Original Article

Pediatric spinal cord injury in Sweden; how to identify a cohort of rare events

M Augutis^{*,1}, H Malker² and R Levi³

¹FoU, Sundsvall Hospital, Sundsvall, Sweden; ²Mid Sweden Research and Development Center, Vasternorrland County Council, Sundsvall, Sweden; ³Karolinska Institute and Frosunda Center, Stockholm, Sweden

Study design: Register study enhanced and verified by medical records and personal interviews and examinations.

Settings: Sweden.

Objectives: To define a method of identifying a study population of rare events. To point out the relative importance of every step, an example is given of identifying persons who sustained traumatic spinal cord injury (SCI) in childhood.

Methods: Cases were identified in seven steps that all needed to be fulfilled, from definition of selection criteria through combination of several data sources, to the use of several verification methods.

Results: Initial screening by registers identified 384 possible cases, which however were found by subsequent analysis to include a large number of incorrect cases. At completion of all analytic steps, 35 living cases could be fully verified and 14 deceased cases could be partially verified.

Conclusions: Registers offer a practical initial source for study population identification. The screening of International Classification of Diseases codes defining SCI only included less than 30% of 'true' SCIs. Subsequently, further refinement and quality control is necessary in order to ensure validity. Such further verification is time-consuming, but nevertheless necessary in order to verify a true cohort.

Spinal Cord (2003) 41, 337-346. doi:10.1038/sj.sc.3101456

Keywords: epidemiologic methods; pediatric; population surveillance; quality control; spinal cord injury

Introduction

Vast amounts of data are available through various administrative information systems for clinical and epidemiological research.

Analyses of population-based administrative registers constitute a practical and cost-effective way to retrieve both background and effect data, as the data already exist.¹ However, there are both strengths and weaknesses to be considered when using such databases for research purposes.² Since the primary purpose of many registers is basically administrative, this inevitably influences data quality.³

Problems particularly arise when identifying cohorts of rare events that require a large background population. This becomes apparent when the diagnosis is not well defined or needs to be more specified than easily available data allow. In the literature there are many cautions about surveys based on register studies, pointing to biases in the reporting methods, data collection systems and classification patterns.⁴ The use of discharge diagnoses for identification of cases may cause problems, such as variation in coding procedures, coding errors, incomplete coding, lack of specificity in available codes and clinical errors in diagnosis.¹

Data must therefore be controlled and a review must be incorporated.^{5–9} The country of Sweden has a tradition of population registration dating back to 1749 when *Tabellverket*¹⁰ started. Hence, a number of nation-wide population-based disease or *effect* registers exist. Quality control of these registers differs depending on how data are processed and linked to other registers.³

Furthermore in Sweden there exists a unique personal identification number for every resident since the late 1940s. This, together with modern computer techniques, np

^{*}Correspondence: M Augutis, RTP, MSc, FoU, Sundsvall hospital, 851 86 Sundsvall, Sweden

facilitates easy linkage of exposure and outcome data at the individual level.³

Other sources of information may also be linked to the population registers. Such sources include data from hospitals, insurance companies, and also from less formal sources such as, for example, special interest organizations.

Finally, making inquiries through the media, for example, television, radio, newspapers and special journals oriented to professionals or patients, can be a valuable tool in estimating the reliability and nonresponse in a specific cohort.

The validity of incidence estimates is obviously dependent on how well true cases are recognized, identified and verified in the use of one or several of the data sources listed above.

Data collection of rare events poses particular problems because of the large amount of data that has to be screened in order to get a study population large enough to enable conclusions. The implication is that the verification of every single case must be made very carefully to avoid misclassifications and differences in classification.

The aim of this paper is to describe a method of identifying and verifying a population-based cohort of rare events, in this case pediatric traumatic spinal cord injury (SCI).

Method

The following method was used for identifying and verifying a study population of chronic rare events (Figure 1).

Definition and selection criteria

The study population is initially defined and criteria for inclusion and exclusion are set. It is desirable to contemplate using previously set definitions and criteria in order to establish a study population comparable with such previous studies.

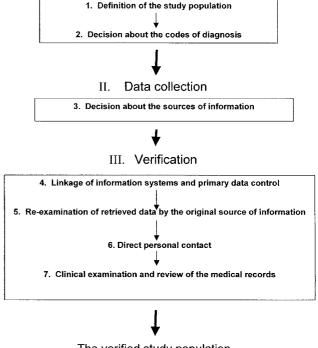
A request of data and codes of data is then performed, including all classification codes that reasonably might be used to label cases belonging to the study population. A fairly broad inclusion of classification codes is recommended in order to minimize under-reporting.

Data collection

A decision must be made regarding the sources of information that could be used in order to minimize under-reporting. In this example, nationwide population registers were used as well as data from specialized institutions and hospitals and other less formal sources.

Verification

Linkage of retrieved data from different data sources is performed in order to avoid duplicates, as well as for identifying and eliminating obvious faults.



Definition and selection criteria

L

The verified study population

Figure 1 Process chart of finding a verified study population of a rare event.

Retrieved data are then sent back to the source where the data originally were reported from, with a request of control of the diagnosis.

Further verification is then made through personal contacts in order to corroborate data from the medical records and *to ascertain the outcome* of cases.

If ambiguity as regards diagnosis and/or outcome remains at this stage.

a clinical examination and/or additional review of medical records can be performed.

Results

The identification of cases of pediatric SCI in Sweden serves as a practical example of application of identifying a cohort of rare event.

Definition and selection criteria

Pediatric SCI was defined and inclusion and exclusion criteria were set and the codes of diagnosis were chosen (Table 1).

Data collection

Three types of sources were used: nationwide population-based registers; patient-based regional registers; and patient-oriented journals/informal sources.

339

Definition	Exclusion criteria	Inclusion criteria	Requested classification codes from the Swedish National Board of Health and Welfare	
Impairment resulting from traumatic lesions of the cauda equina or the spinal cord proper. Lesions with complete sensory-motor restitution were also included, as defined by American Spinal Injury Association (ASIA), impairment Scale grade E, but with documented evidence of previous symptoms and signs of SCI for at least 2 weeks after the injury ⁷	Children with isolated root lesions or no injury to the intraspinal neural elements Children with congenital malformations of the spinal cord Children with acquired SCI of nontraumatic etiology	The injury must have occurred in Sweden during 1-1-1985 until 31-12-1996 At the time of the injury, the child should be between 0 and 15 years of age	The Swedish versi 9th edition of the Classification of D (rev. ICD 9 used i 1987–1995 and the in the 8th edition of this classificatio 806 A-X 952 A-X	International Diseases n the period eir equivalents
the injury' A definition of the period when mortality should be included is not made in this article. Later, when the incidence was estimated, mortality directly related to the traumatic event was operationally defined as death occurring within 1			767E 344 A-X	Birth injuries, injuries to the spine and the spinal cord Other paralyses

Table 1 Definition, exclusion, criteria inclusion criteria and the decision about the codes of diagnosis

In total, about 3.6 million possible events during 1-1-1985 until 31-12-1996 were included in the screening (Table 2). During the 12 years surveyed, the total population of Swedish children below 16 years of age was 20171 823 children, with an average population of 1 680 000 children/year (max/min: 1 767 000/1 613 000).¹⁵

Verification

Verification of cases was achieved by linkage of information from the various data sources (Table 3). Retrieved data were checked against the medical records, by sending it for verification by a senior staff physician at the treating hospital (Table 4) and by direct contact with the cases.

Out of the original 384 cases, 15% (N=59) were either duplicates or did not fulfill the set definition. For children dead prior to hospital admission (N=43) found in the Cause of Death Register, no detectable registration faults were found. Further verification of this group was not feasible. This left a group consisting of 282 cases identified with a possible SCI (Table 3) (Figure 2).

Of the 211 cases registered with classification code 806 (ie *SCI with fracture of the spine*), 68% (N = 144) were found to be incorrect. This first step of verification gave 28% "true" SCI (N = 64) of those registered with the

classification code 806 and 952 (N=232) (Table 4). None of the 42 cases identified by the Medical Birth and Malformation Register (MBR) could be verified as having SCI. The two cases identified through the Habilitation centers were not found in the Hospital Discharge Register (HDR) by any of the defined classification codes. Both these cases had sustained SCI as a peroperative complication at surgery of spinal deformity. Additionally, one case with verified SCI was found through an informal source, that is, through personal contact with hospital staff. In 18 cases, in the HDR and the MBR, further verification of the diagnosis was not possible since their medical records could not be found at the treating hospital (ie missing records). The department heads verified 69 cases as having SCI. Of these, 14 cases had died (Table 4) (Figure 2).

Another 17 children were excluded after controlling the diagnosis by telephone interviews. For five cases, the diagnosis could still not be verified; their medical records were therefore reviewed and a clinical examination to determine sensory/motor/neurological level of the injury was performed according to the set definition. Two of these cases did not have SCI. A third case had a birth injury and was excluded because of atypical symptoms and fell outside the scope of this study (Figure 2).

Table 2 Decisions about the sources of information and the number of cases with SCI found in each source

Sources of information used in the example	Description	Quality control within the data source	Total registered cases (children 0–15 years)	Number of children registered with SCI
Nation-wide population registers The Hospital Discharge Register (HDR) ¹⁶	Data on individual patients who have been treated as inpatients. Compulsory in Sweden since 1984	A check is made on all registrations in the HDR that all data are reported and are valid, for example, personal identification number, hospital and main diagnosis. The dropout rates are very low, usually less than 4.5%. Diagnoses might be wrong for about 10% of the cases with a large variability between different diagnoses	Includes data for around 136000 children/year under the age of 16 years (In total, 1.6 million cases were screened for SCI over a 12-year period)	249 cases
The Cause of Death Register (CDR) ¹⁷	All deaths in Sweden are reported to the CDR. Upon each death, a qualified physician must issue a death certificate before the person can be buried. This means that additional information, that is, a pathology report can be received after the death certificate is issued	Faults by coding are less than 3%, and the under-reporting to the CDR is negligible ¹⁶	The mortality among Swedish children aged 0–15 years is around 600 children/year (Over a 12-year period, around 7200 cases were screened for SCI)	57 cases
The Medical Birth and Malformation register (MBR) ¹⁹	Includes injuries because of birth. The register is based on copies of the medical records used at antenatal care, delivery care and at the examination of the newborn infant	The register information is checked each year with the corresponding information on infants in the civil registration. The under- reporting is about 1% each year. ¹⁸ The accuracy of the diagnosis is not because of any routine quality control or follow-up	Around 100 000 children are born yearly in Sweden and are included in the register (Around 1.2 million cases were screened for SCI during 12 years)	42 cases

Table 2 (continued)

Sources of information used in the example	Description	Quality control within the data source	Total registered cases (children 0–15 years)	Number of children registered with SCI
<i>Local or regional registers</i> Habilitation ²⁰	Children between the age 0 and 18 years with chronic SCI in Sweden may, after the initial care and rehabilitation, be referred to a nation-wide organization of Habilitation centers. Usually, the Habilitation center consists of many small multidisciplinary teams who work on a local basis, close to the child's home	The registers are new and under construction. No quality control is performed	Each county has its own organization and inclusion/exclusion criteria 50 centers were surveyed There is no information found over the total number of cases included in these local 'registers'	26 cases
Informal sources and media Advertisements in three patient-oriented journals Informal sources Total			A circulation of around 31 000 copies/issue	9 cases 1 case
			3.6 million possible events were included in the screening	384 cases

All cases, $N=384$				
Included Further verification	Excluded			
	Obvious faults N=Excluded cases according to the	Duplicates N=Cases registered in the HDR and in an	Deaths prior to hospital admission N=cases found exclusively in the CDR	
	set definition and the inclusion criteria in different sources	additional source		
282 children identified with a possible SCI	HDR, $N = 12/249$ Habilitation, $N = 16/26$ Advertisements, $N = 3/9$	CDR, $N = 14/57$ Habilitation, $N = 8/26$ Advertisements, $N = 6/9$	Of the 57 children with suspected SCI identified through the CDR, 43 children had died prior to hospital admission and subsequently not found in the HDR. In the example no further reviews of their autopsy protocols were obtained	
282 cases	31 cases	28 cases	43 cases	

Table 4 Description of the requested codes of diagnosis in the HDR, the MBR, cases from the Habilitation centers, informalsources and the result after verification by the department heads

Sources and classification codes (rev. ICD9)	Total checked	Other diagnosis	Missing records	Incomplete identification	'True' SCI
HDR					
SCI with fracture 806	211	144*	9	0	58
SCI without fracture 952	21	14	1	0	6
Birth injury 767E	3	1	1	0	1
Other paralyses-344	2	1	0	0	1
MBR					
Birth injury 767E	42	33	7	2	0
Habilitation centers					
Children reported as having SCI	2	0	0	0	2
Informal sources					
Children reported as having SCI	1	0	0	0	1
Total	282	193	18	2	69

*In total, 128 children were verified by the department heads as having sustained code 805, Fracture of the spine without a SCI.

From the original screening of 384 cases, 35 thus had qualified the diagnosis and set definition of SCI and were included in the verified study population of children with SCI. Figure 2 illustrates the algorithm of this process.

Discussion

Definition and selection criteria

Definition of the study population The definition is set to find a specific study population. In order to make comparisons with other studies, the set definition and the inclusion/exclusion criteria must be specified in detail.

To obtain enough cases, the study often has to cover a long period, in this case 12 years. Ideally, changes in time trends should have been taken into consideration, but this was not possible.

The definition of the period when mortality should be included is essential for estimates of mortality in a

A large background population is needed for a population-based study dealing with rare events.

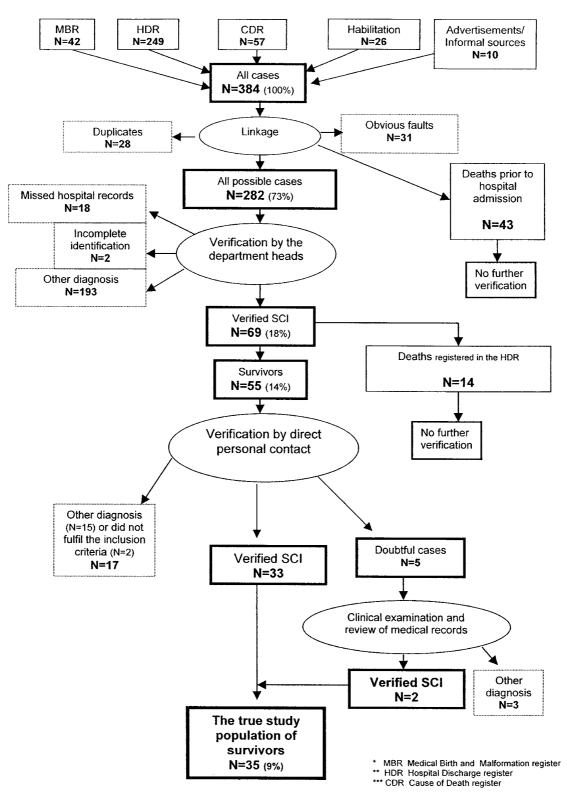


Figure 2 Flow-chart of the process of finding the true study population, quantified with the example of pediatric SCI in Sweden.

specified diagnosis. Later when the incidence was estimated, mortality directly related to the traumatic event was operationally defined as death occurring within 1 year postinjury.¹¹ The time limit was chosen

in order to make comparisons with other studies.¹² This limit can be questioned when a person with SCI is exposed to a higher mortality risk during the entire lifetime.²¹ Another aspect of this problem is to consider

343

the cause of death, that is, in this case, whether the child died of SCI or not. Such definition can lead to subjectivity and bias and might cause problems when comparing studies.

Decision about the codes of diagnosis To ensure comprehensiveness, all codes that might hide the criteria for the set definition should be included in the first screening.

Our example revealed three important problems.

Firstly, the national and the international classification codes are not universally comparable.^{13,14,22} The revised Swedish version of the classification code 806 lacked specificity as compared to the set definition. Cases with *isolated involvement of the spinal roots* could thus be registered as 806, although these cases were excluded in the set definition.

The International Classification of Diseases (ICD) (9th rev.)²² lack even more specificity than the Swedish version^{13,14} since the 806 code also includes *concussion* of the cord. Knowledge about such inconsistencies is essential when comparing studies from different countries.

Secondly, the different revisions may lack comparability. Since 1997, the revised ICD 10 is used in Sweden.²³ In the 10th revision, several additional classification codes have been included that could have hidden the set definition of SCI and thus are necessary to retrieve. This underlines the importance of specifying which revision is used and which differences, if any, are there in a national revision as compared to an international one.

Thirdly, it is almost impossible to cover all codes of diagnoses that could hide the set definition. In the example, two cases found through the Habilitation centers had an SCI as a complication at surgery for spinal deformity, that is, they were registered with the classification code of scoliosis. Also one case, which was found through informal contacts, was registered with another classification code.

Data collection

Decision about the sources of information Nation-wide registers provide a practical way to retrieve data for screening of possible cases. The population-based registers at the National Board of Health and Welfare are examples of such information systems and provide a large source of information open for research.

However, in the present case a problem arose, as the HDR did not cover all county councils, during the first 2 years surveyed. In this particular situation, this was probably negligible, but even so it is a bias to consider. To collect data from several sources is a much more troublesome task, open for nonresponse and bias of different kinds because of misunderstandings, as well as differences in collection and administration of information. In this case, data from the Habilitation centers were difficult to retrieve since several centers had only

recently started to register diagnosis and therefore could fail to report cases with SCI.

Although the public is often a good and helpful source of information, the advertisement in the patientorientated journals together with the Habilitation centers provided less information than the administrative registers. On the other hand, as in the example, using such sources of information can provide information about nonresponse in the registers, especially when the condition under study is chronic in nature.

Verification

Linkage of information systems and primary data control The first screening of the study population from the different sources will include cases that actually do not fulfill the definition of the study population, as well as duplicates retrieved from different sources. Detection of duplicates was possible since every Swedish citizen has a unique personal identification number.

Re-examination of retrieved data by the original source of information In order to verify retrieved cases, department heads at the principal treating clinic were contacted in writing with a request for diagnosis confirmation by crosschecking with medical records.

In this case, verification proved to be very time consuming, especially as register data from the HDR lack precision according to the set definition and the inclusion criteria. Until further verification was made, the re-examination of data by the original source of information revealed that less than 30% of the registered classification code 806 and 952 had a "true" SCI.

It is important to include an informative letter of the project plan to attract the interest of the head of the department. The best way to generate interest in the project is to have personal contact with the verifying physician. Repeated reminders are also required. None-theless, a number of biases must be considered. The quality of the verification depends on factors like time, knowledge, level of care²⁴ and interest of the verifier.

In this study, the most common error in the HDR was because of coding, for example, the case had a spinal fracture without any neurological symptoms. The diagnosis was correctly expressed, but the wrong classification code was chosen.

A common fault was interchanging of classification code digits, for example, 816 or 860. This is an important type of bias to consider as well as using and designing registers.

Difficulties in diagnosing SCI at birth are well documented.²⁵ In this case, none of the cases registered as SCI in the Medical Birth and Malformation Register had sustained an SCI.

According to Swedish law, it is mandatory to save the medical records and store them for easy retrieval. In the study, medical records were nevertheless found to be missing, all from cases with short hospitalization (<10 days).

In a study dealing with serious and chronic events, it could be assumed that such a nonresponse would not extensively influence the result.

Direct contact Information must be checked further and supplemented by other sources, for example, the patient or a relative. Mailed questionnaires are often less efficient because of high nonresponse rates. Verification can instead preferably be done by telephone (if the group is suitably small).

Clinical examination and review of the medical records It may be necessary to review medical records and/or perform a clinical examination if there is still doubt whether the patient should be included in the defined study population.

If further studies on the group are planned, caution must however be taken not to influence the subsequent study.

Ethical considerations In human studies, approval must typically be obtained from an ethics committee before the study can start. Furthermore, according to the Swedish law, persons included in the study must give their informed consent. Before personal contact with a patient is made, permission from the treating physician should also be obtained.

Informed consent from the patients themselves could not be received until the group was identified and verified. Even after verification by the responsible physician, there might still be many doubtful cases, which is an ethical problem, for example, if a patient is contacted and the diagnosis turns out to be wrong.

Conclusions

Registers may be an efficient way to screen for rareevent populations. However, this study shows that the registered ICD-codes 806 (*SCI with fracture of the spine*) and 952 (*SCI without fracture of the spine*) gave less than 30% 'true' spinal cord injuries. Therefore, further refinement and quality control is necessary in order to verify a true cohort. Quality control proved to be time consuming, but all steps were necessary in finding a true study population.

Of the original cases (N=384) identified by the screening procedure, less than 9% (N=35) were found to be correctly classified.

The Ethics Committee at the University in Umeå approved the project (permit no. Um dnr 96-289). Permission was received from the Swedish Data Inspection Board.

Acknowledgements

This work was supported by a grant from *Cancer och Trafikskadades Riksförbund* and Våsternorrland County Council.

References

- 1 Steinberg EP, Whittle J, Anderson GF. Impact of claims data research on clinical practice. *Int J Tech Asess Health Care* 1990; **6:** 282–287.
- 2 Flood AB. Peaks and pits of using large databases to measure quality of care. *Int J Tech Assess Health Care* 1990; **6**: 253–262.
- 3 Malker H. Register epidemiology in the identification of cancer risks. In: *Arbete och Hälsa vetenskaplig skriftserie*. National Board of Occupational Safety and Health: Solna 1988, pp 1–50.
- 4 Dickman CA *et al.* Pediatric spinal trauma: vertebral column and spinal cord injuries in children. *Pediatr Neurosci* 1989; **15**: 237–256.
- 5 Thurman D et al. Surveillance of spinal cord injuries in Utah, USA. Paraplegia 1994; **32:** 665–669.
- 6 Levi R. The Stockholm Spinal Cord Injury Study: medical, economical and psycho-social outcomes in a prevalence population. Doctoral dissertation. Karolinska Institute: Stockholm 1996.
- 7 Johnson R *et al.* Evaluating sources of traumatic spinal cord injury surveillance data in Colorado. *Am J Epidemiol* 1997; **146:** 266–272.
- 8 Surkin J et al. Spinal cord in Mississippi, findings and evaluation, 1992–1994. Spine 2000; 25: 716–721.
- 9 Van Asbeck F, Post M, Pangalila R. An epidemiological description of spinal cord injuries in The Netherlands in 1994. *Spinal Cord* 2000; **7:** 420–424.
- 10 Sköld P. Kunskap och Kontroll. Den svenska befolkningsstatistikens historia. Umeå: Demografiska databasen, 2001.
- 11 Augutis M, Levi R. Pediatric spinal cord injury in Sweden: incidence, etiology and outcome. *Spinal Cord* (in press).
- 12 DeVivo M, Krause J, Lammertse D. Recent trends in mortality and causes of death among persons with spinal cord injury. *Arch Phys Med Rehabil* 1999; **80**: 1411–1419.
- 13 Socialstyrelsen. Klassifikation av sjukdomar 1987. Socialstyr: Stockholm 1986.
- 14 Socialstyrelsen. Klassifikation av sjukdomar 1968: tillrättalagd för sjukhusbruk = International statistical classification of diseases, injuries and causes of death, 1965 revision adapted for indexing of hospital records and morbidity statistics. Systematisk förteckning = Systematic list. 4. uppl. edn. Utg.: Stockholm 1968
- 15 Statistics Sweden, Population Statistics. Part III, Distribution by sex, age, marital status and citizenship by municipality, etc. 1985–1996.
- 16 Statistics NCBo. In-patient statistics 1981 in somatic hospitals. In: *Statistics Sweden*. Stockholm 1984.
- 17 Statistics NCBo. Cause of death 1983. In: *Statistics Sweden*. Stockholm 1985.
- 18 Online Welfare TNBoHa. Welcome to the Centre for Epidemiology. The National Board of Health and Welfare: Stockholm 1998 (World Wide Web URL: /http:// www.sos.se/epc/epceng.htm).

- 19 Ericson A, Källén B, Winberg J. Surveillance of malformation at birth. A comparison of two-record systems run in parallell. *Int J Epidemiol* 1977; **6:** 35–41.
- 20 Sanner G. Habilitering. In: Bille B, Olow I (eds). Barnhabilitering-vid rörelsehinder. Almqvist & Wiksell: Stockholm 1992.
- 21 De Vivo MJ et al. Prevalence of spinal cord injury. Arch Neurol 1980; 37: 307–308.
- 22 The International classification of diseases. In: *The International Classification of Diseases: 9th Revision, Clinical Modification: ICD-9-CM*, 2nd edn. DHHS publication: Washington, DC (PHS) 80-1260; 1980.
- 23 Socialstyrelsen. Klassifikation av sjukdomar och hälsoproblem 1997 Systematisk förteckning. 1. uppl. edn. Socialstyr.(distributör): Stockholm 1996.
- 24 Mahonen M et al. The validity of hospital discharge register data on coronary heart disease in Finland. Eur J Epidemiol 1997; 4: 403–415.
- 25 Perlman M. Neonatal spinal cord injury in the infant: etiology, diagnosis, treatment and outcome. In: R B, M M (eds). *The Child with a Spinal Cord Injury*. American Academy of Orthopaedic Surgeons: Rosemont, IL 1996, pp. 161–167.