that featured social contact had higher mortality in this period. However, a strong socio-economic gradient to excess mortality also existed, even after accounting for exposure in the workplace.

Keywords: excess mortality, 1918-9 influenza pandemic, Spanish flu, socioeconomic health inequality, occupational health risk

JEL Codes: N34, I14.

- We explore the socio-economic gradient in excess mortality during the Spanish Flu.
- We develop a new measure for occupational risk of exposure based on whether occupations involve frequent social contact and on whether work is conducted in indoor environments.
- There was a strong socio-economic gradient to excess mortality during the Spanish Influenza pandemic .
- While we find some evidence that occupational exposure was associated with higher mortality, the status and skill components of occupations had a far stronger effect, suggesting that other aspects of social class were more important predictors of excess mortality. For instance, exposure at home due to crowding or poor existing health.

1 Introduction

In 1918-1919, the Spanish Flu took the lives of an estimated 50-100 million people worldwide (Johnson and Mueller 2002). In the Netherlands alone, over 41.000 persons died, of which about 30.000 perished during the autumn wave of 1918 (CBS 1918; Quanjer 1921). At the time, it was thought that social differences in mortality were relatively small, both in the Netherlands and abroad. For example, the Dutch Health Council reported that there were notable differences in mortality between age groups and regions, but found no reason to conclude that occupation or wealth affected mortality (Quanjer 1921). In recent decades, however, evidence has grown that there existed a socioeconomic gradient in mortality to the Spanish flu (Mamelund 2006). Yet, uncertainty remains, and the mechanisms that linked together socioeconomic factors and mortality risk remain unclear. In this paper, we use new micro-level data containing occupations and dates of death for the provinces of Gelderland, Overijssel, Zeeland, Drenthe, and Zuid-Holland in the Netherlands to further establish whether and how mortality risk to the Spanish Flu was linked to socioeconomic and occupational characteristics.

Recent studies have linked the socioeconomic gradient in mortality rates during the 1918-19 pandemic to a number of pathways. First, crowding in homes among the poor may lead to new infections in the same household (Mamelund 2006). Second, existing health conditions may have exacerbated outcomes after infection with the Spanish flu, and some of these health conditions had an established underlying socioeconomic dimension such as tuberculosis (Janssens and Van Dongen 2018; Mamelund and Dimka 2019). Third, illiteracy was linked to increased infection rates and worse outcomes after infection, possibly due to reduced access to medical care and lower awareness and following of public health guidelines during the pandemic (Grantz et al. 2016). Fourth, access to health care in general may have been limited for disadvantaged social groups, as was for example found for black and coloured residents in South Africa during the Spanish Flu (Fourie and Jayes 2021). Fifth, Bengtsson, Dribe, and Eriksson (2018) suggested that differences in the degree to which people had interpersonal interactions in their daily lives may have contributed to the social gradient in pandemic mortality.

In this work, we show that a social gradient existed in Spanish Flu mortality in the Netherlands, and analyse socioeconomic and occupational factors which may have contributed to risk of infection with the Spanish Flu and risk of lethal outcomes. Particularly, we are interested in the social gradient in Spanish Flu mortality while taking into account occupation-related risk of infection. Such a social gradient may point to the relevance of resources and pre-existing health differences in shaping survival during the pandemic, whereas a relationship with occupational characteristics may point towards the role of infection risk in the workplace.

Our case is the Spanish Flu in the Netherlands, the history of which has largely remained unwritten (for exceptions, see De Melker 2005; Gooyer 1968; Mourits et al. 2021; Quanjer 1921; Vugs 2020). We focus on the second wave of the Spanish Flu, in autumn 1918, which was by far the most deadly phase of the pandemic. By using occupations registered on death certificates, we can measure socioeconomic status, whether work led to frequent social contact and whether it took place in enclosed spaces. With this new occupational classification, we establish whether people worked in conditions related to increased likelihood of viral transmission, and test whether these characteristics account for the social gradient in Spanish Flu mortality in the Netherlands.

2 Theory and literature

2.1 Historical background

The 1918 influenza pandemic has been described as "the mother of all pandemics" (Taubenberger and Morens 2006). After a mild summer wave in 1918, a deadly second wave took hold during the final months of the First World War. Between September and December 1918, the pandemic took more lives than the preceding four years of warfare. The flu was notorious for its lethality, but especially feared as it heavily affected adults between ages 20-40 who are normally least affected by infectious diseases. Especially young men and pregnant women were affected, and among pregnant women the risk of a stillbirth was elevated. The infected initially showed regular flu symptoms, which could quickly develop into pneumonia. Death often followed within days. The sheer horror is perhaps best described in the memoirs of the surgeon general of the US army Vaughan: "I see hundreds of young, stalwart men in the uniform of their country [...] placed on the cots until every bed is full and yet others crowded in. The faces soon wear a bluish cast; a cough brings up the blood stained sputum. In the morning the dead bodies are stacked about the morgue" (Kolata 2001; Vaughan 1926).

In the 20th century, the Spanish Flu was mostly perceived as a socially neutral disease that affected both the poor and the rich. Immediately after the outbreak, many doctors and scientists maintained that there was no socioeconomic gradient in who was affected (Bengtsson, Dribe, and Eriksson 2018; Quanjer 1921). For the Netherlands, the Dutch Health Council (Quanjer 1921) wrote that for influenza-related mortality: "One cannot conclude that there is any distinction, neither by occupation, nor between the more and less affluent.” [“Men kan daaruit niet tot eenige voorkeur, noch voor beroep, noch voor meerdere of minder gegoedheid besluiten”, p. 24]. The idea of the Spanish flu as a socially neutral disease persisted for the remainder of the 20th century. Doctors and scholars thought that the influenza virus infected and killed all classes equally because the disease was so highly transmissible (Mamelund, Shelley-Egan, and Rogeberg 2019).

2.2 Social differences in early 20th-century mortality

The image of Spanish Influenza as a socially neutral disease has changed in the past decades. Statistical evidence has since then shown that mortality to the Spanish Flu was higher in poorer countries (Johnson and Mueller 2002; Mamelund, Shelley-Egan, and Rogeberg 2019), even though there are recent studies arguing that Spanish Influenza was unrelated to pre-pandemic economic variables (Brainerd and Siegler 2003; Crosby 2003). Furthermore, in more developed countries mortality among the poor was higher, even though there was no perfect social gradient in mortality (Bengtsson, Dribe, and Eriksson 2018; Mamelund 2006). This disadvantage of the lower classes was especially notable among men (Bengtsson, Dribe, and Eriksson 2018).

The discussion about the role of socioeconomic status during the Spanish flu pandemic links up to a larger debate on the origins of socioeconomic differences in mortality. Across Europe, the social gradient in health was growing in the first half of the 20th century (Bengtsson and Van Poppel 2011). In industrialized and urbanized England, survival differences by socioeconomic group were already present around 1900 (Antonovsky 1967; Razzell and Spence 2006). In some regions, such as Southern Sweden, a health gradient by social class emerged relatively late (Debiasi 2020). In the Netherlands in 1918, mortality differences after age 50 were only to a very limited extent affected by social class (Mourits 2019), but in younger age groups, which were especially affected by the Spanish Flu, there existed a social gradient in health from the end of the 19th century. Van Poppel, Jennissen, and Mandemakers (2009) found that between ages c. 35–55, the elite in the Netherlands had a survival advantage over farmers and the middle class, whereas the working class had a survival disadvantage. Both existing differences in health as well as other social class characteristics related to knowledge and skills, income, and occupational characteristics may have played a pivotal role in excess mortality during the Spanish Influenza pandemic .

According to the fundamental cause theory (Link and Phelan 1995), a socioeconomic gradient in health persists even when the disease environment changes, as socioeconomic resources are connected to health advantages regardless what the dominant diseases and causes of death are. However, the specific advantages linked to socioeconomic resources are dependent on the local context (Clouston et al. 2016). For diseases that are not well understood or cannot be treated effectively, such as the Spanish Influenza or COVID-19, socioeconomic differences in disease burden or mortality are still expected. For example, resources can help people isolate from sick relatives or co-workers, improve their access to information and understanding of the appropriate behaviour to avoid exposure, and ease access to health care. These factors can mitigate health risks related to infectious disease, regardless what the dominant diseases are.

Recent studies have provided more evidence for socioeconomic contrasts in Spanish Influenza mortality. Herring and Korol (2012) show that the poorer neighbourhoods of Hamilton, Ontario had significantly higher influenza mortality rates than the richer neighbourhoods. Tuckel et al. (2006) and Fanning (2010) found that immigrant groups from Italy and Eastern Europe in Hartford, Connecticut and Norwood, Massachusetts, who often had low socioeconomic status, were harder hit by the Spanish Influenza pandemic than non-immigrants. Similarly, Clay, Lewis, and Severnini (2018) find that the share of foreign-born immigrants and pre-pandemic typhoid rates predict higher all-cause mortality during the pandemic.

Among the factors that may have contributed to higher risk among lower socioeconomic status groups figures risk of exposure prominently. [1] Socioeconomic status is and was closely related to the degree and intensity of social contact, and households living in crowded conditions may have had a larger likelihood of exposure to the Spanish flu, especially in urban slums. Specific factors for socioeconomic differences in Spanish Influenza mortality include hygiene and crowded housing, implying that the sick could not be cared for in a separate room. Such living conditions were also related to existing health before the arrival of the pandemic. Tuberculosis in particular may have been a risk factor, as it may have contributed to the likelihood of a secondary infection with bacterial pneumonia as well as reduced access to health care (Grantz et al. 2016; Mamelund 2006). Not surprisingly, then, the working class and those in small housing had higher death rates in Oslo, Norway (Mamelund 2006).

[1] There are also studies that focus on vulnerability to infection. One way through which vulnerability could lead to higher mortality during the Spanish Influenza was through air pollution, to which the poor were more exposed. Clay, Lewis, and Severnini (2018) show that pollution due to coal-fired electricity generation worsened the impact of Spanish Influenza in US cities, as the air pollution probably made lungs more susceptible to infection. Similarly, Brundage and Shanks (2008) argue that not the virulence of Spanish flu, but the likelihood of secondary infections played a core role in explaining its high death rates and age profile of the deceased as well as differences between occupational groups.

2.3 Occupational exposure to the Spanish flu

Exposure to disease does not only occur at home but also at the workplace. For Chicago it was found that unemployment rates were related to decreased mortality rates and transmission, indicating that social contact and not poverty itself could have been a key factor in mortality to the Spanish flu (Grantz et al. 2016). From the early 20th century, working conditions in factories were recognized as a general risk factor for transmission of infectious diseases (Van Der Woud 2010). In 1898, Dutch socialist politician Domela Nieuwenhuis (1898) wrote about the poor hygiene, and exposure to toxins, dust, and small particles among the working poor. By 1917, critique of labour conditions had become more widely accepted. Statistics Netherlands reported that working men aged 35-54 working outdoors or near furnaces had lower all-cause and tb-related mortality than those who were subjected to organic dust from tobacco, flower, and textiles, who worked with chemicals and toxins, or – even more detrimental – were exposed to grinding dust from glass, metal, porcelain, or stone (CBS 1917). For the same period in Sweden, reports by healthcare professionals addressed how poor air quality and unhygienic conditions in factories made workers more prone to tuberculosis (Sundin and Willner 2007).

Bengtsson, Dribe, and Eriksson (2018) provide evidence that the ability to maintain distance from others (now commonly known as ”social distancing”) affected mortality rates favourably. In southern Sweden, farmers were least affected by the Spanish flu, and low-skilled and unskilled workers most strongly. Although this points towards the importance of solitary work, it should be noted that the protective effect of farming was only found for men, and that white collar workers had no mortality advantage compared to the working class. To understand whether social contact affected Spanish Influenza mortality, social contact at the workplace needs to be measured in more detail. Differences in mortality by social class may have been directly affected by job characteristics and working conditions that are not fully captured by social class schemes (Debiasi 2020). Recent studies have highlighted substantial heterogeneity in mortality rates within social classes but between occupational groups (Debiasi 2020). The decades before 1918 were dominated by infectious disease mortality, and occupational differences in mortality may have been, at least partly, related to exposure to infectious disease. For example, among white collar men in Sweden, especially among health professionals, there was a mortality disadvantage, but not among religious professionals. It is contested whether the clergy may have had a better lifestyle than the general population (Debiasi 2020), as Bavarian monks had no survival advantage over non-cloistered men before the 1950s (Luy 2003). More likely, the difference in mortality rates between doctors and religious professionals can be explained by exposure to infectious disease due to the nature of their occupation.

In this paper, we take an in-depth look at occupational characteristics and mortality during the Spanish Influenza pandemic . Instead of including specific occupations as has been done in earlier studies to all-cause and infectious disease mortality, we address broad categories of occupations defined by occupational risk factors. We supplement existing coding systems for occupational skill and status with a coding for occupational risk of exposure to infectious disease. Specifically, we look into the degree of social contact at work and whether workers worked in indoor environments, at the time often in cramped and poorly ventilated conditions with workers working closely together, conditions under which viral transmission tends to be higher than outdoors.

3 Methods and data

3.1 Death certificates and the civil registry

We use death certificates from the Dutch civil registry for the years 1910-18. The Dutch civil registry was introduced in 1811 and provided legal proof of birth, marriage, and death. By the turn of the twentieth century the procedure was well-established. For everyone who died in the Netherlands a death certificate was issued, in duplicate so that safekeeping was ensured (Mourits, Van Dijk, and Mandemakers 2020; Vulsma 1988). Digitisation of these certificates has been ongoing since the 1990s by local and provincial archives. Death records list the date of death, full name, age, sex, place of residence, and the occupation of the deceased.[2]

These individual-level records allow us to analyse mortality during the Spanish Influenza pandemic in more detail than previously. Other historical sources, such as municipal reports or national health reports, give only aggregated deaths per year and do not split out observations by age, sex, or occupation (see for example, the Historical Database Dutch Municipalities (Boonstra 2020) or (CBS 1918)). The aggregated number of deaths per municipality is available for age groups, but without information on occupation or other indicators of socioeconomic status. Some records of individual-level causes of death data survive, but they are fragmented and scarce. Death certificates are thus the only remaining source that combines key individual-level characteristics with sufficient nationwide coverage to generate meaningful insights into

[2]We used the digitised certificates from https://www.openarch.nl/api/docs/. Data processing scripts are available at https://github.com/CLARIAH/wp4-civreg/tree/master/deathsv1, and the dataset at https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/PCAEGG.

mortality patterns along socioeconomic lines.

Death certificates have been digitised for all Dutch provinces, but not for all municipalities (figure 1, top left panel). Compared to the Historical Database Dutch Municipalities, death certificates are digitised for 85 per cent of the 1,117 municipalities existing in 1918, accounting for an estimated 80.5 per cent of the total Dutch population at the time. A large part of the missing death certificates are from the city of Amsterdam, where historical death certificates have not been digitised.

Some certificates were entered by more than one archive, or were digitised multiple times by the same archive, resulting in duplicates. Duplicates have been removed from the data. [3]. Certificates were digitised on a per-archive basis, and therefore there is small variation in what information from the certificates was digitised (figure 1). We exclude municipalities where the share of reported ages, sex, and precise dates is below 50%.[4] In the case of occupations, missing information often reflects the certificates as many people did not have an occupation, especially in the case of children, the elderly, and women, and thus we exclude municipalities where the share of recorded occupations is below 10%.

The changes in the sample due to dropping municipalities are shown in table **??**, showing that it leaves us with 215 511 certificates for 224 municipalities, mostly from the provinces of Gelderland, Overijssel, Zeeland, Drenthe, and Zuid-Holland. The most important bias introduced this way is that the three largest cities (Amsterdam, Rotterdam, and Utrecht) are missing from the dataset. While our data does contain cities, of cities, including a number of textile and industrial centres, this does mean that our data has a more rural focus than data for the country as a whole.

Table **??** also shows the results of further selection steps, and how each step affects the composition of the included sample. The most important changes occur when we drop certificates without a registered occupation and other missing variables, which strongly changes the sex composition of the sample. Occupations were rarely registered for women, and hence our final data set describes a largely male population.

[3] Duplicates were identified by checking for certificates with an identical death date and location, and dropping certificates with similar full names within these clusters (using string distances from Loo (2014)). Although some remaining duplicates cannot be excluded, our earlier findings demonstrate that any over- or underestimation of deaths is stable over time at the municipal level, and should not affect our results (Mourits et al. 2021)

[4] When sex was not digitised, we inferred this from the first names, using predictions from the names on certificates that did have a reported sex, resulting in low rates of missingness for all municipalities.

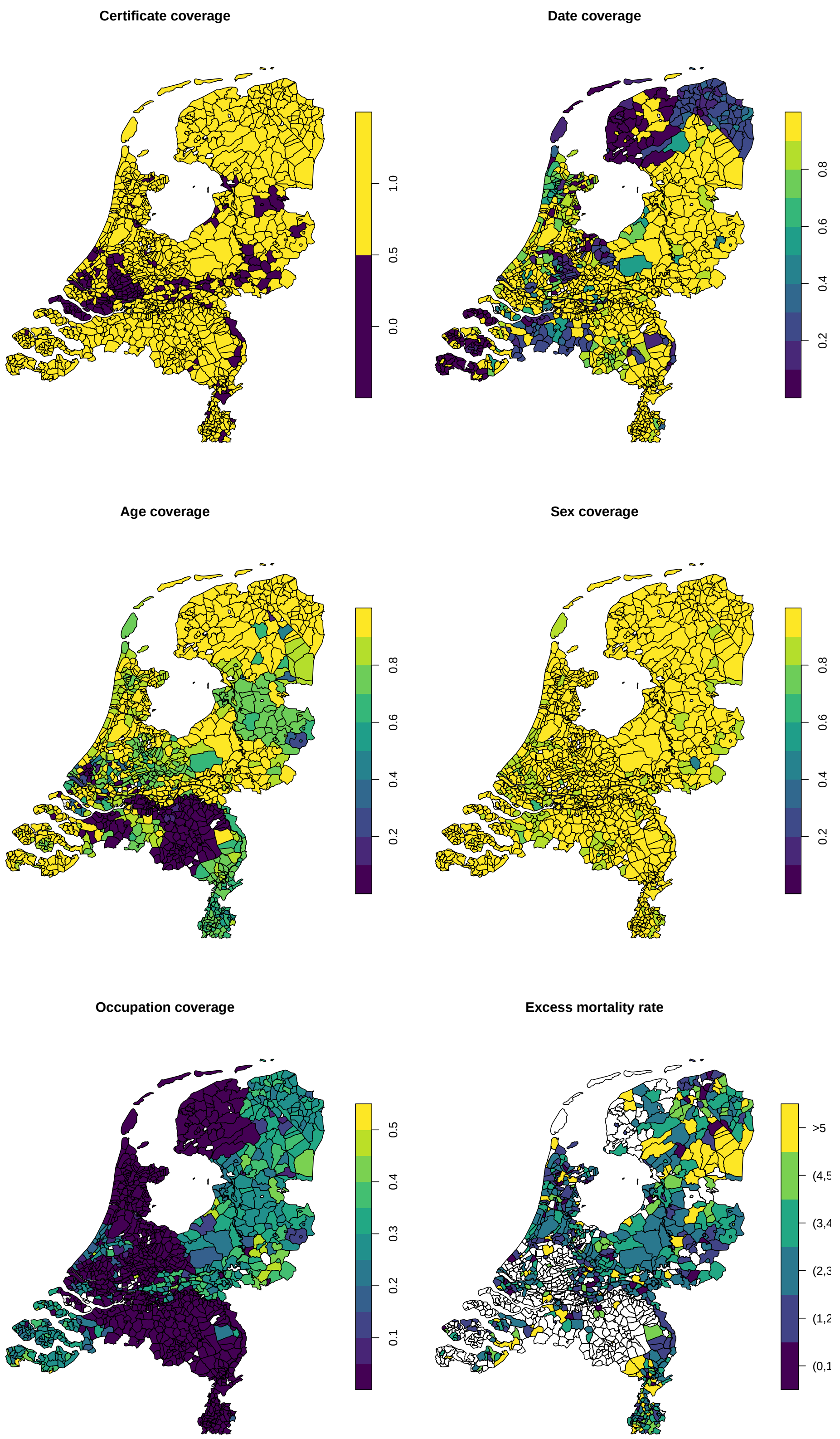

Figure 1: Maps of coverage of certificates dates, age, sex, and occupations on death certificates, and excess mortality rate.

Table 1: Summary statistics and selection steps.

selection	certificates	municipalities	age	% male	% unskilled	% contact
start	741758	1116	43	52	32	27
certificate coverage	715703	946	43	52	33	27
variable coverage	220795	215	42	52	30	26
drop 1914	198557	215	42	52	30	26
Sep-Dec	64365	215	41	51	31	27
$11 < \text{age} < 79$	35077	215	53	51	32	27
drop missing	12625	215	50	91	32	27

3.2 Occupational coding

Standardised occupations of the deceased were matched against standardised occupational titles from the Historical Sample of the Netherlands (HSN) (Mandemakers et al. 2020). This dataset contains over 280 000 Dutch occupational titles, with numerous spelling variations, coded in HISCO (Historical Classification of Occupations) and HISCLASS (Historical International Social Class Scheme) (Van Leeuwen and Maas 2011; Van Leeuwen, Maas, and Miles 2002). Almost 98 per cent of occupational titles on the death certificates could be coded this way.[5] To infer the skill level of each occupation, we used the skill category of its corresponding HISCLASS, ranging from high (HISCLASS 1-2), medium (HISCLASS 3-4, 6-8), low (HISCLASS 5, 9-10), to unskilled (HISCLASS 11-13).

As an alternative to the HISCLASS system, we also use HISCAM occupational status scores as a robustness check (Lambert et al. 2013). This is a non-categorical measure that infers occupational status by estimating the ”social interaction distance” from historical marriage behaviour. The advantage of HISCAM is that it measures how a myriad of implicit social factors affects social stratification, making it a worthwhile addition to the HISCLASS skill dimension. HISCAM also provides a useful alternative coding of farmers. In the HISCLASS skill scheme, farmers are coded as medium skilled workers, while HISCAM views them as a relatively low status profession. Both arguments make sense insofar farmers had to run a complicated business, but the certificates we rely on do not always distinguish between the operators of large farms, smallholders and farm labourers.[6] Since a large share of our observa-

[5]Some minor adjustments to the HSN-HISCO scheme were made. See `https://github.com/rijpma/spanish`

[6]The Dutch term *landbouwer* which is frequently used on the certificates is not specific enough to make this distinction.

tions come from the more rural east of the country, and because these areas are also where the pandemic struck hardest (see figure 1), the measurement of farmers could affect our analyses.

Occupational working conditions were coded on two dimensions: 1) working in an indoor environment (yes/no), and 2) regular social interaction (yes/no). Coding was done at the three-digit HISCO level. For example, an office clerk was considered to be working indoors, but had infrequent social contact, especially strangers. A tailor also worked under a roof but would have had frequent social contact through dealings with customers. Conversely, a farm worker worked neither indoors, nor in close contact with others. Occupations were independently coded by two or more researchers and then compared. Approximately 75 per cent of occupations were coded identically in the first round. Remaining occupations were discussed by the coders until an unanimous decision was reached. Although our scoring omits regional and within-occupation variance, it captures conditions for exposure to and transmission of the virus on the work floor. [7]

3.3 Excess mortality

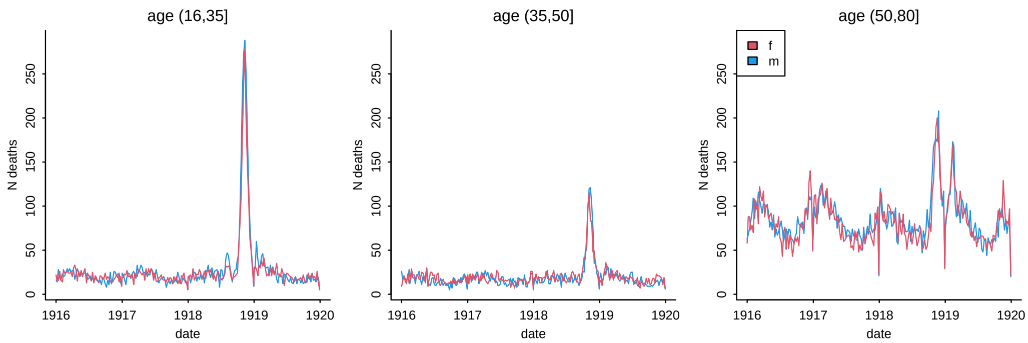

Figure 2: Weekly number of deaths by age and sex in the Netherlands from death certificates , 1917–1919

As death certificates in the Netherlands do not include a cause of death, we work with all-cause mortality. We use excess mortality as our outcome variable: the number of deaths in excess of what would be expected based on previous years. This approach is used as the 1918 municipal population at risk (by age, sex, and occupation) is not available. We calculate excess mortality rates (EMRs) for 1918, comparing the number of deaths between September 1st to

[7] Appendix C provides our coding of the most frequent occupations in the dataset.

December 31st in 1918 per age group, sex, municipality, and occupational cluster to mortality in the preceding years. Baseline mortality was established by calculating the average mortality of the same occupational groups (by sex and age) in autumn 1910-1917 (September through December). The Netherlands remained neutral during the First World War, so that dramatic changes to the population at risk and mortality rates that characterised many other countries at the time were lacking (Colvin and McLaughlin 2021). The exception was the year 1914, when the Netherlands briefly took in 1 million Belgian refugees, who largely returned before the end of the calendar year. We excluded 1914 from the baseline to ascertain that our estimates of excess mortality are not biased by the events of the First World War in the Netherlands.

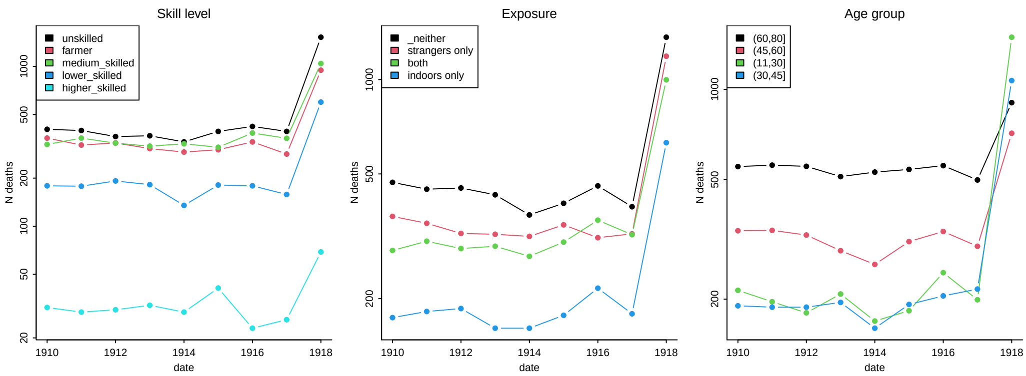

Figure 3: Log annual deaths in September-December, 1910–1918, by occupational skill level (left panel), exposure (middle panel), and age group (right panel)

Our approach relies on assumption that the population at risk is relatively stable in the period leading up to the Spanish Flu. Figure 2 shows that relative to the huge mortality spike of the Spanish Flu, weekly death number were stable in the 2.5 years preceding the pandemic, especially in the age categories that we are interested in. Moreover, figure 3 shows that the annual number of deaths for the main subgroups of interest, in the selection of the data described above, were also stable over the 1910-17 period. In table 2, we calculate excess mortality rates, which confirm in more detail the picture shown in figure 3.

We can also use external demographic datasets to investigate the possibility of strong demographic shifts in the Netherlands in the 1910–17 period. The Historical Database Dutch Municipalities shows that across 1,225 municipalities, the coefficient of variation in population was on average only 0.04 (Boonstra 2020). Using total Dutch population figures from the Human Mortality Database, it is shown that the coefficient of variation in 10-year age groups av-

Table 2: Excess mortality rates in September-December 1918 by occupational skill group, with bootstrapped standard errors.

group	EMR	sd	difference	sd_diff
higher_skilled	2.29	0.32	-	-
medium_skilled	3.02	0.08	0.73	0.33
lower_skilled	3.35	0.17	0.33	0.19
unskilled	3.90	0.12	0.55	0.21

eraged to, again, 0.04 (*Human Mortality Database*). While the total Dutch population in this period grew 1.5 % per year, implying growth in the subgroups, year-to-year shifts in mortality are small compared to the events of 1918.

3.4 Analysis

Mortality differences for between 1918 and the preceding years are modelled using an OLS-regression with the ratio of deaths in September–December 1918 over the average number deaths in these months for the period 1910–17 as the dependent variable. The distribution of these ratios is highly skewed, so we take the logarithm of the dependent variable. [8]

However, as we calculate excess mortality per municipality by sex, age, and social class, for 59% of our observations for 1918 there are no deaths, resulting in a ratio of zero. We do not discard these data because the fact that these groups had zero mortality during the Spanish Flu is an important fact. To include these observations, we add one to to the dependent variable, so that zero deaths have a value of one. This means we estimate excess mortality in location t, month t, occupational group o, age group a, and sex s as follows:

$$\log \frac{deaths_{itosa}^{year=1918}}{baseline\, deaths_{itosa}} + 1 = \alpha + \sum \beta\, skill_o + \sum \gamma\, exposure_o + \delta\, farmer_o + \sum \zeta\, age_a + \sum \eta\, sex_s + \nu_l + \xi_t + \epsilon_{itosa} \quad (1)$$

Standard errors are clustered by location (Zeileis, Köll, and Graham 2020). For locations we use Economic Geographic Regions (EGG): clusters of municipalities defined by Statistics Netherlands below the NUTS-3[9] level. We pre-

[8] This ratio differs by a value of one from the excess mortality rate, which is calculated as the ratio of excess deaths minus expected deaths in the numerator. We do not use this statistic to avoid negative values.

[9] Nomenclature of Territorial Units for Statistics. See: https://ec.europa.eu/eurostat/web/nuts/background

fer to work at this relatively low level of aggregation to be able to control for the spatial patterns of the pandemic as well as the differences between urban and rural regions which the EGG regions capture, but NUTS-3 region in the Netherlands do not.[10]

To verify the validity of our outcomes, we conduct a number of robustness checks. First, we explore whether the level of geographic aggregation affects our results. Larger geographic areas mean that we have relatively fewer cases of zero deaths for 1918, and, more generally, will reduce variance in our data, albeit at the expense of a lower number of observations. In addition to the EGG level, we also estimate our models for municipalities, COROP[11] (NUTS-3) regions, and provinces (NUTS-2). For similar reasons we also provide separate estimates for high and low mortality regions. This allows us to focus on the regions most affected by the Spanish Influenza pandemic , and again, exclude many cases of zero deaths. The cutoff was an overall excess mortality 2.5 times the mortality we would usually expect in September-December. In addition, we also look at a variety of alternative types of models to deal with the zeroes in our data. We also check whether population density affect our estimates due to easier transmissibility in cities compared to the countryside, Finally, we ascertain whether the 1914-1918 mobilisation of troops does not affect our results by controlling for the presence of army bases.

4 Results

Table 3 shows the results of this model, progressively adding more controls to come to our preferred model in the rightmost column. For each estimate we report the coefficient, standard error, and significance level.

In the most basic model, it can be seen that there is a skill gradient in the excess mortality rates of 1918. Compared to higher-skilled workers, medium skilled workers workers had 48 percent higher mortality than they would normally experience ($\exp(0.39) - 1$). For unskilled workers this was higher still: their excess mortality was 63 percent higher than that of higher skilled workers. Lower skilled workers also had substantially higher excess mortality than higher-skilled workers, but lower than unskilled workers and medium-skilled workers. An important driver of this break in the skill gradient are farmers. They are a large group in our dataset with relatively high mortality who are classified as medium-skilled in the HISCLASS scheme. We return to this point

10 An example of our dataset can be found in appendix A.

11 Artificial regions defined for statistical analyses by Statistics Netherlands, named after the COördinatie commissie Regionaal OnderzoeksProgramma

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
medium skilled	0.39^{***}	0.31^{***}		0.30^{***}	0.31^{***}	0.32^{**}	0.39^{***}
	(0.09)	(0.08)		(0.09)	(0.09)	(0.10)	(0.09)
lower skilled	0.21^{***}	0.21^{***}		0.19^{**}	0.18^{*}	0.19^{*}	0.21^{**}
	(0.05)	(0.05)		(0.07)	(0.08)	(0.08)	(0.08)
unskilled	0.49^{***}	0.49^{***}		0.48^{***}	0.49^{***}	0.49^{***}	0.56^{***}
	(0.11)	(0.11)		(0.14)	(0.14)	(0.15)	(0.12)
farmer		0.27^{***}	0.31^{***}	0.40^{***}	0.48^{***}	0.46^{***}	0.45^{***}
		(0.07)	(0.08)	(0.11)	(0.11)	(0.11)	(0.09)
indoors only			0.01	0.16	0.17	0.17^{*}	0.13
			(0.06)	(0.08)	(0.09)	(0.09)	(0.08)
contact only			0.21^{**}	0.22^{**}	0.22^{**}	0.21^{**}	0.18^{**}
			(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
both			0.02	0.11	0.16^{*}	0.15	0.12
			(0.06)	(0.07)	(0.08)	(0.08)	(0.07)
age (30,45]					-0.24^{***}	-0.24^{***}	-0.25^{***}
					(0.05)	(0.05)	(0.05)
age (45,60]					-0.45^{***}	-0.44^{***}	-0.43^{***}
					(0.05)	(0.05)	(0.05)
age (60,80]					-0.49^{***}	-0.50^{***}	-0.47^{***}
					(0.06)	(0.06)	(0.06)
male					0.32^{***}	0.30^{***}	0.37^{***}
					(0.06)	(0.06)	(0.06)
event_month10						0.38^{***}	0.39^{***}
						(0.05)	(0.05)
event_month11						0.77^{***}	0.78^{***}
						(0.06)	(0.06)
event_month12						0.25^{***}	0.27^{***}
						(0.04)	(0.04)
(Intercept)	0.26^{**}	0.26^{**}	0.53^{***}	0.14	0.16	−0.18	-0.79^{***}
	(0.09)	(0.09)	(0.06)	(0.14)	(0.20)	(0.20)	(0.19)
Region FE	No	No	No	No	No	No	Yes
R^2	0.02	0.02	0.01	0.03	0.08	0.16	0.23
Adj. R^2	0.02	0.02	0.01	0.03	0.07	0.15	0.22
Num. obs.	3701	3701	3700	3699	3573	3573	3573

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Regression models of log excess mortality rate. Region-clustered standard errors between parentheses.

in more detail below. Here we observe that if we separately control for farmers (model 2 onwards), the skill gradient is still not strictly a perfect gradient, but closer to it, with medium and lower skilled workers having similar higher excess mortality compared to higher skilled workers, and unskilled workers having the highest excess mortality.

Next, we introduce our occupational exposure variables (model 3). Here, we find that they generally predict higher mortality, specifically, about 20 percent higher in the case of contact occupations compared with occupations that had neither of the exposure characteristics. The other occupational exposure variable, indoor work, is also related to higher excess mortality, but not estimated with sufficient precision to draw firm conclusions. Here again we include the ”farmer” control, as this large group has high excess mortality, but we scored it neither as being indoors nor involving frequent contact with strangers. Of particular interest is what happens to the skill gradient when we introduce these exposure variables (model 4). We find that while controlling for exposure, the skill gradient in excess mortality remains largely unchanged, although the mortality estimate for farmers increases. Overall, we find that working indoors had no effect on excess mortality. However, occupational exposure to infectious disease due to frequent contact predicts somewhat higher excess mortality, though skill remains a more important predictor.

Adding age and sex controls further increases the estimate of excess mortality among farmers, but otherwise does not change this picture. This is important because we know that Spanish influenza affected younger men in particular, and we can expect them to have lower-skilled occupations due to career effects. Indeed, keeping all else constant, we find the highest excess mortality for age groups 12-30 (reference category), with progressively lower mortality going up in age group. Adding month indicators to capture the phase of the epidemic also leaves the estimates for skill level largely unchanged. We finally add region fixed effects to adjust for the fact that some regions were struck harder than others (the rural, poorer North-East in particular, see Mourits et al. 2021). This somewhat strengthens the skill gradient. In our preferred model, including the full set of demographic controls and region and time fixed effects medium skilled workers have 47 percent higher excess mortality compared to high skilled workers; lower skilled workers 23 percent and unskilled workers have 75 percent higher excess mortality. The predicted excess mortality due to the exposure characteristics of the occupations of the deceased is around 18 percent for contact occupations compared to the baseline of occupations characterised by neither exposure risk.

4.1 Alternative occupational coding and farmers

To investigate whether our results are sensitive to the choices in occupational coding schemes, we test a number of alternative models (table 4). In column 2, the preferred model from table 3 is estimated again, but for a dataset where excess mortality was calculated with farmers dropped from the dataset. In the next column, as another alternative, we recode farmers as lower skilled.

We find that omitting farmers from our dataset makes only a limited impact on the social gradient. Recoding farmers as lower skilled does make a difference. The N-shaped gradient, where, compared to higher skilled workers, lower skilled workers had lower excess mortality than medium skilled workers, disappears, and now medium and lower skilled workers have similar excess mortality. In this model, our exposure variables give the counter-intuitive result that working indoors predicts lower excess mortality. Most likely, because lower-skilled labourers who worked indoors still had lower excess mortality than farmers.

We also estimate our preferred model with HISCAM scores instead of HISCLASS. Again, the model based on HISCAM shows that, keeping all else constant, higher occupational status predicts lower excess mortality. A one point increase in HISCAM would result in a halving of excess mortality; since the HISCAM scores in our data range from 0.40–0.99, going from the lowest to the highest status occupation is associated with an decrease in excess mortality of 44%. The fact that HISCAM is a continuous variable also allows us to assess non-linear effects of socioeconomic status on Spanish Flu excess mortality using non-parametric techniques (regression splines Wood 2003). The results are shown in figure 4, where it can be seen that the higher excess mortality is concentrated in the lower status occupations. At a HISCAM score of c. 0.55 or higher (covering 23% of all deaths in our dataset, and 30% of all aggregated cells) and keeping all other variables equal, predicted excess mortality is negative.

4.2 Robustness checks

4.2.1 High and low mortality regions

To verify the validity of our outcomes, we included a number of robustness checks in the appendix. Table 6 in appendix B.1 compares how our preferred model performs in regions with relatively low or high overall mortality. The cutoff was an overall excess mortality rate of 2.5, that is 2.5 times the mortality we would usually expect in September-December. As expected, we find that most estimates in the high mortality areas are higher compared to low mortal-

	all occupations	no farmers	farmers recoded	hiscam
medium skilled	0.39^{***}	0.39^{***}	0.36^{***}	
	(0.09)	(0.09)	(0.09)	
lower skilled	0.21^{**}	0.21^{**}	0.36^{***}	
	(0.08)	(0.08)	(0.09)	
unskilled	0.56^{***}	0.55^{***}	0.45^{***}	
	(0.12)	(0.12)	(0.11)	
farmer	0.45^{***}			0.52^{***}
	(0.09)			(0.08)
hiscam				-0.65^{***}
				(0.14)
indoors only	0.13	0.13	-0.23^{***}	-0.13^{*}
	(0.08)	(0.08)	(0.05)	(0.06)
contact only	0.18^{**}	0.19^{**}	-0.09	0.17^{**}
	(0.06)	(0.06)	(0.06)	(0.06)
both	0.12	0.11	-0.17^{***}	-0.03
	(0.07)	(0.07)	(0.05)	(0.06)
age (30,45]	-0.25^{***}	-0.25^{***}	-0.26^{***}	-0.18^{***}
	(0.05)	(0.06)	(0.05)	(0.03)
age (45,60]	-0.43^{***}	-0.41^{***}	-0.42^{***}	-0.29^{***}
	(0.05)	(0.06)	(0.05)	(0.04)
age (60,80]	-0.47^{***}	-0.45^{***}	-0.47^{***}	-0.32^{***}
	(0.06)	(0.06)	(0.06)	(0.05)
male	0.37^{***}	0.32^{***}	0.33^{***}	0.10
	(0.06)	(0.06)	(0.06)	(0.05)
event_month10	0.39^{***}	0.38^{***}	0.40^{***}	0.22^{***}
	(0.05)	(0.05)	(0.05)	(0.04)
event_month11	0.78^{***}	0.73^{***}	0.79^{***}	0.49^{***}
	(0.06)	(0.06)	(0.06)	(0.05)
event_month12	0.27^{***}	0.24^{***}	0.27^{***}	0.15^{***}
	(0.04)	(0.04)	(0.04)	(0.03)
(Intercept)	-0.79^{***}	-0.52^{**}	-0.37^{**}	0.25^{*}
	(0.19)	(0.17)	(0.14)	(0.10)
R^2	0.23	0.22	0.22	0.17
Adj. R^2	0.22	0.20	0.20	0.16
Num. obs.	3573	3097	3522	4702

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Regression models of log excess mortality rate, exluding selected occupations. Region-clustered standard errors between parentheses.

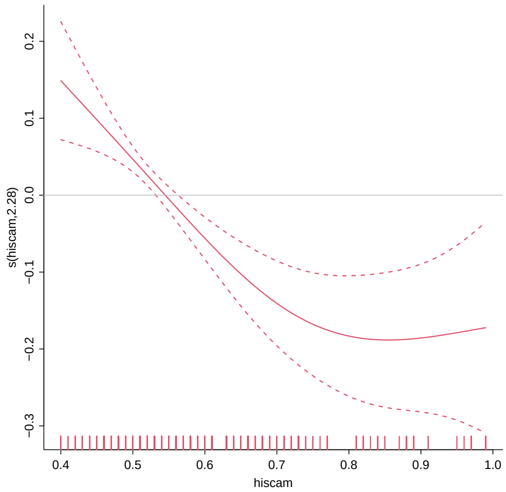

Figure 4: Partial effect of HISCAM on excess mortality

ity areas. But otherwise, the results are similar.

4.2.2 Aggregation level

While we prefer to work at a relatively low level of aggregation, a concern could be that this does result in cells in our dataset with few observations, either for our 1910–17 baseline, or for the Spanish Influenza pandemic deaths. To investigate this, in table 7 we also estimate the preferred model from table 3 at different levels of geographic aggregation: municipalities, EGG regions, Corop (NUTS-3) regions, and provinces (NUTS-2). Overall, the results are similar to the preferred EGG aggregation, though overall effect sizes tend to be somewhat larger at higher levels of aggregation. The one exception is the skill-gradient at the province level (column 4) where excess mortality for unskilled workers is estimated to be lower than at the EGG or Corop level.

4.2.3 Population density and army controls

Important factors in the spread of any infectious disease are population density and geographic mobility, as these make it easier for the disease to move from one person to the next. While the inclusion of region fixed effects in our models should capture any time invariant impact of these factors, we explore this possibility further in table 8, appendix B.3, where we show the effect of including population density and the presence of army bases or hospitals in a municipality. Because these variables are measured at the municipality level, we have to aggregate our data to this level as well. As expected, the overall skill-gradient is the same as these effects would also be captured by the location fixed effects included in the models in table 7. The effect of the population variable itself is positive, as expected. However, population density does not have a significant effect on excess mortality, even if we exclude population size from our model.

We also include a control for municipal military bases in part of our models for more reliable estimates of excess mortality. Although the Netherlands remained neutral during World War I, the Dutch army was mobilised from 1914 to the end of 1918. Living conditions in the barracks and other army camps were poor (Koten and Weel 2014), and young adult men were affected by the Spanish Flu in large numbers. As many soldiers were conscripts, their occupation listed at death would not be 'soldier', but their occupation before the War. Only 0.2% of the death certificates have occupational titles that can be linked to the military between 1910 and 1918, which is surely an underestimation for this period. To control for this potential heavy increase in local excess mortality around military bases, we added a dummy variable to indicate whether divisions of the Dutch army were present in 1918 at the places of death recorded in the certificates. These locations were obtained from combining place names of military bases in 1918 listed in Ringoir (1980) with the locations where Dutch soldiers were admitted to military hospitals in 1913, listed in annual reports (Landmacht 1913). We selected military hospitals for 1913, because no military hospital reports are available between 1914 and 1918, and the recording procedure changed after the War.

Army bases have little predictive value, which is surprising. It should be mentioned, however, that one of the most important military bases in the Netherlands, the main navy base in Den Helder, had some of the highest excess mortality in the country (figure 1), but had to be excluded from our dataset because the occupations for these certificates were not digitised.

4.2.4 Alternative model specifications

As a final check of our results, we investigate whether the way we handle the cases of zero deaths during the Spanish Influenza in our model affects our results. The following approaches are tried here: dropping cases of zero deaths (column 2), not taking the logarithm of the dependent variable (column 3), a quasipoisson model with the ratio of 1918 to baseline deaths as the dependent variable (Silva and Tenreyro 2006, column 4), a quasipoisson model of the number of deaths during the Spanish Influenza pandemic with the log baseline deaths as the offset (column 5), and taking the inverse hyperbolic sine of the ratio (Bellemare and Wichman 2020; Burbidge, Magee, and Robb 1988, column 6).Appendix B.4, table 9 shows that the basic gradient we find in our preferred specification is found in other model forms as well. The main exception here is the model where we drop all the groups with zero deaths during the Spanish Influenza pandemic , in which the patterns we find elsewhere are reversed. As mentioned earlier, absence of Spanish Influenza pandemic deaths is a relevant outcome that should be included in the model. The precision of our estimates is also lower in the two quasipoisson models. Beyond that, what stands out is that the alternative models provide higher estimates of the same pattern.

5 Conclusion

We explored whether there were occupation- and class-related differentials in excess mortality during the Spanish Flu in the Netherlands. Using data from the civil registry, we found that higher social classes had lower mortality. Excess mortality was highest among farmers and unskilled labourers, lowest among higher-skilled labourers, with excess mortality for medium-skilled and lower-skilled workers in between. These findings are in line with evidence from recent studies on Norway and Sweden (Bengtsson, Dribe, and Eriksson 2018; Mamelund 2006), and strongly contrast with the perceptions of contemporaries that the Spanish flu had a relatively egalitarian impact across the society (Mamelund 2006; Quanjer 1921).

We used information from the Dutch civil registry to measure mortality to any cause, rather than disease-specific excess mortality. Death registrations are deemed much more accurate for epidemics, as all-cause mortality rates are not affected by wrong or missing diagnoses of the cause of death, includes individuals who had multiple diseases at the time of death, and includes those who died indirectly from the Spanish Flu (Colvin and McLaughlin 2021). The national death registers included individual information on the age, occupation, and date of death as well as sex and the municipality of residence of the deceased. Because we do not know the population at risk, we could not estimate survival models. However, our robustness checks showed that the data from the Dutch civil registry allows for very robust excess mortality estimates. This makes it possible to estimate the association between social class and excess mortality, while also taking occupational characteristics, age, and regional variance in mortality into account (Colvin and McLaughlin 2021; Mamelund 2011). Moreover, by estimating excess mortality rates we explore whether the Spanish flu increased existing inequalities in mortality in the working population (Van Poppel, Jennissen, and Mandemakers 2009).

In the 1910s, there was a clear social gradient in the mortality rates of the working population. Age, sex, month of infection, and regional differences were the most important predictors of Spanish flu mortality. Yet, the association between socioeconomic status and excess mortality rates was not explained by sex, age or month of death. Moreover, controlling for municipality size and presence of local army bases, including regional fixed effects, or analyzing specific regions with high excess mortality resulted in higher estimates of the magnitude of social class differences in excess mortality during the Spanish flu. In other words, there was a clear social disparity in excess mortality during the autumn wave of the 1918-19 influenza pandemic. These estimates come on top of the existing social gradient in mortality among individuals age 35-55 in the Netherlands (Van Poppel, Jennissen, and Mandemakers 2009), indicating that the Spanish Flu increased existing social inequalities in mortality.

We set out to explore whether the social differences in Spanish Flu mortality could have been caused by occupational characteristics, including social interaction at the workplace and indoor work. We expected that working together with others in an enclosed space or regular social interaction at the workplace might be related to increased chances of infection and death with the Spanish Flu in autumn 1918. To test this hypothesis, a contact index was constructed. The team discussed each instance where there were differences in coding after two separate members coded occupations. After two rounds of discussion, all occupations were coded. We found evidence that social contact at the workplace, as indicated by this occupation-based measurement we developed, was related to increased Spanish flu mortality. However, contrary to our expectations, it did not seem to matter whether individuals worked together indoors or not, as we found similar excess mortality estimates for working indoors, having frequent interactions, and the combination of the two. In unison with our findings for social class, the estimates of our contact index were robust after we controlled for municipality size, local army bases, regional fixed effects, or specifically looked at regions with high excess mortality. Thus, our characterisation of historical occupations in the extent to which individuals worked indoors or social interaction could not explain socioeconomic differences in Spanish flu-related excess mortality, even though contact with people at the workplace, either indoors or frequently outside, increased Spanish Flu mortality.

The increasing social inequalities during the Spanish flu pandemic on top of existing socio-economic differences in mortality are most likely related to the same underlying fundamental causes. Combined, HISCLASS and our contact index measure the importance of income and exposure. The statistically robust and clear mortality gradient ranged from high mortality for unskilled labourers to medium- and lower-skilled labourers with more average mortality rates, and finally the elite with the lowest mortality. This finding indicates that the higher social strata were, to some degree, able to protect themselves against the Spanish flu. It is tempting to think that we are looking at income effects, as educational divides in mortality were uncommon before the second half of the 20th century. Indeed, earlier studies have shown that Spanish flu mortality decreases with each additional room in a house (Mamelund 2006) and that poverty was related to increased Spanish flu mortality across the globe (Murray et al. 2006). In the 1918 Netherlands, lower-skilled and medium-skilled labourers were by no means rich, but they could afford significantly better housing and food than unskilled labourers (Brooshooft 1897; Van Der Woud 2010). However, material resources are unlikely to fully explain the complex and multifaceted advantages that social status provides, as non-material resources - such as existing health conditions, occupational hazards, knowledge about infectious diseases, and access to health care - have been at least equally important in explaining survival differentials in other social contexts (Clouston et al. 2016; Debiasi 2020; Edvinsson and Broström 2012; Elo 2009; Link and Phelan 1995).

An unexpected finding is the high excess mortality rates of farmers. At the turn of the 20th century, farmers were known to outlive their peers in many populations (Ferrie 2003; Gagnon et al. 2011; Mourits 2019; Schenk and Van Poppel 2011; Smith et al. 2009; Temby and Smith 2014; Van Poppel, Jennissen, and Mandemakers 2009). Yet, our findings indicate that farmers had one of the highest excess mortality risks around the Spanish Flu, especially after we control for expects of social contact and age composition. This finding contradicts an earlier study for Sweden (Bengtsson, Dribe, and Eriksson 2018). This is probably because excess mortality rates were especially high in Drenthe and parts of Groningen and Overijssel, rural and poor regions where many people were involved in farming. Contemporaries noted that the poverty in this region might have elevated mortality. Previous to and after the pandemic, farmers often had a survival advantage despite their poverty, as they continued to be physically active at advanced ages and had access to fresh food in a time when refrigeration was not commonplace. However, most studies have found no evidence of a farmer bonus between ages 20 and 40, which was the age group hit hardest by the Spanish flu. It could very well be that poverty protected farmers from man-made diseases caused by idleness and unhealthy eating habits, enabling them to live to old age, but made them vulnerable to outbreaks of infectious disease, as houses were small and the rural population as a whole was malnourished.

Taken together, our results show that the Spanish flu did not hit the population evenly. We find that higher social classes experienced lower excess mortality rates, and these differences were not explained by occupation-specific risk factors included in this work. Working indoors or frequent social contact at the workplace was related to increased Spanish flu mortality, but did not drive social class differentials, and high risk jobs existed across the social spectrum. Overall, the social factors that put populations at higher risk during the pandemic may also have been found in factors that determined existing health differences across the population, exacerbating existing social gradients in mortality.

6 Funding and acknowledgements

Ingrid van Dijk was funded through the "Landskrona Population Study", Swedish Foundation for the Humanities and Social Sciences (RJ).

We would like to thank Svenn-Erik Mamelund, Niels van den Berg, Mar Rodríguez-Girondo, and the participants to OsloMet's PANSOC webinar for their comments on earlier drafts of this paper.

References

Antonovsky, Aaron (1967). “Social class, life expectancy and overall mortality”. In: *The Milbank Memorial Fund Quarterly* 45.2, pp. 31–73.

Bellemare, Marc F. and Casey J. Wichman (2020). “Elasticities and the Inverse Hyperbolic Sine Transformation”. In: *Oxford Bulletin of Economics and Statistics* 82.1, pp. 50–61. DOI: 10.1111/obes.12325.

Bengtsson, Tommy, Martin Dribe, and Björn Eriksson (2018). “Social class and excess mortality in Sweden during the 1918 influenza pandemic”. In: *American journal of epidemiology* 187.12, pp. 2568–2576.

Bengtsson, Tommy and Frans Van Poppel (2011). “Socioeconomic inequalities in death from past to present: An introduction”. In: *Explorations in economic History* 48.3, pp. 343–356.

Boonstra, O (2020). “Historische Database Nederlandse Gemeenten, v3”. In: *IISH Data Collection*.

Brainerd, Elizabeth and Mark V Siegler (2003). *The Economic Effects of the 1918 Influenza Epidemic*. 3791, p. 41.

Brooshooft, P (1897). *Officieele en feitelijke waarheid: Bijdrage tot de kennis van onze arbeiderstoestanden*. ’s-Gravenhage: gebrs. Belinfante.

Brundage, John F and G Dennis Shanks (2008). “Deaths from bacterial pneumonia during 1918–19 influenza pandemic”. In: *Emerging Infectious Diseases* 14.8, p. 1193.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb (1988). “Alternative Transformations to Handle Extreme Values of the Dependent Variable”. In: *Journal of the American Statistical Association* 83.401, pp. 123–127. DOI: 10.1080/01621459.1988.10478575.

CBS (1917). *Statistiek van de sterfte onder de mannen van 18-65 jaar met onderscheiding naar beroep en de positie daarin bekleed in verband met leeftijden en doodsoorzaken in de jaren 1908-1911*. ’s-Gravenhage: gebrs. Belinfante.

— (1918). *Statistiek van de sterfte naar den leeftijd en naar de oorzaken van den dood over het jaar 1918*. ’s-Gravenhage: gebrs. Belinfante.

Clay, Karen, Joshua Lewis, and Edson Severnini (2018). “Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic”. In: *The Journal of Economic History* 78.4, pp. 1179–1209. DOI: 10.1017/S002205071800058X.

Clouston, Sean AP, Marcie S Rubin, Jo C Phelan, and Bruce G Link (2016). “A social history of disease: Contextualizing the rise and fall of social inequalities in cause-specific mortality”. In: *Demography* 53.5, pp. 1631–1656.

Colvin, Christopher L and Eoin McLaughlin (2021). “Death, demography and the denominator: Age-adjusted Influenza-18 mortality in Ireland”. In: *Economics & Human Biology* 41, p. 100984.

Crosby, Alfred W. (2003). *America's forgotten pandemic: the influenza of 1918*. 2nd ed. Cambridge ; New York: Cambridge University Press. 337 pp.
De Melker, RA (2005). "De huisarts en 'de Spaansche griep' 1918-1920". In: *Huisarts en Wetenschap* 48.13, pp. 44–49.
Debiasi, Enrico (2020). "The Historical Origins of the Mortality Gradient: Socioeconomic Inequalities in Adult Mortality over Two Centuries in Sweden". PhD thesis. Lund University.
Domela Nieuwenhuis, Ferdinand (1898). *Een vergeten hoofdstuk. Blanke slaven: Supplement op het historisch gedenkboek "Eene halve eeuw", 1848-1898*. Amsterdam: S.L. van Looy.
Edvinsson, Sören and Göran Broström (2012). "Old age, health, and social inequality: Exploring the social patterns of mortality in 19th century northern Sweden". In: *Demographic Research* 26, pp. 633–660.
Elo, Irma T (2009). "Social class differentials in health and mortality: Patterns and explanations in comparative perspective". In: *Annual review of sociology* 35, pp. 553–572.
Fanning, Patricia J (2010). *Influenza and inequality: One town's tragic response to the great epidemic of 1918*. University of Massachusetts Press.
Ferrie, Joseph P (2003). "The rich and the dead: Socioeconomic status and mortality in the United States, 1850-1860". In: *Health and labor force participation over the life cycle: Evidence from the past*. University of Chicago Press, pp. 11–50.
Fourie, Johan and Jonathan Jayes (2021). "Health inequality and the 1918 influenza in South Africa". In: *World Development* 141, p. 105407. DOI: 10.1016/j.worlddev.2021.105407.
Gagnon, Alain, Marc Tremblay, Helene Vezina, and Jamie A Seabrook (2011). "Once were farmers: Occupation, social mobility, and mortality during industrialization in Saguenay-Lac-Saint-Jean, Quebec 1840–1971". In: *Explorations in Economic History* 48.3, pp. 429–440.
Gooyer, Arie Cornelis de (1968). *De Spaanse griep van'18*. Philips-Duphar.
Grantz, Kyra H et al. (2016). "Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918". In: *Proceedings of the National Academy of Sciences* 113.48, pp. 13839–13844.
Herring, DA and E Korol (2012). "The north-south divide: Social inequality and mortality from the 1918 influenza pandemic in Hamilton, Ontario". In: *Epidemic encounters: Influenza, society, and culture in Canada*, pp. 97–112.
Human Mortality Database. Available at www.mortality.org or www.humanmortality.de (data downloaded on 20-07-2021.

Janssens, AAPO and Elien Van Dongen (2018). "A natural female disadvantage? Maternal mortality and nutrition related causes of death in the Netherlands, 1875-1899". In: *Tijdschrift voor Sociale en Economische Geschiedenis*.

Johnson, Niall P. A. S. and Juergen Mueller (2002). "Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic". In: *Bulletin of the History of Medicine* 76.1, pp. 105–115. DOI: 10.1353/bhm.2002.0022.

Kolata, Gina (2001). *Flu: The story of the great influenza pandemic of 1918 and the search for the virus that caused it*. New York: Simon & Schuster.

Koten, Jan-Willem and André Weel (2014). "Ziekte en dood binnen de gemobiliseerde nederlandse krijgsmacht 1914 - 1918". In: *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 22, pp. 275–279.

Lambert, Paul S., Richard L. Zijdeman, Marco HD Van Leeuwen, Ineke Maas, and Kenneth Prandy (2013). "The construction of HISCAM: A stratification scale based on social interactions for historical comparative research". In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46.2, pp. 77–89.

Landmacht, Geneeskundige Dienst der (1913). *Statistisch overzicht der behandelde zieken van het leger hier te lande in het jaar 1913*. 's-Gravenhage: Algemeene Landsdrukkerij.

Link, Bruce G and Jo Phelan (1995). "Social conditions as fundamental causes of disease". In: *Journal of health and social behavior*, pp. 80–94.

Loo, M. P. J. van der (2014). "The stringdist package for approximate string matching". In: *The R Journal* 6.1, pp. 111–122.

Luy, Marc (2003). "Causes of male excess mortality: Insights from cloistered populations". In: *Population and Development Review* 29.4, pp. 647–676.

Mamelund, Svenn-Erik (2006). "A socially neutral disease? Individual social class, household wealth and mortality from Spanish influenza in two socially contrasting parishes in Kristiania 1918–19". In: *Social science & medicine* 62.4, pp. 923–940.

— (2011). "Geography may explain adult mortality from the 1918–20 influenza pandemic". In: *Epidemics* 3.1, pp. 46–60.

Mamelund, Svenn-Erik and Jessica Dimka (2019). "Tuberculosis as a risk factor for 1918 influenza pandemic outcomes". In: *Tropical medicine and infectious disease* 4.2, p. 74.

Mamelund, Svenn-Erik, Clare Shelley-Egan, and Ole Rogeberg (2019). "The association between socioeconomic status and pandemic influenza: Protocol for a systematic review and meta-analysis". In: *Systematic reviews* 8.1, pp. 1–6.

Mandemakers, K et al. (2020). "HSN standardized, HISCO-coded and classified occupational titles, HSN release 2020.02". In:

Mourits, Rick J. et al. (2021). "Retracing Hotbeds of the 1918–19 Influenza Pandemic. Spatial Differences in Seasonal Excess Mortality in the Netherlands". In: *Historical Life Course Studies* 10.3, pp. 145–150.
Mourits, R.J. (2019). *Exceptional lives, extraordinary families. Familial clustering of longevity in the 19th and early-20th centuries*.
Mourits, R.J., I. Van Dijk, and K. Mandemakers (2020). "From matched certificates to related persons: Building a dataset from LINKS-Zeeland". In: *Historical Life Course Studies* 9, pp. 49–68.
Murray, Christopher JL, Alan D Lopez, Brian Chin, Dennis Feehan, and Kenneth H Hill (2006). "Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis". In: *The Lancet* 368.9554, pp. 2211–2218.
Quanjer, AAJ (1921). *De griep in Nederland in 1918 tot 1920*. 's-Gravenhage: Gezondheidsraad.
Razzell, Peter and Christine Spence (2006). "The hazards of wealth: Adult mortality in pre-twentieth-century England". In: *Social History of Medicine* 19.3, pp. 381–405.
Ringoir, Hendrik (1980). "Vredesgarnizoenen van 1715 tot 1795 en 1815 tot 1940". In: *Bijdragen Van De Sectie Militaire Geschiedenis* 8, pp. 1–136.
Schenk, Niels and Frans Van Poppel (2011). "Social class, social mobility and mortality in the Netherlands, 1850–2004". In: *Explorations in Economic History* 48.3, pp. 401–417.
Silva, J. M. C. Santos and Silvana Tenreyro (2006). "The Log of Gravity". In: *The Review of Economics and Statistics* 88.4, pp. 641–658. DOI: 10.1162/rest.88.4.641.
Smith, Ken R, Geraldine P Mineau, Gilda Garibotti, and Richard Kerber (2009). "Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002". In: *Social Science & Medicine* 68.9, pp. 1649–1658.
Sundin, Jan and Sam Willner (2007). *Social change and health in Sweden: 250 years of politics and practice*. Solna: Alfa Print.
Taubenberger, Jeffery K and David M Morens (2006). "1918 Influenza: the mother of all pandemics". In: *Revista Biomedica* 17.1, pp. 69–79.
Temby, Owen F and Ken R Smith (2014). "The association between adult mortality risk and family history of longevity: The moderating effects of socioeconomic status". In: *Journal of Biosocial Science* 46.6, pp. 703–716.
Tuckel, Peter, Sharon Sassler, Richard Maisel, and Andrew Leykam (2006). "The Diffusion of the Influenza Pandemic of 1918 in Hartford, Connecticut". In: *Social Science History* 30.2, pp. 167–196. DOI: 10.1017/S0145553200013432.
Van Der Woud, Auke (2010). *Koninkrijk vol sloppen: Achterbuurten en vuil in de negentiende eeuw*. Amsterdam: Bert Bakker.

Van Leeuwen, Marco H. D. and Ineke Maas (2011). *HISCLASS: a historical international social class scheme*. Leuven: Leuven University Press.
Van Leeuwen, Marco H. D., Ineke Maas, and A. Miles (2002). *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
Van Poppel, Frans, Roel Jennissen, and Kees Mandemakers (2009). "Time trends in social class mortality differentials in the Netherlands, 1820-1920: An assessment based on indirect estimation techniques". In: *Social Science History*, pp. 119–153.
Vaughan, Victor Clarence (1926). *A doctor's memories*. Indianapolis: Bobbs-Merrill Company.
Vugs, RM (2020). *In veel huizen wordt gerouwd: De Spaanse griep in Nederland*. Soesterberg: Aspekt.
Vulsma, Rudolf Ferdinand (1988). *Burgerlijke stand en bevolkingsregister*. 's-Gravenhage: Centraal Bureau voor Genealogie.
Wood, S. N. (2003). "Thin-plate regression splines". In: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114.
Zeileis, Achim, Susanne Köll, and Nathaniel Graham (2020). "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R". In: *Journal of Statistical Software* 95.1, pp. 1–36. DOI: 10.18637/jss.v095.i01.

Table 5: Example excess mortality dataset

EGG	indoor	strangers	sex	month	agegroup	baseline	flu	emr
71	1	1	m	10	50	0.14	0	0.0
97	0	1	m	11	50	0.29	0	0.0
27	0	0	m	12	60	0.14	0	0.0
38	0	1	m	12	60	0.57	2	3.5
30	0	1	m	12	30	0.14	1	7.0

A Data example

B Robustness checks

B.1 High and low mortality regions

	all	low EM	high EM
medium skilled	0.39^{***}	0.35^{***}	0.45^{**}
	(0.09)	(0.08)	(0.14)
lower skilled	0.21^{**}	0.20^{**}	0.23
	(0.08)	(0.07)	(0.13)
unskilled	0.56^{***}	0.38^{*}	0.73^{***}
	(0.12)	(0.15)	(0.16)
farmer	0.45^{***}	0.42^{**}	0.49^{***}
	(0.09)	(0.16)	(0.11)
indoors only	0.13	0.18	0.10
	(0.08)	(0.11)	(0.12)
contact only	0.18^{**}	0.18	0.19^{*}
	(0.06)	(0.09)	(0.09)
both	0.12	0.16	0.09
	(0.07)	(0.09)	(0.11)
age (30,45]	-0.25^{***}	-0.23^{**}	-0.25^{***}
	(0.05)	(0.07)	(0.07)
age (45,60]	-0.43^{***}	-0.36^{***}	-0.49^{***}
	(0.05)	(0.08)	(0.07)
age (60,80]	-0.47^{***}	-0.34^{***}	-0.57^{***}
	(0.06)	(0.09)	(0.08)
male	0.37^{***}	0.31^{***}	0.44^{***}
	(0.06)	(0.06)	(0.08)
event_month10	0.39^{***}	0.29^{***}	0.47^{***}
	(0.05)	(0.07)	(0.06)
event_month11	0.78^{***}	0.67^{***}	0.88^{***}
	(0.06)	(0.08)	(0.08)
event_month12	0.27^{***}	0.29^{***}	0.25^{***}
	(0.04)	(0.05)	(0.06)
(Intercept)	-0.79^{***}	-0.20	-0.91^{***}
	(0.19)	(0.24)	(0.22)
R^2	0.23	0.20	0.26
Adj. R^2	0.22	0.19	0.24
Num. obs.	3573	1659	1914

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Regression models of log excess mortality rate for low and high excess mortality regions. Region-clustered standard errors between parentheses.

B.2 Different aggregation levels

	municipalities	EGG	COROP	Province
medium skilled	0.23***	0.39***	0.55***	0.42***
	(0.05)	(0.09)	(0.12)	(0.10)
lower skilled	0.15***	0.21**	0.32**	0.24*
	(0.04)	(0.08)	(0.10)	(0.12)
unskilled	0.34***	0.56***	0.62***	0.47**
	(0.07)	(0.12)	(0.16)	(0.16)
farmer	0.44***	0.45***	0.36**	0.23
	(0.07)	(0.09)	(0.12)	(0.13)
indoors only	0.14*	0.13	0.17	0.04
	(0.06)	(0.08)	(0.10)	(0.09)
contact only	0.16***	0.18**	0.17	0.06
	(0.05)	(0.06)	(0.09)	(0.12)
both	0.10*	0.12	0.14	0.11
	(0.05)	(0.07)	(0.10)	(0.09)
age (30,45]	−0.16***	−0.25***	−0.24***	−0.27***
	(0.05)	(0.05)	(0.05)	(0.06)
age (45,60]	−0.26***	−0.43***	−0.56***	−0.67***
	(0.04)	(0.05)	(0.04)	(0.05)
age (60,80]	−0.24***	−0.47***	−0.65***	−0.77***
	(0.05)	(0.06)	(0.06)	(0.07)
male	0.22***	0.37***	0.48***	0.29***
	(0.04)	(0.06)	(0.07)	(0.05)
event_month10	0.24***	0.39***	0.47***	0.52***
	(0.03)	(0.05)	(0.07)	(0.08)
event_month11	0.51***	0.78***	0.95***	1.02***
	(0.05)	(0.06)	(0.09)	(0.12)
event_month12	0.16***	0.27***	0.31***	0.35***
	(0.03)	(0.04)	(0.06)	(0.10)
(Intercept)	−0.55***	−0.79***	−0.33	0.37**
	(0.11)	(0.19)	(0.21)	(0.14)
R^2	0.19	0.23	0.26	0.29
Adj. R^2	0.16	0.22	0.25	0.28
Num. obs.	5169	3573	2385	1261

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Regression models of log excess mortality rate at different levels of aggregation. Region-clustered standard errors between parentheses.

B.3 Population density and army controls

	municipalities	*EGG*	COROP	Province
medium skilled	0.22^{***}	0.38^{***}	0.52^{***}	0.37^{**}
	(0.05)	(0.09)	(0.13)	(0.12)
lower skilled	0.13^{**}	0.19^{*}	0.32^{**}	0.22
	(0.04)	(0.08)	(0.11)	(0.12)
unskilled	0.34^{***}	0.55^{***}	0.59^{***}	0.44^{**}
	(0.07)	(0.13)	(0.17)	(0.16)
farmer	0.45^{***}	0.48^{***}	0.36^{**}	0.26^{*}
	(0.06)	(0.09)	(0.13)	(0.11)
indoors only	0.12^{*}	0.14	0.17	0.05
	(0.06)	(0.08)	(0.10)	(0.09)
contact only	0.16^{***}	0.20^{**}	0.17	0.07
	(0.05)	(0.07)	(0.09)	(0.10)
both	0.08	0.12	0.14	0.12
	(0.05)	(0.08)	(0.10)	(0.09)
age (30,45]	-0.16^{***}	-0.25^{***}	-0.24^{***}	-0.26^{***}
	(0.04)	(0.05)	(0.05)	(0.06)
age (45,60]	-0.26^{***}	-0.44^{***}	-0.58^{***}	-0.66^{***}
	(0.04)	(0.05)	(0.04)	(0.05)
age (60,80]	-0.25^{***}	-0.48^{***}	-0.68^{***}	-0.79^{***}
	(0.05)	(0.06)	(0.06)	(0.06)
male	0.18^{***}	0.33^{***}	0.45^{***}	0.29^{***}
	(0.05)	(0.06)	(0.08)	(0.05)
event_month10	0.24^{***}	0.39^{***}	0.47^{***}	0.51^{***}
	(0.03)	(0.05)	(0.07)	(0.09)
event_month11	0.50^{***}	0.77^{***}	0.94^{***}	1.01^{***}
	(0.04)	(0.06)	(0.09)	(0.12)
event_month12	0.16^{***}	0.26^{***}	0.31^{***}	0.35^{***}
	(0.03)	(0.04)	(0.06)	(0.10)
log Population '18	0.16^{***}	0.23^{***}	0.07	0.03
	(0.02)	(0.04)	(0.09)	(0.05)
(Intercept)	-1.73^{***}	-2.77^{***}	-0.98	-0.04
	(0.19)	(0.55)	(1.15)	(0.69)
R^2	0.14	0.19	0.22	0.26
Adj. R^2	0.14	0.18	0.22	0.25
Num. obs.	5169	3573	2385	1261

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: Regression models of log excess mortality rate including population density controls and no region FE. Region-clustered standard errors between parentheses.

B.4 Alternative zero handling

	log x+1	drop 0	no log	poiss EMR	poiss N	asinh
medium skilled	0.39***	−0.56***	1.18*	0.88*	0.63	0.50***
	(0.09)	(0.07)	(0.48)	(0.39)	(0.35)	(0.11)
lower skilled	0.21**	−0.43***	0.48	0.61	0.44	0.27**
	(0.08)	(0.11)	(0.48)	(0.37)	(0.34)	(0.09)
unskilled	0.56***	−0.71***	1.97**	1.19**	0.86*	0.72***
	(0.12)	(0.09)	(0.60)	(0.45)	(0.40)	(0.15)
farmer	0.45***	−0.49***	1.82***	0.65***	0.54***	0.57***
	(0.09)	(0.10)	(0.49)	(0.18)	(0.13)	(0.12)
indoors only	0.13	−0.35**	0.24	0.13	0.17	0.17
	(0.08)	(0.11)	(0.36)	(0.18)	(0.13)	(0.10)
contact only	0.18**	−0.24*	0.74	0.29*	0.22	0.23**
	(0.06)	(0.11)	(0.38)	(0.14)	(0.12)	(0.08)
both	0.12	−0.37**	0.17	0.07	0.06	0.15
	(0.07)	(0.11)	(0.39)	(0.18)	(0.15)	(0.09)
age (30,45]	−0.25***	−0.08	−1.68***	−0.41***	−0.41***	−0.31***
	(0.05)	(0.04)	(0.31)	(0.06)	(0.06)	(0.06)
age (45,60]	−0.43***	−0.54***	−3.09***	−0.98***	−1.06***	−0.52***
	(0.05)	(0.06)	(0.45)	(0.08)	(0.07)	(0.06)
age (60,80]	−0.47***	−0.80***	−3.50***	−1.23***	−1.36***	−0.57***
	(0.06)	(0.07)	(0.53)	(0.10)	(0.11)	(0.07)
male	0.37***	−0.27*	1.08	0.43	0.26	0.47***
	(0.06)	(0.11)	(0.70)	(0.30)	(0.30)	(0.07)
event_month10	0.39***	0.36***	1.77***	1.15***	1.00***	0.49***
	(0.05)	(0.09)	(0.35)	(0.11)	(0.15)	(0.06)
event_month11	0.78***	0.69***	4.52***	1.90***	1.74***	0.97***
	(0.06)	(0.09)	(0.61)	(0.15)	(0.18)	(0.07)
event_month12	0.27***	0.22**	1.04***	0.81***	0.71***	0.34***
	(0.04)	(0.07)	(0.17)	(0.14)	(0.13)	(0.05)
(Intercept)	−0.79***	3.28***	−1.17	−2.40***	−1.50*	−1.02***
	(0.19)	(0.21)	(1.29)	(0.68)	(0.61)	(0.23)
R^2	0.23	0.34	0.18			0.23
Adj. R^2	0.22	0.31	0.17			0.22
Num. obs.	3573	1236	3573	3573	3573	3573
AIC						
BIC						
Log Likelihood						
Deviance				17681.43	4195.89	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Alternative model forms for regressions of log excess mortality rate. Region-clustered standard errors between parentheses.

C Occupations

Below we tabulate the most frequent occupations in the four exposure categories.

Table 10: Most frequent occupations on death certificates for deceased age 10-70, september-december 1910–1918.

HISCO	emr	occtitle	N	strangers	indoors	both	neither	skill	hisca
61220	3.5	landbouwer	2596	0	0	0	2596	medium	[illegible]
99900	4.9	arbeider	2034	2034	0	0	0	un	[illegible]
99930	4.0	fabrieksarbeider	559	559	559	559	0	un	[illegible]
41025	3.4	koopman	484	484	484	484	0	medium	[illegible]
54020	5.0	dienstbode	382	382	382	382	0	un	[illegible]
62210	4.4	veldarbeider	364	0	0	0	364	un	[illegible]
95410	2.8	timmerman	364	364	0	0	0	lower	[illegible]
41030	2.3	winkelier	247	247	247	247	0	medium	[illegible]
99920	3.0	dagloner	212	212	0	0	0	un	[illegible]
79100	3.6	kleermaker	194	194	194	194	0	medium	[illegible]
93120	2.7	schilder	190	190	0	0	0	lower	[illegible]
83110	3.2	smid	170	0	170	0	0	medium	[illegible]
80110	3.1	schoenmaker	164	164	164	164	0	medium	[illegible]
95120	2.4	metselaar	151	151	0	0	0	medium	[illegible]
39310	8.9	kantoorbediende	131	0	131	0	0	lower	[illegible]
79510	2.8	naaister	130	0	130	0	0	lower	[illegible]
98620	2.9	voerman	127	127	0	0	0	lower	[illegible]
77610	5.6	bakker	116	0	116	0	0	medium	[illegible]
78200	3.2	sigarenmaker	116	0	116	0	0	lower	[illegible]
4217	4.0	schipper	106	NA	NA	NA	NA	medium	[illegible]
64100	8.7	visscher	105	0	0	0	105	medium	[illegible]
37040	2.5	loopknecht	92	92	0	0	0	lower	[illegible]
62105	6.1	boerenknecht	92	0	0	0	92	un	[illegible]
51050	1.5	cafehouder	90	90	90	90	0	medium	[illegible]
81990	3.8	klompenmaker	88	0	88	0	0	lower	[illegible]
81120	2.5	meubelmaker	87	0	87	0	0	medium	[illegible]
61270	3.1	bloemist	84	0	0	0	84	medium	[illegible]
62740	2.2	tuinman	73	0	0	0	73	un	[illegible]
45130	7.0	winkelbediende	70	70	70	70	0	lower	[illegible]
13000	5.3	onderwijzer	68	68	68	68	0	higher	[illegible]