

Datapresentation of death statistics

– Danish unskilled workers die prematurely

Nils Mulvad and Tommy Kaas

The biggest Danish analysis of detailed data on death rates for unskilled workers in Denmark showed exactly where and of what diseases the workers have a higher mortality rate compared to the average dane. See the methods, lessons and the story in English.

See story and data on the web: www.3f.dk/dod

See magazine story in English: www.kaasogmulvad.dk/unv/dod.pdf

The story *3F Members Die Prematurely* ran in the Danish magazine Fagbladet the 26th of June 2009.

This story investigated death records of all members of the union 3F (unskilled workers) comparing it to standard death rates for the whole population. The aim was to see where the mortality rate was higher (and lower) for the members of the union, and which causes of death were higher and lower compared to the whole population.

This is the first time an analysis of Standard Mortality Rates has been done on such a big group and on such a detailed level in Denmark. It's also the first time a web-presentation this ambitious was attempted in Denmark.

Major findings

1. The total number of deaths in the group analyzed was 121,054, compared to the expected figure (according to the standard mortality rate) of 113,200, i.e. 7,854 more deaths than normal. The total group consisted of 1.035.072 persons – then this was the biggest investigations of deaths for unskilled workers in Denmark.

2. Apart from homicide, of which there are relatively few cases, accidents are the primary cause of 3F members' high rate of excessive mortality. Accidents caused 1,242 more deaths than the norm in the period studied. If 3F members had conformed to the national average (i.e. the standard mortality rate), there were would have been 5,224 deaths caused by accidents. But the actual number of fatal accidents was 6,466.

The causes of death vary greatly: falls, toxic substances, and traffic accidents. And these accidents especially involve men who have been members of 3F for a short period of time. The death occurs early in life, i.e. the members who die lose many of their good years of life.

3. The overall excessive mortality rate for 3F is almost solely a problem for male members. They constitute 75% of the persons studied, but account for 97% (7,611) of the excessive 7,854 deaths.

4. When 3F members reach the age of 75, they no longer seemed to be plagued by excessive mortality rates, when looking at the entire period studied.

5. The most important causes of higher mortality rates are lung cancer, chronic obstructive pulmonary disease (COPD), accidents and heart disease. The excessive mortality rates caused by COPD and lung cancer increase throughout the period.

In other words, smoking-related disease is a major cause of excessive mortality rates among 3F members, and in this respect it is actually worst for those who have been members of 3F for a long period of time.

Getting raw data on death records

We have a possibility in Denmark to get the raw data under the regulation of private scientific research. The research plan has to be accepted by the Danish Data Protection Agency (Datatilsynet) and The National Board of Health - the primary Health Care Authority in Denmark (Sundhedsstyrelsen).

This is normally accepted for projects from Kaas & Mulvad, which is data analyzers doing this kind of stuff for the media. Both Nils Mulvad and Tommy Kaas has a background as long term computer-assisted reporters.

When the authorities accept the research plan we can get a copy of all membership data including personal identification numbers (and information on branch, job, date of membership and – for some of the members - date of leaving the union).

Part of the conditions for this is that the findings must be published as a scientific report and then the stories can be done on the basis of this report. Actually the conclusions in the report turned out to be used directly as the angles for the stories. And we publish the report on the web the same day as the stories are published.

As a new way of using this possibility to get data we also published all findings on the web in an interactive graphic, making a limit on at least 10 actually deaths in every group to be sure of not breaking the rules of data protection. If there was less than 10 deaths the numbers were not published for that group.

This was also accepted by the authorities, so we see this as a big step forward to make it easier and better to publish the findings of analysis done under these regulations.

In the material for the award we have included this scientific report in Danish and an abstract in English.

Results

We pointed out a lot of branches and jobs with specific high mortality rate and this is now being used to look for the reason and change the way of working and living for these people.

Especially the branch of seamen, which have a high mortality, has been using this in meetings and discussion between members.

The findings are also in the general debate in the society to point out we still have a high mortality rate of unskilled workers.

Advices

1. Making demographic analysis of so many people using Microsoft Access is very close to the limit of Access. It might be better to use other software for that.
2. We had a very tight connection to the main expert in Denmark in demographic analysis of death records and turning this into Standard Mortality Rates. From him we got the data on the SMR for the general population in the different years and he supervised all the steps in the analysis. This was very, very helpful.
3. We got all the raw data using the method of being accepted as doing a private scientific research. We think this method could be used in many other countries, but it needs to be investigated carefully before you make the first try.

Challenges

The large amount of records for this analysis gave us difficulties. The authority releasing 5 more years of death records during the analysis forced us to redo the analysis.

Then putting all the data on the web in an interactive graphic and with auto generated stories for each branch, job and age was a major challenge. It was necessary to really make this easy to use for the audience, and also respecting the demand on having at least 10 deaths in a group before publishing the story.

The content of the stories on the web and in the magazine are the same, but on the web we added the whole report as an interactive graphic. In fact this is a combined publication in a paper magazine and on the web.

Doing this interactive graphics with data, sparklines and auto generated text involved a cooperation with a good colleague in US. And this is then also an example of working cross-border, not on the story, but on the data presentation.

Social science methods

We got all data of members of the trade union 3F from the day they started with an electronic register in 1987. Data for the first years was not very valid and we decided to began the analysis for people who was a member in 1990 or later.

In the dataset was more than a million members with information on personal identity number, latest branch of the union and latest jobcode.

We delivered all the identification numbers to the Health Care Authorities and got back records on status of the person (dead, alive, disappeared, changed id-number, etc). For all the deaths we also got a code of the primary cause of death.

Because the group of changed personality identification numbers was rather big we run the data on their new number against the death records once more.

Out of these data we worked on a method to define a starting date in the union for each member, a starting day in the analysis and a potential end day in the analysis for people who died or totally disappeared.

We then for every kind of grouping of the data could calculate the actual number of deaths of a specific cause.

From our coach on the methods we then got the numbers of the general death records for every age, sex and period of years. They were grouped in 12 categories.

Out of these data we could calculate the expected number of deaths of the same grouping of data as we made on calculating the actual number of deaths. Dividing the actual deaths with the expected deaths we got the Standard Mortality Rate for each group of age, sex, branch and job of every cause of death.

The mortality rates for all existing 3F members and previous 3F members registered electronically from 1990 to 2006 are then compared to the standard mortality rates for the Danish population in general.

The total number of deaths in the group analyzed was 121,054, compared to the expected figure (according to the standard mortality rate) of 113,200, i.e. 7,854 more deaths than normal. Out of these 7,854 extra deaths, 3,713 were persons who had been members of the union for eight years or more, and 4,140 were persons who had been members of the union for less than eight years. Persons who have been members less than eight years have an excessive mortality rate of 10%, whereas this same rate is only 5% for the rest of the members.

Significant help

We signed up from the beginning with the best Danish expert in making analysis of Standard Mortality Rate at the National Institute for Public Health, a senior scientist Knud Juel. He helped us with the data on death records for the general population and in detailed discussed the methods of defining the start and end date for the group analyzed.

He also was involved in the decision not to make a statistical analysis of the data due to the fact that we actually analyzed the total population of 3F members.

Autogenerated text

Examples of the autogenerated text on the web-page:

Høj dødelighed hos <branch> (where SMR >=120)

3F's afdeling <Branch> har en overdødelighed på <SMR-100> procent i årene 2002-2006. Det farligste job for <branch> er <jobcode> med en overdødelighed på <SMR-100> procent. I alt <sum of actual death> fra <branch> og tidligere medlemmer af fagforeningen er døde i årene 1990-2006.

Størst overdødelighed er der af <diagnosis>, hvor dødelighedsraten er <SMR> (331) – det er <SMR-100> procent højere end for befolkningen generelt.

Dødelighed stiger for <branch> (where SMR >100 AND SMR2002_2006/SMR2000_2001 > 1,05 AND SMR2002_2006/SMR1990_2006 > 1,05)

3F's afdeling <Branch> har en overdødelighed på <SMR-100> procent i årene 2002-2006. Det farligste job for <branch> er <jobcode> med en overdødelighed på <SMR-100> procent. I alt <sum of actual death> fra <branch> og tidligere medlemmer af fagforeningen er døde i årene 1990-2006.

Størst overdødelighed er der af <diagnosis>, hvor dødelighedsraten er <SMR> (184) – det er <SMR-100> procent højere end for befolkningen generelt.

Farligste jobs: (criteria should be SMR>100)

Bygningsstruktorer/Jord Og Betonere	151	212
Lagerarbejdere (Htsa)	139	14
Kl, Specialarbejdere M.Fl.	132	389
Unknown 100 230 (unknown shall not be shown)		
Jern- Og Metal-Virk. Under Di/Co + Uden Arb.	97	85

Vigtigste årsager til høj dødelighed: (criteria should be SMR>100)

Selv mord	184	50
Kronisk obstruktiv lungesygdom	140	88
Lungekræft	137	176
Ulykker	125	87
Alt andet	101	257

Dødelighed falder for <branch> (where SMR <100 AND SMR2002_2006/SMR2000_2001 < 0,95 AND SMR2002_2006/SMR1990_2006 < 0,95)

3F's afdeling <Branch> har mindre dødelighed end befolkningen generelt i årene 2002-2006 og dødeligheden er faldet i forhold til årene før. Det farligste job for <branch> er <job> med en dødelighed på <SMR> procent.

I alt <sum of actual death> fra <branch> og tidligere medlemmer af fagforeningen er døde i årene 1990-2006.

Størst overdødelighed er der af <diagnosis>, hvor dødelighedsraten er <SMR> (190) – det er <SMR-100> procent højere end for befolkningen generelt.

Farligste jobs: (criteria should be SMR>100)

Bygningsstruktorer/Jord Og Betonere	151	212
Lagerarbejdere (Htsa)	139	14
Kl, Specialarbejdere M.Fl.	132	389
Unknown	100	230
Jern- Og Metal-Virk. Under Di/Co + Uden Arb.	97	85

Vigtigste årsager til høj dødelighed: (criteria should be SMR>100)

Selv mord	184	50
Kronisk obstruktiv lungesygdom	140	88
Lungecancerkræft	137	176
Ulykker	125	87
Alt andet	101	257

Dødelighed for <branch> (all the rest)

3F's afdeling <Branch> har ikke set de store udsving i dødelighed de senere år. Det farligste job for <branch> er <job> med en dødelighed på <SMR> procent.

I alt <sum of actual death> fra <branch> og tidligere medlemmer af fagforeningen er døde i årene 1990-2006.

Størst overdødelighed er der af <diagnosis>, hvor dødelighedsraten er <SMR> – det er <SMR-100> procent højere end for befolkningen generelt.

More info:

Kaas & Mulvad
post@kaasogmulvad.dk