ORIGINAL ARTICLE

An administrative data algorithm to identify traumatic spinal cord injured patients: a validation study

B Welk^{1,2}, E Loh³, SZ Shariff², K Liu² and F Siddiqi⁴

Objective: To assess the validity of different administrative data sources available for the identification of traumatic spinal cord injured (TSCI) patients.

Study design: Retrospective validation study.

Setting: Ontario, Canada.

Participants: Adult patients seen in tertiary outpatient spinal cord rehabilitation clinics after 1 April 2002.

Outcome measures: Sensitivity, specificity, positive and negative predicative values of diagnostic ICD10 codes from Canadian Institutes of Health Discharge Abstracts (CIHI-DAD), Rehabilitation Coding Groups (RCG) from that National Rehabilitation System (NRS), and spinal cord injury fee codes from the Ontario Healthcare Insurance Plan (OHIP). Secondary outcome was the agreement between actual lesion level and RCG/ICD10 coded lesion level.

Results: The RCG codes in the NRS have high sensitivity (92%, 95% confidence interval (CI): 87–95%) and specificity (97%, 95% CI: 94–99%) for the identification of true TSCI patients, whereas CIHI-DAD ICD10 codes are highly specific (99%, 95% CI: 95–100) and moderately sensitive (76%, 95% CI: 79–87%). OHIP fee codes had poor sensitivity (64%, 95% CI: 57–71%). Agreement between true lesion level and the NRS and CIHI-DAD coding is good (Kappa of 0.65–0.88 and 0.56–0.70, respectively).

Conclusion: This study demonstrated that the NRS is able to accurately discriminate between patients with and without a TSCI. A large population of incident and prevalent TSCI patients are identifiable using administrative data.

Sponsorship: This study was funded by a grant from the Division of Urology, Western University.
Spinal Cord (2014) 52, 34–38; doi:10.1038/sc.2013.134; published online 12 November 2013

Keywords: validation studies; sensitivity and specificity; International Classification of Diseases; epidemiology; spinal cord injuries

INTRODUCTION

Observational studies allow physicians to answer epidemiologic questions and to study the real-world effectiveness and generalizability of interventions. They are also well suited for measurement of uncommon or long-term outcomes. Administrative data are defined as data that are routinely collected for a purpose other than a specific research question.¹ Canada's universal health-care system provides complete administrative data for the population, and is ideally suited for observational studies.

A high-quality administrative data study requires a valid population or outcome. Selected data elements may or may not accurately identify a diagnosis or outcome.¹ Misclassification of patients can bias the study results and hamper the interpretation of the study conclusions.²

Traumatic spinal cord injury (TSCI) patients are well suited for study with administrative data. TSCIs are rare, making prospective cohort studies challenging. Important secondary health complications occur over a patient's lifetime, and therefore prospective follow-up, or accurate retrospective studies are difficult and costly. Previous validation studies in this population have not included a 'true negative' population of patients to properly assess the specificity, sensitivity and negative predictive value of a data element for the identification of TSCI patients.^{3–5} No previous studies have examined

the validation of the National Rehabilitation System (NRS) or physician billing data.

The primary objective of this study was to evaluate the diagnostic characteristics of data elements for identification of TSCI patients. The secondary objectives were to assess the agreement among lesion levels, and to describe the population of patients currently available in the administrative data of Ontario, Canada.

MATERIALS AND METHODS

We conducted a population-based retrospective validation study, using medical records as the 'gold standard'.² The administrative data originated from Ontario, Canada (population 13 million) and were made available for evaluation at the Institute for Clinical Evaluative Sciences (ICES). Ontario has universal healthcare provided through a single government payer. For this study, a TSCI was defined as an acute injury to the spinal cord or cauda equina caused by an external physical force resulting in neurological impairment.⁴ Data analysis was carried out at ICES. Study reporting and methods conform to recommendations for administrative data validation study reporting.¹ Ethics approval was granted by Western University.

Data sources and quality

Reference standard. We created a validation cohort using individual patient medical records. All patients seen at the tertiary spinal cord clinic serving

¹Department of Surgery, Western University, London, Ontario, Canada; ²Institute for Clinical Evaluative Sciences—Western (ICES Western), London, Ontario, Canada; ³Department of Physical Medicine and Rehabilitation, Western University, London, Ontario, Canada and ⁴Department of Clinical Neurologic Sciences, Western University, London, Ontario, Canada

Correspondence: Dr B Welk, Division of Urology, Western University, St Joseph's Health Care, Room B4-667, 268 Grosvenor Street, London, Ontario, Canada N6A 4V2.

E-mail: blayne.welk@sjhc.london.on.ca

Received 13 August 2013; revised 9 October 2013; accepted 10 October 2013; published online 12 November 2013

Southwestern Ontario (catchment 1.6 million people) after 1 January 2002 were reviewed. Patients were identified as true positive patients (inclusion criteria: TSCI occurring in Ontario between 1 April 2002 and 31 January 2012 and age > 18 years at injury date) or true negative patients (inclusion criteria: non-TSCI or associated neurologic condition such as multiple sclerosis, transverse myelitis, spina bifida, or Guillain Barre syndrome, age > 18 years as of 1 April 2002) based on the hospital chart review. These patients were chosen as true negative patients, as they all had neurologic conditions that could result in admission to hospital or rehabilitation, and could be miscoded as a TSCI.

There were a total of 1203 consecutive patients seen by one of four physiatrists, and 462 patients met our inclusion criteria. The most common reasons for patient exclusion were the patient did not have a TSCI (for the true positive group) or a related condition (for the true negative group) (427), their injury date was before 1 April 2002 (253), or their age at injury was < 18 years of age (53). Prespecified data elements (name, date of birth, Ontario Health Insurance Plan (OHIP) number, gender, diagnosis, and if applicable date and level of spinal cord injury (SCI)) were extracted from the medical records (by Dr Welk) using a template; he was blinded to the administrative data coding. Lesion level was determined based on the assessment made at the time of rehabilitation admission by the treating specialist. The accuracy of the SCI lesion level extraction in this study was verified using data from the Rick Hansen SCI Registry (RHSCIR),⁶ which has prospective data capture of TSCI patients.⁵ Of the 203 patients with an SCI that we identified, 56 were matched to their RHSCIR data record. Lesion level agreement was excellent, with an intraclass correlation coefficient of 0.90, and < 6% of patients had disagreement about a cervical versus thoracic segment lesion.

Administrative data. Since 2002, all rehabilitation centers in Ontario have been required to submit standardized data to the NRS. All patients are assigned a Rehabilitation Client Group (RCG) code, specifying the diagnosis responsible for admission. All hospitals in Canada are required to submit hospitalization details to the Canadian Institutes for Health Information Discharge Abstract Database (CIHI-DAD). The data quality of both the NRS and CIHI-DAD is maintained through variable verification, and audit-feedback; reabstraction studies have shown > 80% agreement with coding elements.⁷ Physician billing occurs through the OHIP, and surgical fee claims are expected to have a high sensitivity and positive predictive value (PPV), as shown with other service payments.⁸

Data analysis. Individuals within the data sources were linked deterministically through unique, encrypted numbers. Data elements from the following sources were evaluated for their diagnostic properties:

1. NRS (RCG codes 04.21x, 04.2xx or 14.1/14.3) and CIHI-DAD (up to 25 ICD10 diagnosis codes may be coded per patient; based on the previous methodology^{5,9} S14.1/S14.1, S24.0/S24.1, S34.0/S34.1/S34.3 and T06.0/T06.1 were used). For patients from our chart abstraction true positive population, we determined if any of these codes were present within 60 days or 180 days of the SCI date as determined from the chart abstraction data. The two different outcome periods (60 versus 180 days) required us to define two TSCI cohorts based on the availability of NRS and CIHI-DAD records up to 31 March 2012: TSCI between 1 April 2002 and 31 January 2012 (for a minimum of 60 days of follow-up data), or TSCI between 1 April 2002 and 30 September 2011 (for a minimum of 180 days of follow-up data). For patients from our chart abstraction true negative population, we determined the presence of these codes anytime within a single observation window (1 April 2002–31 March 2012).
2. OHIP: fee code E383A (introduced in October 2005 as an additional payment for surgical decompression/stabilization of TSCI patients within 6 weeks of injury). A similar process as described for the NRS was used. The observation window was 1 October 2005–31 January 2013 (a longer observation window was possible due to the availability of more recent OHIP records).

Using a diagnostic test framework, we determined sensitivity, specificity and positive/negative predictive values. Our reference standard was the chart

abstraction data, and our primary outcome was the presence of an incident TSCI (binary variable).

Demographic characteristics were summarized for patients with traumatic SCI who were identified using the best performing algorithm. Socioeconomic status is a quintile rank based on the median neighborhood income. Rural residence is measured based on the population of a patient's place of residence and the availability of local health services.¹⁰ Ontario drug benefit database contains the prescriptions for all patients in Ontario who are on disability or social assistance, or are over the age of 65 years.¹¹

Statistical methods

Given our chart abstraction patient cohort with > 200 patients in each group, and using an alpha of 0.05, we have 80% power to detect a diagnostic sensitivity of 90% (with a lower 95% confidence interval (CI) of 80%).¹² Descriptive statistics (median, interquartile range (IQR)) are provided. Standard diagnostic test formulas were used.¹³ The Wilson score method was used to calculate 95% CIs.¹⁴ Unweighted Kappa was calculated to measure the observed agreement between the SCI lesion level as determined in our chart review cohort and the administrative data sources.¹⁵ Statistical analysis was performed with SAS 9.2 (SAS Institute Inc, Cary, NC, USA).

RESULTS

A total of 462 patients with a known diagnosis (221 with a TSCI and 241 without a TSCI) were successfully matched to administrative data records from the NRS, CIHI-DAD and OHIP databases. General characteristics of these patients are reported in Table 1. Each year of the observation period (2002–2012) contributed a median of 20 (IQR 18–25) TSCI patients. The date of the TSCI as extracted from the medical records compared with administrative data sources was very accurate (CIHI-DAD hospital admission date had a median difference of 0 days, 95% interval –2 to 16 days difference; NRS injury date had a median difference of 0 days, 95% interval –9 to 3 days difference).

The diagnostic characteristics of the data elements in each of the administrative data records were determined (Table 2). All data sources were highly specific ($\geq 94\%$), indicating a strong tendency for these data sources to correctly identify patients who do not have

Table 1 Characteristics of the patient cohort of 462 patients (from medical record review)

	TSCI patients (TP)	Non-TSCI patients (TN)
Number	221	241
Percent male	79%	54%
Median age (IQR) at SCI	42 (27–57) years	
Median age (IQR) as of March 2012	46 (33–61) years	51 (35–66) years
<i>Lesion level (Traumatic SCI patients)</i>		
C1–C4	25%	
C5–C8	34%	
T1–T12	38%	
≤ L1	3%	
<i>Diagnosis (non-traumatic SCI patients)</i>		
Congenital abnormalities		39%
Disc herniation		6%
Demyelinating disease		23%
Transverse myelitis		9%
Spinal vascular compromise		6%
Other		17%

Abbreviations: IQR, interquartile range; SCI, spinal cord injury; TSCI, traumatic spinal cord injury.

True positive (TP) patients had a definitive traumatic SCI. True negative (TN) patients had conditions associated with tetra/paraplegia, but definitively not a TSCI.

TSCI. Sensitivity was variable (64–94%); among the three data sources, the NRS RCG codes performed very well, with a sensitivity of 92% (indicating a good ability to correctly identify patients who did have TSCI) and a Kappa (as a measure of agreement) of 0.89. There was a negligible difference among the diagnostic test results when comparing the 60- and 180-day windows. Several combinations of these data elements were also investigated to see whether sensitivity and specificity could be further optimized (Table 2). Among all possible combinations of data elements, the optimal algorithm for specificity was a patient with both NRS and CIHI-DAD data elements present (which had a specificity of 99% (95% CI 97–100%)), and the optimal algorithm for sensitivity was NRS or CIHI-DAD data elements (which had a sensitivity of 94% (95% CI 90–97%)).

The RCG and ICD10 codes also specify the TSCI lesion level. The accuracy of these data elements compared with the true lesion level (based on the chart abstraction data as the reference standard) was determined (Table 3). There was generally moderate to substantial agreement¹⁵ for CIHI-DAD records (kappa = 0.56–0.70), and substantial to almost perfect agreement with NRS records (kappa = 0.65–0.88).

The optimal method to identify incident TSCI patients was through the NRS: a total of 1801 TSCI patients were identified between 1 April 2002 and 30 March 2012 in Ontario, Canada (Table 4). Each year contributed a median of 189 (IQR 186–195) patients. The optimal definition to identify prevalent TSCI patients was the presence of either an NRS or a CIHI-DAD code; 3427 patients

Table 2 Diagnostic performance of data elements from NRS, CIHI-DAD and OHIP databases for the identification of incident TSCI patients (percentage, 95% CI)

	Total TP/TN patients included ^a	Sensitivity	Specificity	Positive predictive value	Negative predictive value	% FP	% FN
<i>NRS RCG codes</i>							
60-day window	203/241	91 (86–94)	97 (94–99)	96 (93–98)	92 (89–95)	3	9
180-day window	200/241	92 (87–95)	97 (94–99)	96 (93–98)	93 (89–96)	3	9
<i>CIHI-DAD ICD codes</i>							
60-day window	203/241	75 (69–81)	99 (97–100)	99 (95–100)	83 (78–87)	1	25
180-day window	200/241	76 (69–80)	99 (97–100)	99 (95–100)	83 (79–87)	1	24
<i>OHIP fee code</i>							
60-day window	160/236	64 (57–71)	94 (90–96)	88 (81–93)	80 (74–84)	6	36
<i>Maximal specificity</i>							
NRS RCG code and CIHI-DAD ICD code	203/241	72 (65–77)	99 (97–100)	99 (95–100)	80 (75–85)	1	28
180-day window							
<i>Maximal sensitivity</i>							
NRS RCG code or CIHI-DAD ICD code	203/241	94 (90–97)	97 (94–99)	96 (93–98)	95 (92–97)	7	4
180-day window							

Abbreviations: CI, confidence interval; CIHI-DAD, Canadian Institutes for Health Information Discharge Abstract Database; FN, false negative; FP, false positive; NRS, National Rehabilitation System; OHIP, Ontario Healthcare Insurance Plan; RCG, Rehabilitation Client Group; TN, true negative; TP, true positive; TSCI, Traumatic spinal cord injury. The 60/180-day window specifies the presence of an administrative data injury date within 60/180 days of the true injury data (based on the chart abstraction data). All combinations of these codes were examined to look at optimal algorithms to maximize sensitivity and specificity. The best performing algorithms for these are included below.

^aThe chart abstraction cohort numbers change because of changes to the observation windows necessary for different coding elements.

Table 3 Diagnostic performance of the individual coding elements for lesion level from the NRS and CIHI-DAD (percentage, 95% CI)

	Number of TP patients (total n = 203)	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Kappa	% FP	% FN
<i>NRS RCG SCI level codes</i>								
C1–4 (RCG 04.2211, 04.2221)	49	76 (62–86)	94 (90–97)	80 (65–89)	93 (88–96)	0.71	6	24
C5–8 (RCG 04.2212, 04.2222)	65	70 (57–80)	93 (88–96)	80 (66–88)	89 (83–93)	0.65	7	30
≤ T1 (4.211, 4.212)	89	94 (87–97)	94 (88–97)	92 (84–96)	96 (91–98)	0.88	6	6
<i>CIHI-DAD level codes</i>								
Cervical (S14.0/S14.1)	114	75 (67–82)	97 (91–99)	97 (91–99)	75 (67–82)	0.70	3	25
Thoracic (S24.0/S24.1)	60	69 (51–75)	98 (94–99)	93 (81–98)	86 (81–98)	0.68	2	36
Lumbar (S34.1/S34.1)	26	50 (32–68)	98 (95–99)	76 (53–90)	93 (88–96)	0.56	2	50

Abbreviations: CI, confidence interval; CIHI-DAD, Canadian Institutes for Health Information Discharge Abstract Database; FN, false negative; FP, false positive; NRS, National Rehabilitation System; RCG, Rehabilitation Client Group; SCI, spinal cord injury; TN, true negative; TP, true positive.

Kappa, as a measure of agreement, is included. No patients with RCG code 4.210 or 4.220 were identified in this sample, and therefore these codes were not included. There were only a small number of patients with cauda equina injuries, and therefore the ICD10 code S34.3 was not included.

Table 4 Characteristics of patient cohorts based on Ontario's administrative data

	Population of incident traumatic SCI patients based on NRS RCG codes	Population of prevalent traumatic SCI patients based on CIHI-DAD or NRS
Number	1801	3427
Date range	1 April 2002–31 March 2012	1 April 2002–31 March 2011
Age (median IQR)	49 (33–65) years	55 (39–72)
Male	74%	71%
<i>Lesion level (NRS RCG codes)</i>		
Quadriplegia (C1–4)	16%	9%
Quadriplegia (C5–8)	24%	15%
Paraplegia ($\leq$ T1)	37%	20%
Unspecified	23%	12%
Missing (no NRS record)	0%	44%
<i>Lesion level (CIHI-DAD ICD10)</i>		
Cervical	37%	47%
Thoracic	16%	15%
Lumbar	6%	9%
Cauda	1%	2%
Unspecified	<1%	1%
Missing (no CIHI-DAD record)	40%	26%
<i>Socioeconomic status</i>		
Quintiles 1 and 2 (low- est SES)	41%	42%
Quintile 3	20%	20%
Quintiles 4 and 5 (highest SES)	39%	37%
Follow-up	Median 4.6 (IQR 2.5–7.1) years after SCI	Observation window available: 1 April 2011–30 March 2012
Death after NRS discharge	14%	17%
Rural residence	16%	17%
Ontario drug benefits usage	Within 2 years of SCI	Within 1-year follow-up
< 65 years of age	44% (588/1336)	32% (710/2223)
$\geq$ 65 years of age	95% (444/465)	69% (833/1204)

Abbreviations: CIHI-DAD, Canadian Institutes for Health Information Discharge Abstract Database; IQR, interquartile range; NRS, National Rehabilitation System; PPV, RCG, Rehabilitation Client Group; SCI, spinal cord injury; SES, Socioeconomic status. An incident cohort, based on the NRS RCG codes (specificity of 99%, PPV of 96%) and a prevalence cohort, based on either an NRS RCG code, or a CIHI-DAD ICD10 code (specificity of 94%, PPV of 96%) are included. The incident cohort has a defined date of SCI, whereas the prevalence cohort has an unknown date of initial SCI, and a uniform 1-year observation window.

were identified (all patients then have a common 1-year observation window from 1 April 2011 to 30 March 2012, Table 4).

DISCUSSION

One of the potential limitations of administrative data is the unknown accuracy of the data elements. This can, however, be mitigated with an appropriate validation study to examine how well a data element identifies the variable it is assumed to represent.

We examined three different data sources for their ability to correctly identify TSCI patients. The NRS (which uses RCG diagnosis

codes) is a Canada-wide rehabilitation database; in this study, it had excellent diagnostic test characteristics, with a high sensitivity (92%) and specificity (97%). The PPV of 96% means that of the TSCI patients who had related NRS RCG codes, 96% were truly TSCI patients. CIHI-DAD ICD10 diagnosis codes had excellent specificity (99%), however, the sensitivity was lower (76%). This is similar to the results from ICD10 validation studies of TSCI patients done in British Columbia, Canada⁵ and Norway.⁴ One reason for this lower sensitivity compared with the NRS is the diagnosis responsible for admission to rehabilitation is likely clearly identified, whereas during a hospital admission a patient may not have complete and accurate documentation of all their injuries. Data coders are also probably more familiar with the limited list of 17 RCG clusters used in the NRS as compared with the 12 000 ICD10 codes. OHIP records had poor sensitivity, which is probably because not all patients with a TSCI would be managed operatively.

The ability of the NRS and CIHI-DAD to identify basic information about lesion level was generally good, with specificities of $\geq$ 93%. The NRS and CIHI-DAD records are able to identify patients with a specific lesion level, but sensitivity is limited (50–94%). These results are similar to those published in a validation study that compared the CIHI-DAD lesion level data with the RHSCIR.⁵ Only the RCG codes for Paraplegia (suggesting a thoracic or lumbar SCI injury) had a low proportion of false negative patients (6%).

There are only a few validation studies available for comparison. Furlan and Fehlings³ found that the ICD10 codes from the Canadian National Trauma Registry were inaccurate in identifying 92 TSCI patients (sensitivity 81%, specificity 7% and PPV 30%). Small, informal validation studies also suggested that there was significant misclassification of patients based on hospital discharge ICD9 TSCI codes.^{16,17} Hagen *et al.*⁴ found that a combination of TSCI ICD10 codes performed better than previous ICD iterations, with a sensitivity of 84%, a specificity of 97% and a PPV of 88%. Finally, Noonan *et al.*⁵ compared the accuracy of level specific information from hospital discharge ICD10 codes with the prospectively collected lesion level in the RHSCIR. The modest sensitivity and high specificity that was demonstrated in their study is very similar to our CIHI-DAD results (Table 3).

Strengths of this validation study include the fact that we used an adequately powered sample to create our validation cohort, and included patients who could act as reference standard true negatives to determine specificity and negative predictive value. We manually reviewed medical records to ensure a correct diagnosis was assigned, and verified the accuracy of this chart abstraction using the RHSCIR.

There are also important limitations that should be acknowledged. This validation study only included patients who survived the initial period between TSCI and their first outpatient follow-up with their physiatrist. Therefore, we do not know whether the administrative data records can accurately measure all TSCIs (such as those who die at the time of the accident, or before hospital discharge). The NRS database would not, under any circumstances, be able to identify patients who died before rehabilitation. This is an important consideration when assessing the incidence of TSCI, or if death is a study outcome. We extrapolated a TSCI level range from the RCG codes, which codes lesion levels together as high/low quadriplegia or paraplegia. The prevalence of TSCI among our data sources is lower than our validation cohort. TSCI is a rare event in all of the data sources, and because of this, the PPV (which is dependent on prevalence) is likely overestimated.¹⁸ Finally, this study was conducted using patients from Southwestern Ontario. The generalizability of this population's administrative data coding may

not extend to other jurisdictions. However, given that our results are generally similar to other regions in Canada,⁵ and Europe,⁴ we feel our ICD10 CIHI-DAD results are broadly applicable.

CONCLUSION

The RCG codes in the NRS have high sensitivity (92%) and specificity (97%) for the identification of true TSCI patients, whereas CIHI-DAD ICD10 codes are highly specific (99%) and moderately sensitive (76%). The RCG and the ICD10 codes have a reasonable ability to discriminate between lesion level, however, there is a moderate amount of misclassification. There is a large population of TSCI patients identified within the administrative data holdings of Ontario. Hopefully future data linkage with prospective data sources, such as the RHSCIR, will expand the ability of administrative data to answer important health services and outcome-based research questions.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was funded by a grant from the Division of Urology, Western University. This study was supported by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The study was conducted through ICES Western which is funded by an operating grant from the Academic Medical Organization of Southwestern Ontario (AMOSO). Dr Amit Garg (ICES Scientist) is thanked for facilitating this project. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES, the Ontario MOHLTC or AMOSO is intended or should be inferred. RHSCIR is a pan-Canadian prospective observational registry at 31 acute and rehabilitation facilities of persons sustaining a TSCI. RHSCIR was initiated as a research study sponsored by the Rick Hansen Institute and was developed to collect data on individuals who sustain spinal cord injuries and link clinicians, researchers and health-care administrators with the goal of improving both

research and clinical practice for individuals with SCI by facilitating the translation of research into clinical practice and promotion of evidence-based practices.

- 1 Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2010; **64**: 1–9.
- 2 van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol* 2012; **65**: 126–131.
- 3 Furlan JC, Fehlings MG. The National Trauma Registry as a Canadian spine trauma database: a validation study using an institutional clinical database. *Neuroepidemiology* 2011; **37**: 96–101.
- 4 Hagen EM, Rekand T, Gilhus NE, Gronning M. Diagnostic coding accuracy for traumatic spinal cord injuries. *Spinal Cord* 2009; **47**: 367–371.
- 5 Noonan VK, Thorogood NP, Fingas M, Batke J, Bélanger L, Kwon BK *et al*. The validity of administrative data to classify patients with spinal column and cord injuries. *J Neurotrauma* 2013; **30**: 173–180.
- 6 Noonan VK, Kwon BK, Soril L, Fehlings MG, Hurlbert RJ, Townson A *et al*. The Rick Hansen Spinal Cord Injury Registry (RHSCIR): a national patient-registry. *Spinal Cord* 2012; **50**: 22–27.
- 7 Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging* 2005; **24**(Suppl 1), 153–170.
- 8 Raina P, Torrance Rynard V, Wong M, Woodward C. Agreement between self-reported and routinely collected health-care utilization data among seniors. *Health Serv Res* 2002; **37**: 751–774.
- 9 Munce SEP, Guilcher SJ, Couris CM, Fung K, Craven BC, Verrier M *et al*. Physician utilization among adults with traumatic spinal cord injury in Ontario: a population-based study. *Spinal Cord* 2009; **47**: 470–476.
- 10 Kralji B. Measuring 'rurality' for purposes of health care planning: an empirical measure for Ontario. *Ontario Med Rev* 2000; 33–52.
- 11 Levy AR, O'Brien BJ, Sellors C, Grootendorst P, Willison D. Coding accuracy of administrative drug claims in the Ontario Drug Benefit database. *Can J Clin Pharmacol* 2003; **10**: 67–71.
- 12 Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005; **58**: 859–862.
- 13 Parshall MB. Unpacking the 2 × 2 table. *Heart Lung* 2013; **42**: 221–226.
- 14 Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998; **17**: 857–872.
- 15 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
- 16 Thurman DJ, Burnett CL, Jeppson L, Beaudoin DE, Sniezek JE. Surveillance of spinal cord injuries in Utah, USA. *Paraplegia* 1994; **32**: 665–669.
- 17 Dahlberg A, Kotila M, Leppänen P, Kautiainen H, Alaranta H. Prevalence of spinal cord injury in Helsinki. *Spinal Cord* 2005; **43**: 47–50.
- 18 Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994; **309**: 102.