Acta Neurochir (2005) [Suppl] 95: 39–41

Printed in Austria

Accurate data collection for head injury monitoring studies: a data validation methodology

J. Barnes, I. Chambers, I. Piper, G. Citerio, C. Contant, P. Enblad, H. Fiddes, T. Howells, K. Kiening, P. Nilsson, and **Y. H. Yau for the *BrainIT Group***

Regional Medical Physics Department, Newcastle General Hospital, Newcastle upon Tyne, UK

Summary

Background. BrainIT is a multi centre, European project, to collect high quality continuous data from severely head injured patients using a previously defined [6] core data set. This includes minute-by-minute physiological data and simultaneous treatment and management information. It is crucial that the data is correctly collected and validated.

Methods. Minute-by-minute physiological monitoring data is collected from the bedside monitors. Demographic and clinical information, intensive care management and secondary insult management data, are collected using a handheld computer. Data is transferred from the handheld device to a local computer where it is reviewed and anonymised before being sent electronically, with the physiological data, to the central database in Glasgow. Automated computer tools highlight missing or ambiguous data. A request is then sent to the contributing centre where the data is amended and returned to Glasgow. Of the required data elements 20% are randomly selected for validation against original documentation along with the actual number of specific episodic events during a known period. This will determine accuracy and the percentage of missing data for each record.

Conclusion. Advances in patient care require an improved evidence base. For accurate, consistent and repeatable data collection, robust mechanisms are required which should enhance the reliability of clinical trials, assessment of management protocols and equipment evaluations.

Keywords: Head injury; data collection; database; validation.

Introduction

The incidence of serious head injury is estimated at 1500 per 100,000 population per annum which equates to over one million head injuries per year in Europe. The long term care of these patients bears a great burden on society and has both social and economic implications for the injured person and carers involved [1]. The incidence and effects of secondary insults play a significant part in the outcome of patients. Prompt, and better medical management of these insults has improved the outcome in these patients [3]. The Traumatic Coma Databank used a common data collection protocol and provided information on patient recovery and outcome [5]. Recent pharmaceutical studies have endeavoured to protect the injured brain from further secondary insults therefore aiming to improve the outcome of a predominately young population. In neither case has high resolution monitoring data been recorded. Therefore the precise relationship between outcome and the deviation of physiological variables made. Clinical Trials of this sort have proved largely inconclusive, and a critical analysis of the possible reasons for this failure is given in an article by Maas [4]. With advances in monitoring and information technology interest has now shifted towards the accurate minute-by-minute monitoring of severely injured patients with a view to observing momentary secondary insults as they occur. Because of the relatively low number of head injuries seen in any one centre, and the number required to produce robust evidence based studies, it is necessary to pool data from several centres. This can only be achieved if the data from different centres can be combined in a standardised format.

The BrainIT Group http://www.brainit.org/ is a collaboration of Healthcare Professionals who came together during the 10th International Symposium on Raised Intracranial Pressure and Neuromonitoring in Brain Injury in Williamsburg, USA in May 1997. From this meeting and subsequent discussions, an EEC grant was obtained to fund the development of a network infrastructure that culminated in the formation of the BrainIT Group. The Group currently consists of 30 European Centres and led by a Steering

Group. The aim is to collect high quality continuous data from severely head injured patients and store it in a database that is available for analysis and review. The group strive to standardise data collection for this group of patients and also to establish and populate data for analysis. Each participating country has been assigned a data validator to help co-ordinate the setting up of centres each in their own country. A core data set has previously been defined [6] that includes minute-by-minute physiological data and simultaneous treatment and management information. Collected data is transferred via the BrainIT website to the Central Database in Glasgow where automated computer tools are used to highlight missing or ambiguous data. Of the required data elements 20% are randomly selected for validation against original documentation along with the actual number of specific episodic events during a known period. This will determine accuracy and the percentage of missing data for each record.

Methods

The defined core data set consists of four aspects of data collection: minute-by-minute physiological monitoring data, demographic and clinical information, intensive care management data and secondary insult management data. All these aspects of data collection are essential for the completion of each patient data file. Eligible patients for this study may be of any age with the inclusion criteria being insertion of an intracranial pressure monitoring device and arterial line. The minimum amount of time a patient may be monitored for is four hours. Data collection in the United Kingdom may begin as soon as the patient is admitted to intensive care as approved by the Multi Research Ethics Committee, Scotland, with written assent being obtained as soon as is practical after admission. In cases where assent is not given for any reason, then the previously collected data will not be retained.

Physiological data collection on such a large scale has never previously been attempted and various methods exist for collection of such data. Some centres use their own specifically devised data collection programmes, whilst others have enlisted commercial help to install and set up equipment to collect the required minute-by-minute data. Data from several areas (Intensive Care Unit, High Dependency Unit, Neurotheatre) may either be collected centrally via a networked system or at each bedside using a computer connected to the physiological monitor. Data from these systems is transferred to Glasgow via the BrainIT website. Collection of physiological data involves very little input from the ICU nursing and medical staff and requires one or two designated persons to be responsible for downloading and sending each patient's data to a dedicated research data server at the BrainIT Co-ordinating Centre in Glasgow.

Collection of demographic and clinical information, and intensive care treatment and management data is by use of handheld devices (PDAs). Specialist software developed by Kelvinconnect Limited enables the user to enter information relating to the patient's head injury and subsequent management whilst the patient is in intensive care. Secondary insult management data is also collected with a list of named target therapies and targets to which these therapies are aimed. All data is collected as long as the patient has arterial and intracranial pressure monitoring. Data is then transferred from the handheld device to a local computer, where it is stored in a database, therefore providing a list of recruited patients for each local centre. Prior to transfer to Glasgow the data collector will obtain a unique eight digit number from the BrainIT website and attach it to each patient's file.

Software on the local computer allows review of the data and all personal details are removed leaving the BrainIT number only as a means of patient identification. Data is then transferred to Glasgow to the central database. Data conversion tools convert data to the BrainIT standard format and it is then added to the central database. A web interface has been designed to allow members of the BrainIT Group access to the database and the information it provides.

Once the data has been received in Glasgow, automated computer tools convert data to a common file format, standardise units, parse the data and highlight missing or ambiguous data. A missing data request is sent to the local centre. Missing data, if available, is entered and the file is returned to Glasgow. This process continues until as much of the missing data as possible is found. Of the required data elements 20% are randomly selected for data validation against original documentation along with the actual number of specific episodic events during a known period. This part of the process requires assistance of the country data validator. A list of requested data for validation is sent to the data validator who will then visit the local centre from where the data was originally sent. Using the patient's notes and charts the data validator will check the list of data against the original documentation. A validation file will be created in the contributing centre using the BrainIT Core Data Collection tool which will then be exported and sent to Glasgow via the internet. Again missing or incorrectly entered data is identified and a new validation request is sent to the data validator. These steps are repeated until all of the missing validation data has been collected. It will be possible to search within the database for both validated and non-validated data. Only validated data should be used in analyses intended for publication or submission, whilst non-validated data may be accessed and used in analyses intended for hypothesis generation.

Data validation procedures are as follows. First of all fully anonymised raw data transferred to Glasgow is kept on the BrainIT group server which is separate from the common database. For all data entered to the common database one of four levels of data validation are progressively applied. Validation level one is a *"well-formed"* data check used to ensure that the data conversion stage functioned correctly, in particular the time-stamp format (YYYY-MM-DD HH:MM) is valid. Data validation level two checks all non-numeric categorical core dataset data for transcribing errors. Validation at this level differs depending on whether monitoring or non-monitoring data is being validated. Validation of monitoring data requires a start and end time for each particular monitoring channel, also that the difference between the range of monitored data values of the two methods (collected versus archived original source) compared with the average of the two methods does not show a significant bias and this will be achieved in a standard means by the application of a Bland and Altman form of statistical analysis [2]. Level three validation involves the conversion of units to BrainIT units if required whilst level four validation involves the data validators to check the accuracy of data collected against source documents. Of the four levels of validation, only the last (level four) requires human resources and so to cope flexibly with periods of low funding, the BrainIT group has also defined three "types" of level four validation: type one or *self validation* where the Principal Investigator validates his own data. Type two is *cross validation* and involves colleagues in nearby BrainIT centres validating each other's data whilst type three is *independent validation* and involves the data validator who has no link with the BrainIT centre. Each patient

Table 1. *Data collected from centres*

Data sent to Glasgow	
Number of centres contributing data	15
Total number of patients recruited	115
Number of patients on database	50
Requests for missing data	21
Average number of requests per patient	2

record in the database is sortable on both *level* and *type* of validation allowing individual investigators to prescribe their own validation security level. For example, for hypothesis generation studies levels 1–4 can be used, but for analysis intended for publication only level four data can be used but with a choice of validation *types* 1–3 (self validation → independent data validators).

Results

The project officially started in September 2002, although staff were not employed until January 2003. The first year of the project was spent developing hardware and software methods and in training staff. Currently 30 centres have agreed to participate and contribute data. Full data recruitment started in January 2004. A total of 115 patients from 15 centres currently have been recruited to the study of which all of the minute-by-minute physiological data, clinical management and demographic raw data have been transferred to Glasgow. To date, 50 patients' data has been converted to the common data format and entered into the project database (Table 1).

Twenty one requests for missing data were sent out from Glasgow to the relevant centre giving an average number of requests per patient of two. Some centres have still not recruited patients and there are several reasons for this. Out of range data will also be highlighted and requests for verification of such data will be sent to the contributing centre.

Discussion

The aim of this study is to standardise the collection of data from head-injured patients in a previously defined set format. This approach ensures that all those involved are collecting data in the same format. Using this approach, data from different centres can be combined together to provide a larger better standardised data set than would normally be available from a single centre. We have set up systems so data transfer can be done over the internet and tools for data validation produced and evaluated. The project is still developing and the techniques still require to be evolved into an automated process. However we have demonstrated that our data transfer process is successful and that missing data can be readily identified and requests made to local centres for such data. The production of a reliable fully populated database from head injured patients provides, for the very first time, a means to access data that can be used for analysis and hypothesis testing. Storage of validated data in this way ensures a firm evidence base which contributors may draw upon and use to ensure standardised practice for head injuries.

References

1. Berkowitz M (1993) Assessing the socioeconomic impact of improved treatment of head and spinal cord injuries. J Emerg Med 11: 63–67
2. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 8(1): 307–310
3. Chesnut R (2004) Management of brain and spine injuries. Crit Care Clin 20(1): 25–55
4. Maas AIR, Steyerberg EW, Murray GD, Bullock R, Baethmann A, Marshall LF, Teasdale GM (1999) Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations. Neurosurgery 44(6): 1286–1298
5. Marshall LF, Becker DP, Bowers SA, Cayard C, Eisenberg H, Gross CR, Grossman RG, Jane JA, Kunitz SC, Rimel R, Tabbador K, Warren J (1983) The national traumatic coma data bank. Part 1: design, purpose, goals and results. J Neurosurg 59(2): 276–284
6. Piper IR, Citerio G, Chambers IR, Contant C, Enblad P, Fiddes H, Howells T, Kiening K, Nilsson P, Yau YH, for the BrainIT group (2003) The BrainIT group: concept and core dataset definition. Acta Neurochir (Wien) 145: 615–629

Correspondence: I. R. Chambers, Regional Medical Physics Department, Newcastle General Hospital, Newcastle upon Tyne, NE4 6BE, UK. e-mail: i.r.chambers@ncl.ac.uk