ORIGINAL ARTICLE

On the need of validating inpatient registers

C Nordgren^{1,2,3}

¹Certec, Department of Design Sciences, Lund University, Lund, Sweden; ²Neurotec Department, Karolinska Institutet, Stockholm, Sweden and ³Spinalis SCI Research Unit, Karolinska Institutet, Stockholm, Sweden

Study design: Register study.

Objectives: To design and implement a validation process to check the completeness of the Hospital Discharge Register (HDR) held by the Swedish National Board of Health and Welfare.

Setting: Sweden.

Methods: An accurate traumatic spinal cord injury prevalence group ($n = 495$) was acquired from the Swedish Spinalis Clinic. A register control was performed on the group by raising three questions to check the validity of the HDR: Is an inpatient stay registered in association with the injury date? Is the reported first length of stay plausible, given the level and extent of injury? Are all the anticipated care and/or rehabilitation providers represented in the HDR?

Results: For 62% (of 413 cases) the first registered hospitalization date correlated with the injury date. For the other 38%, hospitalization was reported to start between 2 and 8651 days after injury. Considering the level and extent of injury, individuals were reported to have unrealistically short initial hospitalization. The prevalence group visited 42 different hospitals and 47 clinics. Five rehabilitation clinics, though, were not reported.

Conclusions: The HDR is a valuable source when conducting epidemiological and health services research. However, using the register without any validation process could, as detected in the investigated diagnosis group, lead to a severe underestimation of the inpatient usage. The study showed that systematic errors could be detected by means of extensive knowledge of the diagnosis group.

Spinal Cord (2008) 46, 748–752; doi:10.1038/sc.2008.42; published online 13 May 2008

Keywords: administrative database; inpatient care utilization; inpatient register; LOS; traumatic spinal cord injury; validation process

Introduction

Understanding the nature and extent of health service utilization is essential for prognosticating future care utilization and for conducting economic evaluations, epidemiological studies and clinical trials. Principally, there are three different ways to collect utilization data: patient supplied, medical records or administrative databases. Several studies have been conducted evaluating the accuracy of the patient reported method, for example.¹ Both over- and under-reporting of hospitalization episodes have been detected. Horwitz *et al.*² assessed the reliability of epidemiological data, in terms of different health conditions, from medical records and proposed strategies for improving the basic quality. Utilizing administrative databases or secondary data involves analysis after collection without a specific research purpose and without interference by the researcher.³

In recent years, data retrieval has been facilitated by the vast amount of health care information collected and stored in databases and registers. The advantages of using these secondary sources are substantial time and cost savings, the size of the sample, the representativeness and the reduced likelihood of bias due to for example, recall, nonresponse and effect on the diagnostic process of attention caused by the research question.³ Despite these advantages, this method has possible shortcomings that must be dealt with. Those mentioned stem from the system *per se* and include structural weaknesses and biases due to incomplete and inconsistent reporting and coding. Furthermore, completeness, accuracy, validity and reliability are critical issues.⁴

A common problem for many administrative database researchers is whether the diagnosis code in question detects all true cases in these databases, for example.⁵ Blomqvist⁴ discusses three methodological approaches for assessing the completeness of registration of cases. The first is to compare the data source with several other independent data sources, the second is review of medical records and the third builds on a comparison between total number of cases in different

Correspondence: C Nordgren, Licentiate of Philosophy, Department of Design Sciences, Certec, Box 118, Lund S-221 00, Sweden.
 E-mail: camilla.nordgren@certec.lth.se
 Received 5 April 2007; revised 27 February 2008; accepted 2 March 2008; published online 13 May 2008

sources. De Vet *et al.*⁶ provide a theoretical background for performing and reading systematic reviews of diagnostic studies by discussing methodological quality in terms of internal and external validity. Rosen⁷ argues that validation has been studied and reported. However, determining true cases is not enough. In analyses examining health care utilization, the data must also be validated. Nevertheless, studies seem to lack descriptions of the validation process and its application to the data included.

Statistics of diseases and surgical treatments of patients have a long history in Sweden. In the 1960s, the National Board of Health and Welfare started to collect data on individual patients who had been treated as inpatients at public hospitals. The county of Stockholm has since 1972 reported all inpatient care to the Hospital Discharge Register (HDR) (www.socialstyrelsen.se/en/Statistics/statsbysubject/The+Swedish+Hospital+Discharge+Register.htm). The register contains different types of information: data on the patient and hospital, administrative data such as date of admission and discharge, and medical data with diagnosis and surgical procedures. For all records reported, data and quality controls are carried out. In an international evaluation of Swedish public health research, it was concluded that Sweden is one of the world leaders in public health research including epidemiology and register-based research.⁸ The use of the HDR has resulted in numerous research articles in different fields.

This paper examines the inpatient care utilization registered in the HDR in an accurate diagnosis group: traumatic spinal cord injury (SCI). While diagnosis detection is a problem for many researchers, this study has the advantage of having access to a population-based cohort of individuals with traumatic SCI through Spinalis, a comprehensive regional SCI outpatient clinic in the greater Stockholm area and the island of Gotland, comprising about 1.9 million inhabitants. The SCI database (Swedish Spinalis Clinic Database, SSCD) was established in the beginning of the 1990s from a survey of regional registers and after individual patient contacts by the Spinalis rehab team and a review of medical records, interviews and physical examinations of subjects. The dropout rate was 6.9%.⁹ This procedure assured the accuracy of the database as an SCI health care database. Spinalis is a part of the established referral procedure, which insures further incidence inclusion in the database.

Methods

This study was a register control, first of the data for the Spinalis group, one by one, to decide whether they were to be included in the prevalence group or not, then of how each of the group members appears in the HDR.

The subject group

In June 1999, the author submitted a retrieval request to the SSCD with the inclusion criteria of living patients with traumatic SCI. The present study is based on the 495 persons who met these requirements. Data about date of birth, gender, injury date, level and extent of injury (ASIA) were

collected for each individual. Since the HDR is complete in the investigated region from 1972 all cases with an earlier injury date were excluded ($n=58$). For the remaining 437 persons, the Swedish national registration office was checked to detect those who had died before 1999 ($n=1$). This case was excluded, as were seven others due to incorrect data (erroneous personal identity number and uncertainty regarding the diagnosis) while one was added. Thus the final total prevalence group consisted of 430 traumatic SCI cases. This study was approved by the Ethics Committee of Karolinska Institutet in Stockholm, Sweden.

The elaborated Hospital Discharge Register

The unique personal identity number assigned to each Swedish resident was used to obtain data from the HDR. Information was collected about each person's registration regarding hospital/institution, clinic, unit, date of admission and discharge. Data was examined from 1972 through 2002.

As the HDR consists of numerous individual pieces of information validating such an extensive register requires investigation on an individual level. Understanding each person's incentive and condition for utilizing inpatient care requires knowledge about individual characteristics. Three fundamental questions were formulated to penetrate the validity of the material:

- (1) Is an inpatient stay registered in association with the injury date?
- (2) Is the reported first hospitalization episode (length of stay, LOS) plausible, given the level and extent of injury?
- (3) Are all the anticipated care and/or rehabilitation providers represented in the material?

Results

Is an inpatient stay associated with the injury date?

For 22 individuals, the exact day and/or month of injury was not specified in their medical records. For those individuals ($n=12$) for whom the month was specified, the first of the month was calculated as the injury day. For those lacking both an injury month and day ($n=10$), the first of January of that year was specified as the injury day and month.

Table 1 shows the number of days between injury date and the first registered inpatient stay. Seventeen persons did not have any inpatient care at all reported after the injury date. Sixteen individuals actually had an inpatient stay already registered the day before the actual injury. About 62% of the group had an inpatient stay reported in direct connection with the injury date ($-1, 0, 1$ days), that is, 38% had their first hospital stay registered 2 days or more after the injury date.

Is the reported first hospitalization episode plausible, given the level and extent of injury?

The number of inpatient days was checked for the individuals with a first hospitalization stay in direct conjunction with the injury date. Included were those with a reported inpatient stay of $-1, 0$ and 1 day in relation to the injury date ($n=257$). Table 2 shows the distribution according to

Table 1 Number of days between injury date and first registered inpatient stay

Number of days between injury date and first registered inpatient stay	Number of persons (N = 413)	
	With exact day and month of injury (N = 391)	Without exact day and month of injury (N = 22)
-1, 0 or 1	257	0
2-10	43	3
11-100	38	6
101-1000	19	3
1001-10 000	34	10

Table 2 Number of individuals distributed according to neurological classification

Neurological level of lesion	ASIA A (complete) N	ASIA B-D (incomplete) N	ASIA E N	ASIA no information N
Cervical	29	86		
Thoracic	36	33		1
Lumbar	3	40		1
Sacral	2	2		
No level			23	
No information				1

Figure 1 Length of stay in the first hospitalization episode after injury date distributed according to level and extent of injury.

level and extent of injury and Figure 1 provides information of LOS for the initial hospitalization. Length of stay ranges from 2-653 days, with a mean of 126 and a median of 108 days.

Initial LOS for some individuals seemed to be interrupted by a holiday such as Christmas or New Year. For 20

individuals this was indeed the case. These inpatient stays had an interruption that lasted between 3-31 days, and by including these individuals' inpatient care after the interruption, the total stay was thus prolonged by 2-171 days.

Are all the anticipated care and/or rehabilitation providers represented in the material?

The inpatient stay prevalence group had utilized care at 42 different hospitals/institutions in the Stockholm area, and 47 different clinics from injury date through 2002.

Discussion

This paper examines the validity of the content in the HDR. The use of a verified population-based SCI health care database including additional controls is supposed to insure the validity of the diagnosis group investigated and the validity process can thus focus on the inpatient data. Considerable knowledge of the diagnosis and its effects in terms of health care utilization proved to be a prerequisite when putting questions to the HDR. Interpreting data and estimating the validity was absolutely necessary to determine under-reporting and lack of reporting.

Extensive investigations were made regarding the patients included in the health care database, for example, if and when a person had died and verification of his/her residential registration. Atypical inpatient entries, such as being admitted to inpatient stay and sustaining a traumatic SCI during that time, were also examined.

To investigate inpatient care in an administrative database presupposes that the disease or condition of the individual has been reported and recorded in the system. Blomqvist⁴ points out three factors that influence this: the care-seeking behavior of the person, the supply of health care and the physician's propensity to admit patients. The nature of an SCI does not leave a person in doubt whether to seek care or not. Nor can the health care organization in Sweden be seen as a hindrance to inclusion in the information system.

Is an inpatient stay associated with the injury date?

A traumatic SCI is an acute and serious condition where immediate care is necessary. The alternative of not seeking care or waiting a couple of days occurs, if ever, in exceptional cases. Some studies that analyze LOS separate acute care hospitalization from rehabilitation hospitalization and stipulate a time limit from the injury date as an inclusion criterion.¹⁰⁻¹³ Others calculate the LOS from the day of injury,¹⁴ include individuals admitted for their initial episode¹⁵ or use date of the index hospitalization as the date of the SCI.¹⁶ All these studies lack further analysis regarding possible divergence between the onset of traumatic SCI and admission date. Putzke *et al.*¹⁷ and Fine *et al.*,¹⁸ on the other hand, include individuals admitted on the first day of injury. In the present study, more than one-third of the prevalence group did not have an initial inpatient registration (-1, 0, +1 day) in conjunction with the injury date. The time lapse between injury date and first inpatient stay should be interpreted as possible days spent in inpatient care.

The prevalence population examined includes persons who moved to the investigated region with a traumatic SCI sustained earlier. This group includes immigrants and residents moving from other parts of the country. The health care system in Sweden allows people to seek health care anywhere in the country. It is most common, however, to utilize the health care system in one's own county. As the patient data originated from one region, utilization of care in other parts of the country was not included. In 2002, controls carried out at the Swedish National Registration Office resulted in finding 31 persons who were registered in a region other than Stockholm. Not being a resident of Stockholm in 2002 does not, however, say anything about place of registration at the time of injury. If a person had moved out of the area and then moved back before April 2002, these movements could not be traced. One explanation for the high figure of divergence besides immigrants could be individuals injured during a temporary stay abroad or in another region in the country. These individuals are typically moved to their residential hospitals after the first trauma period and so, their initial hospitalization would not have been recorded in the investigated register.

Is the reported first hospitalization episode plausible given the level and extent of injury?

Comparing results from other investigations is a way of determining whether the reported inpatient figures are plausible. Other studies on traumatic SCI and initial LOS show a great variation in the number of days spent in hospital/rehabilitation with figures ranging from 1 up to 4742 days.¹⁹ These studies all have different inclusion criteria, terms of description and grouping when it comes to etiology, age, injury year, level and extent of injury and type of care.^{10–13,15–17,20} This impedes comparison. No discussion has been found in any of the published studies about how realistic it is to have very few or many inpatient days registered. SCI patients at high risk of extended LOS, referred to as outliers, are defined in a study of Burnett *et al.*²¹ as patients whose LOS exceed the mean by more than two standard deviations, while Cifu *et al.*¹⁰ use four standard deviations. These studies discuss the outliers in terms of long LOS, but it should also be of considerable interest to focus on outliers in the other direction, in terms of short LOS.

The present findings do not differentiate type of care: acute vs rehabilitation. The data represent days of inpatient stay irrespective of type, in a public hospital. Some researchers do not examine the acute LOS, for example.^{15,16} Several studies^{11–13,20} spring from persons discharged from the Model System with an enrol system of patients being admitted within 60 days of injury. In some studies, 86–93% of all patients were admitted to the investigated clinics within 21 days of injury. It is thereby uncertain how much inpatient time the patients have consumed before the rehabilitation LOS. It can therefore be interpreted that the LOS could in some cases actually be prolonged by as much as 60 days. A lack of differentiation between the acute and rehabilitation organization could perhaps explain some of the differences and the great range of LOS.

Having a complete (ASIA A) cervical injury and utilizing only 2, 7 or 22 inpatient days seems unrealistic (Figure 1). Further investigation through other sources of data would be required to reveal the likelihood of having an extremely short or long initial inpatient stay. Knowledge about medical care organizations in your country, rehabilitation regime and, of course, the investigated diagnosis are of value when determining which cases need more examination.

Are all the anticipated care and/or rehabilitation providers represented in the material?

The HDR contains inpatient data from public hospitals. At least five institutions/rehabilitation units were not represented at all in the register. This was surprising even though these institutions are typically foundations or privately owned, because the county council regularly purchases care and rehabilitation services from these providers, making these rehabilitation beds regarded as public. One of these is also included in the established referral procedure of rehabilitation and a great majority of all incidence patients stay at this rehabilitation unit for a while.

Conclusions

Systematic errors are not easy to discover and the chance of discerning them does not correlate with the size of the sample. When found, it is possible to compensate for them. What this study reveals is that extensive knowledge of the investigated diagnosis group can be a necessity when examining and evaluating the data. Researchers using administrative databases like this one need to validate their data to attain reliable results.

Acknowledgements

This study was made possible by the kind assistance of the Spinalis unit. The author thanks Lars Werhagen, MD, PhD and Peter Kitzing, MD, PhD for valuable help. The financial support for this research, provided by Stiftelsen för bistånd åt rörelsehindrade i Skåne, is greatly appreciated.

References

- 1 Kennedy ADM, Leigh-Brown AP, Torgerson DJ, Campbell J, Grant A. Resource use data by patient report or hospital records: do they agree? *BMC Health Serv Res* 2002; 2: 2.
- 2 Horwitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. *J Chron Dis* 1984; 37: 825–831.
- 3 Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol* 1996; 25: 435–442.
- 4 Blomqvist P. *On the Use of Administrative Databases in Health Care Analyses [Dissertation]*. Karolinska Institutet: Stockholm, Sweden, 1998.
- 5 Augutis M, Malker H, Levi R. Pediatric spinal cord injury in Sweden; how to identify a cohort of rare events. *Spinal Cord* 2003; 41: 337–346.
- 6 De Vet HCW, van der Weijden T, Muris JWM, Heyrman J, Buntinx F, Knottnerus JA. Systematic reviews of diagnostic research. Considerations about assessment and incorporation of methodological quality. *Eur J Epidemiol* 2001; 17: 301–306.

- 7 Rosen M. National health data registers: a Nordic heritage to public health. *Scand J Public Health* 2002; **30**: 81–85.
- 8 Kamper-Jorgensen F, Arber S, Berkman L, Mackenbach J, Rosenstock L, Teperi J. Part 3: international evaluation of Swedish public health research. *Scand J Public Health* 2005; **33**: 46–84.
- 9 Levi R, Hultling C, Nash MS, Seiger A. The Stockholm spinal cord injury study: 1. Medical problems in a regional SCI population. *Paraplegia* 1995; **33**: 308–315.
- 10 Cifu DX, Seel RT, Kreutzer JS, McKinley WO. A multicenter investigation of age-related differences in lengths of stay, hospitalization charges, and outcomes for a matched tetraplegia sample. *Arch Phys Med Rehabil* 1999; **80**: 733–740.
- 11 Eastwood EA, Hagglund KJ, Ragnarsson KT, Gordon WA, Marino RJ. Medical rehabilitation length of stay and outcomes for persons with traumatic spinal cord injury—1990–1997. *Arch Phys Med Rehabil* 1999; **80**: 1457–1463.
- 12 McKinley WO, Seel RT, Gadi RK, Tewksbury MA. Nontraumatic vs traumatic spinal cord injury: a rehabilitation outcome comparison. *Am J Phys Med Rehabil* 2001; **80**: 693–699.
- 13 Seel RT, Huang ME, Cifu DX, Kolakowsky-Hayner SA, McKinley WO. Age-related differences in length of stays, hospitalization costs, and outcomes for an injury-matched sample of adults with paraplegia. *J Spinal Cord Med* 2001; **24**: 241–250.
- 14 Webb SB, Berzins E, Wingardner TS, Lorenzi ME. Spinal cord injury: epidemiologic implications, costs and patterns of care in 85 patients. *Arch Phys Med Rehabil* 1979; **60**: 335–340.
- 15 Tooth L, McKenna K, Geraghty T. Rehabilitation outcomes in traumatic spinal cord injury in Australia: functional status, length of stay and discharge setting. *Spinal Cord* 2003; **41**: 220–230.
- 16 Dryden DM, Saunders LD, Rowe BH, May LA, Yiannakoulis N, Svenson LW *et al*. Utilization of health services following spinal cord injury: a 6-year follow-up study. *Spinal Cord* 2004; **42**: 513–525.
- 17 Putzke JD, Richards JS, Devivo MJ. Gunshot versus nongunshot spinal cord injury—acute care and rehabilitation outcomes. *Am J Phys Med Rehabil* 2001; **80**: 366–370.
- 18 Fine PR, Stover SL, Devivo MJ. A methodology for predicting lengths of stay for spinal cord injury patients. *Inquiry* 1987; **24**: 147–156.
- 19 Samsa GP, Landsman PB, Hamilton B. Inpatient hospital utilization among veterans with traumatic spinal cord injury. *Arch Phys Med Rehabil* 1996; **77**: 1037–1043.
- 20 Greenwald BD, Seel RT, Cifu DX, Shah AN. Gender-related differences in acute rehabilitation lengths of stay, charges, and functional outcomes for a matched sample with spinal cord injury: a multicenter investigation. *Arch Phys Med Rehabil* 2001; **82**: 1181–1187.
- 21 Burnett DM, Kolakowsky-Hayner SA, Gourley EV, Cifu DX. Spinal cord injury ‘outliers’: an analysis of etiology, outcomes, and length of stay. *J Neurotrauma* 2000; **17**: 765–772.