ORIGINAL ARTICLE

Referral bias in hospital register studies of geographical and industrial differences in health

Helle Soll-Johanning, PhD,
Harald Hannerz, Statistician, *& Finn Tüchsen*, Msc

Correspondence: Helle Soll-Johanning, National Institute of Occupational Health, Lersø Parkalle 105, DK-2100 Copenhagen. E-mail:hsj@ami.dk

ABSTRACT

Introduction: The Danish National Hospital Register contains four patient types: full-time inpatients, part-time inpatients, outpatients and emergency ward patients. The aim of the present study was to investigate whether results from comparative hospital register studies depend on which patient types we choose to include in the analysis.

Methodology: The hospital register was linked to the centralised civil register and the employment classification module. All economically active persons in Denmark aged 20-59 years 1st January 1995 (N = 2,281,480) were followed for six years. We calculated SIRs, first by county then by industry and finally by industry adjusted for county, for a variety of diagnostic groups and for each of the following types of cases: A) full-time inpatients, B) all inpatients, C) all inpatients and outpatients, D) all patients. The ratio between the maximum and the minimum of the four types of SIRs was calculated for each combination of the examined population groups and diseases. A max/min ratio was regarded as a sign of referral bias if it was above 1.2 and statistically significant.

Results: When calculating SIRs by county 46.7 percent of the max/min ratios signified referral bias. The percentage was 5.5 when calculating SIRs by industry and only 1.7 when they were calculated by industry adjusted for county.

Conclusions: Estimates of geographical health differences are often distorted by differences in the health care organisation. Estimates of industrial health differences tend to be robust with a few identifiable exceptions. Standardisation for county will eliminate bias.

Dan Med Bull 2004;51:207-10.

INTRODUCTION

For more than a decade The Occupational Hospitalisation Register (OHR) has been an important source of data for analyses of associations between occupation and hospitalisation (1, 2). Former analyses have all been based on inpatient cases. Since 1995 the register also covers the conclusion of outpatient courses and emergency ward visits, and the first study based on these patient types is now in preparation.

Hospital registers are often used to obtain a standardised incidence ratio (SIR) to be used as a proxy for the morbidity in a group of people relative to that in a standard population. Many such studies have been performed (3-7), but only a few studies (8, 9) have evaluated the effect of differential referral practice by health system.

In Denmark the counties own the hospitals with the exception of a national and a few small private ones. The hospital system might therefore differ by county. A treatment given in one county to full-time inpatients may be offered to outpatients in other counties. Geographical comparisons using hospital treatment as a proxy measure of the relative risk ratio may therefore be biased. Because some industries are localised in particular counties one might also expect bias to arise when comparing standardised incidence ratios across industries.

Since 1995 it has been possible to distinguish between four patient types – full-time inpatients, part-time inpatients, outpatients and emergency ward patients. The aim of the present study was to investigate whether results from comparative hospital register studies depend on which patient types we choose to include in the analysis. The first part of the investigation concerned geographical comparisons, the second concerned industrial comparisons and the third and final part concerned industrial comparisons that were to be drawn with adjustment for geographical differences in health care organisation.

MATERIAL AND METHODS

The material consisted of data obtained through a record-linkage between three national registers – the centralised civil register, the national inpatient register, and the employment classification module.

The centralised civil register contains information on gender, addresses and dates of birth, death and migrations for every person who is or has been an inhabitant of Denmark sometime between 1968 and today (10). Since 1975 every person in Denmark older than 16 years has been classified annually according to industry, occupation and employment status in the employment classification module (11, 12). Age and gender are part of the personal identification number (PIN) and are recorded practically without errors. The completeness and accuracy of these data are shown by the fact that matching from different registers on PIN was 100% complete (8). The information on industry in the employment classification module is expected to be of good quality but not without errors (8). A Danish thesis from 1998 showed good agreement for industry code for bus drivers and rather good agreement for the occupational code (13).

All gainfully employed persons in Denmark aged 20-59 years 1st January 1995 (N = 2,281,480) were followed for hospital records in the national patient register and for deaths and migrations in the centralised civil register from 1st January 1995 to 31st December 2000. Cohort members were no longer at risk of becoming a case (censored) from the date of emigration, death or first occurrence of the concerned diagnosis. Prevalent cases at baseline were not excluded because it was only possible for the inpatients.

Information on gender, age and county at baseline was extracted from the centralised civil register while information on employment status and industry at baseline was extracted from the employment classification module in 1994. If an employee had been employed in more than one industry, the industry giving the highest income was chosen.

In the follow-up period we regarded the following types of cases:

A. Full-time inpatient.
B. Full-time or part-time inpatient.
C. Full-time inpatient, part-time inpatient or outpatient.
D. Full-time inpatient, part-time inpatient, outpatient or emergency ward patient.

By full-time inpatient we mean an inpatient at a ward where care is provided around the clock. By part-time inpatient we mean an inpatient at a ward where care is only provided less than twelve hours per day. We included only the principal diagnosis.

For a variety of diagnostic groups and for each type of case we calculated age and gender standardised incidence ratios (SIR) first by county, then by industry and finally by industry with adjustment for county. We used direct standardisation (14).

The following diagnostic groups were included: Infectious and parasitic diseases (ICD-10 = A00-B99), neoplasms (C00-D48), diseases of blood and haemopoetic tissues (D50-D89), endocrine and nutrional diseases (E00-E90), diseases of the nervous system and sense organs (G00-G99), diseases of the circulatory system (I00-I99), diseases of the respiratory system (J00-J99), diseases of the digestive system (K00-K99), diseases of the musculoskeletal system and connective tissues (M00-M99), diseases of the genito-urinary system (N00-N99) and injury and poisoning (S00-T98). Table 1

Table 1. Incident cases by disease and case definition in the Danish population age 20-59 years by January 1, 1995 through December 31, 2000.

ICD-10	Diagnosis	Case_A	Case_B	Case_C	Case_D
A00-B99	Infectious and parasitic diseases	22,456	24,463	33,230	41,140
C00-D48	Neoplasms	79,073	87,173	116,969	117,417
D50-D89	Diseases of blood and haemopoetic tissues	4,989	6,155	8,244	8,579
E00-E90	Endocrine and nutrional diseases	21,531	31,298	46,139	47,527
G00-G99	Diseases of the nervous system and sense organ	23,788	28,453	49,866	54,766
I00-I99	Diseases of the circulatory system	89,855	101,486	132,147	142,065
J00-J99	Diseases of the respiratory system	47,636	51,309	78,624	87,205
K00-K99	Diseases of the digestive system	102,797	129,349	176,908	184,858
M00-M99	Diseases of the musculoskeletal system	83,103	106,891	248,897	294,759
N00-N99	Diseases of the genito-urinary system	89,887	104,359	181,686	188,313
S00-T98	Injury and poisoning	124,592	135,291	298,121	869,941

A: Full-time inpatient.
B: Full-time or part-time inpatient.
C: Full-time inpatient, part-time inpatient or outpatient.
D: Full-time inpatient, part-time inpatient, outpatient or emergency ward patient.

shows the number of cases for each of the case definitions and each main group of diagnosis.

We used the industrial classification of the National Institute of Occupational Health, which contains 58 different industrial groups (see www.ami.dk for details). The classification is an aggregation of the Danish Industrial Classification of All Economic Activities 1993 (15) which is a national version of the European Industrial Classification of All Economic Activities (NACE).

ROBUSTNESS ANALYSIS

Let *SIR_A* be the SIR-value corresponding to cases of type A, *SIR_B* the SIR-value corresponding to cases of type B etc. Let *MM* be the ratio between the maximum and the minimum of the four types of SIR-values, i.e.

$$MM = \frac{\max\ (SIR\_A, SIR\_B, SIR\_C, SIR\_D)}{\min\ (SIR\_A, SIR\_B, SIR\_C, SIR\_D)}$$

The above ratio was used as a measure of robustness and was calculated for each combination of the examined population groups and diseases.

Monte-Carlo simulation was used to estimate the expected 95th percentile of *MM* under the null hypothesis which states that the SIR-value is independent of type of case. A max/min ratio (*MM*) was considered statistically significant if it was higher than the expected 95th percentile. It was arbitrarily considered to be of practical significance if it was higher than 1.2, and it was considered to be significantly high if it reached both statistical and practical significance. A significantly high max/min ratio was regarded as evidence of referral bias.

The general principles of how to use computer aided Monte-Carlo simulation to estimate statistical parameters are well described by, for example, Ross (16) and Morgan (17). The simulations in the present study were performed by use of the statistical analysis software SAS version 8.1 on the basis of the following definitions and assumptions:

For a given diagnosis: Let n_A be the total number of cases of type *A*, and let n_B be the total number of cases of type *B* etc. Let $n_{A,i}$, $n_{B,i}$, $n_{C,i}$ and $n_{D,i}$ be the corresponding numbers in group *i.*

The number of cases of type *A* in group *i* is Poisson distributed with estimated mean

$$\lambda_{1,i} = \frac{n_A n_{D,i}}{n_D}$$

The number of cases of type *B* in group *i* is the sum of two Poisson distributed random variables where the mean of the first is $\lambda_{1,i}$ and the mean of the other is given by

$$\lambda_{2,i} = \frac{(n_B - n_A) n_{D,i}}{n_D}$$

The number of cases of type C in group i is the sum of three Poisson distributed random variables where the means of the first two are given by $\lambda_{1,i}$ and $\lambda_{2,i}$ while the mean of the third is given by

$$\lambda_{3,i} = \frac{(n_C - n_B) n_{D,i}}{n_D}$$

The number of cases of type *D* in group *i* is the sum of four Poisson distributed random variables where the means of the first three are given by $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ while the mean of the fourth is given by

$$\lambda_{4,i} = \frac{(n_D - n_C) n_{D,i}}{n_D}$$

For each diagnosis and for each group we simulated 1000 of the above quartet and calculated the corresponding *MM*-value each time. The 95th percentile was based on the distribution of the 1000 simulated *MM*-values.

RESULTS

Table 2 shows the percentage of significantly high max/min ratios (*MM*) by diagnosis and comparison category. The table is to be understood as follows: When the SIR was calculated for infectious and parasitic diseases by county, 46.7 percent of the counties had a significantly high max/min ratio. When it was calculated by industry, 5.2 percent of the industries had a significantly high max/min ratio, and when it was calculated by industry adjusted for county the percentage was 1.7. As expected, the percentage was always higher when counties were compared than it was when industries were compared, and when industries were compared the max/min ratio tended to decrease when the SIR was subjected to county-adjustment.

The occurrence of referral bias was relatively low for neoplasms and for circulatory diseases, while it was high for diseases of blood and haemopoetic tissues and for diseases of the musculoskeletal system. A flagrant example of non-robust SIR-values was obtained by comparing the SIR for musculoskeletal diseases among people in the county of Sønderjylland with that of people in the municipalities of Copenhagen and Frederiksberg. When only inpatients were included the risk ratio (RR) was 1.88 (95% CI: 1.82, 1.94), and when all cases were included the RR was 0.87 (95% CI: 0.86, 0.89).

When the SIR was calculated by industry adjusted for county, only 11 of the 638 investigated max/min ratios were significantly high. These are shown in Table 3 and are to be understood as follows: When only full-time inpatients were included the standardised incidence ratio (*SIR_A*) for infectious and parasitic diseases among people in the hospital industry was 106. When both full-time and part-time inpatients were included the standardised incidence ratio (*SIR_B*) was 107. When full-time inpatients, part-time inpatients and outpatients were included the standardised incidence ratio (*SIR_C*) was 128, and when all of the last mentioned patients plus emergency ward patients were included the standardised incidence ratio (*SIR_D*) was 128. The maximum of the four SIR-values was 128 and the minimum was 106 that makes the max/min ratio (*MM*) equal to 1.21.

The hospital industry was the only one, of the industries shown in Table 3, whose referral pattern remained the same for all diagnostic

Table 2. Percentage of significantly high* max/min ratios by diagnosis and comparison category.

Diagnosis		County	Industry	Industry with county-adjustment
A00-B99	Infectious and parasitic diseases	46.7	5.2	1.7
C00-D48	Neoplasms	6.7	0.0	0.0
D50-D89	Diseases of blood and haemopoetic tissues	46.7	5.2	3.4
E00-E90	Endocrine and nutritional diseases	80.0	8.6	0.0
G00-G99	Diseases of the nervous system and sense organs	53.3	10.3	6.9
I00-I99	Diseases of the circulatory system	6.7	0.0	0.0
J00-J99	Diseases of the respiratory system	46.7	5.2	1.7
K00-K99	Diseases of the digestive system	53.3	0.0	0.0
M00-M99	Diseases of the musculoskeletal system	80.0	5.2	0.0
N00-N99	Diseases of the genito-urinary system	33.3	1.7	0.0
S00-T98	Injury and poisoning	60.0	19.0	6.9

*) Above 1.2 and statistically significant at a 0.05 level.

Table 3. Significantly high* county-adjusted max/min ratios (MM) and standardised incidence ratios for each of the four case types by diagnosis and industry.

Diagnosis (Code)	Industry (Code)	SIR_A	SIR_B	SIR_C	SIR_D	MM
Infectious and parasitic diseases (A00-B99)	Hospitals (440)	106	107	128	128	1.21
Diseases of blood and haemopoetic tissues (D50-D89)	Printing works and publishing (130)	64	73	89	88	1.40
	Pharmaceutical industry (260)	70	85	103	109	1.56
Diseases of the nervous system and sense organ (G00-G99)	Metal and steelworks, and foundries (010)	82	86	158	153	1.94
	Paper, cardboard and bookbinding industries (140)	106	114	129	124	1.21
	Garage (282)	126	117	97	98	1.30
	Slaughterhouse industry (370)	110	128	137	136	1.25
Diseases of the respiratory system (J00-J99)	Hospitals (440)	105	105	125	128	1.22
Injury and poisoning (S00-T98)	Fire service, lighthouse and salvage corps (180)	111	111	118	139	1.25
	Car dealers (281)	84	86	96	104	1.24
	General practitioners, dentists etc. (471)	92	93	87	72	1.29

*) Above 1.2 and statistically significant at a 0.05 level.

Table 4. Standardised incidence ratios for each of the four case types and max/min ratios (MM) by diagnosis among people employed in Danish hospitals.

Diagnosis (Code)	SIR_A	SIR_B	SIR_C	SIR_D	MM
Infectious and parasitic diseases (A00-B99)	106	107	128	128	1.21*
Neoplasms (C00-D48)	107	109	120	120	1.12*
Diseases of blood and haemopoetic tissues (D50-D89)	88	91	103	103	1.17*
Endocrine and nutritional diseases (E00-E90)	94	97	99	99	1.05
Diseases of the nervous system and sense organ (G00-G99)	108	107	113	112	1.05
Diseases of the circulatory system (I00-I99)	110	115	120	122	1.11*
Diseases of the respiratory system (J00-J99)	105	105	125	128	1.22*
Diseases of the digestive system (K00-K99)	102	103	108	108	1.06*
Diseases of the musculoskeletal system (M00-M99)	114	112	126	128	1.14*
Diseases of the genito-urinary system (N00-N99)	109	110	112	112	1.03
Injury and poisoning (S00-T98)	104	106	119	125	1.19*

*) Statistically significant at significance level 0.05.

groups. The SIR-values for this industry were always higher when outpatients were included in the case definition than they were when only inpatients were included (see Table 4).

DISCUSSION

The results of the present study indicate that the magnitude of bias due to differences in health care organisation would be far from negligible if the Danish hospital register was used to investigate geographical differences in health. This could be due to tradition, the distance to a hospital or which offer one gets in a county: a general practitioner or a hospital visit. The results also indicate that most SIRs calculated by industry are robust when controlling for the effect of county. Our study does, however, not exclude the possibility that some counties are treating fewer cases independently of patient type.

We chose not to regard any particular patient type as a golden standard. Inpatient discharge diagnoses may be better clinically founded, but if people with a certain disease were treated as inpatients in one county and as outpatients in another the least biased measure would be achieved by lumping the patient types together. Other advantages of using a case definition including all patients are: first, a proxy based on all patients may detect an incidence earlier than a measure based on inpatient discharges, and second, on average three times as many cases are available for analysis. We have not excluded prevalent cases at baseline because it was only possible for one patient type. Unpublished analysis on the inpatient cohort shows that exclusion of prevalent cases hardly changes the SIR.

An outpatient treatment course is often associated with lost working hours due to travels between the workplace and the hospital. This inconvenience would not exist if the treatment was delivered at the workplace. It was therefore not surprising to find that people in the hospital sector were more inclined than others to be treated as outpatients. Neither was it surprising to find that the SIR for injuries decreased among people employed at general practitioners when emergency ward visits were included in the case definition. It is well known that general practitioners can handle a large proportion of the injuries that are treated at emergency wards. People working at a general practitioner might be less inclined to visit emergency wards in such instances.

The SIR for diseases of the nervous system and sense organs increased drastically among people employed in metal and steelworks and foundries when outpatients were included in the case definition. Such a pattern could be the result of a detection bias if for instance a screening for toxic encephalophathy took place among people employed in the metal industry but not among people employed in other industries. We have no specific information that can explain the pattern seen in this study.

With the above exceptions, the Danish hospital register seems to be well suited for industrial health analyses. Our study does, however, not exclude the possibility that people in some industries are treated more frequently than people in other industries independently of patient type. The consequences of a cox arthrosis would, for example, differ between a manual construction worker and an office clerk.

There are also a number of other methodological problems one should be aware of when using the register to estimate disease frequencies in the population. For obvious reasons the register can only be used to describe the occurrence of diseases that require hospitalization or outpatient treatment at a hospital. With regard to diagnoses frequently treated in primary health care, analyses based on the hospital register will only provide information on a selected proportion (the most serious cases) of the disease spectrum.

In many studies disease incidences are compared within different geographical regions (3, 18) and within different occupations (19). Some studies report that differences in hospitalization between geographical regions seem to be partly explainable by differences in admission practices (18).

When planning future hospital based studies, as well as in the interpretation of such studies performed in the past, one should be aware of the magnitude of bias that might occur from regional differences in health care organisation.

In conclusion we found: firstly, estimates based on geographical differences reflect the organisation of the health care system rather than differences in disease pattern, and secondly, estimates based on industries are likely to be robust with a few identifiable exceptions and, thirdly, standardisation for county will eliminate bias.

ACKNOWLEDGEMENT

The Danish Health Insurance Fund financed this study: grant 2003/B035. The authors are indebted to Otto Andersen, Statistics Denmark, for help with the access to and linkage of the original data.

REFERENCES

1. Tüchsen F, Bach E, Marmot M. Occupation and hospitalization with ischemic heart diseases: A new nationwide surveillance system based on hospital admissions. Int J Epidemiol 1992; 21:450-9.
2. Baarts C, Mikkelsen KL, Hannerz H, Tüchsen F. Use of a national hospitalization register to identify industrial sectors carrying high risk of severe accidents. A three-year cohort study of more than 90,000 Danish men. Am J Ind Med 2000; 39:619-27.
3. Hammar N, Larsen FF, Faire U de. Are geographical differences and time trends in myocardial infarction incidence in Sweden real? Validity of hospital discharge diagnoses. J Clin Epidemiol 1994; 47:685-93.
4. Merlo J, Lindblad U, Pessah-Rasmussen H, Hedblad B, Rastam J, Isacsson SO et al. Comparison of different procedures to identify probable cases of myocardial infarction and stroke in two Swedish prospective cohort studies using local and national routine registers. Eur J Epidemiol 2000; 16:235-43.
5. Servadei F, Verlicchi A, Soldano F, Zanotti B, Piffer S. Descriptive epidemiology of head injury in Romagna and Trentino. Comparison between two geographically different Italian regions. Neuroepidemiology 2002; 21:297-304.
6. Sheikh A, Alves B. Age, sex, geographical and socio-economic variations in admissions for anaphylaxis: analysis of four years of English hospital data. Clin Exp Allergy 2001; 31:1571-6.
7. Homa DM, Mannino DM, Redd SC. Regional differences in hospitalizations for asthma in the United States, 1988-1996. J Asthma 2002; 39:449-55.
8. Tüchsen F, Andersen O, Olsen J. Referral bias among health workers in studies using hospitalization as a proxy measure of the underlying incidence rate. J Clin Epidemiol 1996; 49:791-4.
9. Mahonen M, Salomaa V, Brommels M, Molarius A, Miettinen H, Pyorala K et al. The validity of hospital discharge register data on coronary heart disease in Finland. Eur J Epidemiol 1997; 13:403-15.
10. Soll-Johanning H, Bach E, Olsen JH, Tüchsen F. Cancer incidence in urban bus drivers and tramway employees: a retrospective cohort study. Occup Environ Med 1998; 55:594-8.
11. Tüchsen F, Bach E. Erhverv og hospitalsindlæggelse 1981-1984. Arbejdsmiljøfondet, Copenhagen, 1992.
12. Statistics, Denmark. Personstatistik i Danmark. Et registerbaseret statistiksystem. Statistics, Denmark, 1994.
13. Bach E. Validering af EHR – Et arbejdsepidemilogisk moniteringssystem. Copenhagen: Arbejdsmiljøinstittut, 1998.
14. Rothman KJ. Modern epidemiology. Boston/Toronto: Little, Brown and Company, 1986.
15. Statistics, Denmark. Danish industrial classification of all economic activities. 2. ed. Statistics, Denmark. 1996.
16. Ross SM. Introduction to probability models. London: Academic Press, 1989.
17. Morgan BJT. Elements of simulation. London: Chapman & Hall. 1984.
18. Wickman M, Farahmand BY, Persson PG, Pershagen G. Hospitalization for lower respiratory disease during 20 yrs among under 5 yr old children in Stockholm county: a population based survey. Eur Respir J 1998; 11:366-70.
19. Leino-Arjas P, Kaila-Kangas L, Notkola V, Ilmo K, Mutanen P. Inpatient hospital care for back disorders in relation to industry and occupation and industry in Finland. Scand J Work Environ Health 2002; 28:304-13.