

ORIGINAL ARTICLE

Incorporation of the International Spinal Cord Injury Data Set elements into the National Institute of Neurological Disorders and Stroke Common Data Elements

F Biering-Sørensen¹, S Charlifue², MJ DeVivo³, ST Grinnon⁴, N Kleitman⁵, Y Lu⁴ and J Odenkirchen⁵

¹Clinic for Spinal Cord Injuries, Neuroscience Centre, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark; ²Craig Hospital, Englewood, CO, USA; ³Department of Physical Medicine and Rehabilitation, University of Alabama at Birmingham, Birmingham, AL, USA; ⁴KAI Research, Inc. (An Altarum Company), Rockville, MD, USA and ⁵National Institutes of Health/National Institute of Neurological Disorders and Stroke (NIH/NINDS), Bethesda, MD, USA

Objectives: To develop consistent variable names and a common database structure for the data elements in the International Spinal Cord Injury (SCI) Data Sets.

Setting: National Institute of Neurological Disorders and Stroke (NINDS) Common Data Elements (CDE) Project and The Executive Committee of the International SCI Standards and Data Sets committees (ECSCI).

Methods: The NINDS CDE team creates a variable name for each defined data element in the various International SCI Data Sets. Members of the ECSCI review these in an iterative process to make the variable names logical and consistent across the data sets. Following this process, the working group for the particular data set reviews the variable names, and further revisions and adjustments may be made. In addition, a database structure for each data set is developed allowing data to be stored in a uniform way in databases to promote sharing data from different studies.

Results: The International SCI Data Sets variable names and database specifications will be available through the web sites of the International Spinal Cord Society (<http://www.iscos.org.uk>), the American Spinal Injury Association (<http://www.asia-spinalinjury.org>) and the NINDS CDE project web site (<http://www.CommonDataElements.ninds.nih.gov>).

Conclusion: This process will continue as additional International SCI Data Sets fulfill the requirements of the development and approval process and are ready for implementation.

Spinal Cord (2011) 49, 60–64; doi:10.1038/sc.2010.90; published online 24 August 2010

Keywords: spinal cord injury; International; Data Set; variable names; database specifications; common data element

Introduction

The purpose of the International Spinal Cord Injury (SCI) Data Sets, to facilitate comparisons of injuries, treatments and outcomes between patients, centers and countries, has been described in previous publications.^{1–8} These data sets appear on the web sites of the International Spinal Cord Society (<http://www.iscos.org.uk>) and the American Spinal Injury Association (<http://www.asia-spinalinjury.org>).

The National Institute of Neurological Disorders and Stroke (NINDS) Common Data Elements (CDE) Project was undertaken to facilitate the development of neurological

data standards and to develop a web site (<http://www.CommonDataElements.ninds.nih.gov>) containing these data standards and accompanying tools. It is intended to help investigators and study staff to collect data with a ‘universal language’ in their clinical studies.

The purpose of the present project is to develop consistent variable names for the data elements included in the International SCI Data Sets and to develop a common database structure. This process will facilitate the adoption of these variables for use by both clinicians and researchers who are in the process of developing research projects or clinical databases. These data set variables have already been through a rigorous consensus, review and approval process within and among individuals and organizations interested in clinical and research work related to SCI.⁹ The free access to these variables will allow researchers and clinicians to avoid the laborious process of defining variables for their

Correspondence: Dr F Biering-Sørensen, Clinic for Spinal Cord Injuries, Neuroscience Centre, Rigshospitalet, TH2091, Blegdamsvej 9, Hornbaek, DK-2100 Copenhagen, Denmark.
E-mail: finbs@rh.regionh.dk

Received 19 April 2010; revised 21 June 2010; accepted 28 June 2010; published online 24 August 2010

questionnaires or databases and should facilitate harmonization across clinical studies.

Materials and methods

Staff members from the NINDS CDE team (NINDS Program Directors along with their contractor, KAI Research, Inc.) approached the Executive Committee of the International SCI Standards and Data Sets committees (ECSCI) after they became aware of the work performed by the various SCI data set working groups. In subsequent discussions between the committee and the NINDS CDE team, a decision was made to cooperate in developing variable names for each variable in the data sets. The ECSCI and the NINDS CDE team decided to assign variable names that were at most eight characters long in order to accommodate a variety of database software/platform options, keeping the simplest type of data system in mind. They were mindful that limiting the length of the variable names to eight characters ensured compatibility with the SAS® Transport format. The SAS XPORT Transport format currently serves as a US Food and Drug Administration standard format for data sets in electronic submissions (<http://www.fda.gov/drugs/developmentapprovalprocess/formsubmissionrequirements/electronic submissions/ucm085361.htm>).

In the autumn of 2008, the NINDS CDE team began the data variable naming process with the International SCI Core Data Set (1). First, variable names of no more than eight characters in length were created for each data element in the International SCI Core Data Set. These variable names were sent by e-mail for review by members of ECSCI. Following the review, a teleconference was held with the involved individuals to discuss possible acceptance or modification of the proposed variable names. After this process was established as acceptable, the International SCI Basic Lower Urinary Tract Data Set (2) was reviewed in the same manner, followed by the International SCI Basic Urodynamic Data Set (3). The process continued with adjustments as the group learned more about how the eight character variable names needed to be structured to be as logical and consistent as possible across the various data sets. The NINDS CDE team subsequently developed a list of conventions to make certain that all variable names were consistently created (Table 1). The NINDS CDE team and the members of ECSCI also made sure that the variable names for all non-key data elements were unique across the data sets. The NINDS CDE team set up a simple database to help them verify the uniqueness of the variable names as the number of data sets they worked with evolved or increased.

While working to assign standard variable names to the data sets, the ECSCI and the NINDS CDE team soon became aware that often the way they assigned the variables for a data set depended upon the structure of the table(s) that would store the information in a relational database. (A relational database is a collection of data items organized as a set of tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Each table (which is some-

Table 1 Conventions adopted during the construction of the eight character variable names for the data elements in the various International Spinal Cord Injury Data Sets (9)

<i>Word or phrase</i>	<i>Abbreviation in eight characters variable</i>
After	AF
ASIA impairment scale	AIS
Anal	ANL
Average	AV
Before	BF
Bladder	BLAD
Blood pressure	BP
Cardiac	CARD, CA
Date	DT
Day(s)	DY
Defecate or defecation	DEF
Discomfort	DCMF
Duration	DUR
Drug(s)	DRG
Dysfunction	DYS
Evaluation	EVAL
Event	EVT
Extended	EX
Findings	FND
Frequency	FRQ
Function	FXN
Gastrointestinal	GI
History	HX
Incontinence	INC
Injury	INJ
Last 7 days	L7DY
Last 3 months	L3M
Last week	LW
Last year	LY
Level	LVL
Left	L
Location	LOC
Lower urinary tract	LUT
Measure	MEAS
Method	MTH
Middle	M
Motor	MTR
Number	NO
Order	OR
Other	OTH
Pain	PN
Perianal	PANL
Physical	PHYS
Problem(s)	PROB
Psychological	PSYC
Quality of life	QOL
Right	R
Sensory	SENS
Specify	SP
Surgery(ies) or surgical procedure(s)	SURG, SR
Symptom(s)	SX
Test	TST
Time	TM
Treatment(s)	TX
Type	TYP
Unrelated	UN
Urinary tract	UT
Volume	VOL

times called a relation) contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. The relational database was invented by EF Codd at IBM in 1970¹⁰). It, therefore, was decided to also propose how the variables

could be stored in an appropriate database structure to facilitate both analysis and sharing of data across studies. The proposed database structure is compatible with various relational database software packages, including Microsoft® Access®, SAS, Microsoft SQL®, Oracle® and so on.

Relational data tables linked by common patient identifiers were established for each data set, which could be used for either cross-sectional or longitudinal studies. With each new data set the ECSCI and the NINDS CDE team defined whether the data set would be captured in a single data table or more than one data table. With this approach, investigators can create limited data subsets of selected variables from multiple data sets for analysis. For example, needed information on patient characteristics could be easily merged with data from the lower urinary tract data set. Moreover, use of common data files will facilitate the combining of data sets collected at multiple locations. Besides determining the number of data tables for each data set, the group also needed to decide whether each data table would have a more horizontal (short and wide) or vertical (tall and narrow) structure. Of note, the proposed database structure offers one way of using the standard variable names in a database, but is not the only structure that could work based on the defined SCI CDE.

The continued process in this project has involved approximately monthly teleconferences and e-mail correspondence for more than one and a half years. In addition, a face-to-face meeting between the NINDS CDE team and members of ECSCI was held during the 35th Annual Scientific Meeting of American Spinal Injury Association, September 2009, in Dallas, Texas.

After there was common agreement with the iterative adjustment process between the NINDS CDE team and the members of ECSCI, the result was presented to the particular working group for the data set. After their review of the eight character variable names and the database structure, suggested final revisions and adjustments were made.

At a later stage in the process of working with the International SCI Data Sets, it was decided to include the International Standards for Neurological Classification of SCI (http://www.asia-spinalinjury.org/publications/2006_Classif_worksheet.pdf),¹¹ so variable names and a database structure were developed in the same way.

Results

The following data sets have been through the complete process described above and will be posted with the eight characters variable names and the suggested relational database structure on the web sites of International Spinal Cord Society (<http://www.iscos.org.uk>) and American Spinal Injury Association (<http://www.asia-spinalinjury.org>) as well as the NINDS CDE project web site (<http://www.CommonDataElements.ninds.nih.gov>):

- International SCI Core Data Set (1)
- International SCI Basic Lower Urinary Tract Data Set (2)
- International SCI Basic Urodynamic Data Set (3)
- International SCI Basic Urinary Tract Imaging Data Set (7)

- International SCI Basic Bowel Function Data Set (5)
- International SCI Extended Bowel Function Data Set (6)
- International SCI Basic Female Sexual and Reproductive Function Data Set
- International SCI Basic Male Sexual Function Data Set
- International SCI Basic Cardiovascular Function Data Set (8)
- International SCI Basic Pain Data Set (4)
- International Standards for Neurological Classification of SCI (http://www.asia-spinalinjury.org/publications/2006_Classif_worksheet.pdf)¹¹

These web sites include an explanation of the purpose of the project and the standard variable names as well as the proposed database structure. The naming conventions described in Table 1 are also provided.

As an example, the original International SCI Core Data Set form (1) is shown in Figure 1 with the eight character variable names included along with notes for the division of the data set into two tables. Those variables that are designed to be collected only once are contained in Figure 1, TABLE #1. The core neurological data are included in Figure 1, TABLE #2 in which each time point of data collection is stored in a separate record to facilitate longitudinal analyses. In fact, this approach would allow more than the collection of admission and discharge data simply by adding additional records reflecting other times post-injury. Each record would be distinguished by its date of data collection, which would be part of the record key.

Discussion

In the process of developing the standard variable names, a priority was to make these as clinically meaningful as possible within the eight character limit, but consideration also was given to making the variable names for similar types of variables as consistent as possible across the various data sets. This process to establish consistency has been lengthy and continues to undergo modification as the authors of each newly reviewed data set experience challenges that need special resolution. This iterative process often requires re-review of data sets for which variable names have already been assigned to ensure full consistency across the entire bank of data sets. As soon as other International SCI Data Sets are completed and approved, these will likewise be added with standard variable names and a proposed database structure.

The ECSCI and the NINDS CDE team gave just as much thought to their work to establish relational data tables for the data sets as they did to developing standard variable names. As previously illustrated with the Core Data Set, the decision to break a data set into more than one data table often was dictated by whether groups of data elements in the data set could be collected at disparate time points from a patient. In general, a horizontal database structure was chosen to facilitate statistical analyses that usually require all variables to be included in a single record with results compared across patients. However, when the unit of analysis would more likely be the individual times of measurement, and multiple measurements could be obtained from each person at potentially inconsistent times

Note: It is recommended that Table #1 and Table #2 have two additional key variables:
 - SITE (to distinguish the location where the data are recorded) and
 - SUBJECT (to distinguish the patient/study participant)

INTERNATIONAL SPINAL CORD INJURY DATA SET
 CORE DATA SET COLLECTION FORM

TABLE #1

Dates (YYYYMMDD)

Birth date BIRTHDT ____/____/____

Injury date INJURYDT ____/____/____

Acute Admission ADMITDT ____/____/____

Final Inpatient Discharge DISCHGDT ____/____/____

Total Days Hospitalized ____ HOSPTLDY

Gender GENDER

Injury Etiology SCIETIOL Vertebral Injury VRTBRINJ Associated Injury ASSOCINJ

Spinal Surgery SPNLSURG Ventilatory Assistance VENTASSI Place of Discharge PLACEDIS

Neurological Data
TABLE #2*

Acute Admission TIMEPT*

Date of Exam ____/____/____ NEUEXMDT

Sensory level
Left Right
SENSLVLL SENSLVLR

Motor level
Left Right
MTRLVLL MTRLVLR

ASIA Impairment Scale
AIS

Final Inpatient Discharge TIMEPT*

Date of Exam ____/____/____ NEUEXMDT

Sensory level
Left Right
SENSLVLL SENSLVLR

Motor level
Left Right
MTRLVLL MTRLVLR

ASIA Impairment Scale
AIS

*TIMEPT is a key variable to distinguish the acute admission and final inpatient discharge records

Figure 1 The International SCI Core Data Set form (1) as it appears with the eight character variable names and division into two tables.

post-study enrollment, a vertical approach was selected with each time of measurement as a separate record to store the data because of its inherent flexibility to accommodate repeated measurements. This approach is similar to the US Model Systems Database in which initial data are contained in a single table while annual follow-ups are in a second table.¹² In the work to develop data tables for the International SCI Data Sets, the ECSCI and NINDS CDE team tried to assign consistent structures across the data sets so as to make it easier to assemble a study database and to share data from multiple sites/studies.

The data collection forms were originally designed to facilitate data collection rather than efficient data storage and analysis. As a result, there is no one-to-one correspondence between the data collection forms and the

database structure. For example, ‘unknown’ may be a single check box on the paper form, but is a choice in multiple code lists in the data table. Rather than creating a unique variable for ‘unknown’, checking the unknown box would result in automatically assigning all appropriate variables the ‘unknown’ response. This explains why the ‘annotated forms’ included on the web sites (<http://www.iscos.org.uk>; <http://www.asia-spinalinjury.org>; <http://www.CommonDataElements.ninds.nih.gov>) have the eight character variable tags superimposed on the form.

Once those responsible for the development of each International SCI Data Set approve and release the variable names and database structures, clinical or research institutions may freely use them to write data entry software programs either for Internet or local data entry. Simple

quality control procedures can also be incorporated into the data entry software or as stand-alone programs.

Although this work will greatly facilitate the combining of data from multiple sites, it is important to understand that data should not be combined without a thorough understanding of their origins. There must be an underlying research design and sampling frame, comparable case ascertainment and data collection procedures, methods to assess data quality at each location, methods to avoid duplicate patient entry and so on. Otherwise, there would be no way to assess representativeness or generalizability of the data, as well as the direction and magnitude of any potential bias that might be present, thereby making results difficult if not impossible to interpret.

Conclusion

Variable names and database structures have now been developed for each published International SCI Data Set and its associated CDEs. This process will continue as additional International SCI Data Sets fulfill the requirements of the development and approval process and are ready for implementation. Additional work is now needed to develop data entry and quality control software that would facilitate the use of these data sets.

Conflict of interest

The authors declare no conflict of interest.

Disclaimer

The views expressed here are those of the authors and do not represent those of the National Institutes of Health (NIH), the National Institute of Neurological Disorders and Stroke

(NINDS), the National Institute on Disability and Rehabilitation Research (NIDRR) or the US Government.

References

- 1 DeVivo M, Biering-Sørensen F, Charlifue S, Noonan V, Post M, Stripling T *et al*. International spinal cord injury core data set. *Spinal Cord* 2006; **44**: 535–540.
- 2 Biering-Sørensen F, Craggs M, Kennelly M, Schick E, Wyndaele JJ. International lower urinary tract function basic spinal cord injury data set. *Spinal Cord* 2008; **46**: 325–330.
- 3 Biering-Sørensen F, Craggs M, Kennelly M, Schick E, Wyndaele JJ. International urodynamic basic spinal cord injury data set. *Spinal Cord* 2008; **46**: 513–516.
- 4 Widerström-Noga E, Biering-Sørensen F, Bryce T, Cardenas DD, Finnerup NB, Jensen MP *et al*. The international spinal cord injury pain basic data set. *Spinal Cord* 2008; **46**: 818–823.
- 5 Krogh K, Perkash I, Stiens SA, Biering-Sørensen F. International bowel function basic spinal cord injury data set. *Spinal Cord* 2009; **47**: 230–234.
- 6 Krogh K, Perkash I, Stiens SA, Biering-Sørensen F. International bowel function extended spinal cord injury data set. *Spinal Cord* 2009; **47**: 235–241.
- 7 Biering-Sørensen F, Craggs M, Kennelly M, Schick E, Wyndaele JJ. International urinary tract imaging basic spinal cord injury data set. *Spinal Cord* 2009; **47**: 379–383.
- 8 Krassioukov A, Alexander MS, Karlsson AK, Donovan W, Mathias CJ, Biering-Sørensen F. International spinal cord injury cardiovascular function basic data set. *Spinal Cord* 2010; **48**: 586–590.
- 9 Biering-Sørensen F, Charlifue S, DeVivo M, Noonan V, Post M, Stripling T *et al*. International spinal cord injury data sets. *Spinal Cord* 2006; **44**: 530–534.
- 10 Codd EF. A relational model of data for large shared data banks. *Commun ACM* 1970; **13**: 377–387.
- 11 Marino RJ, Barros T, Biering-Sørensen F, Burns SP, Donovan WH, Graves DE *et al*. International standards for neurological classification of spinal cord injury. *J Spinal Cord Med* 2003; **26**(Suppl 1): S50–S56.
- 12 DeVivo MJ, Go BK, Jackson AB. Overview of the national spinal cord injury statistical center database. *J Spinal Cord Med* 2002; **25**: 335–338.