New Forms of Data for Research – Three Examples from Social Science

Lisbeth Pedersen, SFI – The Danish National Centre for Social Research, Copenhagen, Denmark, lp@sfi.dk

Abstract

Denmark has a unique and comprehensive collection of register data. In order to facilitate register-based research, Statistics Denmark gives access to de-identified micro-data. This arrangement gives researchers a unique opportunity use micro-data. Over the years SFI – The Danish National Centre for Social Research has accumulated a vast amount of experience regarding the use of these data. The present paper briefly describes three examples of Danish social research projects that involve register data: an analysis of structural trends, a combined survey/register study and an impact study. The first example is a research project that is based on one of the large databases maintained by Statistics Denmark (the IDA database) that contains information on the total population of employees and enterprises. The second example demonstrates how a survey gains valuable information by being combined with register data that could validate the information on changes over time. The third example is a study that is based on a collection of registers containing information on labour market behaviour and characteristics of individuals and families. This last example demonstrates the use of register data in informative impact analysis.

1. Introduction

Denmark has a unique and comprehensive collection of register data covering many aspects of the Danish society. Altogether, data from 250 subject areas are available for research purposes through Statistics Denmark. To facilitate register-based research, Statistics Denmark gives access to de-identified micro-data, i.e. data on person, family, household, workplace and company levels. These data are made available via secure remote access from the researcher's own computer through the Internet to Statistics Denmark's servers. This arrangement gives researchers a unique opportunity to use micro-data in their research. Over the years SFI – The Danish National Centre for Social Research has accumulated a vast amount of experience regarding the use of these data and in working in close collaboration with Statistics Denmark. Authorization is only granted to Danish research environments. Foreign researchers may however gain access if they are affiliated with an authorized Danish environment.

SFI – The National Centre for Social Research deals with research questions concerning structural trends, individuals' attitudes and behaviour and the impact of policy initiatives on the Danish welfare state. The following text presents three examples of studies based on register data. Each example illustrates one of the three types of research questions, and the three studies also demonstrate different uses of register data in social research. The first study uses data from one of the major databases maintained by Statistics Denmark. The second study uses register data in combination with survey data. The third study uses register data to examine the impact of a specific policy initiative by comparing the target group with a control group which is constructed from register data.

2. Register data in research projects

Denmark introduced the Personal Identification Number (CPR) in 1968. In 1970, it was used in a census for the first time. During the 1970s, the first attempts were made to base the production of statistics on registers, but the existing registers were neither sufficiently comprehensive nor sufficiently well-established until 1981. In Denmark, like in all Nordic countries, the person and business registers today cover a very substantial part of the production of statistics. Most of the registers date back to the beginning of the 1980s, which makes them very useful in e.g., longitudinal research projects. The contents of these registers also cover many fields of research such as labour market research, sociology, health and business economics. The strength of this system is that the identification keys (personal identification number, address, central

business register number and property title number) enable correlation of data both within a specific year and, as mentioned, longitudinally across several years.

Most of the data derive from the administrative registers of governmental agencies, and they offer high quality data on entire populations of Danish persons, buildings and companies. Data can be combined in endless ways allowing researchers to produce unique analyses of dynamic processes and fluctuations, using the Danish population as their study population.

Altogether, data from 250 subject areas are available for research purposes through Statistics Denmark. Data from Statistics Denmark can easily be linked to data from other sources, e.g. survey data or data from other governmental agencies.

In order to reduce the cost of data sets for research purposes and to solve specific data issues, Statistics Denmark has set up a number of research databases. These databases could be seen as intermediate products created for the benefit of the research process. The Integrated Database for Labour Market Research (IDA) is one of the frequently used research databases. One purpose for creating this database was to solve a difficult problem of definition, the identity of enterprises over time, a task that individual researchers were unable to fulfil for reasons of both time and funding. Nine to ten man-years were spent developing this database, a work which was jointly funded by the Danish Social Science Research Council and Statistics Denmark. The IDA database contains information on the total population of people and enterprises in Denmark from 1980 and onwards.

3. First example: a structural analysis based on IDA

This first example is a research project based on The Integrated Database for Labour Market Research (IDA). Since the late 1990s, researchers have been able to combine this information with detailed information on hourly wages and give detailed descriptions of factors connected to enterprises, individual characteristics, and the labour market that create wage differentials between individuals.

This study shows how register data can be used to describe important social structures and analyse the causal relations underlying structural trends. The main subject of the study is wage development and wage relations in Denmark with a specific emphasis on the wage gap between women and men. The main purpose of the study is to examine the development of the wage gap between women and men. Secondly, the register data is used to reveal the significant factors explaining the measured difference. The study is based on data from the Wage Statistics Service Registry which are combined with information from the IDA database. The Wage Statistics Service Register is based on reports from companies in Denmark. The companies are required by law to submit information on the wages of every single employee. The register contains information on all employees in private companies with more than 10 employees (with the exception of a few minor industries) and all employees in the public sector (with the exception of a small group of employees with short working hours and employed on an hourly basis. Using these data you can draw a complete picture of wage levels and wage structures in the Danish labour market. The wage register is connected to the IDA database which has the unique property of combining information on each workplace in Denmark with information about every single employee at the workplace. The central observation of the analysis is hourly wage for each job function. For every employee, this observation is combined with information on:

-Level of education
- Employment in public or private sector
- Industry
- Work function
- Marital status
- Number of children and age of youngest child
- Workplace location.

In addition, information on the company size and the composition of skills is tested, but typically, these explanatory variables have no significant importance in analyses of this kind, where the information on individual employees is very detailed. The analysis, which is a so-called Blinder-Oaxaca decomposition analysis (an internationally recognized method) splits the wage gab in one part that is explained by the characteristics of the employee and the employee's work place and an unexplained part that cannot be outlined by the available data.

There are three significant main results of the analysis (Mona Larsen, 2010). Firstly, the wage gap between men and women is remarkably constant over time. From 1997 to 2006 the wage gap remained frozen at a very stable level around 18-20 per cent throughout the period. Secondly, the analysis shows that a significant proportion of the difference can be explained from the available data mentioned above. The wage gap in the public sector is particularly well explained since about 80 per cent of the wage difference is accounted for by the explanatory variables. A somewhat smaller proportion of the wage difference in the private sector can be explained from the data, but overall, approximately 60 per cent of the wage gap between women and men can be explained from the data Finally the decomposition analysis shows that the vast majority of the wage gap can be explained by the location of the workplace in sector and industry and by the job function. In this analysis the individual characteristics of the employees prove to be insignificant, while education and work experience mean very little.

Figure 1 Explanation of the wage gap
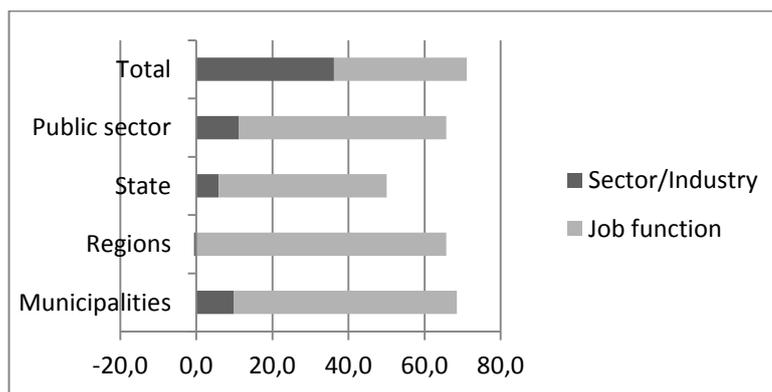
The policy implications drawn from this study are, firstly, that equal pay should be more actively promoted in order to reduce the wage gap in Denmark. Secondly, the reason for the very persistent wage gap is, above all, the labour market segregation and the close relationship between wages and industries, sectors, and job functions which means that the solutions are to be found here.

4. Second example: A combined survey/register analysis of employment of people with disabilities.
Data from Statistics Denmark can easily be linked to data from other sources, e.g. survey data or data from other governmental agencies. This second example demonstrates how a survey examining employment opportunities for people with disabilities gains valuable information by being connected to register data that could validate the information on changes in employment rate over time.

Not all social problems can be explained using registry data. If you want to examine subjects like handicaps and disabilities you must make use of other types of data. A handicap or a disability is not an objectively measurable phenomenon comparable with, for example, hourly wages or employment. The existence and degree of these types of disorders will depend on the individual assessment of, for example, a doctor, a social worker, an employer or the person with the handicap or the disability.

In 2001, the Danish government adopted an employment strategy for people with disabilities. In order to evaluate the success of this strategy SFI, has since 2002, studied the development of the employment of people with disabilities. Among other things, these studies shed light on changes in the public health situation, employment trends, and implications of disabilities for labour market behaviour. The data is collected through surveys among a large sample of the Danish population. In this case, SFI buys access to Statistics Denmark's large labour force survey and adds questions about disabilities and employment. The Labour Force Survey is conducted four times a year on a sample of about 40,000 people. SFI's questions about disabilities are integrated in one of the quarterly surveys every third year. The survey data is supplemented by data from Statistics Denmark's registers containing information on the respondents' education, marital status, and similar information. This increases the quality of the data and reduces the scope of the questionnaire as well as the answering time.

The main purpose of the SFI study is to shed light on the development of employment among people with disabilities. In Denmark, there is a relatively large difference in the rate of employment for people with and without disabilities. Among people without disabilities, about 3 out of 4 in the working age groups are in employment, while this applies to only approximately half of the group of people with disabilities. It has been the government's stated aim to increase employment for this group of citizens.

The latest report from SFI on disabilities and employment was released in January 2013 (Kjeldsen, Houlberg & Høgelund, 2013). This report collates the results from the measurements of employment among people with disabilities that have been made since 2002. The analysis of survey data shows that for people with disabilities, the employment rate is far below the employment rate of other employees without disabilities. The proportion of people with disabilities in employment fluctuates over the years with a noticeable decrease in the employment rate after 2008, when the employment rates began to decrease due to the economic crisis.

The disadvantage of measuring employment trends using survey data is that there are significant changes within the group of people with disabilities over the years. This applies to changes in its composition as well as to the respondents' perception of their own disability. To control for the first element, this study linked survey data with register data from Statistics Denmark. Use of these data made it possible to examine how the level of employment has developed for those who in the 2002 survey reported that they had a disability. The manner in which employment is measured in survey data and in register data is slightly different, and therefore the employment levels shown in the two studies are slightly different (see figure 2). However, figure 2 also

shows that the courses of the employment rates are similar for the two types of data. This indicates that the survey data can indeed be considered as a valid measure of the employment among people with disabilities. Furthermore, both the survey and register data show that employment trends are identical for people with and without disabilities. There is no relative improvement of the employment situation of people with disabilities over the years. On the other hand, people with disabilities are not hit harder by the crisis than other people.

Figure 2. Development in employment for persons with a disability
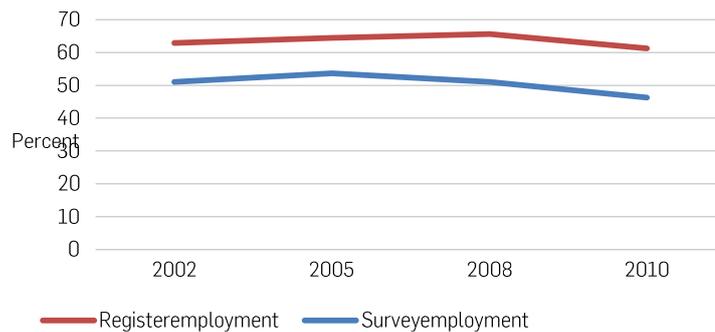
5. Third example: An effect analysis of activation measures.
The last example is a study based on a collection of labour market registers that contains information on people depending on transfer income, measures used by the public authorities, and a large number of data on personal characteristics and family backgrounds of individuals in the Danish population. The following example explains how these kinds of data can be used to make fairly precise evaluations of the employment effect of measures used in Danish government activation policies.

The impact of social policy or employment policy initiatives can be measured in several ways. To be able to isolate the effect of a given measure, it is necessary to compare the outcome for the target group with the course of a control group. This can be done using an experimental design trial or by matching the target group with a control group that is constructed using registry data. The so-called non-experimental design is considered slightly less reliable, but it is less expensive and less problematic to conduct. SFI has conducted a number of non-experimental studies of the effect of employment measures. However, the example below is found in a study published by the National Audit Office of Denmark (Rigsrevisionen) in 2010. The main question is whether the money allocated to the activation of welfare recipients is well spent in the sense that the chosen measures increase the employment rate for the target group.

The methodology applied in this impact study is called "matching". Matching aims to imitate the randomized controlled trial in the sense that for each target group participant, a person who did not participate in activation, but who is similar to the participant, is chosen for the control group. In this way systematic differences between participants and members of the control group are minimized. Data from the registers are used to select individuals in the control group. The analysis is based on register data on transfer income that is combined with a wide range of registry data from Statistics Denmark, Medicines database, the National Patient Registry, the Central Psychiatric Research Register, etc. This combination of register data makes it possible to follow the persons as they move in and out of the labour market as well as the healthcare, education, and activation systems over several decades. Using this method, the study measures the effect of participating in an activation programme rather than

postponing the decision to participate to a later point of time. The effect is measured by the degree of self-support.

Figure 3. The degree of self-support after participation in activation

———	Target group
———	Control group
———	Impact
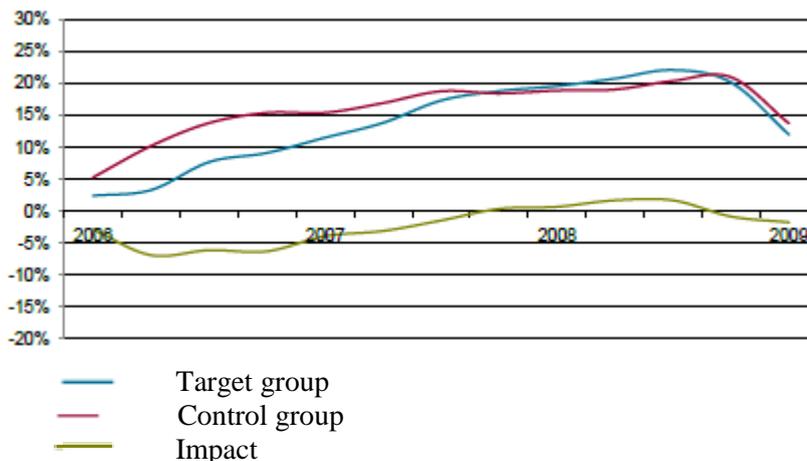
Source: Rigsrevisionen, 2010

Figure 3 shows the effect of participation in specially designed projects for the most disadvantaged social assistance recipients in the first quarter of 2006. The projects involve 2,134 persons altogether. The blue line in figure 3 shows the self-support for participants in activation, the red line shows the self-support rate for the control group, and the green line shows the effect on self-support, i.e. the difference between the values of blue and red lines. The figure shows that the self-support rate for the control group (the red line) is greater than the self-support rate for the group of participants (the blue line) in the period from 2006 to early 2008. In the remainder of the period, the degree of self-support is essentially the same for the two groups. The figure also shows that, essentially, the total effect (the green line) of specially designed projects for social assistance recipients – the effect throughout the entire period – was negative. The analysis thus shows that the activation measure did not meet its objective.

6. Concluding remarks
Register data is a very important data source for research studies in Denmark. I have mentioned three examples of analyses that make use of this type of data, but I could have mentioned many others. All three examples above concern Danish issues, but there are good opportunities to make comparative studies in collaboration with researchers from other countries.

References

Kjeldsen, Marie Møller, Houlberg, Helle Sofie & Høgelund, Jan (2013): Handicap og beskæftigelse. Udviklingen mellem 2002 og 2012. SFI 13:01, København

Larsen, Mona (2010): Lønforskelle mellem kvinder og mænd 2007. Analyser for lønkommissionen. SFI 10:10. København

Rigsrevisionen (2010): Beretning til Statsrevisorerne om effekten af aktivering af ikke-arbejdsmarkedsparate kontanthjælpsmodtagere. København