

## **Research using administrative and hybrid data Some Danish examples**

Tor Eriksson

Danish scholars in social sciences have above all used administrative and hybrid data sets in their research on the functioning of labour markets, but also for investigations of social and educational policies. Another area where there is great potential to use these types of data sources is the analysis of industrial economics and policy. However, in this field they have so far been used considerably less. It should also be mentioned that in Danish research in social medicine administrative data sources have been used quite extensively; internationally famous data sets are cancer and psychiatric registers which contain detailed information about all those individuals suffering from these illnesses to the extent that they have been admitted to a specialist hospital or ward for treatment.

In what follows, I shall briefly describe some examples of research using administrative register-based data. I do not pretend to give a survey or even a representative sample of all the work that has been carried out. My examples reflect my own background as a labour economist, and the examples are from areas of research that I have found interesting to follow (and in some cases also to contribute to) and which have attracted a lot public interest.

### **1. How does it work?**

In Denmark, as in the other Nordic countries, central and local government keeps many registers for administrative purposes. These registers typically identify individuals on the basis of their unique social security numbers. Much of the statistical information produced by Statistics Denmark builds on these registers. Thus, for instance the information about individuals' and households' annual incomes come from tax registers set up on the basis of individuals' and firms' tax forms. Another example is the unemployment and labour market policy statistics which largely emanate from information produced by the employment service and the unemployment insurance systems. In some cases, however, the information comes from non-governmental sources; an example is the wage statistics, which are collected the employers' federations and constitute an important part of the basic information used in wage negotiations. The unique individual identifier, in Denmark called the cepr number, allows you to

follow individuals over time (to create longitudinal data sets) and to merge information from different registers where the individual is in.

## **2. Brief history of the use of labour market register data**

Although I may be slightly biased, I think it is fair to say much of the applied work in Denmark, especially in economics, using large scale data sets constructed from registers, has its roots in Aarhus. As from the mid-eighties a group of labour economics researchers affiliated with the business school and university in Aarhus began in cooperation with Statistics Denmark to build a longitudinal labour market research register data set – the Longitudinal Labour Market Register (LLMR) – and to carry out empirical analyses based on these data. In the beginning, the data sets were typically 5 or 1 per cent samples of the population aged 16 to 74 and the time period covered began in 1976. In fact, Denmark was one of the first countries where researchers had access to have large nationally representative data sets of this type.

The early analyses were predominantly concerned with topics related to unemployment which at that time was the most pressing social problem in Denmark (and many other European countries). Another source of inspiration was more academic: the new microeconomics of the labour markets, notably search theory. Thus, many research papers from the eighties applied the then new methods of duration analysis to the register data. Another major research area was the analysis of taxes and labour supply inspired by the discussions of the need for tax reforms. In the nineties, the work and research on the register data became more formalized within the Centre for Labour Market and Social Research (CLS), which was funded by a large five-year grant from the Danish

National Research Foundation (“Grundforskningsfondet”) and started in 1993. With the establishment of CLS, the scope of research was broadened considerably to also include topics like returns to education, pensions and retirement, effects of labour market programmes and psychiatric disorders and labour market outcomes. The broader research programme also meant that data from several additional registers were added to the core data set, the LLMR. An important change associated with CLS was the establishment of a branch office of Statistics Denmark in Aarhus (physically in the same building, but the branch office serviced researchers in the whole Northern Jutland area), which made access and use of the register data much easier and also facilitated cooperation between the researchers and the statistics producers.

The LLMR consists of various registers supplemented with data from the latest census in 1970. Thus, data on education come from the Census in 1970 and from reports from all educational institutions on their current population of students

and their completion. This means that the educational register describes the situation at the 1970 census and includes all upgrades thereafter.

The wage information is constructed as follows. The point of departure is register data containing tax-based information on the total earnings paid to each individual worker during the year. Earnings may consist of earnings from several employers. The data are considered to be of high quality because the tax authorities use them to assess each employee's earnings. As the wage records also constitute deductible labor costs for the employers, firms have strong incentives to provide accurate and timely information to the tax authorities.

The number of working hours is estimated as follows. The employers' contributions to a comprehensive pension scheme are determined by the number of hours worked as a fraction of normal annual working hours. Thus, for hourly paid workers, i.e., all blue-collar workers, pension contributions were up to 1993 proportional to the number of hours worked. For monthly paid salaried employees the supplementary pension is computed based on the normal length of the working day according to a three-step scale. The IDA makes use of information about the employers' contributions to the pension schemes to compute the annual number of working hours for each individual. It should be emphasized that these are estimates. One problem is that the supplementary pay for overtime hours does not yield additional points for the pension schemes. Hence, overtime hours are not properly accounted for.

Hourly wage rates are calculated by dividing the earnings at a particular employer with the estimated annual working hours at that employer. The estimated hourly wage rates information is most reliable for the hourly paid workers. However, after 1993 pension contributions have gradually also been paid during sickness and unemployment spells. Consequently, as from the mid-nineties, the quality of the hourly wage information is likely to be poorer.

For a long time a specific feature of Danish statistics was that workplaces also had unique identification numbers. Thus, in the early eighties a major project within Statistics Denmark was started up with the aim of building a linked employer-employee dataset, which included all individuals, aged 15 to 74, and all workplaces (note, not firms; but, see below). This data set, which was completed in 1990, has information about each employee's wage and employment situation (at the end of November each year)<sup>1</sup> and a host of demographic and other individual characteristics as well as some information about the workplace (size, location, industry). This longitudinal data set, called

---

<sup>1</sup> The persons and plants are matched at the end of November in each year. Consequently, only changes between ends-of-Novembers are accounted for (not intermittent changes).

IDA, was later in the nineties, thanks to the cooperation of a research team in Aarhus and Statistics Denmark, transformed by aggregating the workplace information to the firm level and by merging IDA with firm information (key accounting numbers) from registers in SD. As a result of these efforts there are now data sets which can link information on individuals, workplaces and firms to each other for a time period beginning with year 1980.

The background data for the IDA consist essentially of the same registers as the LLMR that were described above. Unlike the LLMR, the IDA data were for a long time not available outside Statistics Denmark. Gradually, access became easier; first by SD setting up a branch office in Aarhus to service users outside Copenhagen, and in recent years IDA is available in many research institutes which fulfill Statistics Denmark's requirement regarding data security. Currently, most users access the data from the offices in their normal workplaces.

After the end of funding of CLS, research using the data sets built up has continued in three different research centers. The first was the National Center for Register Research, which is physically located together with Statistics Denmark's Aarhus unit. NCRR has mainly carried out psychiatric research based on internationally unique register data constructed by merging LLMR and/or IDA and the Danish Psychiatric Register which encompasses information about all individuals (and their illness history) who have been admitted to a psychiatric hospital or ward in Denmark since 1970. The second is the Centre for Social Integration and Marginalization (CIM) which was a five-year research project with participants from several academic institutions in Denmark and focused on marginalization issues, in particular the assimilation of immigrants, but also unskilled labour and women with small children into the Danish labour market. The third is the Center for Corporate Performance (CCP) which differs from the other two chiefly because the research here uses linked employer-employee data and deals more with firm performance and employee behavior within firms and organizations.

In parallel the use of register data has also spread to other universities and research institutes. Thus, today at most economics departments the data sets described above are used to some extent and at large research institutes like the Danish Social Research Institute (SFI) and the Local Government Research Institute (AKF) a fairly large number of researchers use them regularly. In Copenhagen the Centre of Applied Microeconometrics (CAM) is another place where large micro data sets are analyzed.

Most of what has been said above pertains to economic research. Due to the very strong influence of Marxist sociology in previous decades, empirical sociology

using large data sets (indeed, any data at all) have been thin on the ground in Denmark. This is slowly changing, especially at the research institutes, but the scarcity of empirical work in sociology is still notable.

Especially the Danish Social Research Institute has carried out several surveys and questionnaire studies, some of which have been done repeatedly, largely with the same questions, albeit not with the same respondents. Lately, some of these surveys have been linked to administrative register data. However, this is not always possible as the respondents were sometimes promised that the information would not be merged with other data sets.

Unlike her neighbouring countries, Norway and Sweden, Denmark has not a strong tradition of utilizing Level of Living Surveys in social research. Such surveys have indeed been carried out, but the results of them seem to receive rather little attention in the public debate and they have not been a major source for empirical studies in economics and sociology.

Increasingly when Danish researchers carry out surveys, they also ask for the cepr-number, and permission to use it, in order to merge the survey information with data from other sources, typically registers. The great advantage of doing this is that you can keep the questionnaire relatively short and, moreover, you do not need to ask questions concerning factual information which the respondents might have difficulties answering. These types of surveys have been done for both firms and individuals. It should be noted, however, that the register information is normally made available by Statistics Denmark with a considerable time lag – 2 to 3 years is the typical delivery lag for labour market register information.

Below I will give a number of examples of research which use different types of register-based data. The first is labour market policy evaluation studies, the second studies matching several registers to examine determinants of suicides, the third immigration analyses, and the fourth investigations of education systems. Finally, I give two examples of analyses based on matched questionnaire-register data sets.

### **3. Labour market policy studies**

One area where administrative register data have proven to be very useful is evaluations of different labour market programmes, such as labour market training programmes, intensified help to target, typically disadvantaged, groups, and unemployment insurance policies. Here especially the longitudinal character of the administrative data combined with information from other registers is a key strength, because it allows for an analysis of both the longer term effects of different policies and of the impact of changes in policies. An

additional strength is the detailed information concerning the programmes and participants. Many studies have been carried out of the effects of Danish labour market policies and not only those that are implemented by the labour market authorities (Ministry of Labour), but also policies implemented by the municipalities (Bolvig, Jensen, and Rosholm (2003)). Although the latter are less well-known, they nevertheless account for a considerable part of the total expenditure spent on labour market policy. Thus, in 2003 about every second participant in active labour market programmes were in programmes implemented by the municipalities. They are typically persons who are not eligible for benefits from unemployment insurance funds.

Alongside the Netherlands, Denmark has had one of the best performances in terms of reduction of unemployment within the EU area since the mid-nineties, and Danish labour market policies and research thereof have therefore attracted some interest as a possible best practice to follow for other countries. The catchword in recent European discussion is “flexicurity” which is considered as a key feature of the Danish labour market. Some of the earlier studies of labour market policies (see Jensen et al. (1993)), and of training programmes in particular, were rather critical of their design and magnitude as the studies did not find that the policies had the intended positive employment and/or wage effects (in this respect Danish research was not unique; studies from other countries also found rather small or no positive policy effects). In view of the large amounts of expenditure devoted to these programmes, the lack of a clear positive impact was rather uncomfortable for policy-makers. Consequently, the research was not well received in policy circles. Nevertheless, as from the mid-nineties, the Danish labour market policies were subject to some major reforms, which typically created stronger incentives for unemployed job seekers to find employment (radically shortening the duration of entitlement to UI benefits and tightening of the criteria for benefits eligibility). This generated a number of studies that asked questions like: was the lower unemployment a consequence of the labour market reforms, and if so, why did these work? Especially the latter question is important, as the design of programmes or reforms typically involves several parameters and for policy purposes it is of course vital to learn what works and what does not.

Thus, in Jensen, Rosholm and Svarer (2003) the authors examine the spectacular decline in the youth unemployment rate in Denmark in the second half of the nineties. Youth unemployment was more than halved in a few years and became lower than the overall unemployment. Thus, unemployment for the age group 20-24 (18-19) has fallen from 15 (6) per cent in 1995 to 5.5 (2.3) in 2004. This is almost unique among OECD countries and therefore merits study. A candidate for explaining the remarkable change in youth unemployment is the implementation of a radical labour market reform, the Youth Unemployment

Programme (YUP), in 1996, which was directed towards unemployed, low-educated youth. The reform meant a significant reduction in eligibility to unemployment benefits to youth and made further entitlement to benefits conditional on participation in training or education. The study analyzes the effects of the implementation of the YUP on the duration of unemployment spells and the transition rates from unemployment to schooling and employment. The longitudinal data set used is constructed by merging several administrative registers with the LLMR. Three effects are analyzed: an announcement effect, a direct programme effect, and a sanction effect. The authors find that the YUP has been partially successful in lowering unemployment. There are no traces of an announcement effect and the sanctions effect was relatively modest. Thus, the lower unemployment rate can mainly be attributed to the direct programme effect.

Rosholm and Svarer (2004) combine two econometric techniques to analyze the incentive and threat effects of active labor market policies on the employment behavior of unemployed job seekers. The earlier literature has chiefly focused on the incentive effects whereas the threat effect has received little, and combined none, attention by previous research. Insured unemployed persons know the rules of the unemployment compensation system and hence know that the longer they remain out of work, the higher is the probability that they will be assigned to an active labor market policy programme. If this is conceived of as stigmatizing, unpleasant or unattractive for other reasons, the risk of being assigned to a programme may lead the unemployed individual to search more actively than before or to be less choosy with respect to the job offers received. This is the threat effect.

The authors use a database called DREAM, which has been constructed by the Labour Market Board by combining several of its administrative registers, to arrive at estimates of the magnitude of the threat effect. They find that a threat effect does not impact the job search behavior of unemployed females. For males there is, however, an effect: the threat effect reduces men's unemployment durations by on average three weeks. The incentive effect, that is, the effect on unemployment duration of actually participating in an active labor market programme, is found to be of minor importance for both genders.

#### **4. Suicides**

A key area of research at the National Center for Register Research (NCRR), which despite its name has had been strongly focused on in particular psychiatry and other areas of medical research, has been the determinants of

suicides.<sup>2</sup> The overall goal is to unravel the relative importance of family background, socio-economic factors, demographics and psychic illness as predictors of suicides; for two overview papers, see Qin et al. (2003), (2005). This research is motivated not only by the fact that, although there has been a steady declining trend since the early eighties, Danish suicide rates are relatively high<sup>3</sup> from an international perspective<sup>4</sup>, but also because this is a question that generally is rather poorly understood. In order to analyze these questions, the researchers are making use of the unique possibilities to link the LLM/IDA data (with the demographic and socio-economic information) to the Danish Psychiatric Register, and the so called Fertility database, which allows the researchers to track an individual's biological parents and siblings, and finally, the register for causes of death for information of suicides. Needless to say, building this giant jigsaw puzzle is a formidable task, but once in place, it constitutes an extraordinary valuable source of information for research on determinants of suicides.

One of the first papers in this research programme is Mortensen et al. (2001), a key finding of which is the important role played by mental illness as a predictor of suicides. The odds ratio of suicide is extremely high in the very first weeks<sup>5</sup> after leaving a psychiatric hospital or ward and while declining remains high during the subsequent months. High odds ratios do not, however, necessarily imply that removal of this cause of suicide would reduce the number of suicides much. But in this case they were: the attributable risk for ever (within previous year) being admitted to a psychiatric ward was 45 (22) per cent. Consequently, the practical importance of this result for the general practice physician is obvious.

Thanks to the unusually rich data set created by combining the above mentioned register, the research group has recently published an impressive series of papers, providing a much more detailed picture of the factors associated with suicides. Thus, for example, Agerbo (2003a, 2003b, 2005a, 2005b) deals with labour market status, including unemployment, and dramatic events in the individual's family (such as spouse's mental illness, or the bereavement of spouse or child) as causal factors. Qin et al. (2003), (2005) are examples of two articles which consider several factors (demographic, family background, socio-economic and psychiatric factors) at the same time.

---

<sup>2</sup> Other central areas of the psychiatric research carried out at NCRR are determinants and consequences of schizophrenia and depression.

<sup>3</sup> Among high GNP countries, Denmark ranks second after Finland. The suicide rate is 20.4 per 100,000 individuals.

<sup>4</sup> Thus, five years ago there was a state commission looking into the causes and remedies of this problem. Some of the center's research had a strong impact on the work and recommendations of the commission.

<sup>5</sup> The odds ratios were around 340 in the first week and 70 in weeks 2 to 4 before the suicide.



## **5. Studies of immigrants**

During the last ten years, the beyond doubt Number One issue in politics and the public debate in Denmark has been the country's immigration policy and the lack of integration of in particular immigrants from Turkey and the Middle East. Thus, for example, the relative unemployment rate of first generation immigrants is highest in the OECD area and the rate for second generation immigrants is also one of the highest. As immigration was a topic that for a long time was not politically correct to discuss, not to speak of questioning its principles and its lackluster success, there was more or less no research based knowledge to inform the discussion when this as a consequence of the great success of an a strongly anti-immigrant party in the last two parliamentary elections could not be held back anymore. As a consequence, the discussion was, and to some extent still is, heavily characterized by prejudice, guesswork and wishful thinking.

It is telling that Statistics Denmark in general had little statistical information about immigrants and had for instance no information about the immigrants' educational background. Thus, in 1999 SD carried out a questionnaire survey that was sent out to all persons living in Denmark (including the present author) but who had not been born in the country. Subsequently, this information is continuously updated for new cohorts of immigrants. These statistics show a striking difference in the educational distribution between immigrant and ethnic Danish youth.

Immigrants make up around 6 per cent of the total population and are heavily concentrated in the capital area and other larger cities. Moreover, due to their clearly higher nativity rate, the immigrant offspring's proportion of the youth cohorts is much higher and is growing faster than that of the native Danes. How immigrant children fare in school and in the transition from school to work is therefore a cause of major concern.

The proportion which begins in a post-compulsory education is almost the same for second generation immigrants and immigrants who came to Denmark when they were aged 0 to 5 and for native Danes. The key problem for all young immigrant groups is, however, that immigrant youth drop out of the education to a much higher extent: almost twice as often as Danes do. This is in particular a problem in vocational educations, where the drop out rate is about 60 per cent among immigrant youth.

The statistical analysis carried out by Nielsen, Rosholm, Smith and Husted (2003) shows that the probability that an immigrant youth starts on an education after compulsory school is higher the higher the parents' level of education and

the more labour market experience the parents have. This is in particular true for college education.

If immigrant parents would have the same level of education and the same amount of labour market experience as the average Danish parent does, the enrollment to college (vocational education) would be 8 (2.5) percentage points higher for immigrant youth and the drop out rate would be 5 percentage points lower. Living in neighborhoods with strong concentration of immigrants was found to be associated with slower labour market assimilation. Although family and ethnical background are important determinants of the immigrant drop out rate, it should be noted that they only explain part of the difference in the drop out rates between native and immigrant youth. Thus, for a more thorough understanding of the gap, other factors are called for.

Husted, Nielsen, Rosholm and Smith (2001) make use of two panel data sets, for the population of immigrants and ten per cent of the Danish population, respectively, to examine the assimilation of immigrants into the Danish labour market. More precisely, the authors estimate a joint model for wages and individual employment probabilities and find that assimilation takes place, that is, the probability of employment increases with number of years spent in Denmark. Differences in wages reflect above all differences in accumulated employment experience. A notable finding is a significant difference in the speed of the assimilation process between refugees and non-refugees: even after controlling for differences with respect to human capital, for the latter, the process is clearly slower.

## **6. Studies of the education system**

Not only second generation immigrants have problems in school, but also a considerable proportion of the children of ethnic Danish parents are performing unsatisfactorily. Thus, according to the latest PISA study (Programme for International Student assessment) from 2003, pupils in Danish schools perform substantially poorer than in most other European countries, and in particular in comparison to the other Nordic countries. This has, of course, generated a lively debate about the causes of the poor performance despite the fact that Denmark spends more money per pupil than any other OECD country in her basic education system. In recent years this debate has been increasingly informed by research based on administrative register data and/or questionnaires linked to register data. Below two examples of such studies are briefly described.

In Rangvid (2003, 2004), the focus is on the relation between school factors and differences in students' subsequent performance.<sup>6</sup> The study is concerned with the question whether students from private schools perform better because private schools are more free to choose plans of study and how to teach, and are therefore more likely to provide a superior teaching environment, or whether the better performance of private school students reflects differences in selection of students to private and public schools, respectively.

The study shows quite clearly that the main part of the difference in performance can be attributed to the sorting effect. The students in private schools are found to have better conditions in terms of e.g., family background for continuing to a higher education. Thus, they would have fared well also in a public school. The pure "private school" effect is found to be rather small.

The PISA study mentioned above has also shown that a more equal distribution of students with different social background across schools leads to a higher average level of reading skills. Denmark, like the other Nordic countries, is characterized both by a internationally fairly equal income distribution, and by an internationally low level of segregation in society. However, when it comes to social segregation between schools, Denmark stands out as considerably more segregated than for instance Finland and Sweden. As children's reading ability in the two latter countries is clearly better than that of Danish children, it suggests segregation could have something to do with it. Thus, the question Rangvid tries to answer is: how much would a more even distribution of pupils with different social background increase the average reading ability among Danish children?

Rangvid (2004) shows that a reduced segregation would indeed lead to an improvement in the reading abilities of the weaker students, while there would no significantly negative effects on the good students' reading ability if they were in socially less segregated classes. Consequently, the results of the study imply that a policy to reduce the socio-economic segregation of school classes would have a positive impact on the average reading ability of Danish children.

## **7. Matching questionnaires with register data**

As mentioned above, in recent years Danish researchers have increasingly started to link questionnaire or survey data to registers. The main disadvantage with register data is of course that as the data collection was not designed for research purposes. Consequently, the operationalisation of the variables, independent as well as dependent ones, is frequently far from perfect.

---

<sup>6</sup> In addition to the PISA data, also good domestic registers on student performance exist, but have been relatively little used outside economics. The reasons are partly the same as those pertaining to the dearth of quantitative sociological research.

Sometimes the variable one would like have is not recorded by the administrative system. One example is wage information emanating from tax assessments. This does not for instance distinguish between the different pay schemes (salary, hourly pay, bonus, commission, stock or stock options) used by the individual's employer. To study them, other sources of data are needed. These can in turn be made much richer by linking them to registers. Let me give you two examples.

In 1999 a group in Aarhus carried out a survey study (a questionnaire) directed at firms. The survey was administered by Statistics Denmark as a mail questionnaire survey in May and June 1999, which was sent out to 3,200 private sector firms with more than 20 employees. The firms were chosen from a random sample, stratified according to size (as measured by the number of full time employees) and industry. The survey over-sampled large and medium-sized firms: all firms with 50 employees or more were included, and 35 per cent of firms in the 20-49 employees range. The response rate was 51 per cent, which is relatively high for the rather long and detailed questionnaire of the type that was used.

The survey represents a unique source of information on Danish firms' internal labour markets and changes therein. In addition to some background information about the firm, each firm was asked about its work organisation, compensation systems, recruitment, training practices and how the firm evaluates its employees. Using unique firm identification numbers from Statistics Denmark, the survey data has been supplemented with information about the firms as well as about their workforces. This information is taken from the augmented IDA described above.

This data set has e.g. been used to examine the effects of new workplace practices, such as self-managed teams, TQM, quality circles, job rotation<sup>7</sup>, etc on firms' productivity and wage costs. Also the impact of different pay practices (individual and team bonuses, stock and stock options, profit sharing) has been studied. A commonly heard argument is that work organisations have changed; hierarchies have become flatter and more flexible and jobs are characterised by broader tasks and more responsibility delegated to individuals or teams. These changes, it is claimed, are associated with higher productivity and profitability.

Most previous studies of whether this is the case have, however, either been based on cross-sections or have lacked detailed information about the employees in the firms. Eriksson (2003) distinguishes between practices adopted for salaried

---

<sup>7</sup> For an in-depth study of job rotation schemes using the linked survey-register data, see Eriksson and Ortega (2006).

employees and those paid by the hour and investigates how their impact differs between early and late adopters as well as between the short- and the medium-term. He shows that it is crucial to control for the skill structure of firms' workforces. Otherwise the returns to practice adoption are substantially overestimated. Late adopters are shown to benefit less, while for early adopters the short-run (2-3 years) gains seem to be quite persistent.

Datta Gupta and Eriksson (2006), examine how the new work organizations affect the gender wage gap at the level of firm by estimating a differences-in-differences model for firm average wages for men and women and using the fact that they can observe firms' wages before and after the implementation of the new workplace practices.<sup>8</sup> The estimation shows that wage gains from the introduction of new workplace practices seem to accrue mainly to salaried men, and in fact, wage losses accrue to women so that so that the gender gap in pay widens at the level of the firm. Thus, the new economy is not the great equaliser.

A second example of a merge of a firm survey and register data is one carried out and analysed in Bolvig (2005). She examines firm provided concerns toward employees with small children, senior employees, and employees with health problems. The key question studied is whether these policies reflect firms being socially responsible, or if they merely work as a substitute for wages.

For this purpose she makes use of a survey of more than 2,000 firms concerning their policies towards different groups of employees which is linked to administrative data for all employees in these firms. The relation between the level of firms' concerns and wages is estimated by means of a wage equation model using IV estimation techniques to deal with possible endogeneity of the concern variables. Bolvig finds that the employees' wages are significantly affected by the level of concern. All effects are rather small, however. Moreover, they differ significantly between white-collar and manual workers. The evidence lends support to the non-wage compensation theory, i.e., social concerns seem to substitute for money wages at the firm level.

## **8. What next?**

During the previous four-five years access from secured offices at universities and several large research institutes has increased and as a consequence the use of register data has spread quickly and they are now widely used in applied research in social sciences. They are also increasingly used in ministries and the for instance by the secretariat of the Economic Council. This is important as it

---

<sup>8</sup> In fact there is only one previous study looking into this issue by Drolet (2002) using less rich data from Statistics Canada.

contributes to an enhanced understanding of the value of microeconomic data sets among decision makers.

Since 2001, the Ministry of Science, Technology & Innovation has supported register-based research in the state budget with an annual payment to Statistics Denmark and the National Board of Health. The use of these means is administered by a board consisting of leading scholars from different fields. It should also be pointed out that the prices for accessing as well as using the register data have fallen significantly during recent years. As a consequence the number of users has increased considerably during the last five years.

A recent initiative from the research community to for a more efficient and more cost-effective use of registers (above all by reducing the overlap in data purchases) is the proposal to build a common longitudinal research database for social sciences research, which would contain the 300 most used variables in Statistics Denmark for the whole population. This would be validated and general samples would be drawn from it on the basis of consensus recommendations from the researchers. Moreover, documentation, which in the past has been something of an Achilles heal , would be elevated to a higher level directly comparable to that of large datasets in Europe and Northern America. Another weakness of current register-based research in Denmark would be helpful in improving, is the little replication of each others findings.

At the time of writing, the initiative described is included in the budget proposal the government has sent to the Parliament.

## References

Agerbo E., (2003a), "Unemployment and suicide", *Journal of Epidemiology and Community Health* (57), 560-561

Agerbo E., (2003b), "Risk of suicide and spouse's psychiatric illness or suicide: nested case-control study", *British Medical Journal*, 327: 1026

Agerbo E., (2005a), "The effect of psychiatric illness and labour market status on suicide: a healthy worker effect?", *Journal of Epidemiology and Community Health* (59), 598-602

Agerbo E., (2005b), "Midlife suicide risk, partner's psychiatric illness, spouse and child bereavement by suicide or other modes of death: a gender specific study", *Journal of Epidemiology and Community Health* (59), 598-602

Bolvig I. (2005), "Firm-provided social concerns - just another compensating wage differentials story?", *International Journal of Manpower* (26), 673-704

Bolvig I., P. Jensen and M. Rosholm (2003), The employment effects of active social policy. IZA discussion paper No. 736

Datta Gupta N. and T. Eriksson (2006), New Workplace Practices and the Gender Wage Gap: Can the New Economy be the Great Equalizer? IZA Discussion paper No. 2038. IZA, Bonn

Drolet M. (2002), Can the Workplace Explain Canadian Gender Pay Differentials? Paper presented at the NZ Conference on Database Integration and Linked Employer-Employee Data in Wellington, New Zealand

Eriksson T. (2003), "The Effects of New Work Practices — Evidence from Employer-Employee Data", in T. Kato and J. Pliskin (eds), *The Determinants of the Incidence and the Effects of Participatory Organizations*. Advances in the Economic Analysis of Participatory and Labor-Managed Firms, Vol. 7, JAI Press, 1-30

Eriksson T. and J. Ortega (2006), "The Adoption of Job Rotation Practices: Testing the Theories", *Industrial and Labor Relations Review* (59), 653-666

Husted L., H. S. Nielsen, M. Rosholm and N. Smith (2001), "Employment and wage assimilation of male first generation immigrants in Denmark", *International Journal of Manpower* (22), 39-68

Jensen P., P. J. Pedersen, N. Smith and N. Westergård-Nielsen (1993), "The effects of labour market training on wages and unemployment: Some Danish results"

in: H. Bunzel, P. Jensen and N. Westergård-Nielsen (eds.), *Panel data and labour market dynamics*. North-Holland, Amsterdam

Jensen P., M. Rosholm and M. Svarer (2003), "The response of youth unemployment to benefits, incentives and sanctions", *European Journal of Political Economy* (19), 301-316

Mortensen P. B., E. Agerbo, T. Eriksson, P. Qin and N. Westergård-Nielsen (2001), "Risks of suicide in relation to income level in people admitted to hospital with mental illness: nested case-control study", *British Medical Journal* (322), 2001 February 10, 334-335 (also republished in *British Medical Journal USA* (1), 2001, 259-260)

Nielsen H. S., M. Rosholm, N. Smith and L. Husted (2003), "The school-to-work transition of 2<sup>nd</sup> generation immigrants in Denmark", *Journal of Population Economics* (16), 755-7876

Rangvid B. S. (2003), Do schools matter? The influence of school inputs on student performance and outcomes. PhD thesis, Aarhus School of Business

Rangvid B. S. (2004), "Elevernes hjemmebaggrund, skolen og undervisningsmiljøet", in: Jan Mejdning (ed.), *PISA 2003 - Danske unge i en international sammenligning*, DPU

Rosholm M. and M. Svarer (2004), Estimating the threat effect of active labour market programmes. Aarhus University department of economics working paper 2004-06

Qin P., E Agerbo, T. Eriksson, P. B. Mortensen and N. Westergård-Nielsen (2000), "Gender Differences in Risk Factors for Suicide in Denmark", *British Journal of Psychiatry* (177), 546-550

Qin P., E. Agerbo and P. B. Mortensen (2003), "Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: A national register-based study of all suicides in Denmark, 1981-1997", *American Journal of Psychiatry* (160), 765-772

Qin P., E. Agerbo and P. B. Mortensen (2005), "Factors contributing to suicide: the epidemiological evidence from large-scale registers", in K. Hawton (ed.), *Prevention and treatment of suicidal behaviour*. Oxford: Oxford University Press, 11-28