UvA ECONOMETRICS

Discussion Paper: 2013/09

# Mortality by occupation in the Netherlands in the 19th century: a re-examination that failed

J.S. Cramer

**www.ase.uva.nl/uva-econometrics**

**Amsterdam School of Economics**
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

# Mortality by occupation in the Netherlands in the 19th century: a re-examination that failed

*Second edition, with additions*

J.S. Cramer

January 1, 2014

**Summary** Earlier analyses of the birth cohorts from 1850 to 1922 from the *Historical Sample of the Dutch Population* by van Poppel and van Gaalen and Schenk and van Poppel conclude that there is no significant variation of mortality by occupation in those generations. A re-examination of their material with special attention to possible changes over time confirms this conclusion for the birth cohorts from 1850 to 1890. For later cohorts the results are anomalous and perverse. In this second edition I consider some explanations that have been put forward, but so far all fail to account for the results.

**Acknowledgement.** I have benefited from helpful comments on the earlier version of this report by Hans van Kippersluis, Liesbeth van Wely and Kees Mandemakers; see Section 4.

## 1. Introduction and conclusion

The HSN Historical Sample of the Dutch Population is a vast and complex database that is constantly revised, extended and improved. Its origin is a random sample from the births recorded in Dutch population registers in the years from 1812 to 1922. The life course of the participants is followed through their subsequent appearances in these registers which record marriage, parenthood, change of address and ultimately death. Ancillary files are opened for parents and siblings of the initial participants.

This material permits a study of the relation of mortality to occupation, which is frequently mentioned in the population registers. Two such analyses have been published, by van Poppel and van Gaalen (2008) and by Schenk and van Poppel (2011); both use subsamples covering the birth cohorts from 1850 to 1922. The authors employ two occupational classifications, HISCLASS and SOCPO, both based on the HISCO scheme (*Historical International Standard Classification of Occupations).* Both studies conclude that there is no clear occupational gradient of mortality. Van Poppel and van Gaalen write (p.226) "Overall we conclude that both occupational classifications employed show only very rarely significant differences in mortality, and that there is no clear evidence of a social gradient of mortality." Three years later Schenk and van Poppel again found little evidence of a general effect of social class on mortality (abstract, p.401).

Schenk and van Poppel have very generously provided me with their database for a secondary analysis. This is of course based on HSN data of a couple of years ago, since superseded by new releases. I wished to introduce a finer set of birth cohort periods to see if the occupational gradient of mortality varied over time, and I here report some simple analyses along these lines, altogether less detailed and sophisticated than the earlier studies, and confined to men. My results confirm their finding that there is no clear overall occupational gradient in mortality in the Netherlands for the birth cohorts from 1850 to 1890. For the birth cohorts after 1890 the results are anomalous, with the lowest occupation having the lowest mortality. This is a perverse result. In section 4 I consider some possible explanations that have been put forward: none is confirmed by the evidence. The result for the later years remains unacceptable and inexplicable.

## 2. Data and Method

The data set I received from Schenk and van Poppel refers to 21 267 individuals. In order to prepare for the estimation of Cox proportional hazards each record has been put into the standard format (see Cleves et al, 2008):

$t_0$ starting date of observation (individual attains the age of 18 years)
$t_1$ date of ending observation (date of death or date of last observation)
outcome 1 if died, 0 if censored ($t_1$ date of last observation)

$t_0$ and $t_1$ are expressed in elapsed time (days from January 1, 1960). The covariates are all dummy indicator variables for a limited number of categories.

Following the practice of the previous analyses all individuals who are younger than 18 years of age at $t_0$ have been omitted; $t_0$ is in effect the date of the individual's $18^{th}$ birthday. This reduces the sample size to 14 864 – 7 497 men and 7 367 women. The present analysis is moreover restricted to men. Estimates of hazard ratios have been obtained by the **stcox** Stata **p**rogramme.

### 3 Seven birth cohorts

While the two earlier analyses distinguish three birth cohorts – 1850-82, 1883-99, 1900-22 – I consider seven ten year periods from 1850-59 to 1910-22. Table 1 shows some descriptive statistics.

Table 1. Descriptive statistics for seven birth cohorts (men)

| cohort | N | no occupation | Unknown Occupation | censored | mean age (yrs) death | censored |
|---|---|---|---|---|---|---|
| 1850-59 | 511 | 4,5% | 4,9% | 20,4% | 61,9 | 39,6 |
| 1860-69 | 574 | 4,5% | 4,7% | 21,4% | 66,0 | 42,2 |
| 1870-79 | 550 | 2,4% | 4,4% | 18,9% | 68,2 | 39,8 |
| 1880-89 | 978 | 2,3% | 2,8% | 17,0% | 68,9 | 43,3 |
| 1890-99 | 1386 | 2,0% | 1,5% | 17,8% | 70,0 | 39,1 |
| 1900-09 | 1545 | 1,2% | 3,5% | 18,6% | 71,1 | 32,3 |
| 1910-22 | 1954 | 8,3% | 21,4% | 24,0% | 69,6 | 22,9 |
| total | 7 498 | 3,9% | 8,4% | 20,0% | 69,1 | 33,8 |

A profound change in the administrative system of population record-keeping on which the HSN is based was introduced in 1938 and implemented in the following years. This made it difficult to follow the sample individuals, and information became sparse. This has affected the later cohorts of the present data set, which dates from several years ago; since then additional information has been obtained and incorporated in the HSN. But here there is a sharp rise in the portion of men with no or unknown occupation in the last birth cohort – by 1938 they were on average 22 years old, and many were simply too young to have left a record of their occupation in the population files before that date. There is also an increase in the share of censored observations, with censoring at an earlier age. Also note the decline in duration of life in the last birth cohort.

I have omitted men with no or unknown occupation according to the HISCLASS classification; their records contribute little information to an analysis of occupational effects. This further reduces the sample size to 6 596 men.

Among this reduced sample there are 1069 censored observations, or 16%. Figure 1 shows the frequency distribution of 1051 of these observations by the date of censoring (in 18 cases this was unknown). The huge peak of 559 observations - over half - occurs between June 1939 and December 1941. It reflects the final transition to the new system of registration, when individuals were observed one last time and thereafter lost from view.

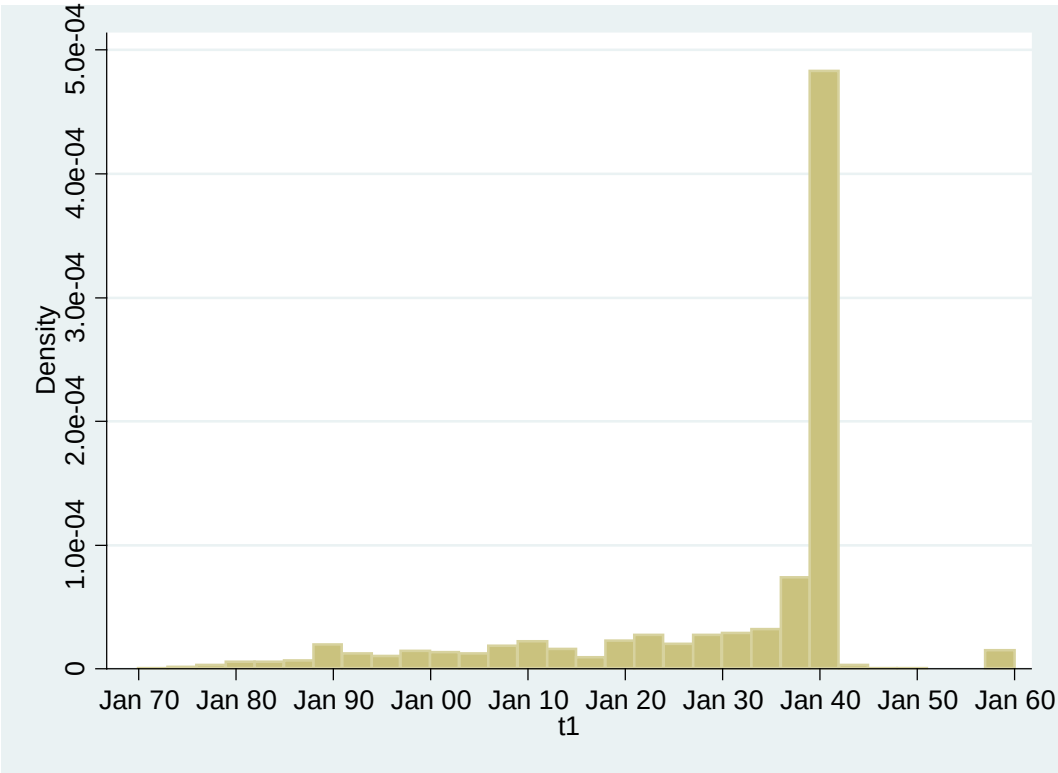


Figure 1. Frequency distribution of the date of last observation
for censored observations (n=1051)

Table 2 show that this 1939-41 censoring is fairly evenly spread over the occupational classes but concentrated in the last three birth cohorts, as stands to reason.

Table 2. Percentage of observations censored in 1939-41
by occupation and by period

| occupation | percentage | period | percentage |
|---|---|---|---|
| day labour | 7,4 % | 1850-59 | 1,7 % |
| unskilled | 12,9 % | 1860-69 | 3,3% |
| low skilled | 7,3 % | 1870-79 | 3,5% |
| farmers | 8,5 % | 1880-89 | 7,5% |
| foremen | 9,2 % | 1890-99 | 8,6% |
| managers | 7,4 % | 1900-09 | 11,1% |
| | | 1910-22 | 12,1% |

The HISCLASS classification by occupation in six broad classes, already employed in Table 2, represents a huge effort of identifying and sorting the various occupations named in the 19th century population records in the HISCO nomenclature, and then regrouping these in a smaller number of classes. Table 3 shows the percentage composition of the seven birth cohorts by these occupational classes. The bottom class is *agricultural unskilled labour*, the traditional *day labourer* who lives by his spade. At first the share of this class declines steadily over time, as does the share of the next class, *unskilled labour,* but after 1890 the share of the bottom class remains almost constant, and *unskilled labour* even shows an unexpected revival in the last cohort, 1910-22. Note that the periods denote *birth* cohorts and that adult life starts approximately twenty years later (as does our mortality analysis). The 1910-22 birth cohort was therefore grown up in 'thirties, and the increase in unskilled labour may possibly reflect the use of unemployed men regardless of their skills in land reclamation projects. - At the other end of the scale, the two top classes account for about half of the observations in all but the first cohort. This permanent uneven distribution reduces the information content of the occupational classification.

Table 3. Percentage composition of birth cohorts by HISCLASS classes (men)

| birth cohort | day labour | unskilled | low skilled | farmers | foremen | managers |
|---|---|---|---|---|---|---|
| 1850-59 | 12,7 | 13,8 | 17,9 | 13,2 | 10,8 | 31,5 |
| 1860-69 | 14,4 | 11,9 | 14,0 | 11,9 | 9,8 | 38,0 |
| 1870-79 | 13,3 | 10,1 | 18,3 | 11,5 | 10,7 | 36,1 |
| 1880-89 | 6,4 | 9,9 | 20,1 | 14,2 | 13,2 | 36,2 |
| 1890-99 | 7,0 | 9,3 | 18,0 | 13,8 | 13,2 | 38,8 |
| 1900-09 | 7,4 | 9,3 | 17,5 | 13,0 | 13,2 | 39,7 |
| 1910-22 | 7,7 | 12,7 | 18,2 | 9,6 | 12,6 | 39,1 |

## 3. Mortality hazard ratios for occupations and cohorts

Within each birth cohort hazard ratios have been estimated for the six HISCLASS groups of Table 3, with the lowest class - day labour - the reference group, and controlling for a threefold regional classification that proved highly relevant in the earlier analyses. The results are shown In Table 4.

Table 4. Hazard ratios for five occupations by birth cohort[i]

| cohort | N. | unskilled | low skilled | farmers | foremen | managers | N-W region | S-E region |
|---|---|---|---|---|---|---|---|---|
| 1. | 463 | 1,08. | 0,84. | 0,67. | 0,99. | 0,79. | 0,73* | n.a. |
| 2.. | 521 | 0,93. | 1,15. | 0,92. | 0,97. | 1,02. | 0,75* | 3,14. |
| 3. | 513 | 0,64* | 0,75. | 0,53*. | 0,62* | 0,79. | 0,83. | n.a. |
| 4. | 929 | 0,66* | 0,71. | 0,54* | 0,70. | 0,72. | 0,91. | 1,08. |
| 5. | 1324 | 1,24. | 1,31. | 1,12. | 1,56* | 1,40* | 1,03. | 1,10. |
| 6. | 1475 | 1,17. | 1,14. | 1,03. | 1,00. | 1,24. | 0,95. | 1,12. |
| 7. | 1373 | 1,15. | 1,06. | 1,06. | 1,05. | 1,30* | 1,05. | 1,11. |

Reference groups: unskilled agricultural labour, west region. In the first three periods there were zero, five and zero observations in the S-E region.

In the first four periods the hazard ratios (while seldom significant) give most occupations an increasingly lower mortality than the reference group. In cohort 5 (1890-1900) the rôles are however reversed and day labour is the healthiest occupation, and this perverse result persists in the cohorts 6 and 7.

The contrast between the two periods – before 1890 and after 1890 – can again be demonstrated by estimating common hazard ratios for occupations and cohorts (always controlling for region) for the two periods, pooling four respectively three ten year birth cohorts. The results are shown in Tables 5 and 6.

Table 5. Hazard ratio's for occupation, period and region (men), birth cohorts from 1850 to 1889

| | | | |
|---|---|---|---|
| day labour | *(1,00)* | 1850-59 | *(1,00)* |
| unskilled | 0,81* | 1860-69 | 0,95 |
| low skilled | 0,85 | 1870-79 | 0,81* |
| farmers | 0,65* | 1880-89 | 0,77 |
| foremen | 0,82* | region W | *(1,00)* |
| managers | 0,84* | region NW | 0,82* |
| | | region SE | 1,03 |

Table 6. Hazard ratio's for occupation, period and region (men), birth cohorts from 1890 to 1922

| | | | |
|---|---|---|---|
| day labour | *(1,00)* | 1890-99 | *(1,00)* |
| unskilled | 1,19* | 1900-09 | 1,04 |
| low skilled | 1,16 | 1910-22 | 1,49* |
| farmers | 1,06 | region W | *(1,00)* |
| foremen | 1,16 | region NW | 1,00 |
| managers | 1,31* | region SE | 1,11* |

If we accept that only few estimates are significant we find a consistent and acceptable pattern of the hazard ratios in the first four decades, from 1850 to 1889, shown in Table 5. All other occupations have a lower mortality than day labour; there is a coarse occupational gradient of three groups, with day labour at the bottom, farmers at the top, and the four other occupations lumped together in between. There is also a steady decline in overall mortality over time. In contrast, Table 6 shows for the period from 1890 to 1922 all occupations having higher mortality rates than day labour, and overall mortality rising by almost half in the last period. These results are anomalous, even perverse, and I do not believe they give a true picture of the period. Some possible explanations are considered in Section 4, but to no avail. I have no explanation for this result, and I do not know what to make of it.

What about time effects in the two earlier studies ? They distinguish only three birth cohorts, 1850-1879, 1880-1899 en 1900-1922. van Poppel and van Gaalen employ these as control variables, and their effect shows a neat decline (see column 2 of their Table 3, p.221); but then the deficiencies of the later years have been corrected by a substantial ad hoc adjustment of censored observations (op. cit. p.216). Schenk and van Poppel provide separate estimates for the three periods in Tables 4a and 4b and (like me) find that in the later periods the higher HISCLASS occupations have a higher mortality than the reference group of day labour, though not so extreme as here. They pay no particular attention to this finding – their major concern is the lack of significance.

## 4. Possible explanations

Various explanations of the curious and unacceptable result for the period 1890-1922 have been put forward. On closer inspection none holds.

*Effect of censoring.* As has been shown in Figure 1 and Table 2 there has been massive and indiscriminate censoring around 1940-41, due to the final implementation of a major administrative change. This might affect the estimates, but it does not, as I found out by the simple expedient of re-estimating the hazard ratios of Table 6 after omitting all censored observations. This reduces the number of observations from 4170 to 3500. The results are not identical, but very nearly so: with one exception (the hazard ratio for the last period, 1910-1922) they differ by less than 1%, coinciding after rounding to two decimals. Apparently **stcox** is not very sensitive to censored observations.

*Effect of particular events.* Two events may have intervened with mortality patterns in the second period, namely World War I and the Spanish Influenza pandemic of 1918-19, and World War II and the famine of 1944-45. Figure 2 show that these events indeed show up by peaks in the distribution of dates of death, but their overall effect is small – there are only 30 deaths, or 1% of the total of 3500 iof the secon birth cohort, in the famine period from September 1944 to June 1945.

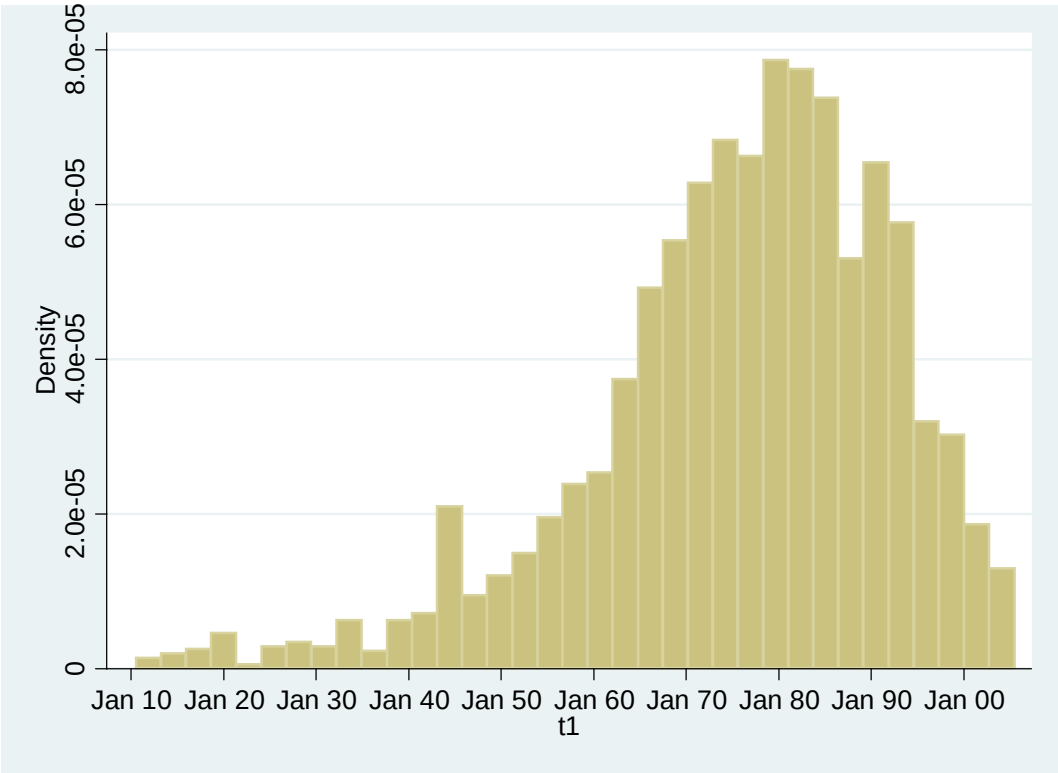


Figure 2. Frequency distribution of date of death, birth cohorts 1890-1922
(men, n-3500)

*Effect of uneven occupational classification.* As we have seen in Table 3 the distribution of observations over the six HISCLASS occupational classes is rather uneven, with many observations concentrated in the higher occupations of foremen and managers. The data set put at my disposal by van Poppel and Schenk does include a more detailed classification that has not been used in their previous studies. It is an extension of HISCLASS, with three of the larger groups subdivided, and I shall call it HISCLASS2. The relation of the two classifications is shown in Table 7.

Table 7. Percentage composition of sample by HISCLASS and HISCLASS2 (men)

| HISCLASS category | percentage | HISCLASS2 category | Percentage |
|---|---|---|---|
| day labour | 8,6 | ibidem | 8,6 |
| unskilled | 10,7 | ibidem | 10,7 |
| low skilled | 17,5 | low skilled farm workers | 1,7 |
| | | other low skilled | 16,2 |
| farmers | 12,4 | ibidem | 12,4 |
| foremen | 12,5 | foremen | 1,0 |
| | | other skilled labour | 11,4 |
| managers | 37,9 | lower clerical, sales | 9,8 |
| | | lower professional, clerical, sales | 14,9 |
| | | lower managers | 8,1 |
| | | higher professional | 3,0 |
| | | higher managers | 2,3 |

Although HISCLASS2 suffers from the defect of two very small classes – low skilled nonfarm workers and foremen – it is an improvement on HISCLASS since the largest class is now divided up. Tables 8 and 9 show the estimates of hazard ratios in this classification along the lines of Tables 5 and 6.

Table 8. Hazard ratio's for HISCO2 classification, period and region, (men), birth cohorts from 1850 to 1889

| | | | |
|---|---|---|---|
| day labour | (1,00) | 1850-59 | (1,00) |
| unskilled labour | 0,80* | 1860-69 | 0,95 |
| low skilled farm | 0,73 | 1870-79 | 0,81* |
| other low skilled | 0,87 | 1880-89 | 0,77* |
| farmers | 0,65* | | |
| skilled worker | 0,74 | region W | (1,00) |
| foremen | 1,42 | region NW | 0,83* |
| lower clerical, sales | 0,94 | region SE | 1,04 |
| lower professional | 0,80 | | |
| lower managers | 0,86 | | |
| higher professional | 0,75 | | |
| higher managers | 0,89 | | |

Table 9. Hazard ratio's for HISCO2 classification, period and region, (men), birth cohorts from 1890 to 1922

| | | | |
|---|---|---|---|
| day labour | (1,00) | 1890-99 | (1,00) |
| unskilled labour | 1,19* | 1900-09 | 1,04 |
| low skilled farm | 1,28 | 1910-22 | 1,49* |
| other low skilled | 1,15 | | |
| farmers | 1,07 | | |
| skilled worker | 1,11 | region W | (1,00) |
| foremen | 1,68* | region NW | 1,00 |
| lower clerical, sales | 1,32* | region SE | 1,10* |
| lower professional | 1,30* | | |
| lower managers | 1,31* | | |
| higher professional | 1,29* | | |
| higher managers | 1,14* | | |

The extreme estimates for foremen are due to their very low incidence – there are only 15 (out of 2426) in the first period, and 53 (out of 4170) in the second. Apart from this, the estimated hazard ratios for the first period again show a reasonable pattern, with all professions having a lower mortality than day labour; among the others the lowliest clerks do worst and farmers best. Overall mortality declines nicely over time. But in the second period this pattern is once more reversed, with day labour having the lowest mortality, and overall mortality rising sharply after 1910. Thus the earlier unacceptable results persist with a more detailed occupational classification.

## 5. References

Van Poppel, Frans, en Ruben van Gaalen (2008) Sociale klasse, sociale mobiliteit en sterfte in Nederland, 1850-2007. Pp.203-236 in: *Honderdvijftig jaar levenslopen, De Historische Steekproef Nederlandse bevolking, red: I. Maas, M.H.D. van Leeuwen, K. Mandemakers.* Amsterdam University Press, 2008.

Schenk, Niels, & Frans van Poppel (2011) Social class, social mobility and mortality in the Netherlands, 1850-2004. *Explorations in Economic History*, vol.48, pp. 401-417.

Cleves, Mario A., William W. Gould, Roberto G. Gutierrez and Yulia U. Marchenko (2008) An Introduction to Survival Analysis Using Stata (second edition). Stata Press, 2008.