Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data

ANTHONY STELL[1], LAURA MOSS[1, 2, 3], IAN PIPER[1, 3]
[1]Dept of Clinical Physics, University of Glasgow, Glasgow, UK
[2]Dept of Computing Science, University of Aberdeen, Aberdeen, UK
[3]Dept of Clinical Engineering & Bioengineering, NHS Greater Glasgow & Clyde, Glasgow, UK
a.stell.1@research.gla.ac.uk

Abstract

An informatics issue common for many fields of medical research is the poor standardisation of baseline clinical management data, which can have a large negative impact on the statistical power of drug studies making use of that data. This baseline variation can be for many reasons – e.g. rarity of the condition – but, despite the development of standardised medical guidelines in many areas, it is still often the case that study data is affected by the "real world" differences in treatment protocols. To improve understanding of that management baseline in general, this paper describes work that builds up an empirical treatment pattern from retrospective intensive care unit (ICU) data. The ultimate goal is to build protocol "objects" that can be compared between specialist centres or "gold standard" guidelines. Variation and differences between these objects can then be quantified - and potentially mitigated - to allow a more standardised comparison of data for studies, as well as providing information on audit and guideline adherence. From a combination of event detection from high-resolution physiological output and association of those detected events with annotated treatment information, an empirical *data-driven* notion of treatment protocols across specialist centres can be built. Using data drawn from Traumatic Brain Injury (TBI) studies, the initial steps of this technological work – including the algorithms and assumptions of these two key functions – are presented. The results when applied to a specific TBI data-set (Piper et al 2010) show how the event numbers vary when key parameters are changed (e.g. the hold-down time) and how this impacts clinical decisions and trial conduct.

Keywords: treatment-protocol, study data, event-detection

1 Introduction

Certain areas of medical research suffer from low statistical power – the ability to identify a statistically significant result – in drug and intervention trials and studies. There are many reasons that can contribute to this: the rarity of the medical conditions; the complexity of the organ affected; etc. Often the result of this low power is that trials must recruit from larger population distributions and necessarily involve more centres that specialise in the condition in question.

When this occurs, a large source of variation now includes the differences in treatment protocol between those specialist centres. A move to standardise the administration of treatment procedures has gained ground over recent decades, resulting in the widespread development and adoption of clinical guidelines (Woolf et al 1999). Following a clinical guideline allows clinicians to follow reproducible treatment procedures in a standard manner. Whilst providing the best available information on reproducible care, this standardisation in treatment and care also helps progress medical research using the bedrock of the scientific method: refine and improve treatment by understanding the current environment then vary one parameter at a time and monitor the effects.

Despite this significant advance in medical treatment provision, it is still the case that guidelines are not always followed. This disparity can occur for many reasons but a common and important issue that has a subtle impact on the inputs to trial and study data, is the difference between *reported* and *actual* treatment or care actions.

It is this difference that the overall goal of this work will attempt to capture technologically, by analysing retrospective study data in the Traumatic Brain Injury domain (TBI). TBI is a prime example of a medical research area with an abundance of low-power trial data. It is widely acknowledged that the progress in understanding putative drug treatments for TBI has been greatly hindered by this lack of useful trial data (Lu et al 2012). The data provides little insight into effective treatment data because of the complex medical processes involved in TBI, which means that any trial must recruit large patient numbers, and therefore has to recruit from widely distributed areas. The result is an overall lack of confidence in treatments and interventions for TBI – most explicitly noted in the lack of certainty underpinning the recommendations made by the authoritative guidelines in the space, made by the Brain Trauma Foundation (BTF) (Bullock et al 1996). It should also be noted that these issues of poor trial data are certainly not exclusive to TBI and exist in other areas such as adrenal cancer where the rarity of the condition means that trials must recruit globally and similarly try to use novel analysis techniques to gain useful insight (Schteingart 2005).

To pursue a solution to this general issue, a worthwhile endeavour is to analyse the data-sets of clinical trials and other studies that already exist and attempt to build patterns of treatment protocols that can then be re-used for future trials. The technological proposal outlined here is an approach for the compilation

and subsequent comparison of clinical workflows between specialist centres, clinicians and patients. This approach can be broken down into the following stages:

1. Detection of events from physiological ICU time-series data
2. Association of those events with treatment information
3. Compiling these associations into a pattern of treatment (a "protocol object") that can be expressed in a standard manner
4. Comparison of these protocol objects between centres within study data-sets (and outside those data-sets for validation)
5. Quantifying the differences that are detected
6. Mitigating or accounting for these differences so that the inputs to trial or studies can be more fairly understood. This is the critical aim of the work – mitigating these differences will potentially allow trial data to show more statistically significant findings.

This paper describes the development and implementation of an algorithm to detect events and associate corresponding treatment from patient data (steps 1 and 2 in the list above).

2 Background

2.1 Trial data

Poor trial data is an issue that potentially affects all areas of medical research. In the specific domain of TBI, there is a general acknowledgement that trial data lacks the statistical significance to move the understanding of treatments forward (Lu et al 2012). Several Cochrane reviews (systematic reviews that analyse a collection of studies to draw additional insight) have been conducted and their findings are inconclusive (often contradicting the original "assumed" clinical finding, such as the use of barbiturates therapy in TBI (Roberts and Sydenham 1999)). It is the case that meta-analyses are only as effective as the studies that they collectively review. Therefore, if those studies suffer from bad design or poor numbers and detail, then a meta-analysis won't highlight anything new. The IMPACT project is an example initiative that attempts to solve the resulting problem using statistical techniques (Maas et al 2010), by modifying the outcome information into more detailed categories and specifically surveying the strictness of enrolment criteria. The results of this are that statistical efficiency is improved by 40%. However it is acknowledged that further validation of these results and an investigation of alternative methods is required. Differences in the baseline clinical management have been quoted as a primary concern in the lack of significant study output (Lingsma et al 2011), so attempts to analyse the nature of standardisation in this area is worth pursuing.

To investigate differences in baseline data it is instructive to first understand the work conducted so far in standardising general treatment and other study protocols. This is best done by looking at the authoritative clinical guidelines in the domain. In TBI, these are the guidelines compiled and maintained by the Brain Trauma Foundation (BTF), which cover all types of situations including intensive care stays, emergency accident-scene care and other specific situations such as trauma sustained whilst in military combat (Bullock et al 1996). In recognition of the varying certainty of the evidence behind their effectiveness, the BTF guidelines provide a tabulation of the confidence level behind a specific recommendation. Three broad classifications of guidelines are published by the BTF (in decreasing order of certainty): Standards, Guidelines and Options. The classification of a specific guideline is based on the classification of the supporting evidence: level 1, 2 and 3 treatment recommendations, supported by class 1, 2 and 3 evidence respectively (and again in decreasing order of certainty of efficacy). Surveying the TBI online searchable guidelines, it appears to be the case that there are very few level 1 treatment recommendations, and therefore a corresponding lack of standards (Shafi et al 2008). This appears to be especially the case for ICP monitoring, a particularly invasive treatment, which unfortunately has been identified as one of the primary avenues of potential progress in the treatment of TBI.

The use of these guidelines has been demonstrated to be clinically effective (Faul et al 2007), but controversy on their utility still exists (Pascual et al 2011) and much work remains in providing broad agreement in their recommendations. A recent alternative viewpoint to guideline-based treatment is a move to set up research infrastructures that support personalized medicine (Saatman et al 2008). This approach attempts to embrace the variation in data, dealing with patient information on a case-by-case basis, though it is hard to see how this can be expanded to more generalized solutions without establishing what how the individual differences between patients arise.

2.2 Technical background

The technical details of what is proposed in this work require the understanding of events in the context of physiological monitoring. Work in this area has focused on the definition of a physiological event through the Edinburgh University Secondary Insult Grade (EUSIG) (Jones et al 1994). These focus on the specific details of what physiological values should be used for threshold crossing (e.g. a value of greater than 100 beats per minute (bpm) for heart rate) and hold-down times (e.g. greater than 100 bpm for 10 minutes), but do not necessarily cover all the clinical definitions that are accepted (Donald et al 2012). Hence a valuable area to investigate is to look at the spread of these definitions and how they are represented in output ICU data. Detecting such physiological monitoring events in the context of high-resolution ICU data is a concept well-represented in the commercially available systems that can be found in modern ICU centres. Systems such as ICM+ (Smielewski et al 2005), Philips CareView (Philips 2013) and Datex Ohmeda (GE Healthcare 2013) are built upon algorithms similar to the event representations referred to above, some of which have the ability to actively vary threshold warnings in response to what the favoured clinical

definition is at a particular centre (Otero et al 2009). Therefore, the data analysed using these event definitions and sourced from these and similar systems, are highly relevant to the analysis.

In terms of associating treatment annotations to detected events, causal association is a very difficult problem that requires much contextual information to unambiguously establish. High specification of the data-set used is the ideal method (e.g. a clinician directly highlighting what event they are administering the treatment for) but it is often the case that such specification is not available (Enblad et al 2004). Work has been conducted to attempt to mathematically attach a treatment to a particular event, but these necessarily have an element of probability, and hence a confidence measure (Sackarelles et al 2010). Therefore, any method that tries to establish this association can only do so to a certain degree of certainty.

In regard to the later steps outlined in section 1, there are many technological initiatives that have a potential impact on this area of evaluating data profile and clinical work-flow differences. Those that most closely link to the idea proposed (of building complex protocol objects) are clinical work-flow systems such as ProForma (Sutton and Fox 2003), which allow guidelines to be compiled and "run" (or enacted) in a programming environment. Similarly, various standards exist for describing the entities that would be required for electronic representations of clinical guidelines (Boxwala et al 2004). The work presented in this paper uses the terminology of object-oriented programming ("instantiation", "attributes", etc) but literature in this domain appears to have no analogy to such an approach. The ultimate goals of creating complex protocol objects also require further investigation of the best ways to parameterise, compare and measure the similarities of such objects. Initial searches in this area have highlighted the possibilities of using measures of semantic similarity (Pederson et al 2007) or using a case-based reasoning approach to establishing patient data profile similarity (Kumar et al 2009).

3 Key technical requirements

The first two functions detailed in the list in section 1 – event detection and association of treatment information – have their own set of specific requirements that are now outlined below.

3.1 Event detection

The requirements to detect events in a set of high-resolution time-series data are 1) an understanding of the structure of the event that is being detected (i.e. the object), and 2) the specific numerical knowledge that populate that structure (i.e. the instantiation of an object for a particular domain). In the case of an event, the key structural characteristics are:

- Event threshold for event start and finish – the value which acts the trigger for knowing when an event may have started
- Event hold-down – the time beyond the threshold that indicates when an event has unambiguously occurred
- Clear hold-down – the time below the threshold that indicates when an event has unambiguously finished)
- Duration
- Value range

Figure 1 shows a schematic of a single physiological monitoring event, with a time-window for treatment overlaid (see section 3.2 for discussion of time windows).

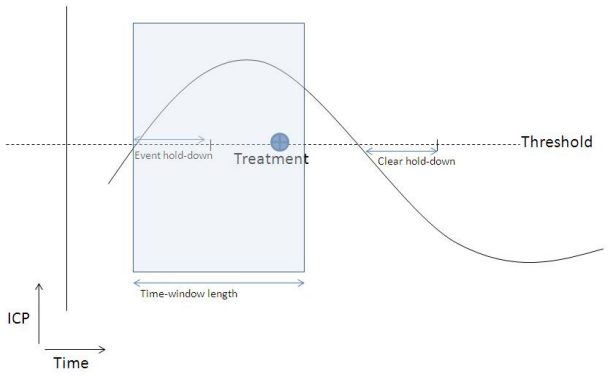

Figure 1: event definition for a given time-series physiological output. A threshold crossed for a specific period (the hold-down) indicates that an event has started. Clear hold-down indicates that the event has finished. Also shown are a treatment at a specific time-point and a time window overlaid for association of that treatment with the event.

To use the object-oriented programming analogy, this event object can be thought of as a complex structure with various attributes. In the context of this work, it is very likely that the structural details of the defined object will never change, as this is a generally accepted definition of an event throughout medical literature (Donald et al 2012). This makes it a re-usable pattern ideal for use in searching physiological time-series data.

In terms of the numerical content for the TBI domain, preliminary analyses suggest that when looking at key medical parameters – intracranial pressure (ICP) and cerebral perfusion pressure (CPP) – optimum values of 20mmHg and 60mmHg respectively serve as the most popular thresholds for specialist neurosurgical centres (Jones et al 1994).

With these key pieces of structural and numerical information about event definition, an analysis program has been built that detects this event pattern within the data-set and compiles related metrics, such as event numbers, distribution, and duration. By varying the key numerical input information (e.g. move event threshold value up or down), these metrics will change and provide information about the overall clinical situation.

Though not addressed in this paper, a further significant refinement to this pattern-matching requirement is to recognise the common characteristic of extended periods of volatility in monitored physiological

output (groups of events). This is discussed further in section 6.

3.2 Association of treatments with events

The second key function is the association of treatment information with the events detected from the physiological data. A feature common to nearly all modern high-resolution ICU data-sets is the annotations of treatments administered to a patient during their stay in intensive care, such as a nurse administering analgesics to provide pain relief or a ventilator machine being attached to a patient to allow steady assistance of breathing.

As shown by those two examples, the structure of a treatment object varies depending on the nature of the treatment and can be simple or complex. The simplest form of representing a treatment would be a timestamp, and a dosage of a certain amount of drug. A more complex object would be the attachment of the ventilator, which has start and end points, duration, and a range of values depending on the breathing assistance given. Other features could also be added to these lists (thereby increasing the complexity).

For the purposes of the analysis described here, the treatment information has been reduced to the simplest point-like structure possible, consisting of only a timestamp, a value and a label. Where the treatments have more complex structures, the treatment information has been deconstructed to use the start and end points as the individual timestamps. The long-term view however is to develop the analysis to incorporate more complex definitions of treatments.

Therefore, the method of association presented involves noting a simple treatment point in the timeline associated with the physiological data. The event detection algorithm is applied to the physiological dataset and for each event detected, several time-windows differing in length (30, 60, 90 and 120 minutes) are explored around the event to identify a corresponding treatment. The first treatment found in this time window is assumed to be in response to the event that the time window is associated with. Limitations to this approach include: that there may be more than one treatment falling within the time window, relating to that event; or overlapping event time-windows may confuse the particular association of a treatment with an event. These limitations are deemed to be acceptably negligible for the analysis run so far (estimated to be less than 1% of the overall event numbers), but must be incorporated as the work progresses.

Association between two events and actions can be calculated in many ways (see section 2). In this context, the association being discussed is causal (we are attempting to establish where a treatment was applied in response to a particular event). If the data is well annotated with treatment information then an explicit parameter target (though not necessarily the exact event) will have been noted, but this level of detailed treatment information is not always available in medical data-sets. Also, the primary goal of the work is to establish treatment protocol independent of the input from the clinician themselves, so the ideal situation is where the treatment target is established without explicit direction from the clinician. However as discussed in section 2, association is an open (and largely unsolved) research question – in this context, the assumptions made to establish the association between treatment and events are as good as can be enacted in this context and time. This is an area of work that could be followed up as a separate avenue of research.

4 Method

This section describes the relevant clinical input used and the algorithms constructed to detect events and associate treatments.

4.1 Numerical instance input

As described in the previous section, for the purposes of detecting physiological monitoring events, the structural information describing such an event will remain unchanged. To implement the methods on real data-sets, the attributes listed in section 3.1 were populated with varying numerical information depending on the instantiation of the event objects. By varying the metrics in this way, information can be derived about what definitions of ICP and CPP events are most commonly used by clinicians.

Following from the most clinically relevant definitions of ICP and CPP events (see section 2), eight parameter definitions are used to cover the most likely definitions. The key point about these parameters is that the threshold value and directions indicate when an event has started or finished (noted in italicized brackets). For instance, the first definition indicates that a raised ICP event will have considered to be started once the ICP goes above 10 mmHg.

- Raised ICP #1 (> *10 mmHg*)
- Raised ICP #2 (> *15 mmHg*)
- Raised ICP #3 (> *20 mmHg*)
- Raised ICP #4 (> *25 mmHg*)
- Raised ICP #5 (> *30 mmHg*)
- Lowered CPP #1 (< *50 mmHg*)
- Lowered CPP #2 (< *60 mmHg*)
- Lowered CPP #3 (< *70 mmHg*)

The other required instance data points are the hold-down and clear hold-down times of an event. These represent the time period after a threshold-crossing where the output continues to remain above that threshold and therefore an event can be considered to have unambiguously occurred. The same method (with opposite polarity) applies at the end of an event – known as the clear hold-down time – to indicate when an event has unambiguously ended. Again, this is an attribute that remains structurally constant, but the numerical value of which can be (and is) varied according to different clinical specialists. In a similar fashion to the threshold value definitions – beginning with the most clinically relevant definitions – four values are applied representing the differences in hold-down and clear hold-down times.

These are 5 mins, 10 mins, 15 mins and 20 mins. Therefore, there are a total of 32 (8 * 4) ways that a physiological monitoring event can be detected in a data-set.

The last attribute requiring numerical variation is the time-window that provides the period over which the detected event can be associated with an annotated treatment.

The basis for the size of the time-window has been the specified reaction times for clinicians in an ICU setting (i.e. administering a drug in response to an event can be reasonably expected to be around 30 minutes). However, a large uncertainty occurs in this variable as the actual administration time can vary to a great degree from the administration reporting time – a doctor saving a patient's life was too busy saving their life, rather than reporting and annotating the treatment). Therefore, the time-window post-event can be varied anywhere from 30 minutes to 2 hours, a value established by surveying the clinicians that contributed to the data-set (Enblad et al 2004). Extending the time-window before the event is a possibility considered due to reporting discrepancies, but the same survey established that pre-event administration reporting was unlikely. Therefore the four time-window definitions (starting at the point of event start) are: 30 mins, 60 mins, 90 mins and 120 mins. With the four time-windows, the total number of analyses for every pass of the data becomes 128 (32 * 4).

The results of this association approach are simple number counts of the treatments that fall within those time windows. Other metrics that can be compiled using this approach include measuring the time to an associated treatment. The time of all the treatments from their associated event start are listed, and the mean and median values are calculated. Also the annotated treatment types are noted ("analgesia", "sedation", etc) so that an understanding of what treatments are administered and in which centres can be built into a definitive list. The treatment target is noted to match the treatment to the correct physiological event (i.e. only a treatment with a target of "CPP" is counted in response to a CPP physiological event).

It is noted here that in the implementation of this work (using the Brain-IT data-set – see introduction to section 5), the criteria are run against a total of 262 patients. The maximum number of analysis runs therefore becomes 33536 (128 * 262). It is beyond the scope of this current paper, but this points to further work required in the optimisation of the analysis based on growing data-set patient numbers.

4.2 Algorithms

The algorithms that drive this method are detailed in this sub-section. The output of the full process requires careful analysis as the structure of the association data is listed per patient, however the number count totals will be compiled as the code traverses the 262 patients (e.g. we want to know the total number of events that have associated treatments for ICP > 20 mmHg, with a hold-down of 10 minutes, within a time-window of 60 minutes. This needs to be totalled up from each patient, then distributed throughout the final totals).

4.2.1 Event-detection algorithm

To detect events from a physiological output stream, the following algorithm is used.

1) Compile the list of parameter objects ahead of processing. This is a list of the eight different definitions of ICP and CPP. A parameter in this context represents a physiological data stream – the parameter referring to a physical measurement of the patient's brain. A representative parameter object is shown in figure 2.

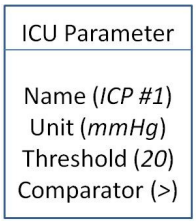

Figure 2: ICU parameter object with the values required to define an event within the data stream (example values in brackets)

This list of parameter objects constitutes part of the minimum required domain knowledge to allow event detection in a physiological data stream to occur. This can be read in from any persistent data store, such as a database, a properties file or an XML ontology file.

2) Querying the patient database: for each patient:
   - Read the patient data into an *"n x n"* vector of vectors (i.e. a matrix).
   - Each line in the object is a time-point (as the sampling rate is minute by minute, therefore each line increments by a minute) and each column is a particular parameter feed
   - The header line is used to identify the column index for the parameter that is of particular interest (ICPm, CPPm, etc)

3) For each parameter in the list compiled in step (1):

   - Retrieve all of the parameter information for that indexed parameter object (name, unit, threshold, etc).
   - For each hold-down definition:
     - Read in the line, timestamp and value (from step 2) and check the time between this timestamp and the last
       - if (gap > 1min)
         - Reset all event metrics and jump to end of the entire checking loop
       - if (event is in progress)
         - Is value still above threshold?
         - if (no)
           - Is the clear condition met?
           - if (yes)

        - if (*potentialClear* option is false)
            - Set the *potentialClear* variable to true
            - Increment the clear hold-down count
        - if (clear hold-down count equals the hold-down definition)
            - Note the event end time and add to event object
            - Add the event to the list of events
            - Increment the event index
            - Add value and timestamp to the event list
- if (event is not in progress)
    - Is value still below threshold?
    - if (no)
        - Is the event condition met?
            - if (yes)
                - if (*potentialEvent* option is false)
                    - Set *potentialEvent* to true
                    - Set event hold-down count to zero
                - else
                    - Increment the event hold-down count
            - if (event hold-down count equals hold-down definition)
                - Note the event start time and add to event object
            - if (event condition NOT met)
                - Reset *potentialEvent* to false
                - Reset event hold-down count

Using this algorithm, the events are extracted for all 32 definitions.

4.2.2 Event/treatment association algorithm

For each patient an association object is instantiated, shown in figure 3.

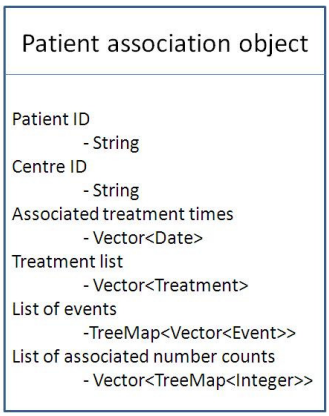

Figure 3: a patient association object containing identifiers, a list of associated treatment times, treatment values (e.g. sedation etc), events and association numbers. A tree-map structure is used to store the indexed information so that the data can be retrieved in order when traversed for output.

To associate the events and treatments for each patient, all treatment information is retrieved, then for each parameter (defined in step (1) of event detection), for each hold-down definition, and for each time-window definition, the following association algorithm is run.

For each event:

- Get the event start time
- Define a time-window instance that begins at the event start and lasts for e.g. 30 minutes
- For each treatment:
    - Get the treatment time
    - Get the treatment target, description and value
    - Isolate all treatment instances that have the tags “cpp”, “icp” or “hypotension” anywhere in the three string values
    - Isolate all the treatments that are end tags
    - If the treatment time is within the time-window bounds:
        - If the treatment is not an end tag and the event does not already have an associated treatment:
            - Add the treatment to list of treatments associated with this event
            - Set the Boolean flag indicating the event now has an associated treatment
            - Get the time to this treatment
- Add the list of associated treatments to the time-window object for this event
- If the associated treatment list is greater than zero:
    - Increment the associated event counter
- Add all time and treatment data gathered to the patient’s association data object and return this to the calling function.

The patient’s association data is then used to output the details to a text file and illustrative charts. As the association data is per-patient, the information must still be traversed for all summary counts and centre-specific information to be output.

These algorithms and processes form the core of the key functions required by the work so far. In the next section, the results achieved when applied to the real-world Brain-IT traumatic brain injury data-set are shown.

5 Results

The Brain-IT core data-set is a repository of 262 patients drawn from specialist neurological centres around Europe, collected with a view to enabling follow-on post-hoc analyses. One of the most comprehensive collections of high resolution TBI data with treatment and surgery

annotations to date, it forms a detailed retrospective view of physiological and treatment data that is well suited to analyses such as the one described in this paper. For more information about the specific composition of this data-set see Piper et al 2010.

Although the BrainIT dataset contains a high number of neurological ICU parameters (e.g. surgery, neurological response, demographics, etc) the work described in this paper focuses on the high-resolution physiological data and the annotated treatment data.

5.1 Data coverage

Preparatory to the analysis, the coverage of physiological parameters in the database is summarised in table 1. Coverage is defined by dividing the number of data points that are not "null" or blank by the overall number of data points for that physiological stream, and calculating the resulting percentage. Parameters with coverage less than 10% are omitted as contributing negligible information.

ICU Parameter	Coverage
RR	26%
HRT	87%
BPs	84%
BPd	84%
BPm	96%
ICPm	84%
CVPm	20%
CPP	82%
TC	70%
SaO2	82%
SaO2pls	23%
ETCO2	19%

Table 1: physiological parameter coverage in the Brain-IT database

By inspecting the coverage for the data points used, a level of initial confidence can be gained to see how well represented the data-points are. If the parameter stream is well covered, it is a reasonable expectation that the event detection algorithm will produce useful information.

From the results shown in table 1, we can see that the blood pressure measures are all at least above 80%, including those of particular interest: mean intra-cranial pressure (ICPm) and cerebral perfusion pressure (CPP).

5.2 Event detection

The event detection algorithm is run across the 32 definitions referred to in section 4.1. Figure 4 shows the relevant results for ICP.

The bar chart in figure 4 shows the total counts of events as detected by the algorithm outlined in section 4. The graph shows 20 columns (the five ICP parameter definitions multiplied by the four hold-down values). The interpretation of the cyclical shape for every four bars makes intuitive sense: as the hold-down value for that particular definition increases, the number of events captured goes down (e.g. an event with a hold-down value of 20 minutes will be less common that one with 5 minutes hold-down). However, slightly less intuitively, the overall numbers of these four-bar cycles will vary according to definition.

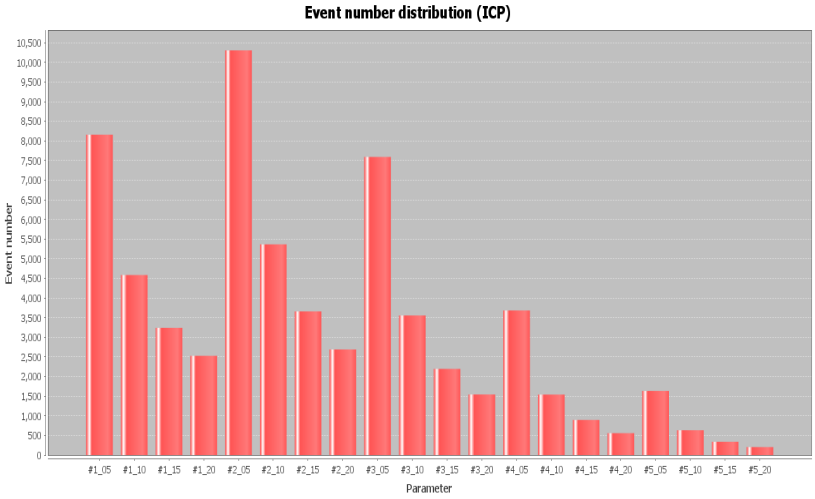

Figure 4: event number distribution for ICP.

From figure 4 we can see that the most populous number of events in ICP come from definition #2 (a crossing threshold of > 15 mmHg). This represents a minima inflection point that will inform the next steps of treatment association. It is believed that definition #3 (< 70 mmHg) of CPP also represents this inflection point, but analysis time ran out before the writing of this publication and must be investigated further.

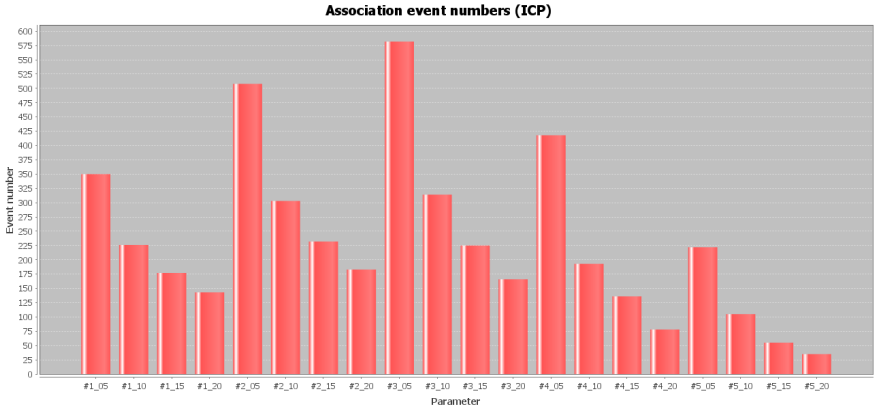

Figure 5: number of ICP events that have treatment associations when a time window of 30 minutes is applied.

5.3 Treatment association

The overall number of treatments annotated in the data-set is 19175. The association of a treatment with an event provides a modifying parameter to the overall event number counts discussed in section 5.2. The inference

that can be made is that general clinical definitions of events dictate that an ICP event occurs when a patient's ICP crosses a threshold of 20 mmHg. The effects of this modification can be most clearly seen in the bar chart that represents the number of events *with an associated treatment* per definition per hold-down value, with a time-window of 30 minutes (time window #1), shown in figure 5.

The graph shape in figure 5 is evidently different from the event count numbers in isolation in figure 4. It is now definition #3 (> 20 mmHg) of ICP with hold-down of 5 minutes that appears to be most numerous, which would suggest that this definition input is triggering a larger number of clinical responses in terms of administrated treatments. Again, the intuitive physical interpretation of these graphs can be seen in the shape of the distribution as the time-window increases towards an asymptote of infinite time. According to the association algorithm presented in section 4, the number of events with associations will approach the total as the time-window approaches infinity (with a sufficiently large time-window, every event will have an associated treatment). Figure 6 (at the end of paper) demonstrates this progression of shape of the ICP events with treatment associations as they move through the other three time-window definitions (gradually definition #2 predominates again).

5.4 Treatment composition

For every combination of parameter definition, hold-down and association time-window, a composition of the actual treatments included in the list can be constructed. It is this information that will eventually inform the construction of a predominating treatment protocol. Figure 7 shows the treatment distribution for ICP definition #3 (> 20 mmHg) with a hold-down value of 5 minutes and a time-window for association of 30 minutes (the predominating EUSIG definition of an ICP event and "most reasonable" association time – see section 2).

The top three treatments applied in this instance are paralysis (18.2%), sedation (17.2%) and osmotic therapy (16.2%), from an overall number of 582 events with treatment associations (7.7% of the total event number for this ICP definition).

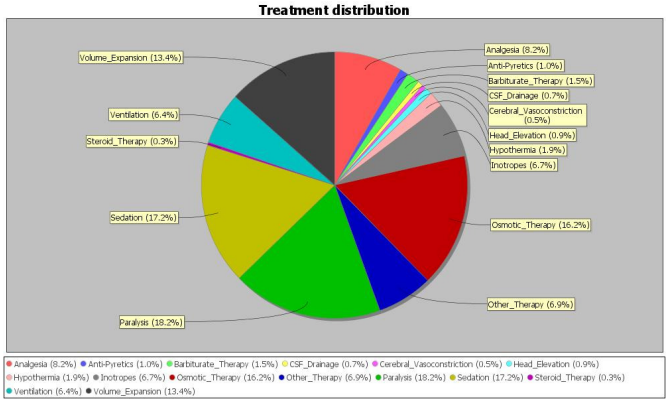

Figure 7: treatment distribution for ICP definition #3 (> 20 mmHg) with a hold-down value of 5 minutes and a time-window of association of 30 minutes.

5.5 Centre-specific information

Using the unique centre reference identifier, event counts, associated treatment counts and treatment composition and times can be applied for each individual centre. For the purposes of discussing the technological ability to derive a centre-specific protocol, the results for the centre in Uppsala, Sweden are as follows:

The top three treatments are paralysis (32.9%), analgesia (13.4%) and a joint third place (11.4%) for ventilation, volume expansion and sedation, from an overall number of 373 events with treatment associations (3.1% of the total event numbers for this centre and ICP definition).

Combined with the metrics for treatment times (measured as well but not presented in this paper), a profile can be built up for a specific centre detailing frequency, response time, and how this profile compares to the guideline-mandated responses or specific study protocols.

6 Discussion and development

The results presented in section 5 describe the first steps in drawing physiological event and treatment information out of high resolution ICU data and using this to form an empirical treatment protocol. In this example we have been able to extract from the results a predominantly accepted clinical definition of an ICP event (> 20 mmHg, 5 minute hold-down).

There are several limitations in this analysis. One is that causality and association of two objects is a difficult process to accomplish. A major assumption has been the one-to-one relationship between an event and an annotated treatment. An estimated measurement of many-to-one (treatment to event) instances has been made and seems to be negligible but requires more exhaustive investigation.

Other limitations are that the structural information used to define a physiological monitoring event object may require modification. The structural description of a single event is well described in this paper, however as mentioned previously, a further extension of this definition, is to describe multiple events as a more complex structure. A collection of events within a specified period, with a specified gap between them, would indicate an "episode", which would have further characteristics likely to be different in nature to an isolated event (for instance, does the notion of a hold-down time still apply to a collection of events?) Other types of events not captured in this work – such as very long-term "events" (e.g. a patient who has ICP just over the threshold for many days) or refractory events (those that end in patient death) – need to be considered. It should also be noted that the data-set used has been compiled by specialist centres that are known to focus on ICP and CPP measurements, so the detail in these readings may be unusually high.

Despite these limitations the initial results do show promise: - an accepted clinical definition has appeared

and a usable treatment pattern has been derived, which, after further validation, can now be expressed in a work-flow language to progress this research.

The next steps in validation are to run the same processes against different data-sets. There are several that have been identified as potentially useful in this space. First is the MIMIC II data-set at MIT (Saeed, Singh and Sanyal 2002). This is a publicly accessible collection of high-resolution ICU data. It is ideal in that it captures similar data-points to that of the Brain-IT dataset but is not specifically targeted to traumatic brain injury. If similar findings in ICP and CPP can be found then this would be a highly significant validation of the event detection and treatment association processes.

Second is the ICM+ (Smielewski et al 2005) data-stream based at the neuro-trauma unit at the Alfred Hospital in Melbourne, Australia. This is very high resolution data (waveform) that is targeted towards traumatic brain injury. The challenge, and independent validation, here would be condensing the information down to the same sampling rate (minute-by-minute) and understand if the processes produce the same result.

The third major data-set sourced to date that would be applicable, is the next generation of the Brain-IT data: the Avert-IT data repository (Stell et al 2009) with data capture tools (Philips CareView and Rhapsody products) which are in place at the Southern General Hospital in Glasgow, UK.

The question of data quality should also be addressed here – the treatment annotations of the Brain-IT dataset have been noted to not have a high degree of (temporal) accuracy. This was largely due to the manual methods of treatment information input that were required. As this is such an important issue in establishing the validity of association, these other data-sets have been specifically inspected to make sure the accuracy of the treatment time-points are as high as possible. Examples of this include the connection of intubation pumps to the bedside monitoring system (allowing the immediate recording of an automatic or manual drug administration) and touch-sensitive detection of hypothermic induction blankets, connected to the Philips Rhapsody system.

As patient information capture technology improves, it is assumed to be the case that the data quality and accuracy improves commensurately. However, another possible avenue to this work is to attempt to infer where a treatment has been administered by analysing the physiological output in a way opposite to the methods described in this paper. The potential value of "repairing" a patchy data-set such as this could be very high indeed, however the number of variables and uncertainty in this procedure would also be high. Either way, these two parallel strands (better capture technology and data improvement algorithms) will likely result in the same outcome: high quality data that provides more useful insight on the state of a patient's physiology, which can contribute to the overall goal of improving study data. The numbers of associated events shown in section 5 are of low representation (7.7% overall and 3.1% for the specific centre), so all work to improve this would be wise to investigate and pursue.

If these validation procedures give rise to results similar to those described in this paper, it will be a significantly general technological solution to extracting treatment protocol information from widely available ICU data. However, in pursuit of the more ambitious goals of this overall work, the next step is to translate the treatment protocols derived into a form of abstract expression (referred to in the introduction as an object). A variety of possibilities exist in this space – for instance ProForma (Sutton and Fox 2003) is a clinical work-flow project that expresses guidelines as actions, conditions and states. But the primary characteristics that must be captured are a more detailed exploration of the nature of the treatment and event objects, and their individual and aggregate relationships. The key relationship measures will be the medical impact of the treatment, the temporal distance and causal characteristics between the treatment and event. In compiling a more complex protocol object that encapsulates this relationship, a key mathematical requirement will be the ability to parameterise such an object and allow it to be compared against others of a similar nature.

The ultimate goal of the work remains to analyse the data variation itself and find a way to compare this against other sets of data (either for trials and studies or drawn from general ICU data collection). The steps described in this paper are a successful attempt to draw out the initial information but many more challenges lie ahead to make this approach to standardising trial data a reality.

7 References

Piper, Chambers, Citerio, et al (2010): The brain monitoring with Information Technology (BrainIT) collaborative network: EC feasibility study results and future direction, *Acta Neurochirurgica*, **11**(152):1859-71

Woolf, Grol, Hutchinson, et al (1999): Potential benefits, limitations and harms of clinical guidelines, *British Medical Journal*, **318**:527-530

Lu, Gary, Neimeier, et al (2012): Randomized controlled trials in adult traumatic brain injury, *Informa Healthcare*, **26**:1523-1548

Bullock, Chesnut, Clifton, et al (1996): Guidelines for the management of severe head injury, *European Journal of Emergency Medicine*, **2** (3):109-127

Schteingart (2005): Management of patients with adrenal cancer: recommendations of an international consensus conference, *Endocrinology related cancer*, **12**:667-680

Roberts, Sydenham (1999): Barbiturates for acute traumatic brain injury (Review), *Cochrane Database Systematic Review*

Maas, Steyerberg, Marmarou, et al (2010): IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury, *NeuroTherapeutics: the journal of the American Society for Experimental NeuroTherapeutics*, **1**(7):127-34

Lingsma, Roozenbeek, Perel, et al (2011): Between-centre differences and treatment effects in randomized controlled trials: a case study in traumatic brain injury, *Trials*, **1**(12):201-206

Shafi, Diaz-Arrastia, Madden, et al (2008): Intracranial Pressure Monitoring in Brain-Injured Patients is Associated with Worsening of Survival, *Trauma and Acute Care Surgery*, **2**(64):335-340

Faul, Wald, Rutland-Brown, et al (2007): Using a cost-benefit analysis to estimate outcomes of a clinical treatment guideline: testing the brain trauma foundation guidelines for the treatment of severe traumatic brain injury, *Trauma and Acute Care Surgery*, **6**(63):1271-1278

Pascual, Georgoff, Maloney-Wilenksy, et al (2011): Reduced Brain Tissue Oxygen in Traumatic Brain Injury: Are Most Commonly Used Interventions Successful, *Trauma and Acute Care Surgery*, **3**(70):535-546

Saatman, Duhaime, Bullock, et al (2008): Classification of Traumatic Brain Injury for Targeted Therapies, *Journal of Neurotrauma*, **25**:719-738

Jones, Andrews, Midgeley et al (1994): Measuring the burden of secondary insults in head-injured patients during intensive care, *Journal of Neurosurgical Anesthesiology*, **1**(6):4-14

Donald, Howells, Piper, et al (2012): Trigger characteristics of EUSIG-defined hypotensive events, *Acta Neurochirugia Supplement*, **114**:135-9

Smielewski, Czosnyka, Steiner, et al (2005): ICM+: software for online analysis of bedside monitoring data after severe head trauma, *Acta Neurochirurgica Supplementum*, **95**:43-49

Philips: Philips Healthcare Carevue/Rhapsody www.healthcare.philips.com. Accessed on 28th October 2013.

GE Healthcare: Datex Ohmeda, www.gehealthcare.com. Accessed on 28th October 2013.

Otero, Felix, Barro, et al (2009) Addressing the flaws of current critical alarms: a fuzzy constraint satisfaction approach, *Artificial Intelligence in Medicine*, **47**:219-38

Sackarelles, Shiau, Kammerdiner, et al (2010): Seizure Monitoring and Alert System for Brain Monitoring in an Intensive Care Unit, *Computational Neuroscience*, **1**:357-369

Sutton, Fox (2003): The syntax and semantics of the ProForma guideline modelling language, *Journal of American Medical Informatics Association*, **10**:433-443

Boxwala, Peleg, Tu, et al (2004): GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines, *Journal of Biomedical Informatics*, **37**:147-161

Pedersen, Pakhomov, Patwardhan, et al (2007): Measures of semantic similarity and relatedness in the biomedical domain, *Journal of Biomedical Informatics*, **40**:288-299

Kumar, Singh, Sanyal (2009): Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU, *Expert Systems with Applications*, **36**:65-71

Enblad, Nilsson, Chambers, et al (2004): R3-survey of traumatic brain injury management in European Brain IT centres year 2001, *Intensive Care Medicine*, **30**:1058-65

Saeed, Lieu, Raber, et al (2002): MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring, *Proceedings of Computers in Cardiology*, **1**:641-644

Stell, Sinnott, Jiang, et al (2009): Federating distributed clinical data for the prediction of adverse hypotensive events, *Philosophical Transactions of the Royal Society*, **367**:2679-2690

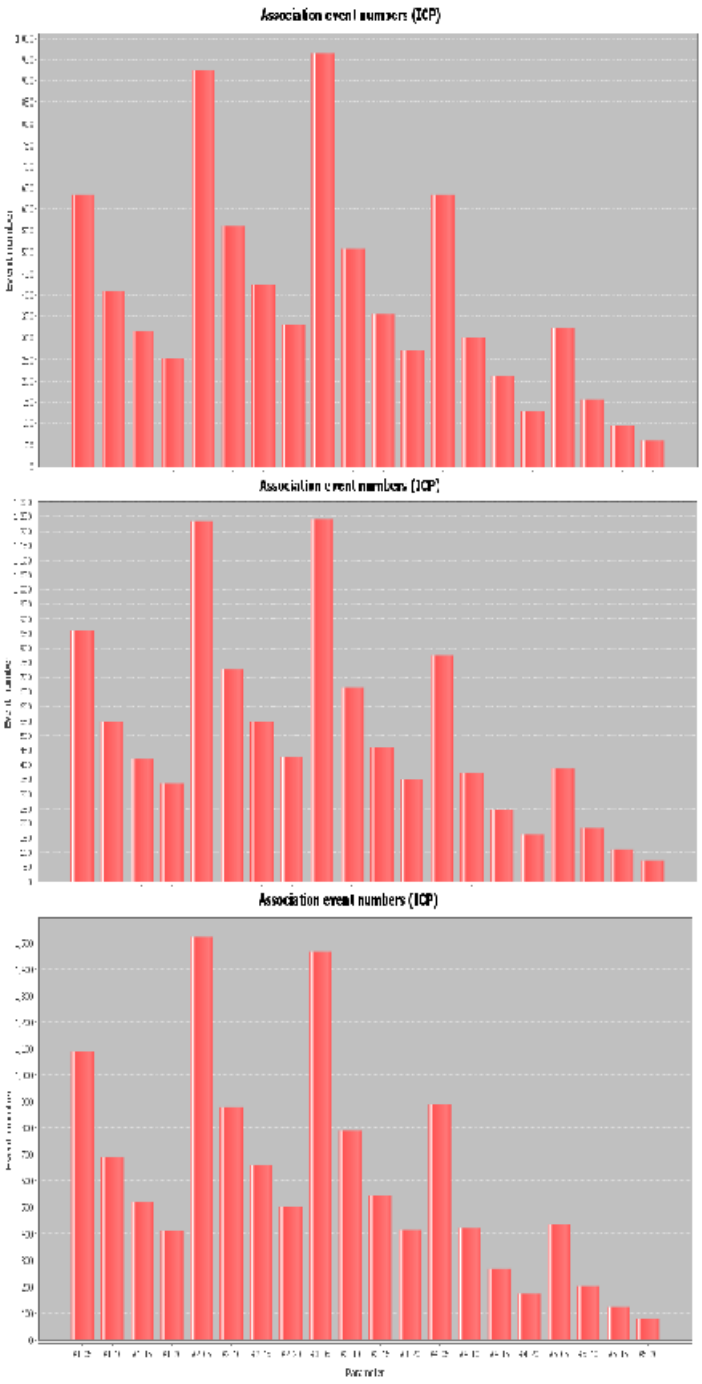

Figure 6 (left): the three other time-window definitions for ICP event and treatment association. As the time-window increases (vertically downwards), the distribution shape reverts back to that of the event count without treatment association (note the transposition of the two largest columns in particular). These provide the transition from figure 4 to figure 5.