



Unequal excess mortality during the Spanish Flu pandemic in the Netherlands

Auke Rijpma^{a,*}, Ingrid K. van Dijk^{b,c}, Ruben Schalk^a, Richard L. Zijdeman^{d,e}, Rick J. Mourits^{c,d}

^a Utrecht University, The Netherlands

^b Lund University, Sweden

^c Radboud University, The Netherlands

^d International Institute of Social History, The Netherlands

^e University of Stirling, United Kingdom

ARTICLE INFO

JEL classification:

N34

I14

Keywords:

Excess mortality

1918-9 influenza pandemic

Spanish flu

Socioeconomic health inequality

Occupational health risk

ABSTRACT

A century after the Spanish Flu, the COVID-19 pandemic has brought renewed attention to socioeconomic and occupational differences in mortality in the earlier pandemic. The magnitude of these differences and the pathways between occupation and increased mortality remain unclear, however. In this paper, we explore the relation between occupational characteristics and excess mortality among men during the Spanish Flu pandemic in the Netherlands. By creating a new occupational coding for exposure to disease at work, we separate social status and occupational conditions for viral transmission. We use a new data set based on men's death certificates to calculate excess mortality rates by region, age group, 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 among men in the Netherlands in the autumn of 1918. We find some evidence that men with occupations that featured high levels of social contact had higher mortality in this period. Above all, however, we find a strong socioeconomic gradient to excess mortality among men during the Spanish Flu pandemic, even after accounting for exposure in the workplace.

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, estimates of contemporaries suggest that 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 to the disease 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 Spanish Flu mortality (Mamelund, 2006; Mamelund and Dimka, 2021; Clay et al., 2019). Yet, uncertainty remains, a large part of the literature is based on aggregate data, and the mechanisms that linked together socioeconomic factors and mortality risk to the Spanish Flu remain unclear. In this paper, we use new micro-level data containing occupations and dates of death for five provinces in the Netherlands to further establish whether and how mortality risk during 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 (Mamelund and Dimka, 2019; Janssens and Van Dongen, 2018). Third, access to health care 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). However, these studies rest on analyses of whole populations, cities or census districts, while evidence from individual-level studies remains scarce. An exception is work by Bengtsson et al. (2018) who suggested a fourth mechanism: differences in the degree to which people had interpersonal interactions in their daily lives, that may have contributed to the social gradient in pandemic mortality.

In this work, we use micro-level data from men's death certificates to show that among men, a social gradient existed in excess mortality during the second and most severe wave of Spanish Flu in the Netherlands. We analyse socioeconomic and occupational factors which may

* Corresponding author.

E-mail address: a.rijpma@uu.nl (A. Rijpma).

<https://doi.org/10.1016/j.ehb.2022.101179>

Received 24 March 2022; Received in revised form 19 July 2022; Accepted 23 August 2022

Available online 13 September 2022

1570-677X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

have contributed to risk of infection with and risk of mortality during the Spanish flu. Particularly, we are interested in the social gradient in mortality while taking into account occupation-related risk of infection. Such a social gradient may point to the relevance of resources and existing health differences in shaping mortality 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 Gooyer, 1968](#); [De Melker, 2005](#); [Quanjer, 1921](#); [Vugs, 2020](#); [Mourits et al., 2021](#)). We focus on the second wave of the Spanish Flu, in autumn 1918, which was by far the deadliest phase of the pandemic. We use occupations registered on death certificates to measure socioeconomic status of the deceased and create a new occupational classification whether work led to frequent social contact and whether it took place in enclosed spaces. We use this to isolate socioeconomic status from occupational exposure to viral transmission.

We find evidence for a social gradient in mortality among men both in strongly and more mildly affected places. Our findings suggest that part of the socioeconomic patterns in Spanish flu mortality identified by previous studies exist independently from occupational differences in the frequency of social contact with others and conditions for viral transmission at work. Other factors related to socioeconomic status are thus likely to have been important and may have included the opportunity to isolate and viral transmission away from work, and existing health differences in the population.

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 young adults between ages 20–40, who are normally least affected by infectious diseases. Especially young men and pregnant women were affected. The infected initially showed common symptoms of the flu, which could quickly develop into pneumonia. Death often followed within days. The 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” ([Vaughan, 1926](#); [Kolata, 2001](#)).

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 et al., 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”.¹ The idea of the Spanish Flu as a socially neutral disease persisted in 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 et al., 2019](#)).

2.2. Social differences in early 20th-century mortality

The image of Spanish Flu 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](#); [Murray et al., 2006](#); [Mamelund et al., 2019](#)), even though there are recent studies arguing that Spanish Flu was unrelated to pre-pandemic economic characteristics ([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 et al., 2018](#); [Mamelund, 2006](#)). Using micro-level data from the Swedish census, ([Bengtsson et al., 2018](#)) show that this disadvantage of the lower classes was especially notable among men and less pronounced among women.

The discussion about the role of socioeconomic inequalities 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 industrialised and urbanised 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 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 et al. \(2009\)](#) 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 role in excess mortality during the Spanish Flu 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 provide health advantages regardless of 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 Spanish Flu, socioeconomic differences in disease burden or mortality are still expected. For example, resources can help people isolate from sick relatives or co-workers, improve access to information and understanding of the appropriate behaviour to avoid exposure, and ease access to health care.

The topic of the Spanish Flu has seen renewed interest by historians and social scientists in recent years ([Boberg-Fazlic et al., 2021](#); [Colvin and McLaughlin, 2021](#); [Devos et al., 2022](#); [Gallardo-Albarrán and de Zwart, 2021](#)), though few studies provide in-depth analyses of socioeconomic differences in mortality. Evidence for socioeconomic contrasts in Spanish Flu mortality comes from a number of studies, predominantly at the aggregate level and showing correlations between local conditions and excess 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, often of low socioeconomic status, were harder hit by the Spanish Flu pandemic than non-immigrants. Similarly, [Clay et al. \(2018\)](#) find that the local share of foreign-born immigrants and pre-pandemic typhoid rates predict higher all-cause mortality during the pandemic. [Clay et al. \(2019\)](#) also find that high illiteracy rates and pre-pandemic infant mortality predict higher mortality during the Spanish Flu. Studies on socioeconomic gradient in Spanish Flu mortality based on individual-level data remain scarce, however.

¹ “[...] niet tot eenige voorkeur, noch voor beroep, noch voor meerdere of minder goedheid besluiten”, p. 24.

Among the factors that may have contributed to higher risk among lower socioeconomic status groups, higher exposure to disease figures prominently.² 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 chance of exposure to the Spanish flu, especially in urban slums. Specific factors for socioeconomic differences in Spanish Flu 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 chance of a secondary infection with bacterial pneumonia as well as reduced access to health care (Grantz et al., 2016; Mamelund, 2006). Not surprisingly, then, working class neighbourhoods and areas with small housing had higher death rates during the Spanish Flu in Oslo, Norway (Mamelund, 2006).

2.3. Occupational exposure to the Spanish flu

Exposure to disease does not only occur at home but also at the workplace. Evidence for a relation between occupational characteristics and Spanish flu mortality largely stems from macro-level studies. For Chicago it was found that local 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 recognised 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 tuberculosis-related mortality than those who were subjected to organic dust from tobacco, flower, and textiles, 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 susceptible to tuberculosis (Sundin and Willner, 2007).

In a study of the Swedish case using individual level data, Bengtsson et al. (2018) provide evidence that the ability to maintain distance from others (now commonly known as “social distancing”) may have affected mortality rates favourably. In southern Sweden, farmers were least affected by the Spanish flu, and low-skilled and unskilled workers the most. Although this points towards the importance of solitary work, the protective effect of farming was only found for men, and white collar workers had no mortality advantage compared to the working class. To understand whether social contact affected Spanish Flu 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 (Debiassi, 2020). Recent studies have highlighted substantial heterogeneity in mortality rates within social classes but between occupational groups (Debiassi, 2020). The decades

² There are also studies that focus on vulnerability to infection. One way through which vulnerability could lead to higher mortality during the Spanish Flu was through air pollution, to which the poor were more exposed. Clay et al. (2018) show that pollution due to coal-fired electricity generation worsened the impact of Spanish Flu in US cities, as the air pollution probably made lungs more susceptible to infection. Similarly, Brundage and Shanks (2008) argue that secondary infections, rather than the virulence of Spanish flu itself, played a key role in the high death rates and age profile of the deceased as well as differences between occupational groups.

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 (Debiassi, 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 men’s occupational characteristics and mortality during the Spanish Flu pandemic. We overcome some of the shortcomings of earlier studies by studying individual-level data for the entire population of part of the Netherlands, including social and occupational characteristics. Our primary variable of interest is socio-economic status. However, instead of analysing specific occupations, as has been done in earlier studies to all-cause and infectious disease mortality, we also measure occupational risk factors for broad categories of occupations. 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 to calculate excess mortality in 1918 in comparison to deaths in the period 1910–1917. 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 et al., 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.³

These individual-level records allow us to analyse mortality during the Spanish Influenza pandemic in more detail than before. 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 Mourits et al., 2016 or CBS, 1918). The aggregated number of deaths per municipality is available by age group, but without information on occupation or other indicators of socioeconomic status. Some records of individual-level causes of death data survive in the Netherlands, but they are only available for isolated periods and municipalities. Death certificates are thus the only remaining source that combines individual-level characteristics with sufficient nationwide coverage to generate meaningful insights into mortality patterns along socioeconomic lines.

Dutch death certificates contained no cause-specific mortality information. However, total numbers of deaths from death registration are sometimes deemed the most accurate way to address the impact of an

³ 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>, the dataset at <https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/PCAEGG>, and further reproduction scripts at <https://github.com/rijpma/spanish>.

Table 1

Summary statistics and selection steps. Source: Openarch Death Certificates. The table shows the stepwise selection of cases for analysis. The database contains a total of 741 758 death certificates. We select only death certificates from municipalities fully available in the database (715 703 cases) and where there is sufficient coverage for our variables (220 795 cases). We also drop certain periods, age groups, and women (16 786), and finally select complete cases (11 469).

	Selection	Certificates	Municip.	Mean age	% unskilled	% contact
1	Start	741 758	1116	43	32	27
2	Certificate coverage	715 703	946	43	33	27
3	Variable coverage	220 795	215	42	30	26
4	Drop 1914	198 557	215	42	30	26
5	Sep-Dec	64 365	215	41	31	27
6	15 < age < 79	34 111	215	54	31	27
7	Men only	16 786	215	54	30	29
8	Complete cases	11 469	215	51	30	29

epidemic. Shifts in overall mortality rates do not depend on potentially wrong or missing diagnoses of the cause of death, includes individuals who had multiple diseases at the time of death, and includes indirect deaths from the Spanish Flu (Colvin and McLaughlin, 2021).

Death certificates have been digitised in all Dutch provinces, but not for all municipalities (Fig. 1, top left panel): the digitised part covers 85 per cent of the 1118 municipalities existing in 1918. A large part of the missing death certificates are from the city of Amsterdam, where digitisation has not yet begun.

Some certificates were entered by more than one archive, or were digitised multiple times by the same archive, resulting in duplicates. These have been removed from the data.⁴

Certificates were digitised on a per-archive basis, and therefore there is variation in what information from the certificates was digitised (Fig. 1), and we have to drop further municipalities for this reason. We can see clear bimodal distributions of missing data, with municipalities clustering at low or high percentages. We therefore exclude municipalities where the share of reported ages, sex, and full dates is below 50%.⁵ In the case of occupations, many of the deceased would not be expected to have an occupation, especially in the case of children, the elderly, and women. Because nearly half the deaths are among women, and the very young and the elderly dominate among those who died in the years 1910–1918, the local share of recorded occupations can be low. We exclude municipalities where the share of recorded occupations is below 10%. We explore the impact of using stricter thresholds in Appendix A.

The changes in the sample due to excluding further municipalities are shown in Table 1, showing that it leaves us with 220 795 certificates for 215 municipalities, concentrated in the provinces of Gelderland, Overijssel, Zeeland, Drenthe, and Zuid-Holland. The most important bias is that three of the four largest cities are missing from the data: Amsterdam, Rotterdam, and Utrecht. However, our data does contain large and medium-sized cities (The Hague, Arnhem, Nijmegen, Leiden, and Apeldoorn), including a number of textile and industrial centres. Because of this and because very small municipalities are underrepresented, the average size of a municipality in our final dataset is 7850 inhabitants, compared to 6015 for the country as a whole. Comparing the average excess mortality rate (EMR) of municipalities that were dropped from our sample due to missing variables (excluding Amsterdam as there is no data available), we find that the difference is negligible: 1.38 outside our sample, and 1.40 in our sample.

⁴ 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 calculated using the ‘stringdist’ library (Van Der Loo, 2014)). Although some remaining duplicates cannot be excluded, earlier work shows that excess mortality estimates are not affected (Mourits et al., 2021).

⁵ 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 very low rates of missingness.

Table 1 also shows the results of further selection steps, and how each step affects the composition of the included sample. As explained below, we calculate excess mortality based on the months September–December, and exclude the year 1914. We only keep certificates where the deceased is aged 16–78.

We only include men in the analyses. Female occupations were not always recorded due to norms about female labour force participation (especially in marriage records Walhout and van Poppel, 2003). Consequently, for every 11 men with a recorded occupation, we only have information about the occupation of one woman. Moreover, recorded occupations for women fall in a small number of categories, such as maids, seamstresses, and nurses. The high percentage of maids in particular makes the women in our data less skilled than the male population. In part, these are real effects, as labour market opportunities for women were limited, especially in more skilled and high-status occupations. At the same time, we believe the data contain insufficient detail about women’s occupations to include them in the current work. We therefore omit women from our analyses. As we discuss below, our results are robust to including women.

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).⁶ 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 et al., 2002). Almost 98 per cent of occupational titles on the death certificates could be coded this way.⁷ 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). We decided to aggregate to these skill level groupings to avoid ending up with very small totals in some of the HISCLASS categories, the higher skilled occupations in particular.

As an alternative to the HISCLASS system, we also use HISCAM occupational classification scores (Lambert et al., 2013). This is a continuous measure that infers distance between occupations in an occupational hierarchy from the social interaction of the occupations incumbents, specifically historical marriage behaviour. The advantage of HISCAM is that it is based on actual behaviour, indirectly accounting for many implicit social factors that affect 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 in HISCAM they are viewed as a relatively low status profession. Both arguments are valid insofar farmers had to run a complicated business,

⁶ <https://iisg.amsterdam/en/hsn/data/occupations>.

⁷ Some minor adjustments to the HSN-HISCO scheme were made. See <https://github.com/rijpma/spanish>.

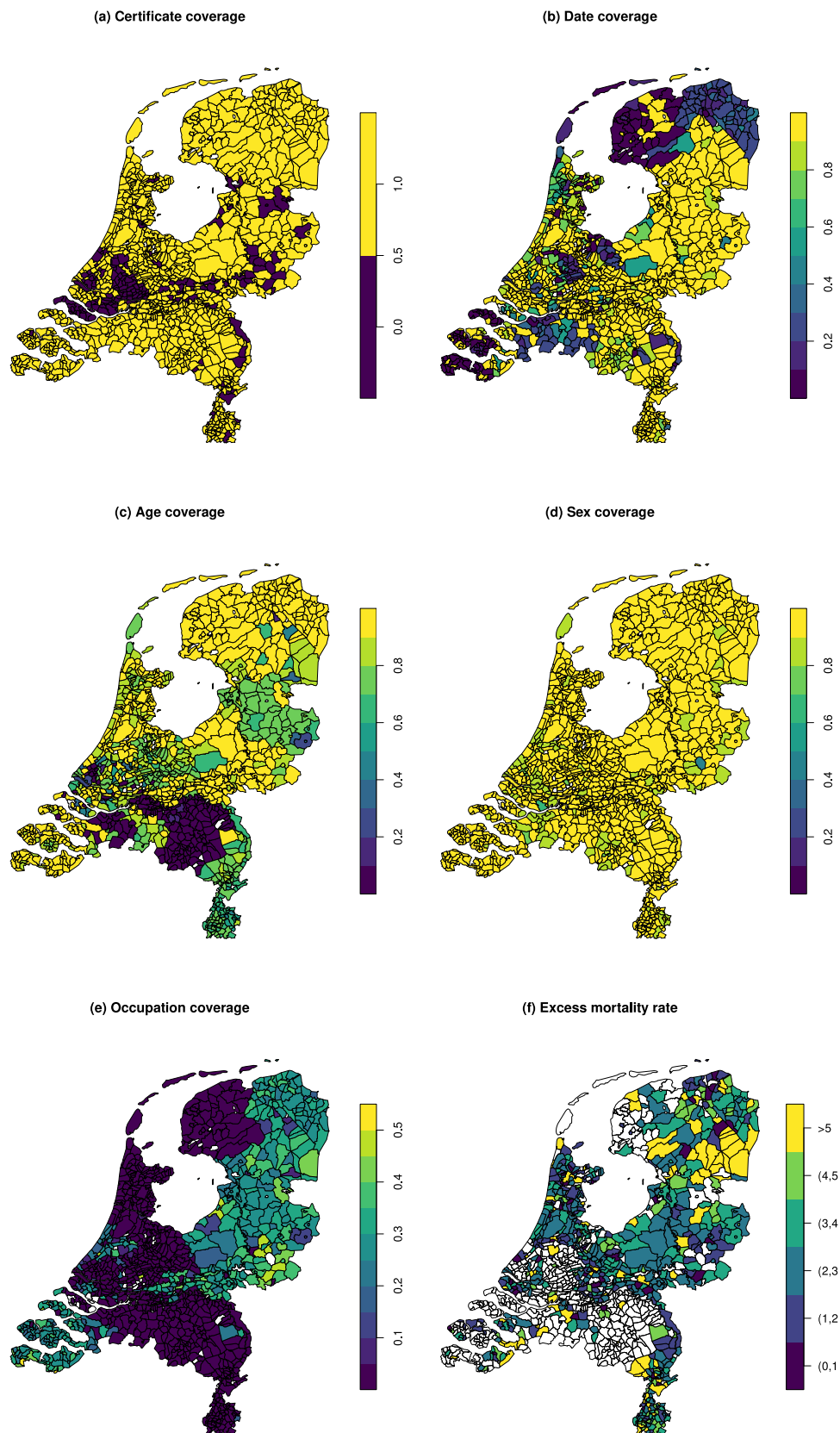


Fig. 1. Maps of geographic coverage of certificates of death (panel a), certificate dates (panel b), coverage of individuals' age (panel c), coverage of sex (panel d), and occupations (panel e) on death certificates. Panel f shows municipal excess mortality rates during the Spanish Flu pandemic (September–December 1918). Source: Openarch Death Certificates. The maps represent the completeness of digitisation of death certificates (panel a) and specific information from certificates (panel b-e). Differences between municipalities exist due to differences in digitisation procedures.

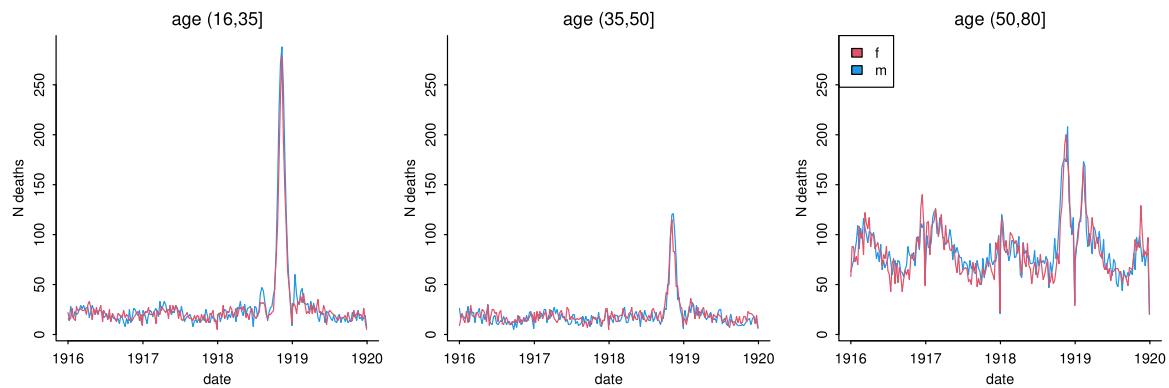


Fig. 2. Weekly number of deaths by age and sex in the Netherlands from death certificates, 1916–1919.
Source: Openarch Death Certificates.

but the certificates we rely on do not always distinguish between the operators of large farms, smallholders and farm labourers.⁸

To analyse exposure to infectious disease in the workplace, we used the HISCO classification to create an occupational indicator for exposure. Comparing occupations with higher and lower excess mortality could lead to ad-hoc explanations, whereas creating this new indicator allows us to systematically account for characteristics in occupations that lead to risk of exposure. Moreover, like the skill levels derived from HISCLASS above, this scheme avoids very low regional counts for certain occupations.

Exposure to infectious disease due to working conditions was 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 with strangers. A tailor also worked indoors 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 at least two researchers and then compared. Approximately 75 per cent of occupations were coded identically in the first round, and the remaining occupational groups were discussed by the coders until an unanimous decision was reached. The remaining differences were often due to differing interpretation of the underlying occupations in the HISCO unit or the actual tasks performed in an historic occupation, which could often be solved by consulting the documentation of HISCO.⁹ Although our scoring omits regional and within-occupation variance in occupational characteristics, it captures perceived conditions for exposure to and transmission of the virus on the work floor. Moreover, the scores measure something different from the skill levels we also use in our analyses. Correlations between these dummy variables range from -0.27 to $+0.32$ (see Table 13 in Appendix B).

3.3. Excess mortality

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.¹⁰ This approach is used because

⁸ The Dutch term *landbouwer* which is frequently used on the certificates is not specific enough to make this distinction.

⁹ Appendix B provides our coding of the most frequent occupations in the dataset. The full coding scheme is described in Schalk et al. (2022), and the data can be found at <https://hdl.handle.net/10622/TFG1DQ>.

¹⁰ Unless stated otherwise, excess mortality in the results and conclusion section is used to refer to excess all-cause mortality during the second Spanish Flu wave.

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 December 31st in 1918 by age group, municipality, and occupational cluster to mortality in the preceding years.¹¹ Baseline mortality was established by calculating the average mortality of the same occupational groups (by age group) in autumn 1910–1917 (September through December). The Netherlands remained neutral during the First World War, so dramatic changes to the population at risk and mortality rates that characterised many other countries at the time are not expected (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 make sure that our estimates of excess mortality are not biased by the events of the First World War in the Netherlands.

Our approach relies on the assumption that the population at risk is relatively stable in the period leading up to the Spanish Flu. Fig. 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, Fig. 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.

We also use external demographic data to investigate the possibility of strong demographic shifts in the Netherlands in the 1910–17 period. The Historical Database Dutch Municipalities shows that across all municipalities, the coefficient of variation in population was on average only 0.04 (Mourits et al., 2016). Total Dutch population figures from the Human Mortality Database show that the coefficient of variation in population by 10-year age groups averaged to, again, 0.04 (Human Mortality Database, 0000, July 2021 version). 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.

In Table 2, we show excess mortality rates by occupational skill level. Less skilled people had higher excess mortality. We now turn to analysing these differences further.

3.4. Analysis

Excess mortality is modelled using OLS-regressions with the ratio of deaths in September–December 1918 over the average number deaths

¹¹ We estimate excess mortality rates for autumn 1918, as it was by far the most deadly wave of the Spanish Flu pandemic and had a distinct geographic pattern. Moreover, to estimate the local EMRs reliably high levels of excess mortality are required, which were not present in the summer of 1918 and winter of 1919, as seen in Fig. 2 and discussed in Qvanjer (1921) and Mourits et al. (2021).

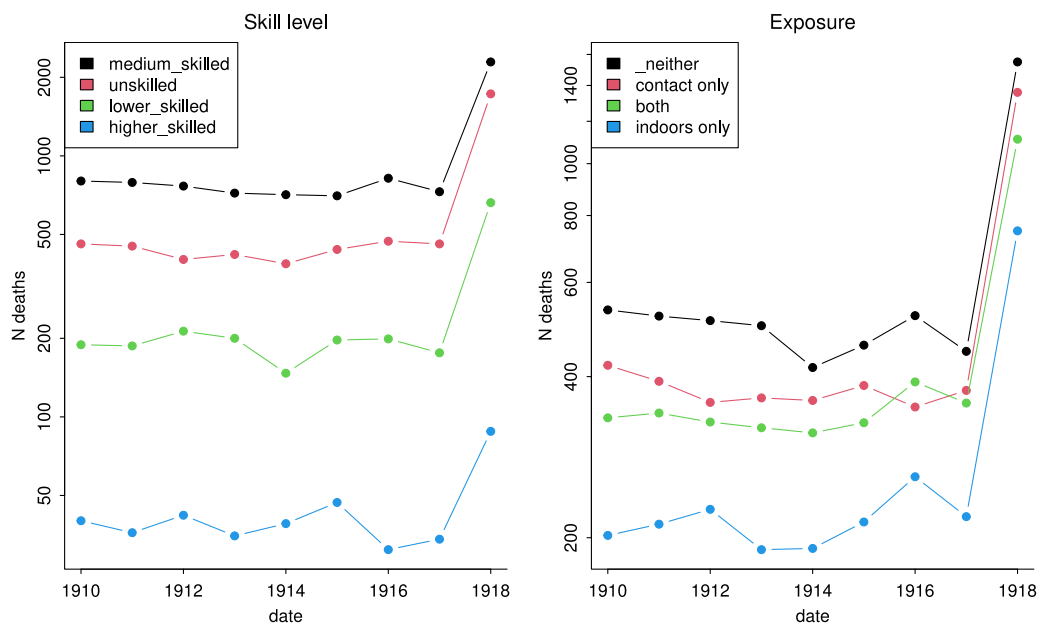


Fig. 3. Line graph of annual deaths in the months September–December, 1910–1918, broken down by occupational skill level (left panel) and exposure (right panel). Number of deaths is on a semilogarithmic scale.

Source: Openarch Death Certificates.

Table 2

Excess mortality rates in September–December 1918 by occupational skill group and their differences, with bootstrapped standard errors between parentheses. Difference refers to the difference between the EMR in that row and the previous.

Source: Openarch Death Certificates.

	Group	Excess mort.	Difference
1	Higher Skilled	2.23 (0.31)	–
2	Medium Skilled	3.04 (0.08)	0.81 (0.31)
3	Lower Skilled	3.38 (0.18)	0.34 (0.19)
4	Unskilled	3.53 (0.13)	0.15 (0.22)

in these months for the period 1910–17 (excluding 1914) as the dependent variable. The distribution of these ratios is highly skewed, so we take the logarithm of the dependent variable.¹² However, as we calculate excess mortality per municipality by 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 important. To include these observations, we add one to the dependent variable. This means we estimate excess mortality in region i , month t , occupational group j , and age group a as follows:

$$\log \left(\frac{deaths_{itja}^{year=1918}}{baseline\ deaths_{itja}} + 1 \right) = \beta_0 + \sum_j \beta_{1,j} skill_j + \sum_j \beta_{2,j} contact_j + \beta_3 farmer_j + \sum_a \beta_{4,a} age_a + \delta_i + \gamma_t + \epsilon_{itja} \quad (1)$$

Standard errors are clustered by region (Zeileis et al., 2020). For regions we use Economic Geographic Regions (EGG): clusters of municipalities defined by Statistics Netherlands below the NUTS-3 level.¹³ We prefer to work at this relatively low level of aggregation to be able to control for the spatial patterns of the pandemic and differences between

¹² 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.

¹³ <https://ec.europa.eu/eurostat/web/nuts/background>.

urban and rural regions which the EGG regions capture, but NUTS-3 region in the Netherlands do not.

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 there are fewer regions without deaths in 1918, and, more generally, will increase precision 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, COROP¹⁴ (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 Flu pandemic, and again, exclude many cases of zero deaths. Additionally, this provides a check for the importance of high-mortality outliers. 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 zeros in our data. We also check whether population density affect our estimates due to easier transmissibility in cities compared to the countryside. We also check whether the 1914–1918 mobilisation of troops affects our results by controlling for the presence of army bases. We also include alternative specifications for our occupational exposure measures and estimate the models including women. Finally, we estimate our models again but using stricter thresholds in the data selection process.

4. Results

Table 3 shows the results of this model, with progressively 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 men had 55 percent higher mortality than they would experience in a normal year ($\exp(0.44) - 1$). For unskilled

¹⁴ Artificial regions defined for statistical analyses by Statistics Netherlands, named after the Coördinatie Commissie Regionaal Onderzoeksprogramma.

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

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Medium skilled	0.44*** (0.09)	0.35*** (0.08)		0.35*** (0.10)	0.30** (0.10)	0.31** (0.10)	0.38*** (0.09)
Lower skilled	0.25*** (0.06)	0.25*** (0.06)		0.24** (0.08)	0.17* (0.08)	0.18* (0.08)	0.21** (0.08)
Unskilled	0.52*** (0.12)	0.52*** (0.12)		0.54** (0.17)	0.45** (0.16)	0.46** (0.16)	0.53*** (0.14)
Farmer		0.33*** (0.08)	0.38*** (0.09)	0.47*** (0.12)	0.46*** (0.12)	0.46*** (0.12)	0.47*** (0.10)
Indoors only			0.03 (0.08)	0.18 (0.10)	0.14 (0.10)	0.14 (0.09)	0.12 (0.09)
Contact only			0.18* (0.08)	0.19* (0.08)	0.18* (0.07)	0.18* (0.07)	0.17* (0.07)
Both			0.01 (0.08)	0.14 (0.09)	0.11 (0.09)	0.10 (0.09)	0.07 (0.08)
Age (30,45]					-0.21*** (0.05)	-0.21*** (0.05)	-0.22*** (0.05)
Age (45,60]					-0.41*** (0.06)	-0.40*** (0.06)	-0.39*** (0.06)
Age (60,80]					-0.46*** (0.07)	-0.47*** (0.07)	-0.43*** (0.06)
Month_10					0.40*** (0.05)	0.40*** (0.05)	0.41*** (0.06)
Month_11						0.78*** (0.06)	0.80*** (0.06)
Month_12						0.28*** (0.04)	0.30*** (0.04)
(Intercept)	0.26** (0.10)	0.26** (0.10)	0.55*** (0.06)	0.11 (0.16)	0.49** (0.17)	0.12 (0.18)	-0.45** (0.17)
Region FE	No	No	No	No	No	No	Yes
R ²	0.02	0.03	0.02	0.03	0.07	0.15	0.22
Adj. R ²	0.02	0.03	0.02	0.03	0.06	0.14	0.21
Num. obs.	3116	3116	3115	3114	3114	3114	3114

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

workers this was higher still: their excess mortality was 68 percent higher than that of higher skilled workers. Lower skilled workers also had substantially higher excess mortality than higher-skilled workers, but not as high as 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 in more detail below. Here we observe that if we separately control for farmers (model 2 onwards), excess mortality is still not strictly increasing with occupational skill, but closer to a full social gradient.

Next, we introduce our occupational exposure variables (model 3). Here, we find that contact occupations had nearly 20 percent higher excess mortality compared with occupations that had neither of the exposure characteristics. In this specification with limited controls, the other occupational exposure variable, indoor work, is not associated with higher excess mortality. Once further controls are added it also shows a positive association, but it is not estimated with sufficient precision to draw firm conclusions. However, an F-test does show that the exposure variables are jointly significant (model 7 has the lowest F-statistic at 4.05, $p = 0.006$).

Of particular interest is what happens to the skill gradient when we introduce these exposure variables (model 4). We find that when controlling for exposure, the skill gradient in excess mortality remains largely unchanged, although the mortality estimate for farmers increases.¹⁵ Overall, we find that working indoors does not predict higher excess mortality. However, occupational exposure to infectious disease

due to frequent contact predicts somewhat higher excess mortality, though skill remains a more important predictor.

Our next model adds demographic controls. Adding age controls reduces the social gradient in excess mortality describe above somewhat, but it is still clearly present. 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 16–30 (reference category), with progressively lower mortality predicted for older age groups. Adding month dummies 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 increases the skill gradient again. In our preferred model, including the full set of demographic controls and region and time fixed effects medium skilled workers have 46 percent higher excess mortality compared to high skilled workers; lower skilled workers 23 percent and unskilled workers have 70 percent higher excess mortality. The predicted excess mortality due to the exposure characteristics of the occupations of the deceased is 19 percent for contact occupations compared to the baseline of occupations characterised by neither exposure risk.

To further analyse the socioeconomic gradient in excess mortality during the Spanish Flu, we turn to the role of farmers in our results and an alternative measure of social status — HISCAM. These alternative models are tested in Table 4. In column 2, the preferred model from Table 3 is estimated again, but for a dataset where excess mortality was calculated without farmers. In the next column, as another alternative, we recode farmers as lower skilled, rather than medium skilled as in the HISCLASS scheme.

Omitting farmers from our dataset makes only a limited impact on the social gradient we find. Recoding farmers as lower skilled does make a difference. The N-shaped gradient, where, compared to higher

¹⁵ We also note that the increase in standard errors is limited compared to model 2 and 3, suggesting the moderate correlations between these two sets of variables discussed in Section 3.2 are not an issue in the model.

Table 4
Regression models of log excess mortality rate, with different treatments of farmers, and with HISCAM scores rather than HISCLASS-based skill levels. Region-clustered standard errors between parentheses.

	Farmer control	No farmers	Farmers recoded	HISCAM
Medium skilled	0.38*** (0.09)	0.38*** (0.09)	0.34*** (0.09)	
Lower skilled	0.21** (0.08)	0.22** (0.08)	0.36*** (0.09)	
Unskilled	0.53*** (0.14)	0.54*** (0.14)	0.42*** (0.12)	
Farmer	0.47*** (0.10)			0.58*** (0.09)
Hiscam				-0.66*** (0.17)
Indoors only	0.12 (0.09)	0.12 (0.09)	-0.22*** (0.06)	-0.13* (0.06)
Contact only	0.17* (0.07)	0.17** (0.07)	-0.10 (0.06)	0.16** (0.06)
Both	0.07 (0.08)	0.08 (0.08)	-0.22*** (0.06)	-0.08 (0.06)
Age (30,45]	-0.22*** (0.05)	-0.21*** (0.06)	-0.23*** (0.05)	-0.14** (0.04)
Age (45,60]	-0.39*** (0.06)	-0.38*** (0.06)	-0.39*** (0.06)	-0.24*** (0.04)
Age (60,80]	-0.43*** (0.06)	-0.42*** (0.06)	-0.44*** (0.06)	-0.27*** (0.05)
Month_10	0.41*** (0.06)	0.40*** (0.06)	0.43*** (0.06)	0.22*** (0.05)
Month_11	0.80*** (0.06)	0.73*** (0.06)	0.80*** (0.06)	0.49*** (0.05)
Month_12	0.30*** (0.04)	0.27*** (0.04)	0.30*** (0.04)	0.15*** (0.03)
(Intercept)	-0.45** (0.17)	-0.22 (0.16)	-0.06 (0.12)	0.29** (0.10)
R ²	0.22	0.21	0.21	0.17
Adj. R ²	0.21	0.19	0.19	0.16
Num. obs.	3114	2723	3064	4195

****p* < 0.001; ***p* < 0.01; **p* < 0.05.

skilled workers, lower skilled workers had lower excess mortality than medium skilled workers, disappears. Medium and lower skilled workers now have similar excess mortality. In this model, our occupational exposure variables show the counter-intuitive result, where working indoors now predicts lower excess mortality. We surmise that this is because the composition of the reference group is changed strongly by farmers group who were classified as having neither an indoors nor a contact occupation.

We also estimate our preferred model with HISCAM scores instead of HISCLASS. Again, we find a socioeconomic gradient in excess mortality. The model using HISCAM shows that, keeping all else constant, higher occupational status predicts lower excess mortality during the Spanish Flu. The HISCAM scores in our data range from 0.40–0.99, so going from the lowest to the highest status occupation predicts a decrease in excess mortality of 23%.

To assess if high-mortality regions drive our results, Table 5 compares how our preferred model performs in regions with relatively low and high overall mortality. Making this distinction is a more direct assessment of the impact of regional Spanish Flu severity versus occupational or demographic characteristics. 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 mortality areas, and we lose some precision on the estimates due to the lower number of observations. Otherwise, the overall results are similar, suggesting that the basic occupational patterns in excess mortality during the Spanish Flu were present regardless of the presence of high-mortality clusters.

Table 5
Regression models of log excess mortality rate for low (EMR below 2.5) and high (above 2.5) excess mortality regions. Region-clustered standard errors between parentheses.

	All	Low EM	High EM
Medium skilled	0.38*** (0.09)	0.33*** (0.09)	0.41** (0.13)
Lower skilled	0.21** (0.08)	0.22** (0.08)	0.18 (0.12)
Unskilled	0.53*** (0.14)	0.30 (0.21)	0.70*** (0.14)
Farmer	0.47*** (0.10)	0.22 (0.17)	0.65*** (0.10)
Indoors only	0.12 (0.09)	0.04 (0.11)	0.19 (0.11)
Contact only	0.17* (0.07)	0.03 (0.07)	0.28** (0.10)
Both	0.07 (0.08)	-0.01 (0.10)	0.14 (0.11)
Age (30,45]	-0.22*** (0.05)	-0.18** (0.07)	-0.23** (0.08)
Age (45,60]	-0.39*** (0.06)	-0.29*** (0.08)	-0.45*** (0.08)
Age (60,80]	-0.43*** (0.06)	-0.28*** (0.08)	-0.54*** (0.08)
Month_10	0.41*** (0.06)	0.30*** (0.09)	0.49*** (0.07)
Month_11	0.80*** (0.06)	0.65*** (0.10)	0.90*** (0.07)
Month_12	0.30*** (0.04)	0.29*** (0.05)	0.30*** (0.06)
(Intercept)	-0.45** (0.17)	0.04 (0.22)	-0.59*** (0.17)
R ²	0.22	0.18	0.26
Adj. R ²	0.21	0.16	0.24
Num. obs.	3114	1340	1774

****p* < 0.001; ***p* < 0.01; **p* < 0.05.

4.1. Robustness checks

To verify the robustness of our results, we included a number of alternative models in the Appendix. In Table 6 we estimate the preferred model from Table 3 at different levels of geographic aggregation: municipalities, EGG regions, Corop (NUTS-3) regions, and provinces (NUTS-2). While we prefer to work at a relatively low level of aggregation, a concern could be that this results in cells in our dataset with few observations, either for our 1910–17 baseline, or for the Spanish Flu pandemic deaths. Overall, the results are similar to the preferred EGG aggregation level, though overall effect sizes tend to be larger at the higher aggregation COROP (NUTS-3) level, and we lose some precision at higher levels of aggregation.

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 aspects of these factors, we explore this possibility further in Table 7, Appendix A, 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 aggregate our data to this level as well for these checks. As expected, the overall socioeconomic gradient is the same as these effects would also be captured by the location fixed effects included in the models in Table 6. The effect of the population variable itself is positive. However, population density does not have a significant effect on excess mortality, even if we exclude population size from our model.

In the same table, we also include a control for municipal military bases in part of our models for more reliable estimates of excess mortality. Troop movements and crowded conditions at army bases could be suspected to be important contributors to the spread of the Spanish Flu (Flecknoe et al., 2018). 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. Given that over 200,000 men out of 6.5 million inhabitants were conscripted, this is surely an underestimation of deaths among conscripts. To control for this potentially strong increase in local excess mortality around military bases, we add a set of dummy variables 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 a list of military bases in 1918 compiled by Ringoir (1980) and from the locations where Dutch soldiers were admitted to military hospitals in 1913, listed in annual reports (Landmacht, 1913). We used military hospitals for 1913 because no military hospital reports are available between 1914 and 1918, and the recording procedure changed after the War. The presence of army bases shows a negative, though statistically insignificant association with excess mortality, which is surprising given positive results found by other studies (Clay et al., 2019). 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 (see Fig. 1), but had to be excluded from our dataset because the occupations for these certificates were not digitised.

There are a number of ways in which the occupational exposure variables could have been used in the models, and in Table 8, Appendix A we explore a number of these specifications. We first look at a model that omits the interaction between indoors and contact occupations. This reduces the size of the effects we find and they are no longer statistically significant. While we think the combination of the two kinds of exposure is important to include in the model, we also note that this further shows that the occupational exposure effect is not consistent in our data. We next look at an additive measure of occupational exposure, by adding up the two dummy variables so that no risk equals zero, one of both indoors or contact risk results in a score of one, and both being present gets a score of two. We find that when we enter this variable as a linear predictor it does not have a statistically significant association with excess mortality. We prefer the specifications where occupational exposure enters as separate dummies because it does not require us to make assumptions about the size of the two variable's effect sizes as adding them up would.

In Table 9 in Appendix A we check whether the thresholds we used in our data selection procedure impact our results. Specifically, we check whether using higher coverage thresholds (at least 80% of age, sex, and date observations in a municipality have to be present rather than 50%). For presence of occupations, we use a threshold of 20% rather than 10%. The results are very similar, the main differences being that the estimates on our exposure variables have become stronger.

We also check whether including women alters our results (Table 10), applying the same model as in Table 3, with the addition of a dummy variable for sex. We find that men in this data indeed had higher excess mortality during Spanish Flu. Beyond that, our main findings are similar across these two models.

As a final check of our results, we investigate whether the way we handled the cases of zero deaths during the Spanish Flu in our model affects our results. The following alternative approaches are explored: 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 Tenreiro, 2006, column 4), a quasiPoisson model of the number of deaths during the Spanish Flu 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 et al., 1988, column 6).

Appendix A, Table 11 shows that the socioeconomic gradient in excess mortality and the higher excess mortality for contact occupations we find in our preferred specification is found in the alternative models as well. The main exception here is the model where we exclude groups with zero deaths during the Spanish Flu pandemic, in which the patterns we find elsewhere are reversed. As mentioned earlier, we think the absence of deaths during the Spanish Flu pandemic 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

This paper explored whether there was a socioeconomic gradient in excess mortality among male workers during the Spanish Flu in the Netherlands. Using individual level occupational information and all-cause mortality data, we calculated excess mortality and found it to be highest among farmers and unskilled labourers and lowest among higher-skilled labourers. Excess mortality for medium-skilled and lower-skilled workers was in between these groups. These findings are in line with evidence from a recent study on Sweden (Bengtsson et al., 2018) and earlier regional evidence, e.g. for Kristiania (Oslo, Norway) (Mamelund, 2006), and strongly contrast with the perceptions of contemporaries that the Spanish flu had a relatively egalitarian impact across the society (Mamelund, 2006; Qvanjer, 1921).

The national death registers we relied on included individual information on the age, occupation, and date of death as well as sex and the municipality of residence of the deceased. This information allowed us to estimate the association between social class and excess mortality, while also taking occupational characteristics, age, and regional variation in excess mortality into account (Colvin and McLaughlin, 2021; Mamelund, 2011). Moreover, controlling for municipality size and presence of local army bases, or analysis of specific regions with high excess mortality resulted in similar estimates of the magnitude of socioeconomic differences in excess mortality during the Spanish flu. In other words, there was a clear socioeconomic gradient in excess mortality during the autumn wave of the 1918–19 influenza pandemic. As we study excess mortality, these estimates come on top of the existing social gradient in mortality among individuals age 35–55 in the Netherlands (Van Poppel et al., 2009), indicating that the Spanish Flu increased existing social inequalities in mortality.

Our analyses take into account whether the social differences in excess mortality during the Spanish Flu could have been driven by occupational characteristics, including social interaction at the workplace and indoor work. For COVID-19, it has been shown that occupational conditions are related to workplace infections (Godefroy and Lewis, 2022). Similarly, working together with others in an enclosed space or regular social interaction at the workplace could have been related to increased chances of infection and death with the Spanish Flu in autumn 1918. To test this hypothesis, a contact index was constructed. We found some evidence that social contact at the workplace predicted higher excess mortality during the Spanish Flu pandemic. However, contrary to our expectations, it did not seem to matter whether individuals worked together indoors or not. While the estimates of our contact index were generally robust for controlling for demographics, municipality size, local army bases, regional fixed effects, or specifically looked at regions with high excess mortality, it should be noted that the effect we find for social contact depends on how these exposure variables are included in the model.

Our findings indicate that the higher social strata were, to some degree, able to protect themselves against the Spanish flu. It is tempting to think that such differences result from differences in income, as educational divides in mortality were uncommon before the second half of the 20th century. Earlier studies show inequalities in Spanish flu at the macro level. Neighbourhoods with larger housing had lower

Spanish flu mortality (Mamelund, 2006) and, across the globe, poverty was related to higher Spanish flu mortality (Murray et al., 2006). In the Netherlands in 1918, 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 account for the 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 by social status in other social contexts (Link and Phelan, 1995; Clouston et al., 2016; Debiasi, 2020; Elo, 2009; Edvinsson and Broström, 2012) including the ongoing COVID-19 pandemic (Drefahl et al., 2020).

We found that farmers had higher excess mortality during the Spanish Flu, especially after we control for occupational skill levels, occupational social contact and age composition. This contradicts earlier work on mortality during the Spanish Flu in Sweden (Bengtsson et al., 2018). More generally, 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). In the Netherlands, excess mortality rates during the Spanish Flu were especially high in Drenthe and parts of Groningen and Overijssel – rural and poor regions where many people were involved in farming (Mourits et al., 2021). Locals noted that the poverty, bad living conditions, and lack of medical care might have elevated mortality in this region (Vugs, 2020). In normal circumstances, farmers were able to outlive their peers even if they were poor, as they continued to be physically active at advanced ages and had access to fresh food in a time when refrigeration was not commonplace (Mourits, 2019; Pes et al., 2013). However, there is no evidence that farmers in the Netherlands lived longer than the middle class or elite between ages 35 and 55 (Van Poppel et al., 2009), which was part of the age group hit hardest by the Spanish flu. It could be that poverty protected farmers from lifestyle diseases, enabling them to live to old age, but made them vulnerable to outbreaks of infectious disease due to poor housing, malnutrition, and lack of medical care.

Some drawbacks of our approach and the data used should be mentioned. We analyse all-cause excess mortality, while ideally we would use sufficiently detailed and reliable cause-specific mortality. High excess mortality during the Spanish Flu means we should mostly pick up Spanish Flu related deaths, but uncertainty remains. Second, incomplete digitisation of death certificates as well as our data selection steps mean we lose some of the large cities in the Netherlands, most notably Amsterdam. While our sample showed no large biases on the basis of observable characteristics, differences in the spread of the virus in high population density settings are possible, and could impact our results. Third, the specific age structure within occupational groups could affect results, if relatively more young men – who were affected

hardest by the Spanish Flu – were unskilled labourers, for example. Finally, in this work we have only included men. Occupational characteristics of women were hard to reconstruct from the death certificates we use. Since both mortality and occupational structure are known to have differed between men and women, this is an important point for future research.

The Spanish flu did not hit the Dutch population evenly. Higher social classes experienced lower excess mortality rates. These differences between socioeconomic groups were not explained by occupation-specific risk factors included in the analyses in this work. Frequent social contact at the workplace was associated with increased excess mortality during the Spanish Flu pandemic, but cannot account for the socioeconomic gradient in excess mortality. Overall, the social factors that put populations at higher risk during the pandemic may also have been factors that determined existing health differences across the population, exacerbating existing social gradients in mortality.

CRediT authorship contribution statement

Auke Rijpma: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content. **Ingrid K. van Dijk:** Conception and design of study, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content. **Ruben Schalk:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content. **Richard L. Zijdemans:** Acquisition of data, Analysis and/or interpretation of data, Revising the manuscript critically for important intellectual content. **Rick J. Mourits:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content.

Funding and acknowledgements

We would like to thank Svenn-Erik Mamelund, Mar Rodríguez-Girondo, Niels van den Berg, participants to OsloMet's PANSOC webinar and two anonymous referees for their comments on earlier drafts of this paper. Any remaining mistakes are our own.

Auke Rijpma, Ruben Schalk, Richard Zijdemans and Rick Mourits acknowledge financial support from the Netherlands Organization for Scientific Research (NWO) for the CLARIAH-PLUS project (grant no. 184.034.023). Ingrid van Dijk was funded through the research programme “Landskrona Population Study” and project “An Age Old Advantage?”, Swedish Foundation for the Humanities and Social Sciences (Riksbankens Jubileumsfond, RJ). All authors approved the final version of the manuscript.

Appendix A. Robustness checks

See Tables 6–9.

Table 6

Regression models of log excess mortality rate at different levels of aggregation: EGG (Economic Geographic Regions), COROP: NUTS-3 regions, Provinces: NUTS-2 regions. Standard errors between parentheses, clustered at the relevant regional level.

	Municipalities	EGG	COROP	Province
Medium skilled	0.21*** (0.05)	0.38*** (0.09)	0.51*** (0.12)	0.38*** (0.10)
Lower skilled	0.15** (0.05)	0.21** (0.08)	0.32** (0.11)	0.22* (0.11)
Unskilled	0.33*** (0.09)	0.53*** (0.14)	0.55** (0.18)	0.41* (0.17)
Farmer	0.46*** (0.07)	0.47*** (0.10)	0.42** (0.14)	0.30 (0.18)
Indoors only	0.13* (0.06)	0.12 (0.09)	0.15 (0.11)	0.04 (0.09)
Contact only	0.15** (0.05)	0.17* (0.07)	0.14 (0.09)	0.04 (0.12)
Both	0.07 (0.05)	0.07 (0.08)	0.06 (0.12)	0.05 (0.11)
Age (30,45]	-0.13** (0.05)	-0.22*** (0.05)	-0.21*** (0.05)	-0.19** (0.06)
Age (45,60]	-0.22*** (0.05)	-0.39*** (0.06)	-0.53*** (0.04)	-0.60*** (0.05)
Age (60,80]	-0.21*** (0.05)	-0.43*** (0.06)	-0.64*** (0.05)	-0.73*** (0.05)
Month_10	0.25*** (0.03)	0.41*** (0.06)	0.51*** (0.09)	0.50*** (0.11)
Month_11	0.51*** (0.04)	0.80*** (0.06)	0.98*** (0.09)	1.00*** (0.12)
Month_12	0.16*** (0.03)	0.30*** (0.04)	0.35*** (0.06)	0.35*** (0.08)
(Intercept)	-0.35*** (0.10)	-0.45** (0.17)	0.25 (0.19)	0.62*** (0.16)
R ²	0.19	0.22	0.25	0.29
Adj. R ²	0.15	0.21	0.24	0.27
Num. obs.	4650	3114	2002	1003

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 7

Regression models of log excess mortality rate including controls for presence of army base and military hospitals as well as population density controls. Data is aggregated to the municipal level. Region-clustered standard errors between parentheses.

	None	Base	Hosp.	Both
Medium skilled	0.21*** (0.05)	0.21*** (0.05)	0.21*** (0.05)	0.21*** (0.05)
Lower skilled	0.14** (0.04)	0.14** (0.04)	0.14** (0.04)	0.14** (0.04)
Unskilled	0.32*** (0.08)	0.31*** (0.08)	0.32*** (0.08)	0.32*** (0.08)
Farmer	0.46*** (0.07)	0.45*** (0.07)	0.45*** (0.07)	0.45*** (0.07)
Indoors only	0.12 (0.06)	0.11 (0.06)	0.12 (0.06)	0.11 (0.06)
Contact only	0.15** (0.05)	0.14** (0.05)	0.15** (0.05)	0.14** (0.05)
Both	0.06 (0.05)	0.06 (0.05)	0.06 (0.05)	0.06 (0.05)
Age (30,45]	-0.13** (0.05)	-0.13** (0.05)	-0.13** (0.04)	-0.13** (0.05)
Age (45,60]	-0.22*** (0.04)	-0.22*** (0.04)	-0.22*** (0.04)	-0.22*** (0.04)
Age (60,80]	-0.21*** (0.05)	-0.21*** (0.05)	-0.21*** (0.05)	-0.21*** (0.05)
Month_10	0.25*** (0.03)	0.25*** (0.03)	0.25*** (0.03)	0.25*** (0.03)
Month_11	0.50*** (0.04)	0.50*** (0.04)	0.50*** (0.04)	0.50*** (0.04)

(continued on next page)

Table 7 (continued).

	None	Base	Hosp.	Both
Month_12	0.16*** (0.03)	0.16*** (0.03)	0.16*** (0.03)	0.16*** (0.03)
Log population '18	0.16*** (0.02)	0.18*** (0.02)	0.17*** (0.02)	0.18*** (0.02)
Army base '18		-0.08 (0.07)		-0.14 (0.10)
Army hosp. '13			-0.03 (0.06)	0.09 (0.09)
(Intercept)	-1.57*** (0.19)	-1.70*** (0.16)	-1.62*** (0.17)	-1.65*** (0.17)
R ²	0.14	0.14	0.14	0.14
Adj. R ²	0.14	0.14	0.14	0.14
Num. obs.	4650	4650	4650	4650

***p < 0.001; **p < 0.01; *p < 0.05.

Table 8

Regression models of log excess mortality rate, with alternative specifications for the exposure variables. Region-clustered standard errors between parentheses.

	Preferred model	No interaction	Additive exposure
Medium skilled	0.38*** (0.09)	0.39*** (0.09)	0.40*** (0.09)
Lower skilled	0.21** (0.08)	0.24** (0.08)	0.25*** (0.07)
Unskilled	0.53*** (0.14)	0.54*** (0.14)	0.57*** (0.12)
Farmer	0.47*** (0.10)	0.40*** (0.09)	0.41*** (0.09)
Indoors only	0.12 (0.09)		
Contact only	0.17* (0.07)		
Both	0.07 (0.08)		
Indoors		-0.02 (0.07)	
Contact		0.06 (0.04)	
Exposure (additive)			0.02 (0.04)
Age (30,45]	-0.22*** (0.05)	-0.22*** (0.05)	-0.22*** (0.05)
Age (45,60]	-0.39*** (0.06)	-0.39*** (0.06)	-0.39*** (0.05)
Age (60,80]	-0.43*** (0.06)	-0.44*** (0.06)	-0.43*** (0.06)
Month_10	0.41*** (0.06)	0.42*** (0.06)	0.42*** (0.06)
Month_11	0.80*** (0.06)	0.79*** (0.06)	0.79*** (0.06)
Month_12	0.30*** (0.04)	0.30*** (0.04)	0.30*** (0.04)
(Intercept)	-0.45** (0.17)	-0.38* (0.16)	-0.39** (0.15)
R ²	0.22	0.22	0.22
Adj. R ²	0.21	0.21	0.20
Num. obs.	3114	3114	3114

***p < 0.001; **p < 0.01; *p < 0.05.

Table 9

Robustness checks for regressions of log excess mortality rate, using stricter data selection steps: municipalities with higher occupation coverage (20 percent), and higher (80 percent) coverage of sex, age, and date variables. Region-clustered standard errors between parentheses.

	Pref.	Occ. ≥ 20 pc	Occ. ≥ 20 pc and rest ≥ 80 pc
Medium skilled	0.38*** (0.09)	0.45*** (0.10)	0.37*** (0.08)
Lower skilled	0.21** (0.08)	0.29** (0.10)	0.23** (0.07)
Unskilled	0.53*** (0.14)	0.74*** (0.12)	0.60*** (0.07)

(continued on next page)

Table 9 (continued).

	Pref.	Occ. \geq 20 pc	Occ. \geq 20 pc and rest \geq 80 pc
Farmer	0.47*** (0.10)	0.74*** (0.09)	0.69*** (0.11)
Indoors only	0.12 (0.09)	0.29** (0.10)	0.24** (0.09)
Contact only	0.17* (0.07)	0.24** (0.08)	0.25** (0.09)
Both	0.07 (0.08)	0.19* (0.09)	0.17 (0.10)
Age (30,45]	-0.22*** (0.05)	-0.22** (0.08)	-0.19* (0.09)
Age (45,60]	-0.39*** (0.06)	-0.42*** (0.07)	-0.33*** (0.08)
Age (60,80]	-0.43*** (0.06)	-0.44*** (0.09)	-0.36*** (0.10)
Month_10	0.41*** (0.06)	0.41*** (0.06)	0.34*** (0.06)
Month_11	0.80*** (0.06)	0.75*** (0.07)	0.71*** (0.08)
Month_12	0.30*** (0.04)	0.30*** (0.06)	0.30*** (0.06)
(Intercept)	-0.45** (0.17)	-0.92*** (0.15)	-0.86*** (0.15)
R ²	0.22	0.25	0.21
Adj. R ²	0.21	0.23	0.19
Num. obs.	3114	2021	1682

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 10

Regressions of log excess mortality rate, excluding and including women. Region-clustered standard errors between parentheses.

	Pref.	Include women
Medium skilled	0.38*** (0.09)	0.39*** (0.09)
Lower skilled	0.21** (0.08)	0.21** (0.07)
Unskilled	0.53*** (0.14)	0.56*** (0.12)
Farmer	0.47*** (0.10)	0.45*** (0.10)
Indoors only	0.12 (0.09)	0.13 (0.08)
Contact only	0.17* (0.07)	0.18** (0.06)
Both	0.07 (0.08)	0.11 (0.07)
Age (30,45]	-0.22*** (0.05)	-0.22*** (0.05)
Age (45,60]	-0.39*** (0.06)	-0.40*** (0.05)
Age (60,80]	-0.43*** (0.06)	-0.44*** (0.06)
Male		0.37*** (0.06)
Month_10	0.41*** (0.06)	0.39*** (0.05)
Month_11	0.80*** (0.06)	0.78*** (0.06)
Month_12	0.30*** (0.04)	0.27*** (0.04)
(Intercept)	-0.45** (0.17)	-0.81*** (0.19)
R ²	0.22	0.23
Adj. R ²	0.21	0.21
Num. obs.	3114	3551

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 11
Regressions of log excess mortality rate, using alternative model forms to deal with observations of 0 excess mortality. Region-clustered standard errors between parentheses.

	Log x + 1	Drop 0	No log	Poiss. EMR	Poiss. N	Asinh
Medium skilled	0.38*** (0.09)	-0.55*** (0.07)	1.27** (0.41)	0.91* (0.38)	0.64 (0.35)	0.48*** (0.11)
Lower skilled	0.21** (0.08)	-0.42*** (0.13)	0.59 (0.36)	0.64 (0.35)	0.49 (0.32)	0.27** (0.10)
Unskilled	0.53*** (0.14)	-0.72*** (0.09)	1.96** (0.60)	1.16* (0.46)	0.88* (0.41)	0.68*** (0.17)
Farmer	0.47*** (0.10)	-0.51*** (0.11)	1.96*** (0.47)	0.69*** (0.17)	0.57*** (0.12)	0.60*** (0.13)
Indoors only	0.12 (0.09)	-0.33** (0.11)	0.20 (0.35)	0.11 (0.17)	0.20 (0.12)	0.15 (0.11)
Contact only	0.17* (0.07)	-0.25* (0.10)	0.54 (0.31)	0.24 (0.13)	0.19 (0.12)	0.21* (0.09)
Both	0.07 (0.08)	-0.38** (0.12)	0.11 (0.33)	0.06 (0.15)	0.07 (0.11)	0.09 (0.10)
Age (30,45]	-0.22*** (0.05)	-0.07 (0.05)	-1.42*** (0.34)	-0.35*** (0.08)	-0.35*** (0.07)	-0.26*** (0.06)
Age (45,60]	-0.39*** (0.06)	-0.55*** (0.07)	-2.80*** (0.42)	-0.91*** (0.09)	-1.01*** (0.08)	-0.47*** (0.07)
Age (60,80]	-0.43*** (0.06)	-0.83*** (0.06)	-3.19*** (0.46)	-1.13*** (0.10)	-1.29*** (0.09)	-0.52*** (0.08)
Month_10	0.41*** (0.06)	0.34*** (0.08)	1.85*** (0.33)	1.16*** (0.12)	0.98*** (0.14)	0.52*** (0.07)
Month_11	0.80*** (0.06)	0.65*** (0.08)	4.29*** (0.45)	1.86*** (0.13)	1.68*** (0.16)	0.99*** (0.07)
Month_12	0.30*** (0.04)	0.17* (0.07)	1.10*** (0.18)	0.84*** (0.13)	0.69*** (0.13)	0.37*** (0.05)
(Intercept)	-0.45** (0.17)	3.09*** (0.16)	-0.42 (0.70)	-2.06*** (0.54)	-1.30** (0.49)	-0.59** (0.21)
R ²	0.22	0.35	0.20			0.22
Adj. R ²	0.21	0.32	0.18			0.21
Num. obs.	3114	1129	3114	3114	3114	3114
AIC						
BIC						
Log likelihood						
Deviance				14 755.65	3586.59	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 12
Most frequent occupations on death certificates for deceased age 16–79, September–December months in 1910–1918.

HISCO	Label	N	Contact	Indoors	Skill	HISCAM	EMR
61220	Farmer	2924	0	0	Medium	54	3.0
99900	Labourer	1860	1	0	Unskilled	49	3.7
41025	Dealer	421	1	1	Medium	66	2.8
62210	Farm worker	368	0	0	Unskilled	51	3.8
95410	Carpenter	344	1	0	Medium	53	2.6
99930	Factory worker	306	1	1	Unskilled	50	3.4
79100	Tailor	205	1	1	Medium	51	3.4
41030	Merchant	199	1	1	Medium	63	2.0
99920	Day labourer	187	1	0	Unskilled	42	3.3
93120	Painter	174	1	0	Lower	55	2.7
83110	Blacksmith	155	0	1	Medium	53	3.1
80110	Shoemaker	154	1	1	Medium	51	2.8
95120	Bricklayer	154	1	0	Medium	48	2.6
98620	Coachman	111	1	0	Lower	49	3.1
64100	Fisherman	110	0	0	Medium	52	8.1
77610	Baker	105	0	1	Medium	59	4.9
78200	Sigar maker	103	0	1	Lower	49	3.2
39310	Office clerk	102	0	1	Lower	65	7.9
61270	Horticulturist	96	0	0	Medium	61	2.9
81990	Clog maker	95	0	1	Lower	47	4.1
37040	Messenger	87	1	0	Lower	53	2.7
4217	Freighter	86	0	1	Medium	55	3.6
51050	Coffeehouse keeper	86	1	1	Medium	57	1.6
62105	Farm hand	82	0	0	Unskilled	49	4.5
81120	Cabinet maker	75	0	1	Medium	53	2.5

Table 13

Pearson correlation coefficients between occupational variables at the EGG (Economic Geographic Regions) level.

	Both	Indoors only	Contact only
Higher skilled	0.26	0.05	-0.14
Lower skilled	-0.05	0.24	0.01
Medium skilled	0.06	0.02	-0.25
Unskilled	-0.15	-0.25	0.34
Farmer	-0.22	-0.19	-0.24

Appendix B. Occupational coding

See Tables 12 and 13.

References

- Antonovsky, Aaron, 1967. Social class, life expectancy and overall mortality. *Milbank Meml. Fund Q.* 45 (2), 31–73. <http://dx.doi.org/10.2307/3348839>.
- Bellemare, Marc F., Wichman, Casey J., 2020. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bull. Econ. Stat.* 82 (1), 50–61. <http://dx.doi.org/10.1111/obes.12325>.
- Bengtsson, Tommy, Dribe, Martin, Eriksson, Björn, 2018. Social class and excess mortality in Sweden during the 1918 influenza pandemic. *Am. J. Epidemiol.* 187 (12), 2568–2576. <http://dx.doi.org/10.1093/aje/kwy151>.
- Bengtsson, Tommy, Van Poppel, Frans, 2011. Socioeconomic inequalities in death from past to present: An introduction. *Explor. Econ. Hist.* 48 (3), 343–356. <http://dx.doi.org/10.1016/j.eeh.2011.05.004>.
- Boberg-Fazlic, Nina, Ivets, Maryna, Karlsson, Martin, Nilsson, Therese, 2021. Disease and fertility: Evidence from the 1918–19 influenza pandemic in Sweden. *Econ. Human Biol.* 43, 101020. <http://dx.doi.org/10.1016/j.ehb.2021.101020>.
- Brainerd, Elizabeth, Siegler, Mark V., 2003. The economic effects of the 1918 influenza epidemic. CEPR working paper 3791, p. 41.
- Brooshoof, P., 1897. *Officieele en feitelijke waarheid: Bijdrage tot de kennis van onze arbeiderstoestanden*. gebrs. Belinfante.
- Brundage, John F., Shanks, G. Dennis, 2008. Deaths from bacterial pneumonia during 1918–19 influenza pandemic. *Emerg. Infect. Diseases* 14 (8), 1193–1199. <http://dx.doi.org/10.3201/eid1408.071313>.
- Burbidge, John B., Magee, Lonnie, Robb, A. Leslie, 1988. Alternative transformations to handle extreme values of the dependent variable. *J. Amer. Statist. Assoc.* 83 (401), 123–127. <http://dx.doi.org/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.
- CBS, 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, Lewis, Joshua, Severnini, Edson, 2018. Pollution, infectious disease, and mortality: Evidence from the 1918 Spanish influenza pandemic. *J. Econ. Hist.* 78 (4), 1179–1209. <http://dx.doi.org/10.1017/S002205071800058X>.
- Clay, Karen, Lewis, Joshua, Severnini, Edson, 2019. What explains cross-city variation in mortality during the 1918 influenza pandemic? Evidence from 438 U.S. cities. *Econ. Human Biol.* 35, 42–50. <http://dx.doi.org/10.1016/j.ehb.2019.03.010>.
- Clouston, Sean A.P., Rubin, Marcie S., Phelan, Jo C., Link, Bruce G., 2016. A social history of disease: Contextualizing the rise and fall of social inequalities in cause-specific mortality. *Demography* 53 (5), 1631–1656. <http://dx.doi.org/10.1007/s13524-016-0495-5>.
- Colvin, Christopher L., McLaughlin, Eoin, 2021. Death, demography and the denominator: Age-adjusted Influenza-18 mortality in Ireland. *Econ. Human Biol.* 41, 100984. <http://dx.doi.org/10.1016/j.ehb.2021.100984>.
- Crosby, Alfred W., 2003. *America's forgotten pandemic: The influenza of 1918, second ed.* Cambridge University Press, <http://dx.doi.org/10.1017/CBO9780511586576>.
- De Gooyer, Arie Cornelis, 1968. *De Spaanse griep van '18*. Philips-Duphar.
- De Melker, R.A., 2005. *De huisarts en 'de Spaansche griep' 1918–1920*. Huisarts en Wetenschap 48 (13), 44–49.
- Debiasi, Enrico, 2020. *The historical origins of the mortality gradient: Socioeconomic inequalities in adult mortality over two Centuries in Sweden* (Ph.D. thesis). Lund University.
- Devos, Isabelle, Bourguignon, Mélanie, Debruyne, Emmanuel, Doignon, Yoann, Eggerickx, Thierry, Greefs, Hilde, Hanus, Jord, Ronsijn, Wouter, Sanderson, Jean-Paul, Soens, Tim, 2022. The Spanish Flu in Belgium, 1918–1919. *A state of the art.* 33, pp. 251–283. <http://dx.doi.org/10.12759/hsr.suppl.33.2021.251-283>.
- Domela Nieuwenhuis, Ferdinand, 1898. *Een vergeten hoofdstuk. Blanke slaven: Supplement op het historisch gedenkboek "Eene halve eeuw", 1848-1898*. S.L. van Looy, Amsterdam.
- Drefahl, Sven, Wallace, Matthew, Mussino, Eleonora, Aradhya, Siddhartha, Kolk, Martin, Brandén, Maria, Malmberg, Bo, Andersson, Gunnar, 2020. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nature Commun.* 11 (1), 5097. <http://dx.doi.org/10.1038/s41467-020-18926-3>.
- Edvinsson, Sören, Broström, Göran, 2012. Old age, health, and social inequality: Exploring the social patterns of mortality in 19th century northern Sweden. *Demogr. Res.* 26, 633–660. <http://dx.doi.org/10.4054/DemRes.2012.26.23>.
- Elo, Irma T., 2009. Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annu. Rev. Sociol.* 35, 553–572. <http://dx.doi.org/10.1146/annurev-soc-070308-115929>.
- 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.
- Flecknoe, Daniel, Charles Wakefield, Benjamin, Simmons, Aidan, 2018. Plagues & wars: the 'Spanish Flu' pandemic as a lesson from history. *Med. Confl. Surv.* 34 (2), 61–68. <http://dx.doi.org/10.1080/13623699.2018.1472892>.
- Fourie, Johan, Jayes, Jonathan, 2021. Health inequality and the 1918 influenza in South Africa. *World Dev.* 141, 105407. <http://dx.doi.org/10.1016/j.worlddev.2021.105407>.
- Gagnon, Alain, Tremblay, Marc, Vezina, Helene, Seabrook, Jamie A., 2011. Once were farmers: Occupation, social mobility, and mortality during industrialization in Saguenay-Lac-Saint-Jean, Quebec 1840–1971. *Explor. Econ. Hist.* 48 (3), 429–440. <http://dx.doi.org/10.1016/j.eeh.2011.05.008>.
- Gallardo-Albarrán, Daniel, de Zwart, Pim, 2021. A bitter epidemic: The impact of the 1918 influenza on sugar production in Java. *Econ. Human Biol.* 42, 101011. <http://dx.doi.org/10.1016/j.ehb.2021.101011>.
- Godefroy, Raphael, Lewis, Joshua, 2022. What explains the socioeconomic status-health gradient? Evidence from workplace COVID-19 infections. *SSM - Popul. Health* 18, 101124. <http://dx.doi.org/10.1016/j.ssmph.2022.101124>.
- Grantz, Kyra H., Rane, Madhura S., Salje, Henrik, Glass, Gregory E., Schachterle, Stephen E., Cummings, Derek A.T., 2016. Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. *Proc. Natl. Acad. Sci.* 113 (48), 13839–13844. <http://dx.doi.org/10.1073/pnas.1612838113>.
- Herring, D.A., Korol, E., 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. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 20-07-2021).
- Janssens, A.A.P.O., Van Dongen, Elien, 2018. A natural female disadvantage? Maternal mortality and nutrition related causes of death in the Netherlands, 1875-1899. *Tijdschr. Soc. Econ. Gesch.* 14, 84–115. <http://dx.doi.org/10.18352/tseg.988>.
- Johnson, Niall P.A.S., Mueller, Juergen, 2002. Updating the accounts: Global mortality of the 1918–1920 "Spanish" Influenza pandemic. *Bull. Hist. Med.* 76 (1), 105–115. <http://dx.doi.org/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*. Simon & Schuster, New York.
- Koten, Jan-Willem, Weel, André, 2014. *Ziekte en dood binnen de gemobiliseerde Nederlandse krijgsmacht 1914 - 1918*. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 22, 275–279. <http://dx.doi.org/10.1007/s12498-014-0118-2>.
- Lambert, Paul S., Zijdemann, Richard L., Van Leeuwen, Marco H.D., Maas, Ineke, Prandy, Kenneth, 2013. The construction of HISCAM: A stratification scale based on social interactions for historical comparative research. *Hist. Methods: J. Quant. Interdiscip. Hist.* 46 (2), 77–89. <http://dx.doi.org/10.1080/01615440.2012.715569>.
- Landmacht, Geneeskundige Dienst der, 1913. *Statistisch Overzicht der Behandelde Ziekten van Het Leger Hier te Lande in Het Jaar 1913*. Algemeene Landsdrukkerij.
- Link, Bruce G., Phelan, Jo, 1995. Social conditions as fundamental causes of disease. *J. Health Soc. Behav.* 80–94. <http://dx.doi.org/10.2307/2626958>.
- Luy, Marc, 2003. Causes of male excess mortality: Insights from cloistered populations. *Popul. Dev. Rev.* 29 (4), 647–676. <http://dx.doi.org/10.1111/j.1728-4457.2003.00647.x>.

- Mamelund, Svann-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. *Soc. Sci. Med.* 62 (4), 923–940. <http://dx.doi.org/10.1016/j.socscimed.2005.06.051>.
- Mamelund, Svann-Erik, 2011. Geography may explain adult mortality from the 1918–20 influenza pandemic. *Epidemics* 3 (1), 46–60. <http://dx.doi.org/10.1016/j.epidem.2011.02.001>.
- Mamelund, Svann-Erik, Dimka, Jessica, 2019. Tuberculosis as a risk factor for 1918 influenza pandemic outcomes. *Trop. Med. Infect. Dis.* 4 (2), 74. <http://dx.doi.org/10.3390/tropicalmed4020074>.
- Mamelund, Svann-Erik, Dimka, Jessica, 2021. Social inequalities in infectious diseases. *Scand. J. Public Health* 49 (7), 675–680. <http://dx.doi.org/10.1177/1403494821997228>.
- Mamelund, Svann-Erik, Shelley-Egan, Clare, Rogeberg, Ole, 2019. The association between socioeconomic status and pandemic influenza: Protocol for a systematic review and meta-analysis. *Syst. Rev.* 8 (1), 1–6. <http://dx.doi.org/10.1186/s13643-018-0931-2>.
- Mandemakers, K., Hornix, J., Mourits, R.J., Muurling, S., Boter, C., van Dijk, I.K., Maas, I., Van de Putte, B., Zijdemans, R.L., van Leeuwen, M.H.D., van Poppel, F., Miles, A., 2020. HSN standardized, HISCO-coded and classified occupational titles, HSN release 2020.02. In: IISH Data Collection. <https://hdl.handle.net/10622/88ZXD8>.
- Mourits, R.J., 2019. Exceptional lives, extraordinary families. Familial clustering of longevity in the 19th and early-20th Centuries. <https://hdl.handle.net/2066/213756>.
- Mourits, R.J., Boonstra, O., Knippenberg, H., Hofstee, E.W., Zijdemans, R.L., 2016. Historische database nederlandse gemeenten, v5. In: IISH Data Collection. <https://hdl.handle.net/10622/RPBVK4>.
- Mourits, Rick J., Schalk, Ruben, Meroño-Peñuela, Albert, Raad, Joe, Rijpma, Auke, van den Hout, Bram, Zijdemans, Richard L., 2021. Retracing hotbeds of the 1918–19 influenza pandemic. Spatial differences in seasonal excess mortality in the Netherlands. *Hist. Life Course Stud.* 10, 145–150. <http://dx.doi.org/10.51964/hlcs9584>.
- Mourits, R.J., Van Dijk, I., Mandemakers, K., 2020. From matched certificates to related persons: Building a dataset from LINKS-Zeeland. *Hist. Life Course Stud.* 9, 49–68. <http://dx.doi.org/10622/23526343-2020-0006>.
- Murray, Christopher J.L., Lopez, Alan D., Chin, Brian, Feehan, Dennis, Hill, Kenneth H., 2006. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *Lancet* 368 (9554), 2211–2218. [http://dx.doi.org/10.1016/S0140-6736\(06\)69895-4](http://dx.doi.org/10.1016/S0140-6736(06)69895-4).
- Pes, Giovanni Mario, Tolu, Francesco, Poulain, Michel, Errigo, Alessandra, Masala, Speranza, Pietrobello, A., Battistini, Nino Carlo, Maioli, Margherita, 2013. Lifestyle and nutrition related to male longevity in Sardinia: An ecological study. *Nutr. Metab. Cardiovasc. Dis.* 23 (3), 212–219. <http://dx.doi.org/10.1016/j.numecd.2011.05.004>.
- Quanjier, A.A.J., 1921. *De Griep in Nederland in 1918 Tot 1920*. Gezondheidsraad, 's-Gravenhage.
- Razzell, Peter, Spence, Christine, 2006. The hazards of wealth: Adult mortality in pre-twentieth-century England. *Soc. Hist. Med.* 19 (3), 381–405. <http://dx.doi.org/10.1093/shm/hkl048>.
- Ringoir, Hendrik, 1980. Vredesgarnizoenen van 1715 tot 1795 en 1815 tot 1940. *Bijdragen Sect. Mil. Geschiedenis* 8, 1–136.
- Schalk, Ruben, Zijdemans, Richard L., Rijpma, Auke, Van Dijk, Ingrid K., Mourits, Rick J., 2022. HSNDB index for exposure to infectious disease. In: IISH Data Collection. <https://hdl.handle.net/10622/TFG1DQ>.
- Schenk, Niels, Van Poppel, Frans, 2011. Social class, social mobility and mortality in the Netherlands, 1850–2004. *Explor. Econ. Hist.* 48 (3), 401–417. <http://dx.doi.org/10.1016/j.eeh.2011.05.010>.
- Silva, J.M.C. Santos, Tenreiro, Silvana, 2006. The log of gravity. *Rev. Econ. Stat.* 88 (4), 641–658. <http://dx.doi.org/10.1162/rest.88.4.641>.
- Smith, Ken R., Mineau, Geraldine P., Garibotti, Gilda, Kerber, Richard, 2009. Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Soc. Sci. Med.* 68 (9), 1649–1658. <http://dx.doi.org/10.1016/j.socscimed.2009.02.010>.
- Sundin, Jan, Willner, Sam, 2007. *Social change and health in Sweden: 250 years of politics and practice*. Alfa Print, Solna.
- Taubenberger, Jeffery K., Morens, David M., 2006. 1918 Influenza: the mother of all pandemics. *Rev. Biomed.* 17 (1), 69–79. <http://dx.doi.org/10.3201/eid1201.050979>.
- Tuckel, Peter, Sassler, Sharon, Maisel, Richard, Leykam, Andrew, 2006. The diffusion of the influenza pandemic of 1918 in hartford, connecticut. *Soc. Sci. Hist.* 30 (2), 167–196. <http://dx.doi.org/10.1017/S0145553200013432>.
- Van Der Loo, M.P.J., 2014. The stringdist package for approximate string matching. *R J.* 6 (1), 111–122.
- Van Der Woud, Auke, 2010. *Koninkrijk vol sloppen: Achterbuurten en vuil in de negentiende eeuw*. Bert Bakker, Amsterdam.
- Van Leeuwen, Marco H.D., Maas, Ineke, 2011. *HISCLASS: A Historical International Social Class Scheme*. Leuven University Press.
- Van Leeuwen, Marco H.D., Maas, Ineke, Miles, A., 2002. *HISCO: Historical International Standard Classification of Occupations*. Leuven University Press.
- Van Poppel, Frans, Jennissen, Roel, Mandemakers, Kees, 2009. Time trends in social class mortality differentials in the Netherlands, 1820–1920: An assessment based on indirect estimation techniques. *Soc. Sci. Hist.* 33, 119–153. <http://dx.doi.org/10.1017/S0145553200010932>.
- Vaughan, Victor Clarence, 1926. *A doctor's memories*. Bobbs-Merrill Company, Indianapolis.
- Vugs, R.M., 2020. In veel huizen wordt gerouwd: De Spaanse griep in Nederland. *Aspekt*, Soesterberg.
- Vulsma, Rudolf Ferdinand, 1988. *Burgerlijke stand en bevolkingsregister*. Centraal Bureau voor Genealogie.
- Walhout, Evelien, van Poppel, Frans, 2003. De vermelding des beroeps: eene ijdele formaliteit? Twee eeuwen vrouwelijke beroepsarbeid in Nederlandse huwelijksakten. *Tijdschrift Soc. Geschiedenis* 29 (3), 301–332.
- Zeileis, Achim, Köll, Susanne, Graham, Nathaniel, 2020. Various versatile variances: An object-oriented implementation of clustered covariances in R. *J. Stat. Softw.* 95 (1), 1–36. <http://dx.doi.org/10.18637/jss.v095.i01>.