Evaluating Clinical Variation in Traumatic Brain Injury Data

Anthony James Stell

MPhys, MSc

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Department of Clinical Physics and Bioengineering

College of Medical, Veterinary & Life Sciences

University of Glasgow

Abstract

Current methods of clinical guideline development have two large challenges: 1) there is often a long time-lag between the key results and publication into recommended best practice and 2) the measurement of adherence to those guidelines is often qualitative and difficult to standardise into measurable impact. In an age of ever-increasing volumes of accurate data captured at the bedside in specialist intensive care units, this thesis explores the possibility of constructing a technology that can interpret that data and present the results as a quantitative and immediate measure of guideline adherence.

Applied to the Traumatic Brain Injury (TBI) domain, and specifically to the management of ICP and CPP, a framework is developed that makes use of process models to measure the adherence of clinicians to three specific TBI guidelines. By combining models constructed from physiological and treatment ICU data, and those constructed from guideline text, a distance is calculated between the two, and patterns of guideline adherence are inferred from this distance.

The framework has been developed into an online application capable of producing adherence output on most standardised ICU datasets. This application has been applied to the Brain-IT and MIMIC III repositories and evaluated on the Philips ICCA bedside monitoring system. Patterns of guideline adherence are presented in a variety of ways including minute-by-minute windowing, tables of non-adherence instances, statistical distribution of instances, and a severity chart summarising the impact of non-adherence in a single number.

Acknowledgements

I would like to thank my supervisors Dr Ian Piper and Dr Laura Moss for their guidance, advice and unending patience throughout this degree. I would also like to thank Dr Martin Shaw, Dr Chris Hawthorne and Dr Roddy O'Kane for their support in providing clinical and statistical data and insights, which have contributed strongly to the outcomes of this work.

I would also like to acknowledge the work of the Brain-IT group investigators and participating centres to the Brain-IT data-set without whom this work could not have been conducted.

On a personal note I would like to thank my parents, Geoffrey and Evelyn, whose love, support and advice was invaluable in navigating the challenges of academia. I'd like to thank my two sons, Joaquin and Alonso, who bring joy to my life every day. And finally, I would like to thank my beautiful and beloved wife, Maria. I could not have achieved this without the sacrifices you made, and the support and love you gave to help me through the most difficult moments. Te amo mi amor.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Contents

Abbreviations

ABP - Arterial Blood Pressure

AIME - Artificial Intelligence in Medicine

AMITIE - Automated Medical Intervention and Treatment Inference Engine

ANOVA - Analysis of Variance

API - Application Programming Interface

ARDS – Acute Respiratory Distress Syndrome

BEST-TRIP – Benchmark Evidence from South American Trials: Treatment of Intracranial Pressure

BP - Blood Pressure

BPMN - Business Process Modelling Notation

BTF - Brain Trauma Foundation

CBF - Cerebral Blood Flow

CBMS - Computer-Based Medical Systems

CDE - Common Data Elements

CPP - Cerebral perfusion pressure

CPPopt - Optimum cerebral perfusion pressure

CRASH – Corticosteroid Randomisation After Significant Head injury

CSF - Cerebrospinal fluid

CSO – Chief Scientist Office

CT - Computed tomography

CVP - central venous pressure

EBM - Evidence Based Medicine

ETCO2 - End-tidal Carbon Dioxide

EUSIG - Edinburgh University Secondary Insult Grade

fMRI – Functional Magnetic Resonance Imaging

GCS - Glasgow Coma Scale

GCSm - Glasgow Coma Scale (motor)

GLIF – Guideline Interchange Format

GOS - Glasgow Outcome Scale

GOSe - Glasgow Outcome Scale (Extended)

HRT - Heart Rate

ICCA – Intellispace Critical Care and Anaesthaesia

ICD9 – International Classification of Diseases 9th revision

ICH - Intracranial Hypertension

ICP - Intracranial Pressure

ICU - Intensive Care Unit

InTBIr – International Initiative for TBI Research

JSP - Java Server Page

LSR - Living Systematic Reviews

MAP - Mean Arterial Pressure

MIMIC – Medical Information Mart for Intensive Care

NeCTAR - National eResearch Collaboration Tools and Resources

NHS - National Health Service

NICE - National Institute of Clinical Excellence

NSH - Neuro-Surgical Hospital

OWL – Web Ontology Language

PDA - Personal Digital Assistant

PET – Positron Emission Tomography

PrOM - Process Miner

PRx - Pressure-Reactivity Index

RCT - Randomised Controlled Trial

RR - Respiration Rate

SAH - Sub-Arachnoid Haemhorrage

SaO2 – Oxygen Saturation

SaO2pls – Pulse Oximetry Oxygen Saturation

SBP - Systolic Blood Pressure

SIGN - Scottish Intercollegiate Guideline Network

SNOMED-CT – Systemised Nomenclature of Medicine – Clinical Terms

SPARQL – SPARQL Protocol and RDF Query Language

SpO2 - Pulse Oximetry Oxygen Saturation

SWRL – Semantic Web Rules Language

TBI - Traumatic Brain Injury

TC - Temperature

Temp - Temperature

UML - Unified Modelling Language

XES – eXstensible Event Stream

List of figures

List of tables

List of Publications

The following publications have been written as part of the work of this thesis.

"Automated Measurement of Adherence to Traumatic Brain Injury (TBI) Guidelines using Neurological ICU Data", Stell A, Piper I and Moss L (2018)

- 11th International Joint Conference on Biomedical Engineering Systems and Technologies, Funchal, Madeira, Portugal, 19-21 Jan 2018, pp. 135-146. ISBN 9789897582813

This paper was shortlisted for the best student paper award at the conference and was invited for publication in the Communications in Computer and Information Science journal series, published by Springer.

"Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data", Stell A, Moss L and Piper I (2014)

- Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management, Auckland, New Zealand, 20-23 Jan 2014, pp. 79-88 (Volume 153). ISBN: 9781921770357

"Knowledge-Driven Inference of Medical Interventions", Stell A, Moss L and Piper I (2012)

- 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), Rome, Italy, 20-22 Jun 2012, DOI: 10.1109/CBMS.2012.6266389

Additional Resources

Application

The following application was created in support of this thesis and is available at the website link below.

www.tbi-guidelineadherence.org

For the purposes of access by examiners, two accounts have been created with the credentials below. A PDF user guide is available for download at the home page of the application.

Username: **examiner_external**
Password: aeNei4ph
Username: **examiner_internal**
Password: Teic4oph

Code repository

The code written to support the work performed in this thesis is available at the following repository link. A “Readme” guide is provided detailing the language and environment requirements for download and execution.

www.github.com/astell/tbi_guidelineadherence

Summary

Current methods of clinical guideline development have two large challenges: 1) there is often a long time-lag between the key results and publication into recommended best practice and 2) the measurement of adherence to those guidelines is often qualitative and difficult to standardise into measurable impact. In an age of ever-increasing volumes of accurate data captured at the bedside in specialist intensive care units, this thesis explores the possibility of constructing a technology that can interpret that data and present the results as a quantitative and immediate measure of guideline adherence.

Clinical scope

Though with potentially general application, the domain chosen was traumatic brain injury (TBI), specifically the threshold monitoring guidelines of ICP, CPP, and BP management. They were chosen due to the complexity and uncertainty inherent in TBI guidelines, combined with the availability of high-volume ICU data in the field.

Hypotheses

1) *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*
2) *Having extracted treatment processes, one is able to develop a method to compare those against other treatment processes to establish the degree of similarity between them*
3) *One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence*

Methodology

The main technological concept in this thesis is that of process models – a construct used in corporate and business domains to model time-varying processes and identify efficiencies. The process models were used to measure the adherence of clinicians to three TBI guidelines (ICP/CPP/BP monitoring thresholds) using physiological and treatment data from bedside machines in neurological ICUs.

Similarly, the relevant guideline texts from the Brain Trauma Foundation (BTF) were represented using Business Process Model Notation (BPMN) so that a comparable process model could be constructed. Building on previous comparison work between process models (Dijkman, Dumas and García-Bañuelos, 2009), a "distance" between the two models was then calculated and applied as a metric of guideline adherence, along with the qualitative components of that metric.

This model was developed into a web-enabled application that can readily feed-back the non-adherence measurements in a clinical environment for any given cohort of patients with standard physiological and treatment output.

Evaluation of the system included:

- Individual unit tests of general adherence cases (e.g. treatment not present), and cases specific to the individual BTF guidelines (e.g. presence of mass lesion/diffuse injury when following the ICP guideline).
- Processing of guideline adherence output on three patients in the Philips ICCA system at the Queen Elizabeth University Hospital, Glasgow, and compared against the patient notes provided by the supervising neuro-intensivists.
- Accuracy of treatment annotation timing – a key component of the system – was evaluated by running a comparison of timing in a "live annotated" ICU dataset, against one produced in a regular ward shift.
- A relationship between guideline adherence and patient outcome was investigated using logistic regression between the instances of non-adherence and the 6-month Extended Glasgow Outcome Score (GOSe).

The system was then applied to large-scale ICU datasets to further explore individual and aggregate information. One was a neurological specialist dataset (Brain-IT) and one a general non-specialist ICU dataset (MIMIC III).

Non-adherence "distance" and duration was presented in a variety of ways to communicate as efficiently as possible how patient management is affected by guideline adherence. These included minute-by-minute windowing output (single number each minute, with component reasons viewable if desired), list per-patient of all non-adherence instances (also with component reasons) and a summary view

using inter-quartile range tables and box-plots (to understand the spread of non-adherence durations).

Results

The following results were obtained from the four evaluations:

1) For the unit tests with artificial data, the framework produced adherence output conforming to expected outcome
2) For the investigation of timing accuracy, on the "live observed" data, 24 events out of 32 across four patients were closely matched, with a mean distance of 3 minutes and a median of 1 minute. The "non-live" timings had no events matched within the asserted time limit (15 minutes)
3) On the patient data with domain expert notes, 80% of treatment annotations were associated with EUSIG events and adherence output could be reasonably matched to the patient notes on two patients out of three.
4) No statistically significant correlation was found between the guideline adherence output and 6-month patient outcome.

From the large-scale datasets:

1) Brain-IT had 17% of treatment annotations associated with EUSIG events, with instances of non-adherence detected according to all cases listed in the unit tests (with the exception of CPP pressor/fluid balance). Severity was reported as "mid-range" for nearly all patients.
2) MIMIC III had 7% of treatment annotations associated with EUSIG events. Some instances of non-adherence patterns were detected with severity also reported as "mid-range".

Conclusions

The conclusions relating directly to the three original hypotheses were:

1) A treatment process for the management of ICP and CPP can indeed be derived from the analysis of physiological and treatment data
2) This process can be compared against other processes of similar nature (in this, the BTF guideline represented in BPMN) to produce adherence output

3) The output of this comparison can be constructed into a clinically accessible tool – in this case a web-enabled application

The overall achievement has been to provide a quantitative and standardised structure for the measurement of guideline adherence, using data from the ICU bedside and the guideline texts.

Original contributions to research

Technical

- The application of process models to neuro-intensive data
- The expression of the Brain Trauma Foundation guidelines as process models
- The application of process model distance calculations to neuro-intensive data (and their use as a guideline adherence measure)
- A novel method of presentation of guideline adherence results
- A novel technological framework: the conversion of text guidelines and clinical data into comparable objects, the implementation of distance calculations to run the comparisons, the implementation of novel presentation techniques

Clinical

- A technological solution to provide direct and detailed information on guideline adherence and clinical management processes of ICP and CPP in neurological ICU data

1. Introduction

Chapter summary

The rise of clinical guidelines from evidence-based medicine is briefly described along with the availability of ICU data, which is often under-utilised. The general clinical goal of this thesis – improvement of knowledge about guideline adherence and the guideline themselves – is described, as well as the key requirements for a technological solution, which are:

- *High resolution physiological data*
- *Comprehensive treatment data*
- *The ability to combine these into a formal process-based expression*
- *The ability to compare this formal expression against other similar process-based entities (e.g. study protocols or local best practices)*

From this, three hypotheses have been developed:

1) *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*
2) *Having extracted treatment processes, one is able to develop a method to compare those against other treatment processes to establish the degree of similarity between them*
3) *One can develop a computerised tool that readily quantifies and displays a metric of actual ICP/CPP management protocol adherence*

The work conducted to address these hypotheses has the following original contributions to the field:

Technical

- *The application of process models to neuro-intensive data*
- *The expression of the Brain Trauma Foundation guidelines as process models*
- *The application of process model distance calculations to neuro-intensive data (and their use as a guideline adherence measure)*
- *A novel method of presentation of guideline adherence results*

- *A novel technological framework: the conversion of text guidelines and clinical data into comparable objects, the implementation of distance calculations to run the comparisons, and the implementation of novel presentation techniques*

Clinical

- *A technological solution to provide a direct and detailed link between guideline adherence and clinical management processes of ICP and CPP in neurological ICU data*

Finally, the methodology used to achieve these goals is described in summary along with a brief outline of the rest of the thesis.

1.1. Background and rationale

Evidence-based medicine (EBM) can be defined as *"the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients"* (Sackett *et al.*, 1996). Though it has been practised as a method of applying medicine throughout history, it has only been formally recognised as a specific methodology in the latter half of the 20th century and has helped progress many fields of medical research (Howick, 2011). An important component in the development of a clinical evidence-base is the creation of clinical guidelines, which provide a standardised description of the current best practice in a particular field.

Across many fields of clinical medicine, guidelines are used to inform and develop best practice. In order to understand whether these guidelines are being followed effectively, there are a variety of methods to monitor compliance. Common current methods to do this include post-hoc surveys or regular meetings after a hospital shift (or similar) to discuss different cases where perhaps the guideline was not adhered to, or negative outcomes were potentially avoidable (Levy *et al.*, 2010).

Nearly all current methods have two features: 1) they are a qualitative evaluation and 2) there is often a long time-lag between the publication of survey or discussion of results, and their submission into either local best practice or to multi-centre evaluations for the further development of the guidelines. Whilst useful, these methods often do not make full use of the data and technology that is now available to many fields of clinical medicine. A potential advantage of using such data and technology would be quantitative evaluations (i.e. understanding the degree to which a guideline has been adhered to) and rapid feedback of non-adherence to guidelines.

The work presented in this thesis attempts to exploit those advantages by providing an ability to monitor clinical guideline adherence, as well as providing measurable quantitative feedback. Using data and technology currently available, the goal of the research is to express the structure of physiological and treatment patient data in such a way that can be immediately compared against best-practice clinical (text) guidelines. The output of the research is to observe

adherence to best-practice guidelines over a study group, with a view to providing additional information to the clinical bedside.

1.2. Scope – clinical and technical

In critical care medicine – traumatic brain injury (TBI) as an example – technology has advanced throughout the late 20th and early 21st centuries to the point where nearly every modern intensive care unit (ICU) in the developed world has a multitude of high frequency data streams available, which can closely capture the application of clinical interventions and the physiological response of patients.

The technologies that enable this output of raw data are well established, and the economics of data storage make retention of large volumes for extended periods a feasible option. However, the key to establishing the integrity of that data for a specific purpose – whether it is a multi-centre randomised controlled trial (RCT) or an audit of local clinical practices – is to monitor that raw data and understand the relationships between clinical treatments and physiological output.

This process involves understanding that relationship at a level above the raw data output from bedside monitors. This could also be considered as observing data at a higher "layer of abstraction". The raw physiological output consists of a series of numbers, which on their own mean very little, but with clinical context can be formed into structures that do have clinical meaning (for example an "adverse event" such as a sudden spike in blood pressure). When this is combined with clinical treatment information (such as the time and dose of a bolus of Noradrenaline) then patterns of clinical behaviour and patient response can be formed.

If the algorithms used to extract and infer these patterns are valid, then it is very likely that this representation will be a highly accurate description of what actually happens in an ICU, due to the proximity to the actual data source, rather than having gone through several layers of interpretation in a qualitative survey or statistical analysis. In theory, it would be possible for such a system to work out – empirically from source – whether a specific process in the ICU has been followed, and if not, by how far it had deviated.

For specific processes, significant secondary inferences can also be made, which would fall into two categories depending on whether the output or the input of the process is being studied. For instance:

1) Does a particular guideline recommendation actually work (viewing adherence output against patient outcome)?
2) Has a particular protocol or guideline been applied correctly (viewing adherence input against the mandated guideline)?

In practical terms, a requirement of monitoring guideline compliance in real-world clinical processes would be for that activity to inhabit a clinical work-flow with minimal impact (i.e. its presence must require no input from a clinician or obstruct clinical treatment). To achieve this, the concept of deviation from a process (or expressed conversely: the similarity of two processes) would have to be measurable using only data that is currently available. It would also have to be measured, calculated and expressed in a manner that would make clinical sense, using an interface that clinicians would be comfortable using and confident of its clinical efficacy.

From all of these considerations, it is suggested that the development of a tool at the patient's bedside to establish actual clinical practice, would help establish the integrity of protocol adherence in general. This in turn would provide stronger validation for clinical recommendations and guidelines, and also provide strong support for current techniques of analysing treatment effectiveness.

To this end, the research described in this thesis aims to build an automated procedure that will evaluate the physiological and treatment information in several ICU datasets, extract the required clinical processes, then compare these processes with others of a similar nature (such as the recommended guidelines from literature). A quantitative measure of similarity, and therefore adherence, will then be available that can provide information on guideline effectiveness and compliance. This will in turn be available as a measure of baseline information that can be incorporated into subsequent studies.

The technical requirements to achieve this are as follows:

- High resolution physiological patient data
- Accurate and comprehensive treatment data
- The ability to combine these into a formalised process expression
- The ability to compare this formalised expression with other similar entities (such as guidelines, study protocols, institutional procedure, etc)

Although the approach presented is general, the methodology must first be applied to a specific clinical domain to test its validity. The chosen area for this evaluation is traumatic brain injury (TBI), and specifically the management of intracranial pressure (ICP) and cerebral perfusion pressure (CPP). TBI has many features that make it a good candidate for study: the condition is complex and certainty in the guideline compilation and compliance is variable (Bullock, Chesnut and Clifton, 1996), it is an environment that heavily uses modern technology that provides high-resolution ICU physiological and clinical treatment data streams (Shaw *et al.*, 2009), and the seriousness and prevalence of the condition means that any advances in the field have the potential to make large and positive impact on the population.

1.3. Hypotheses

In support of the discussion above, the following hypotheses have been formulated:

1. **In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients**
   a) Treatment processes for ICP and CPP management in TBI can be expressed by a work-flow data structure, comprised of “primitive” objects (a simple point value and time stamp) and “complex” objects (many values with interacting sub-structures).

   b) The treatment processes that are extracted are clinically meaningful and accurately reflect clinical management in a neurological ICU environment.

2. **Having extracted treatment processes, one is able to develop a method to compare those against other treatment processes to establish the degree of similarity between them**
   a) Treatment processes and other types of relevant process (e.g. protocols, guidelines, institutional policies) can be expressed in a standard form.

   b) This standard form expression can be compared in a measurable way.

   c) This standard form expression can be written in a computer-interpretable format.

3. **One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence**
   a) The standard form can be implemented in an application that can be integrated into a modern neurological ICU.

   b) The implementation provides meaningful and clinically useful feedback to clinicians.

By addressing these specific hypotheses, it is believed that the answers will inform and progress the knowledge contributing to the collection and analysis of ICU data to support evidence-based tools in critical care medicine.

1.4. Methodology and contributions

There are several technological steps that are involved in this work:

- The classification of events in physiological output known as EUSIG events (Edinburgh University Secondary Insult Grade), and compilation of an event log from this
- The expression of those event logs as process models
- The extraction of clinical guideline texts into process models
- The comparison of two process models using complex similarity/distance algorithms

Together, these steps form the framework through which the possibilities of quantitative, real-time guideline adherence monitoring can be explored. Other technological and computational methods have been explored throughout the course of the research, and they will be discussed in the appropriate sections of the thesis. However, though there are always merits and drawbacks in the use of different technologies, the overall goal remains the establishment of such a system in principle.

Event detection and representation from time-series data are common methods of data analysis in medicine. However, the automated semantic analysis of textual guidelines is an approach that has largely, to date, stayed in the realm of medical informatics research. Process models, at a general level, are processes of the same nature classified together into one model: a single process can be considered as an instantiation of that model. The use of process models and comparison/similarity calculations are borrowed from the field of business process management – most commonly used to describe real-world problems of project management and corporate efficiency (Panagacos, 2012). It has been used in isolated instances of medical problems, but again mainly in the logistical administration of hospitals and other large-scale corporate structures (where the fact that these structures are medical in nature is incidental) (Perimal-lewis *et al.*, 2012).

Figure 1.1 shows a simplified schematic of how the proposed framework in this thesis is used to convert ICU data and guideline text into comparable datasets. Step 1 is the translation of ICU data to an event log; step 2 in the conversion of this to a process model; step 3 is the formulation of the clinical guideline (Brain Trauma Foundation – BTF Guidelines) into a process model of similar format; and step 4 is the comparison between the two process models.

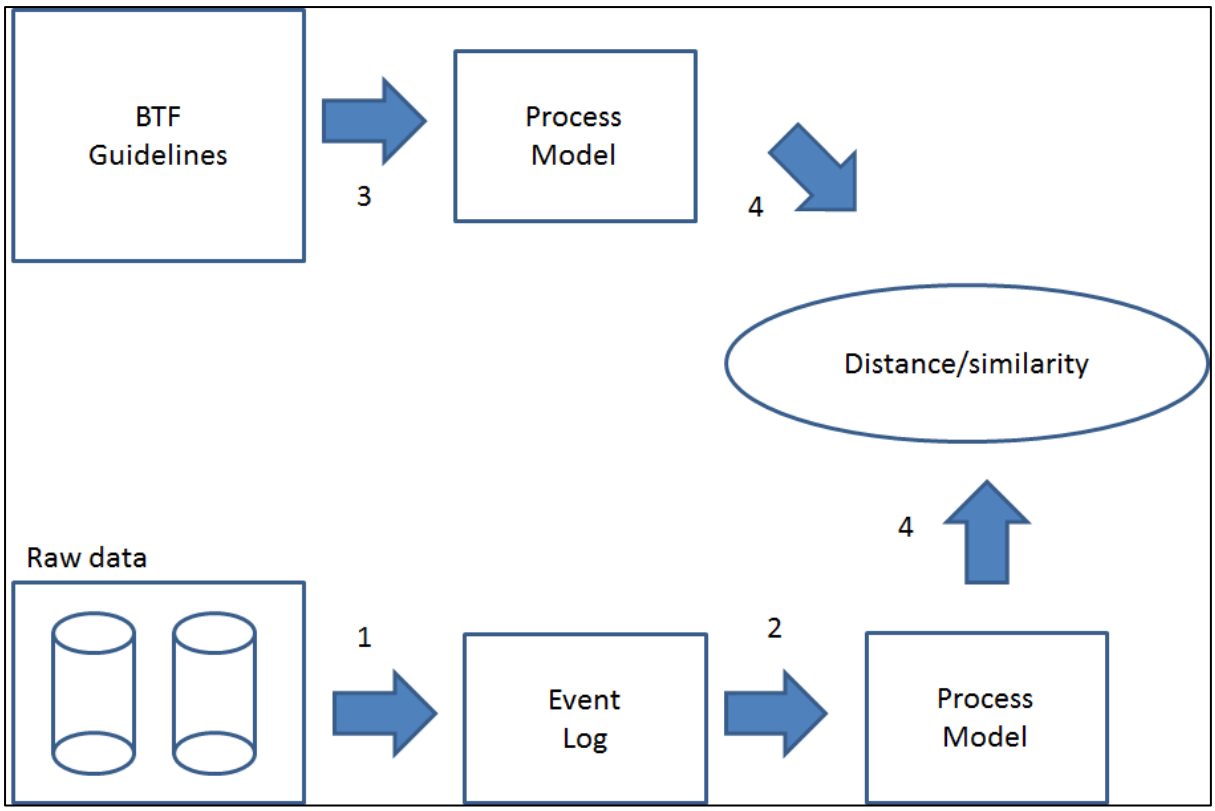

Figure 1.1: simplified schematic of the architectural process underpinning the proposed research

From the hypotheses and methodology described, the specific original contributions of this thesis are:

Technical

- The application of process models to neuro-intensive data
- The expression of the Brain Trauma Foundation guidelines as process models
- The application of process model distance calculations to neuro-intensive data (and their use as a guideline adherence measure)
- A novel method of presentation of guideline adherence results
- A novel technological framework: the conversion of text guidelines and clinical data into comparable objects, the implementation of distance calculations to run the comparisons, the implementation of novel presentation techniques

Clinical

- A technological solution to provide direct and detailed information on guideline adherence and clinical management processes of ICP and CPP in neurological ICU data

1.5. Chapter outline

The rest of this thesis will be divided up as follows:

- Chapter 2: Background – TBI, clinical technology and data
- Chapter 3: Literature review
- Chapters 4 – 7: Method
    - Chapter 4: Expression of clinical guidelines as process models
    - Chapter 5: Translation of ICU data to process models
    - Chapter 6: Calculating distance and similarity of process models
    - Chapter 7: Framework implementation
- Chapters 8 – 10: Results
    - Chapters 8: Evaluation
        - Unit testing of general and individual guideline cases
        - Treatment annotation timing verification
        - Domain expert information from a real clinical setting
        - Logistic regression of guideline adherence against 6-month GOS
    - Chapter 9: Application of framework to the Brain-IT dataset
    - Chapter 10: Application of framework to the MIMIC III dataset
- Chapter 11: Discussion and future work
- Chapter 12: Conclusion

Chapters 2 and 3 will provide a more in-depth background to the work in both the clinical and technological fields, providing background and literature on the broad clinical issue as well as the issues specific to the domain under study (TBI). Chapters 4, 5, 6 and 7 detail the specific technological methods developed and their implementation. Chapter 8 describes the evaluation work of the framework and its application against various benchmarks such as the accuracy of treatment annotations, statistical relationship to 6-month patient outcome, and the experience of domain experts in the field. Chapters 9 and 10 detail the results of the system when run against two different large-scale ICU datasets (Brain-IT and MIMIC III). Finally, chapters 11 and 12 outline the discussion, conclusions and avenues for potential future work. Each chapter is preceded with a small summary of the chapter contents.

2. Background

Chapter summary

The clinical mechanisms underpinning traumatic brain injury (TBI) – in particular intracranial pressure (ICP) and cerebral perfusion pressure (CPP) – are described, providing background to the necessary data required for collection in this domain.

The methods of data collection and synthesis for the development of TBI guidelines by the recognised leader in the field – the brain trauma foundation (BTF) – are then described. These are arrived at through meta-analyses of qualifying studies and trials and ascribed a confidence level out of three categories: option (level 3), guideline (level 2), or standard (level 1). A notable feature of the TBI domain is the lack of level 1 and level 2 recommendations.

The theme of addressing uncertainty is further developed with a discussion of studies and novel methods that have been attempted to improve confidence in TBI guidelines. Notable are contradictory findings in various studies, the negative results (i.e. unable to ascertain clinical significance) in Cochrane reviews, and a novel attempt to improve study power post-hoc by the TBI-IMPACT group.

Finally the current state-of-the-art in TBI data representation is discussed, with a view to exploiting this data to improve certainty in the output of TBI studies and trials. EUSIG (Edinburgh University Secondary Insult Grade) events are described as a pattern of representing physiological patient events – particularly relevant to ICP and CPP. Systems to interpret this are described, such as ICM+ and CareScape, and data repositories to build on this further are introduced, such as Brain-IT and CENTER-TBI.

The research questions addressed in this work are issues which are believed to be applicable to many clinical fields. However, as previously mentioned, in order to test the validity of the proposed approach a specific domain has been chosen: traumatic brain injury (TBI).

2.1. Traumatic brain injury and intracranial pressure

Traumatic Brain Injury (TBI) is defined as the damage sustained by the brain resulting from an external mechanical force, and is one of the leading causes of hospitalization, with almost 349,000 admitted to hospital in the UK in 2013-14 (Headway, 2018).

When the brain or skull is damaged due to an external force, there are many primary injuries that can result, including (but not limited to): hydrocephalus, brain oedema, tearing of axons, tearing of blood vessels resulting in formation of "mass" lesions (in this case, collections of blood) such as extradural, intradural or parenchymal haematomas.

The management of these primary conditions is often the immediate concern of the physician, but there are also secondary injuries that can result, which involve the more-subtle relationship between brain volume and brain pressure. This relationship comes about because of the equilibrium that exists between the various components of the skull, based mainly on the assumption that the skull is an incompressible structure filled with fluid and different types of neuronal and vascular tissue. The formal expression of this is known as the Monro-Kellie hypothesis:

"The sum of volumes of brain, CSF, and intracranial blood is constant. An increase in one should cause a decrease in one or both of the remaining two." (Mokri, 2001)

Part of the outcome of this doctrine can be shown graphically as the relationship between intracranial pressure and intracranial volume (figure 2.1). As can be seen there is a mainly linear relationship until a critical intracranial volume is reached – the volume where "compensatory craniospinal volume" becomes exhausted (i.e. most of the available space within the skull-spinal axis is filled). Above this point,

the intracranial pressure then increases dramatically to any further increase in added volume (Reilly & Bullock, 2005).

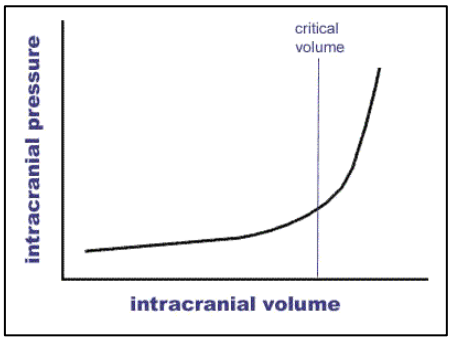

Figure 2.1: graph of cranio-spinal volume vs pressure, exemplifying the Monro-Kellie doctrine (Trauma, 2014)

Focusing on the cranial component and following from this equilibrium relationship, are a variety of pressures and pressure gradients that exist within the skull. The most important function of these is to maintain constant cerebral blood flow (CBF), providing oxygen and nutrients to the brain. The brain is – unusually amongst human organs – highly sensitive to critically low blood flow and can quickly become ischaemic if deprived of that flow. This flow is maintained by a net pressure gradient across the cerebral vasculature, known as cerebral perfusion pressure (CPP), which is dependent upon intracranial pressure (ICP) and mean arterial pressure (MAP):

$$CPP = MAP - ICP$$

This pressure is important when looking at the process of cerebral auto-regulation. This is the process where – to protect the brain from ischaemia – a physiological mechanism balances small perturbations in the pressure-volume interdependence, to maintain the optimum blood flow (Reilly & Bullock, 2005). Auto-regulation is achieved primarily by small arteries that expand or contract depending on the chemical messages received from other physiological control systems. These methods of physiological control include nitric oxide as a chemical messenger, which affects the proteins that control the blood vessel wall response (also known as endothelins). Recent work in (Payne, 2016) also suggest contributory mechanisms of capillaries via pericytes may be involved.

Figure 2.2 shows the relationship between blood flow and pressure, where the auto-regulation mechanism is active and most effective. Auto-regulation is a biological feedback mechanism that works to adjust a system's response to stimuli. Cerebral auto-regulation aims to maintain an adequate and stable cerebral blood flow. The edges of this curve indicate where the pressure is either too high or too low for auto-regulation to work effectively and where the disruption of regular organ function begins to occur.

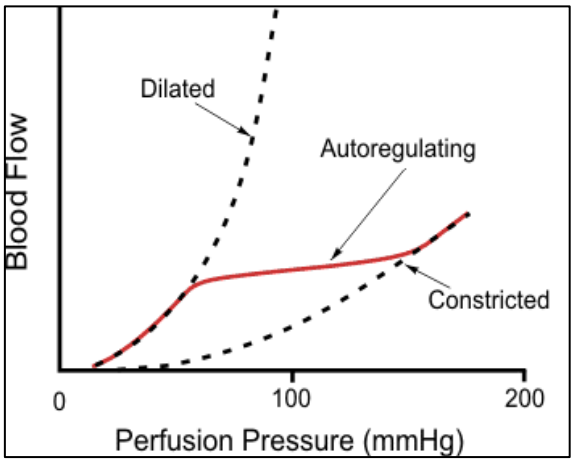

Figure 2.2: a graph of perfusion pressure versus blood flow, illustrating the curve of auto-regulation (CVPhysiology, 2017)

Measurement of the forces acting across the cerebral vasculature such as the ICP and CPP can be, and often are, used as secondary measures to indicate the status of the CBF. However, they can also indicate the presence of complicating secondary injury factors in themselves. For instance, hypotension – an abnormally low blood pressure – can be detected by monitoring an increase in ICP, which can indicate an impending potentially fatal drop in cerebral blood flow.

In terms of general treatment approaches for brain injury, it has traditionally been the case that clinical management has focused solely on the primary insult (the initial impact and injury). As such, consensus tended towards the idea that patient outcome was closely tied to the severity of these initial injuries in isolation. However, research over the last two to three decades suggests that secondary insults – such as altered CBF, cerebral edema, ischemia, hypoxia or hypotension – which may result from either developing patho-physiology from the primary injury or as an iatrogenic consequence of therapy, also have a significant impact on patient outcome, more than was previously thought (Jones *et al.*, 1994).

When the primary injury results in a space occupying lesion that can be removed surgically, the primary management is surgical. In terms of treatment for these secondary insults, general medical opinion appears to be that for the non-surgical management there is no first-line therapeutic approach, due to the complex interplay of factors. However, there are principles of treatment that can aid the management of a patient suffering from TBI as shown in table 2.1 (Reilly & Bullock, 2005).

Aim	Therapy
Lower intracranial volume: maintain or lower ICP: Brain volume, CSF volume, Blood volume	Mannitol, steroids, surgery, CSF drainage, diuretics, controlled ventilation
Correct gases: Hypoxia, Hypercapnia	Controlled ventilation
Improve brain perfusion: Blood flow, Blood pressure, Blood viscosity	Calcium antagonists, Maintain blood volume, Haemodilution

Table 2.1: treatment therapy and the immediate physiological target

An important principle in the case of TBI is the reduction of intracranial pressure (ICP), which would require a reduction of the intracranial volume, by various means (e.g. reduction of blood volume, management of autoregulation, reduction of CSF volumes). The understanding of this approach is the pursuit of many TBI research communities throughout the world – for instance the three-yearly ICP conference (ICP 2019 Congress) – and motivates the specific gathering and analysis of physiological and treatment information relating to ICP and CPP.

With this overview of the physiological mechanisms of TBI and the main treatment principles, specifically for issues in ICP management, we now look at the formulation and maintenance of the leading guidelines in TBI hosted by the Brain Trauma Foundation (BTF).

2.2. TBI guidelines

Clinical treatment guidelines provide a method to bridge the gap between evidence and clinical practice, standardize treatment practices, and improve

patient care. The compilation of guidelines is based upon a review of existing literature which can range from multi-centre randomized controlled trials to small scale studies. To reflect the range in evidence base, guidelines are often accompanied by an associated level of confidence which reflects the quality of the literature that led to a particular recommendation. As medical literature is constantly being updated, a natural feature of medical guidelines are for them to be constantly developed and refined as medical research and science progresses; in essence they can be considered as a "living document" (Kaiser and Miksch, 2009).

There are various sources of guidelines for the treatment of TBI, but those endorsed by the national health services of many developed countries (e.g. Scottish Intercollegiate Guideline Network (SIGN) and National Institute Centre of Excellence (NICE), for Scotland and England respectively) very often focus upon the immediate emergency triage that must be administered to a brain-injured patient before they are admitted to the intensive care unit (Harbour and Miller, 2014). The most comprehensive guidelines in the area of TBI are published by the Brain Trauma Foundation (BTF), which cover all types of situations including intensive care stays, emergency accident-scene care, specific situations such as trauma sustained whilst in military combat, and of particular interest: long-term treatment and outcome effects (Bratton and Chestnut, 2006).

The guidelines published by the BTF provide a common benchmark against which an institution can compare their procedures for the treatment of TBI. They are generally accepted as providing the best "gold standard" in TBI care, and have been associated with the development of TBI care management over the last two decades (Faul *et al.*, 2007) However, despite the advances in the standardisation of TBI treatment through this guideline development process, non-adherence is still relatively commonplace, for a variety of reasons such as lack of awareness, agreement and familiarity with the guidelines (Hesdorffer and Ghajar, 2007).

In recognition of the varying evidence used to generate guidelines and the effectiveness of those guidelines, the BTF provide a tabulation of the confidence level behind a specific treatment or guideline. Three broad classifications of guidelines exist in the BTF (in decreasing order of certainty): Standards, Guidelines

and Options. These classifications of specific guidelines are based in turn, 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).

Level 1	None
Level 2	Intracranial pressure (ICP) should be monitored in all salvageable patients with a severe traumatic brain injury (TBI; Glasgow Coma Scale [GCS] score of 3-8 after resuscitation) and an abnormal computed tomography (CT) scan. An abnormal CT scan of the head is one that reveals hematomas, contusions, swelling, herniation, or compressed basal cisterns.
Level 3	ICP monitoring is indicated in patients with severe TBI with a normal CT scan if two or more of the following features are noted at admission: age over 40 years, unilateral or bilateral motor posturing, or systolic blood pressure (BP) < 90 mm Hg.

Table 2.2: BTF guideline recommendations for Indications for Intracranial Pressure Monitoring

As a relevant example the recommendations for **"Indications for Intracranial Pressure Monitoring"** are detailed in table 2.2 (Braintrauma, 2018)[1].

As is standard throughout the BTF guidelines, the text supporting these recommendations include an overview of the medical issue being discussed, more in-depth information on the scientific foundation for arriving at the conclusions that they have, key issues for future investigation, and a summary. Of particular relevance to this work is the outline of the literature review process that sourced the evidence-base for this information, which in this case is:

[1] The text for table 2 had been originally written in mid-2014. Between then and late 2017 the BTF had significantly updated their website, along with much of the latest evidence supporting current recommendations from the 3rd to the 4th editions. Next to the ICP monitoring example there is now the following qualifying text, which neatly exemplifies the exact issue discussed in this chapter – i.e. that even though their criteria had become more strict between editions, they still did not have enough information to make strong recommendations:

"The Level II and III recommendations from the 3rd Edition of these guidelines are not supported by evidence meeting current standards because they were derived from descriptive studies, or from studies that do not meet the current inclusion criteria for this topic. While no evidence is available from comparative studies to support a formal recommendation, the Committee chose to re-state here the 3rd Edition recommendations. The rationale for doing so is to maintain sufficient recognition of the patient characteristics associated with risk of increased intracranial pressure."

"For this update, Medline was searched from 1996 through July of 2004 (see Appendix B for search strategy), and results were supplemented with literature recommended by peers or identified from reference lists. Of 36 potentially relevant studies, 12 were added to the existing table and used as evidence for this question (Evidence Tables I, II, and III)"

The evidence tables themselves are divided into three, one each to support the three questions posed about intracranial hypertension (ICH) in the scientific discussion section:

1. Which patients are at risk for ICH?
2. Are ICP data useful?
3. Does ICP monitoring and treatment improve outcomes?

As examples, table 2.3 shows the most recent contributing study for each question (all supporting level 3 recommendations only).

Overall, this process is an example of a meta-analysis of all available information, one of the most effective tools available for building an evidence-base of knowledge from study information. But there are many areas where the procedure can be significantly improved which is discussed in more detail in the next section.

Question	Reference	Description	Conclusion
1	(Miller *et al.*, 2004)	82 severe TBI patients were retrospectively analyzed regarding initial CT findings relative to ICP	CT findings regarding gray/white differentiation, transfalcine herniation, size of ventricles, and basilar cistern sulci are associated with, but not predictive of, intracranial hypertension
2	(Servadei et al., 2002)	ICP ranges assessed in patients with traumatic subarachnoid hemorrhage to determine if there were any identifiable changes predictive of worsening CT findings	ICP monitoring was the first indicator of evolving lesions in 20% of patients. However, in 40% of patients, CT worsening was not associated with ICP elevations, thus ICP monitoring alone may be inadequate to follow CT abnormalities
3	(Aarabi *et al.*, 2006)	Prospective observational study of 50 severe TBI patients, 40 with intractable ICH whose ICP was measured before decompressive craniectomy	Of the subgroup of 40 whose ICP had been measured before decompression, the mean ICP decreased after decompression from 23.9 to 14.4 mm Hg ($p < 0.001$). Of the 30-day survivors of the total original group of 50 (n = 39), 51.3% had a GOS score of 4 or 5

Table 2.3: most recent studies contributing to evidence supporting recommendations for "Indications for Intracranial Pressure Monitoring"

2.3. TBI studies and trials addressing evidence-base uncertainty

Clinical studies and trials are one of the most effective and well-known methods for compiling a clinical evidence-base, such as those used to support clinical guidelines. There are varying degrees of accuracy and knowledge that can be extracted depending on the methods used (e.g. a randomised-controlled trial is considered the "gold standard" but can be prohibitively expensive to conduct, so in some areas less accurate study methods are acceptable). In TBI there is a lack of strong evidence to support current treatment practices, the main indicator of which is the lack of level 1 recommendations ("standard") in the BTF guidelines. Currently the only Level 1 recommendation in the "Inhospital Severe TBI Guidelines" is against the use of steroids as a treatment, and throughout the guidelines the overwhelming majority of recommendations are of level 3 certainty ("option"), rather than level 2 ("guideline") (Haddad and Arabi, 2012).

Throughout the literature – discussed further in section 3.2 – there are many studies that support and also contradict the individual recommendations of the BTF guidelines. Examples of support include treatment using hyperventilation (Neumann *et al.*, 2008), treatment through patient cooling (Harris *et al.*, 2012), or treatment through the administration of barbiturates (Morrow and Pearson, 2010)). Examples that contradict the BTF, include (Pascual *et al.*, 2011) where oxygenation treatments are evaluated and suggest that the BTF guideline-mandated method may not be ideal. These studies and papers do make contributions to the evidence base, but it is notable that they are often smaller longitudinal studies, with confidence not sufficient to significantly influence the guidelines (i.e. with level 1 confidence).

The accuracy of studies to build the evidence base in TBI can also be confounded by ethical considerations, which are particularly significant due to the complexity and sensitivity of the brain. The CRASH trial (Edwards *et al.*, 2005) discovered a marked negative treatment effect in the administration of corticosteroids for traumatic brain injury, and was therefore halted on ethical grounds (Czekajlo and Milbrandt, 2005). That the use of steroids was detrimental was a surprising result in itself, as it had been used for years as standard treatment without the overall knowledge that it was doing more harm than good. As such the CRASH trial can be

considered as a successful scientific outcome. However, the trial was also (appropriately) stopped before completion, so the data also stopped being collected. It is likely that further information would help inform the community, and the clinical evidence-base, in even greater detail.

Another related example is that of ICP monitoring which is an important potential avenue of treatment study, as it provides near-direct clinical insight into a patient's ICP/CPP status. This in turn informs the administration (or not) of remedial treatments to manage ICP/CPP-related secondary insults, which are increasingly shown to affect the long-term patient outcome.

This tension between clinical importance and invasiveness was present in the BEST-TRIP trial, a study conducted in Ecuador and Bolivia directly investigating the importance of ICP monitoring (Chesnut *et al.*, 2012). The investigators of this study concluded that the study indicated no treatment effect in ICP monitoring, which would require a new assessment of future ICP monitoring methods and the ethical implications of whether this intervention would be administered. However, these conclusions have been challenged in the expert community as an over-interpretation of results, with questions about the strict adherence to TBI guidelines in the study – given the resourcing of the participating centres – and the lack of data on arterial hypotension. The particular issue of ICP monitoring appears to be far from settled (Härtl and Stieg, 2013) and uncertainty in the evidence-base remains.

Another method of mitigating against random effects of statistical uncertainty – or put a different way, understanding the significance of a single result against the overall landscape – is to analyse groups of studies systematically. These systematic reviews (or meta-analyses) combine the results of a number of similar studies to increase the "power" of the analysis. These techniques are well regarded, but can still be subject to issues arising from their constituent RCT analyses; issues such as publication, selection and agenda bias (Eysenck et al, 1994). Therefore, if these methodology issues exist in the trials used for the meta-analysis, the results of the meta-analysis will have corresponding low certainties.

One of the most comprehensive contributors to the field of evidence-based medicine is the Cochrane Collaboration (Cochrane, 2018), which provides a database of well-conducted RCTs and systematic reviews. The entry criteria are strict and therefore the reputation of information from a "Cochrane Review" is considered highly reliable. If a trial or study has a methodology which is at odds with the requirements for inclusion, they will simply not be included. In terms of TBI, there are various entries in the Cochrane database, such as (Harris *et al.*, 2012) (patient cooling), (Sahuquillo et al, 2006) (decompressive craniectomies) or (Roberts and Sydenham, 2009) (barbiturate therapy), but nearly all produce negative results, indicating no ability to comment on a treatment's effectiveness (in this regard the clear outcome of the CRASH trial is unusual). So, although TBI research is represented with trials that are well conducted and follow sound methodological principles, the statistical significance of the results are often low. Therefore, the baseline of clinical understanding remains undetermined and the uncertainty in TBI treatment guidelines remains.

A leading research group in TBI based in Antwerp, Belgium have attempted to address these issues using post-hoc statistical analysis. This analysis is known as the IMPACT project (McHugh *et al.*, 2007) and attempts to extract further information from the low-power TBI datasets already in existence by modifying the statistical analyses of these data. They include the use of broad enrolment criteria; changing the outcome analysis from dichotomous ("good" and "bad") to a finer-grained five-ordinal state (three "good" and two "bad", with varying degrees of severity); and a covariate adjustment to baseline patient characteristics. Results from this approach appear to be inconclusive, but the possibility of manipulating the analysis in this way has yet to be ruled out as unviable (Maas *et al.*, 2010).

In summary, despite the available tools and other novel attempts to establish greater certainty in the formulation of TBI guidelines, much uncertainty remains. However, a key avenue that shows promise in reducing this uncertainty is the advance of technology and data in TBI. In the next section, the nature and representation of that data and technology is described in more detail.

2.4. Physiological and treatment representation using TBI data

A modern intensive care unit is very much a "data rich" environment. Signal processing technology and the advance of Moore's Law (an approximate doubling of transistor capacity every two years) have resulted in the feasibility of data collection and storage at an unprecedented scale. This in turn has led to ICUs (and neurological ICUs) that have many machines monitoring patients and producing large amounts of clinical information to aid with healthcare delivery.

There are many types of data that come from this environment covering a wide spectrum of information and clinical behaviours: physiological output, drug/intervention treatment information, surgeries, routine and exceptional clinical events, and more. Arguably the most important are those of patient physiological signals (indicating their status and health), and clinical treatment. With these two types, it is possible to monitor – either in real-time or retrospectively – the status of a patient and their reaction to those clinical drugs/interventions.

In terms of data structure, an important piece of related work was the identification of a "physiological monitoring event" from routinely collected physiological data, which includes the Edinburgh University Secondary Insult Grade (EUSIG) (Jones *et al.*, 1994), the standard outline of which can be seen in figure 2.3. It shows various structural features such as the event/clear hold-down (length of time for an event to have officially started/finished) and the threshold crossing value. This model can be used for data represented as a continuous time-series, which in the neurological ICU would include (amongst others): intracranial pressure (ICP), cerebral perfusion pressure (CPP), blood pressures (BP), central venous pressure (CVP), heart rate (HRT), temperature (Temp), respiration rate (RR), pulse oxygenation (SpO2), end-tidal carbon dioxide (ETCO2).

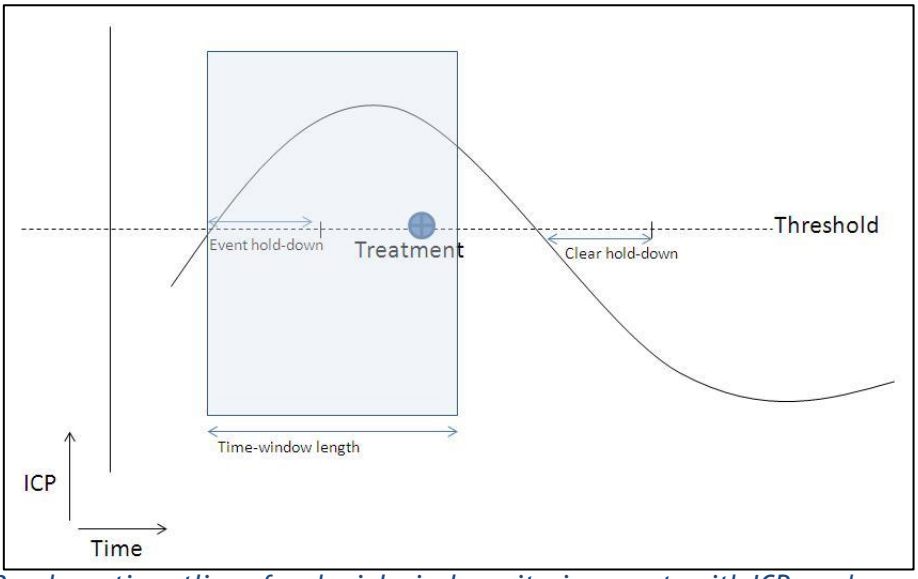

Figure 2.3: schematic outline of a physiological monitoring event, with ICP used as an example

When analysing physiological data using these structural definitions of events, one of the key clinical questions is the specific physiological values that 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). This is often a source of debate within the expert community. For instance, the numerical definition of a hypotensive event – one defined (semantically) as an abnormal drop in blood pressure – can be many and varied, involving not only different values (e.g. diastolic BP drops below 40 mmHg), but different parameters (e.g. mean BP or in the neurological case, cerebral perfusion pressure) (Eastridge *et al.*, 2007).

In the case of ICP – which, as discussed in sections 2.2 and 2.3, is believed to be a useful indicator of brain status, but is difficult to monitor practically – the uncertainty is such that there is even debate not only on the value of thresholds, but also the nature of how the data has been summarised (minute-by-minute or averaged wave-form values) (Shaw *et al.*, 2009), or whether patient-specific thresholds can be identified by charting individual information on pressure-reactivity index (PRx) versus ICP (Lazaridis *et al.*, 2014).

Systems such as ICM+ (Smielewski *et al.*, 2005), Philips CareVue (CareVue, 2018) and Carescape monitors (GEHealthCare, 2018) are built to detect events as defined above. To enable decision support for clinicians, the focus of these systems is to

use this information to trigger real-time predictive alarms over a useful timescale. They often have the sophisticated ability to vary a threshold warning (e.g. heart rate goes above 100 bpm) in response to whatever the favoured clinical event definition is at a particular centre (Donald *et al.*, 2012). The level of detail of raw data captured is very high (e.g. millisecond wave-form data), and again this contributes to the discussion in the expert community about the benefits of what level of detail is optimal for informing clinical behaviour (Hemphill *et al.*, 2005).

Although patient monitoring systems routinely capture physiological parameters, treatment annotations are much harder to capture, partly due to their nature, which can often be both irregular and highly dependent on human intervention (e.g. a clinician making a judgement call to administer a drug).

Systems such as Philips CareVue have only recently moved forward in this regard by connecting drug pumps directly to the integrated data system. If this option is not available, then an alternative is manual clinical entry. This is an example of where older clinical practice in an ICU is "playing catch-up" to the technological environment (i.e. the physiological data streams) that surrounds it. A discussion is required, similar to that of data resolution and beyond the scope of this work, to understand if the wealth of information in a certain ICU environment, is actually a help or a hindrance to clinicians. Some studies have been conducted looking at the rates of annotation within an ICU to aid neurological studies, such as (Enblad *et al.*, 2004) and have found that the numbers of actual treatments recorded could still be greatly improved. In fact, a significant proportion of treatments are delivered by hand rather than through drug pump infusion and thus remain dependent upon accurate timing event annotation by medical staff.

To utilise such data and technology more broadly, information platforms have been developed a layer "above" the data coming from bedside machines. The Brain-IT consortium documented a full specification of data parameters within a neurological ICU (Nilsson *et al.*, 2005), with a view to supporting future neurological monitoring projects. As data collection and storage technologies have advanced over the course of the last two decades, the outline of these schemas would often act as a roadmap of the ideal data to collect, whilst the implementation of collection would itself take many years to become an

achievable goal - relating to the Brain-IT schema, this type of implementation has included in projects such as AVERT-IT (Stell, Sinnott and Jiang, 2009)[2]. A separate initiative, focusing more intensively on robust data collection in the TBI field, is CENTER-TBI (CENTER-TBI, 2018). The project comprises many work packages covering both clinical and technological areas of research. One of the main technological goals is to make sure that the quality and coverage of data collected is as great as possible given the abilities that are now available from bedside machines in many neurological ICUs.

It should be noted that such platforms are often driven by domain experts, and hence a possible source of bias is the clinician's particular field of expertise. A key point in improving the process of understanding patient response is to use data that represents a "real" ICU as closely as possible and the barrier to using this data is often due to the problems involved in the methods of data collection. A pioneering project that has recognised this and built a dataset to support the use of this in future studies, is MIMIC III (Saeed, 2007), a repository of over 30,000 individual, anonymised patient data records collected since 2003 from a series of ICUs. The combination of physiological and irregular treatment information together in one uniformly accessible and generalised data store is highly unusual – this is based on their findings but is also corroborated by the survey of literature in this thesis (chapter 3).

Finally, a note on the inherent uncertainty in measurements of the pressures involved. ICP monitoring is most often achieved using ventricular and parenchymal catheters, inserted into the cranium. There are various brands of these catheters, and some ongoing work on non-invasive methods of measuring ICP (Ragauskas, 2012), though these methods are yet to be widely adopted, likely due to their lower accuracy than direct methods. There is a reported "zero drift" on the latest parenchymal catheters (Brattan, 2007), however this has been tested in (Citerio, 2008) and found to be not necessarily true under demanding clinical conditions (the conclusion is that this brand of catheter – the Neurovent-P – is as good as but not better than other catheters). The reported uncertainty in using such catheters

[2] Disclosure of interest: the author did a significant amount of work for the AVERT-IT project

is around +/- 1.1 mmHg (Citerio, 2008), which provides a context for understanding the reality of threshold crossing in the earlier discussion.

2.5. Discussion

This section has covered the mechanisms of TBI, TBI guideline formulation, relevant TBI trials and studies, and some of the predominant data representations in TBI.

One important and recurring issue that is directly relevant to this work is that of the uncertainty in the findings of TBI studies. There are many possible reasons for this, such as resourcing, legal issues, or socio-economic factors. But these are not necessarily specific to TBI and often apply to all medical domains. Exceptional medical conditions can be hampered by the lack of patients and data to work with (for instance, in a rare but life-threatening condition such as adrenal cancer). However, with TBI, there appears to be an inter-play between the complexity of the condition, the life-threatening severity and the required invasiveness of treatments, all of which combine to confound the establishment of clear base-lines of information that further research can be built upon.

There are other general issues in trial and study reporting, which also contribute to this lack of knowledge. These include the communication of results and their subsequent formulation into guidelines. As mentioned in section 2.2, guidelines and other evidence repository tools are “living documents” requiring constant maintenance and update, and periods of years can go past before the central evidence-base is updated in a meaningful way. This can have a cyclical negative effect if practices and treatments don’t reflect current research knowledge.

Another general issue is that of access to the raw trial data which can often be very limited. The output of the analysis is what is published and fed back into the scientific/medical community, and there is little actively-enforced regulation on providing the raw data. Though the greater awareness of initiatives such as clinicaltrials.gov, demanding clarity and openness of study/trial data (ClinicalTrials, 2018), is having a progressive influence on the transparency of data used for trials and studies. With the limited certainty that is a theme of the

guidelines on brain trauma, it is likely that this is a source of issue with the analyses performed on TBI data, and one that is generally under-reported.

Finally, we see that due to the advances in technology, increasing amounts of representative data can be gathered from the bedside, interpreted and sent to larger data platforms to give an increasingly accurate view of patient information. As more and more advanced repositories develop, pioneering efforts to understand that data and technology may aid in establishing greater certainty to the formulation of TBI guidelines.

3. Literature review

Chapter summary

Literature in the following areas were reviewed and discussed in order to present a fuller understanding of the background and rationale behind this thesis, to survey similar work conducted in the field, and to identify the most relevant technological solution:

- *Issues of adherence to clinical guidelines in general*
- *Issues of adherence specific to clinical guidelines in TBI*
- *Novel attempts to improve adherence in general without technology (in general and TBI)*
- *Novel attempts to improve adherence using state-of-the-art technology*

In general, the main issues preventing adoption of clinical guidelines involved methods of dissemination, the authority of guideline publishers, and the rise of personalised medicine as a counter-point to population-wide guidelines.

Specific to TBI, the issues raised were more numerous but essentially the same as those mentioned in chapter 2. Namely: the low power of studies leading to uncertainty about recommendations, and contradictory findings for specific treatments, despite large-scale attempts at well-conducted meta-analyses.

Two examples of non-technological attempts to improve adherence were discussed: the Surviving Sepsis campaign and the CENTER-TBI project. Both projects curate data manually with the intention of constantly updating large-scale information for the improvement of source information.

The review of state-of-the-art technology included decision support systems, smart-phone apps, clinical guideline formalisms and process models. Decision support was reviewed as an end-point goal of the work of this thesis, whilst the review of smart-phone apps addressed innovations at the point of information delivery.

Representations of the nature of clinical guideline information were covered by guideline formalisms and ontologies (providing comprehensive domain context).

Though the full feature-set required when adopting clinical guideline formalisms and ontologies was deemed unnecessary for the work of this thesis, a sub-set of the features was found to be uniquely useful: the process-oriented nature of guidelines. This feature is well represented by the concept of process models, a technology commonly applied to business processes and logistical administration.

Throughout the review, analysis of guideline adherence at a detailed level of clinical management appeared to be missing in general, and in TBI in particular. Traditional studies have a "low resolution" view of the details of clinical management, and technology that leverages these studies lack the detailed combination of physiological and treatment data.

A section describing the aims of this thesis is outlined at the end of this chapter. It presents a framework proposal based on the review of the clinical domain and the surveyed technologies, to address the gap between clinical management and guideline adherence.

3.1. Adherence to clinical guidelines

Clinical guidelines have contributed to standardized clinical practice and have advanced the quality of patient treatment for many decades. In more recent years, their development has followed a more rigorous process, using evidence-based techniques to avoid bias of either specialisation or agenda (Watters, 2008). However, despite these systematic attempts at improvement, guidelines are not always followed, and this can be for a variety of reasons.

3.1.1. Dissemination of clinical guidelines

The basic methods of communication of clinical guidelines can often be a barrier to the adoption of the guideline procedure. Even if the information contained in the guideline has consensus in the community of experts that it serves, sometimes a lack of awareness of the official guideline can adversely affect adherence.

As an example, in the provision of care for osteoarthritis (Nelson *et al.*, 2013) findings indicated that there was relative agreement between the centres involved, on what treatments should be provided. This was despite large variation in familiarity and adherence with the official guidelines (79% of those surveyed said they were aware of the management guidelines, whilst 54% adhered). The study concluded that guidelines in this area were effective, but the methods of dissemination required improvement. So participating centres were indeed following broadly the same principles, but this was only partially due to the influence of the official guideline.

Issues of dissemination appear as a common thread when evaluating adherence. (Ansari *et al.*, 2003) looked at beta-blocker use in heart failure and showed various methods and channels of disseminating the guideline information. These were to use a nurse facilitator (direct intervention by trained specialist), general education (documents, leaflets, etc) and clinical reminders (automated interventions). These all had different effects on adherence, with the nurse facilitator being the most successful. (Rood *et al.*, 2005) indicated that a study of glucose measurement and regulation improves greatly when dissemination is provided through computer-assisted, rather than through paper-based, means.

A systematic review of guideline dissemination strategies (Prior, Guerin and Grimmer-Somers, 2008) showed that the (non-) effectiveness of passive dissemination is a significant result. Similar to the (Ansari *et al.*, 2003) study, where direct intervention is taken by a person or automated method, the adherence rate is markedly better than if the guideline document and information is published passively (e.g. using conferences, websites or didactic lectures).

Other studies (Grol, 2001) and (Azocar *et al.*, 2003) show that targeted and behaviourally "disruptive" methods are best for disseminating information and influencing clinical practice. Similarly (Grol and Grimshaw, 2003) and (Almatar *et al.*, 2016) have shown that only comprehensive interventions on all levels of input and with specific targets and barriers identified stand a chance of influencing behaviour. Therefore, understanding the effectiveness of these different methods of dissemination is an important factor in developing tools to improve awareness and adherence (related to the third hypothesis of this thesis).

3.1.2. Authority of guideline publisher

An implicit assumption in the use of clinical guidelines is that they represent the most up-to-date knowledge in terms of clinical interventions, and the perceived quality and trustworthiness of the guideline itself can often go un-questioned.

This authority is an important aspect when considering adherence and contributes to the many other human characteristics that mark whether a clinical guideline – which is not a legal mandate – should be followed. Examples of systematic reviews that have been conducted into the question of guideline authority in general health-care, includes: the evaluation of attitudes towards guidelines (Farquhar, Kofa and Slutsky, 2002), examination of bias in self-reporting and awareness of guidelines (Steinman *et al.*, 2004), discussion of the positive and negative effects on clinical practice of guidelines (Grimshaw and Russell, 1993), and a more recent article that highlights the negative impact of the "evidence-based" movement, such as the volume of guideline evidence, and the mechanical implementation of the guideline in patient care (Greenhalgh *et al.*, 2014).

The results of these studies highlight issues that challenge conventional clinical wisdom – such as the large percentage of clinicians not meeting hypertension

guideline standards (medication prescription 67%, blood pressure management 42%) despite these guidelines being largely unchanged for 30 years (Steinman *et al.*, 2004). However, as highlighted in (Farquhar, Kofa and Slutsky, 2002), there is a general feeling that while clinical guidelines are a useful source of information, the associated administrative overhead that official compliance requires can cause institutions and clinicians to be tentative when considering full-scale adoption. (Grimshaw and Russell, 1993) conclude that, as a general rule, there is an unmistakeable improvement in clinical practice when clinical guidelines are adopted but their conclusions also showed vast variation in that improvement depending on other factors such as dissemination, education and resourcing (similar to the issues discussed in section 3.1.1).

As discussed in (Watters, 2008) the move from general clinical practice guidelines to those underpinned by a systematic evidence base, does also vastly improve the quality of the guideline. This is due to the transparency of the guideline development process, along with the primary focus of removing systematic bias (compared to previous approaches which, with less transparency, often tended to reflect the treatment or economic goals of those developing the guideline).

A final consideration on the authority of guideline developers is the possibility of competing sources of guideline publication, due to cultural, organisational or political reasons. One study looked into the nature of guideline development in general (Fervers *et al.*, 2006). It concluded that there was a marked lack of trans-contextual adaptation of guidelines – i.e. guidelines defined in one cultural and organisational setting were rarely, if ever, considered for other settings. This leads to an "organic" development of expert communities, the possibility of guidelines developing in parallel with – and isolation from – each development group and missing critical translational developments. Although this situation is not prevalent, where it does exist these factors undermine the systematic and evidence-based nature of clinical guidelines. Another study (Kearns, Moss and Kinsella, 2013) examined a specific case – management of patients with a fractured neck of femur – and found that the recommendations spanned many guidelines from across a spectrum of guideline-producing institutional bodies, with

some similarities but also conflicts that could lead to inconsistent patient care if followed to the letter.

3.1.3. Personalized medicine

A view-point that is gaining ground in general health-care and which can act as a counter-point to the use of medical guidelines is the notion of personalized medicine. This is where the priority is put on tailoring treatment for a patient's specific case instead of abstracting the treatment process into something more general. (Goldberger and Buxton, 2013) provide a discussion about personalised medicine versus clinical guidelines, though they do not appear to provide strong arguments for evidence-based guidelines, choosing to hold guidelines as the authoritative position to be argued against, and provide counter arguments espousing the benefits of personalised medicine. Indeed, when reviewing this theme in literature, feelings on either side of the debate apparently run high, suggesting a level of "evangelism" when defending one of the two sides (Miles, Loughlin and Polychronis, 2008)[3].

However, more reasoned arguments have also been made, particularly involving the discussion of gene therapy, which is arguably where the link to individual personalisation is strongest. (Hamburg, 2010) discusses the potential for gene therapy as the field of translational research develops, where the identification of genetic markers in individuals will likely play a more prominent role in the development of clinical treatments at the phenotypic level.

Despite the apparent "natural" opposition of personalised and evidence-based medicine, the overall argument made in papers such as (Goldberger and Buxton, 2013), is one that this research work attempts to address: that there is a need for cautious interpretation of large-scale random controlled trials and studies. This is an over-arching problem that does not mean that the two approaches to treatment (guidelines vs personalized) necessarily have to be mutually exclusive.

[3] For example, from the referenced paper: *"… no author has been able to convincingly show the superiority of the Evidence-Based Medicine 'approach' [original quotation marks] and such assertions […] remain what they originally were: expressions of bald rhetoric and intellectually bankrupt hyperbole"*

3.1.4. Summary

From the areas covered in this section, we can see that a number of issues exist in general that are impediments to the successful adoption of clinical guidelines. Various methods are used to improve these adoption rates but with differing levels of success, and these issues are relevant to medicine in general. In the next section we look at the issues in adherence to guidelines specifically in the domain of TBI.

3.2. TBI specific guidelines

In 1994 the Brain Trauma Foundation (BTF) began an initiative to formulate treatments for brain injury into standardised, internationally-recognised guidelines. Since then studies have been conducted that show dropping mortality rates and improved long-term outcomes due to the adoption of these guidelines. One example is (Bratton and Chestnut, 2006), and another is (Tarapore *et al.*, 2016), both showing the continued trend of improvement through the last two decades.

Evidence for this level of confidence in the guidelines is available: a survey of TBI management in 1995 (Ghajar et al, 1995), was one of the original studies that the BTF guidelines were formed in response to. When the numbers in that study are compared to those collected seven years later (Fakhry *et al.*, 2004), an undoubted improvement in patient outcome is shown, largely attributed to the gradual adoption of BTF guidelines by many centres over this time period.

However, adherence to the BTF guidelines is not universal – many studies outline their potential deficiency in various aspects such as hypothermia (Clifton *et al.*, 2001), intubation (Franschman *et al.*, 2009) and the need for ICP monitoring (Chesnut *et al.*, 2012). In one study (Lee *et al.*, 2015), the investigation has focused on whether it is feasible, or even possible, to adhere to all 15 of the BTF guidelines when treating brain-injured patients. Their conclusions are that it is indeed a difficult objective to achieve and in many cases is also unnecessary. And in contradiction to (Tarapore *et al.*, 2016), another study conducted by (Dawes *et al.*, 2015) concludes that patient outcomes are in fact not statistically affected by strict adherence to the BTF guidelines.

However, whilst guideline non-adherence is a common issue across medicine for reasons of lack of awareness, familiarity, agreement, or outcome expectancy (Cabana *et al.*, 1999), it is the case that studies investigating non-adherence to BTF guidelines are very focused in their rejection of the specific guideline. The NICE guidelines concerning treatment of TBI treatment in the UK (NICE, 2014) suggested that the key recommendation of *"transfer the TBI patient to a hospital with a specialist neuro-trauma centre"* is a grey area that causes many clinicians to reject the mandated guideline (though this is an issue affected by resourcing as well – if every hospital had a specialist neuro-trauma centre, the problem would not occur, a finding also supported by (Ghajar, 2000)). In another example, (Pascual *et al.*, 2011) discusses the interventions mandated by the BTF guidelines when oxygenating the brain blood flow and conclude that those recommended actually worsen survival rates. ICP monitoring in particular, is an area that highlights the divided opinion of the expert community about the most appropriate treatment, demonstrating that much more work is required to underpin this particular BTF guideline. For instance, (Dawes *et al.*, 2015) conclude that compliance with the BTF guideline on ICP monitoring and craniotomy has "minimal association with risk-adjusted outcomes in patients with severe TBI" whereas (Talving *et al.*, 2013) conducting a very similar study conclude that "Patients managed according to the BTF ICP guidelines experienced significantly improved survival". Evidently, uncertainty and disagreement surrounding the validity of many of the BTF guidelines still exists.

Various studies have been conducted that investigate the adherence to BTF guidelines in regard to particular treatments. (Neumann *et al.*, 2008) look at the administration of hyperventilation; (Griesdale *et al.*, 2010) quantified the adherence when applying an external ventricular drain to TBI patients; and (Griesdale *et al.*, 2015) examined the association of CPP being maintained within the guideline range, with patient outcome. All report overall adherence to the BTF guidelines – associated with positive outcomes - but with notable exceptions largely delimited by geographical areas (e.g. in Europe, recommendations on early prophylactic hyperventilation and cerebral oxygenation monitoring are not followed in the majority of TBI centres (Neumann *et al.*, 2008)). It is proposed that the overall adherence of specialist centres could be improved, but a critical point

is that all of these studies accept *a priori* that the BTF guidelines *should* be adhered to.

Finally, the issue of regional and national differences is one that has received attention from studies in TBI, especially as the search for significant findings push clinicians to collaborate over wider areas, with potentially greater differences in treatment patterns. On one hand (Lingsma *et al.*, 2011) conclude that differences in outcome between centres do not affect the treatment effect in TBI RCTs, as is commonly considered to be the case. On the other (Hukkelhoven *et al.*, 2002) run a comparison of two RCTs for the same drug, Tirilizad (a drug to treat acute ischaemic stroke) and conclude that the differences in treatment patterns between centres, countries and continents are significant and do affect RCT outcomes.

As already noted in chapter 2, there is large scope for uncertainty in the development of TBI guidelines, and this inevitably brings an unwillingness to fully adhere to a guideline. However, when discussing with a clinician (Dr Chris Hawthorne, University of Glasgow, 31st July 2018, pers. comm) in terms of specific parameter targets, they made the following two points when considering adherence, illustrating the pragmatic relationship developed with guidelines in the course of day-to-day routine:

1. *"There may become a point in patient care when the likelihood of survival is minimal and further aggressive treatment is felt to be futile. In these cases, non-adherence with guidelines may represent a decision to focus on palliative care rather than on targeting physiological parameters such as ICP and CPP."*
2. *"The BTF Guidelines are a very balanced and pragmatic set of guidelines that are extremely helpful to clinicians. However, little (if any) of the evidence is "level 1" and the guidelines themselves allow (indeed encourage) clinicians to consider the individual patient. This means that non-adherence in terms of targeting a specific CPP parameter may be common but may be acceptable within the scope of the guidelines."*

With this is mind, we now survey the areas where clinical guideline adherence has been improved using novel techniques, in a non-technological capacity and specifically for TBI.

3.3. Novel attempts to improve adherence

Evident from this review so far is that guideline adherence is subject to great variation. When considering how to improve adherence, the reasons outlined above can be broadly categorised into the result of one of two prime causes: being *unwilling* to adhere to a guideline and being *unable* to adhere. Whilst techniques to address the first category include improved dissemination, communication and various long-term social methods, improvements in the second category, which is usually functional in nature (e.g. lack of resources/time), can be approached using "behaviourally disruptive" methods.

3.3.1. Data collection approaches

Most attempts to improve adherence to guidelines in the medical domain involve a direct change or implementation of a care procedure. In these cases, the evidence-base for a guideline comes from a panel of experts in the field that have reached a consensus for various treatments. The novel attempts then concern the implementation of that guideline in patient care in a standardised and accountable way.

A campaign that exemplifies this approach is "Surviving sepsis", which targeted improvement of patient care by specifically supporting guideline adherence through the identification of resuscitation and management "bundles". Part of this was an intensive data collection arm, which – in real-time – forced clinicians to systematically add data as part of clinical routine (Levy et al. 2010). The results of this work have shown a marked improvement in adherence to the guidelines, but an emergent complication was the lack of ability to stay current with the latest guidelines and update procedures to reflect this over a feasible timescale. Feedback from the first four years of this project into the re-development and improvement of sepsis guidelines has been cautiously optimistic (Dellinger *et al.*, 2013). And whilst not specifically providing a new type of analysis, a side-effect of

the rigorous collection is that it does provide a large repository of sepsis data that is potentially useful for future studies[4,5].

3.3.2. TBI-specific

As mentioned in section 2.4, CENTER-TBI comprises many work packages covering both clinical and technological areas of research in TBI. Much of the work conducted by this group has established beyond doubt that large variation exists in the implementation and adherence rate to the TBI guidelines, even in large-volume studies conducted in well-resourced specialist centres (Cnossen *et al.*, 2016a). A sub-set of this group, as part of the TBI-IMPACT project – also mentioned in section 2.3 – had attempted to address variation using post-hoc statistical techniques, but the results had been inconclusive (Maas *et al.*, 2010).

A central component of the CENTER-TBI initiative, extending on work begun in the TBI-IMPACT project, is the use of competitive effectiveness research (CER), which is a broad definition of various analytical tools. One of the main facets of CER is the use of "Living Systematic Reviews" (LSRs) – systematic reviews which are conducted with the express purpose of being updated at regular intervals, as opposed to being done "once and never again". The cost of conducting such a review can be high, so keeping the study as a constantly updated document is a good way to maintain constancy without having the initial set-up costs of the study each time. As the CENTER-TBI project progresses, individual LSRs have yet to be developed, but evidence of their use and requirement are now starting to be published (Cnossen *et al.*, 2016b), (Synnot *et al.*, 2016).

The use of the CENTER-TBI repository in this way, does leverage technology and data but still largely involves the manual collection and curation of that data. In the next section we now discuss direct technological (i.e. automated and semi-automated) attempts to improve guideline adherence in general medicine.

---

[4] The author of this thesis also witnessed a ward review meeting at the Glasgow Royal Infirmary, which was one of a set of regular sessions that are now in place due to the Surviving Sepsis campaign. The process was well established and made a comprehensive and detailed review of outcomes that week in the ward (death or discharge), how they related to the mandated guideline, and noted for feedback into the appropriate administration where any deviations had occurred.

[5] Since submission of this thesis, an app related to the Surviving Sepsis campaign has been released which neatly combines the output from this section (data collection) and 3.4.2 (apps) (Surviving Sepsis, 2019).

3.4. Technology – state-of-the-art

When surveying the state-of-the-art in the context of improving clinical guideline adherence, there is a large landscape of technology to consider. However, to focus the work to the most relevant areas the review has been constrained to: a general (and brief) consideration of the role of decision support technology; the use of smart-phones and apps; clinical guideline formalisms and representations; and process models.

3.4.1. Decision support

Central to understanding the availability and utility of data in the ICU is being aware of the advances of ICU technology, particularly in the past four decades. One paper written in 1987 (Shortliffe, 1987) describes the concerns within the medical community that computerized support technology would outgrow its supporting role and end up replacing clinical decision-making. Nearly three decades on, this outcome has yet come to pass, and familiarity with techniques of artificial intelligence have eased these concerns[6], with clinical judgement by a trained human still a required part of the interface with technology. However, the focus of research still remains upon understanding expert medical knowledge and problem-solving skills, and attempting to reproduce them with technology, in order to free resources for pursuing more advanced medicine. The pursuit of accuracy in this latter point of judgement is one of the main goals of the science underpinning decision support.

As the mechanics of decision support strive for accurate automation, one way to measure progress is to look for metrics of efficiency. Hospital administrators may look at cost-benefits within a particular hospital, government health policy-makers will be concerned about cost-benefit to society as a whole, or the individual clinicians may look at overall patient outcome improvement. Whatever the targeted improvement is, there will be some metric associated with it, such as a protocol refinement of adult respiratory distress syndrome (ARDS) (East *et al.*, 1992) or enforcing adherence to diabetes guidelines (Lobach and Hammond, 1997).

[6] More accurately: the concerns have eased “somewhat”. Debate in the area is still controversial.

Studies in other areas as diverse as chronic kidney disease (Ennis *et al.*, 2015) and thyroid cancer management (Likhterov *et al.*, 2016) support the conclusion that improvements in guideline adherence in these respective areas, improve similar metrics and – by extension – patient outcomes. It is in studies like these latter two that the connection between guideline adherence and decision support is made most explicit: in both, the concept of a retrospective assessment of protocol adherence helps understand why decisions in an ICU are taken, and the output allows later prospective decisions to be guided by understanding the effectiveness of the given protocol (also related to the third hypothesis of this thesis).

3.4.2. Smart-phones and apps

One method that has increased in step with the utility of personal smart-phones is the electronic application (or "app"). The physical proximity to a clinical professional and the increasingly-reliable connection to other digital assets over the Internet make this one of the most convenient technological interventions.

This shift towards the use of smart-phones can be immediately seen in their adoption by major guideline developers in national health-services (e.g. the NICE guidelines for the NHS in England (NICE, 2014)). These provisions allow quick and easy access to guidelines for immediate consultation – useful in the clinical environment – but have yet to provide reliable dynamic interactivity in the update of the guidelines involved (the app is largely "broadcast only").

However, as fast-moving app development continues, various studies and pilot tests are now being carried out to include greater functionality that would allow more interactive access to a knowledge-base of clinical expertise. Some are specifically designed – similar to the NICE apps – to provide guideline expertise to professionals at the point-of-care, such as the "Sidelines Guidelines" app (Lee, Struik and Ahmed, 2016) which helps medical practitioners administer first-aid correctly during high-contact sports. Others focus on point-of-care help for a specific condition, such as a fitness and health app designed to provide aid in the case of seizure (Pandher and Bhullar, 2014). In this latter study, the focus was often on the development of the app and its ability to disrupt the behaviour of an epilepsy sufferer by providing immediate seizure management advice.

A further addition to this provision of clinical information at the point-of-care, is the integration of clinical guideline software into smart-phone apps, which is currently being conducted by the European MobiGuide project (MobiGuide, 2014). The anticipated outcome of this project is a direct stream of standardised guidelines, combined with personal digital health records, which interacts directly with patients (Peleg, Shahar and Quaglini, 2013). The interactions take the form of alerts and notifications with suggested corrective actions, so that adherence to the relevant protocol is maintained as closely as possible. Measurements of adherence improvement through this system can be seen at (Peleg, Shahar, Quaglini, Broens, *et al.*, 2017a).

Finally, another useful feature of apps is their ability to directly connect to medical registries, which focuses the input of information (usually in the form of a daily electronic diary) directly from the patient and synchronises with the central registry to allow accurate communication with their consulting clinician. Example apps of this type have been developed for rare diseases such as Niemann-Pick (Sinnott *et al.*, 2015) or for public health challenges such as alcohol consumption (Zheng, Z, Bruns, L, Jr, Li, J, Sinnott, 2017). The improvement of adherence occurs as a third-party (the consulting clinician) can immediately see if the patient is non-adherent and take correcting action if appropriate [7].

3.4.3. Clinical guideline formalisms and representations

The behavioural disruption that the smart-phone app provides is almost always due to the convenient proximity of the knowledge base (the guideline data repository) to the end-point of information delivery (e.g. the patient or clinician). In this thesis however, the focus of the work is not only upon delivery but on the representation and processing of that information. The area relevant to this thesis is the use of electronic clinical guidelines, their formal specification as (realistic) medical processes, and their ability to be interpreted by both humans and computers.

To illustrate clinical guideline formalisms, an example is the ProForma technology, a project developed by the COSSAC group (Fox, 2017). ProForma is a language

[7] Disclosure of interest: the author works for the Melbourne eResearch Group, developing the registries that the apps referenced in this paragraph upload to.

specification designed to *"support decision making and plan execution"* for the authoring and execution of clinical guidelines. In order to build applications around this specification, the TALLIS software is required (COSSAC, 2014), around which various tools and tutorials are offered to build a complex clinical guideline object, which can then be "enacted", and output received as to what a specific clinical decision should be, given the context of the guideline.

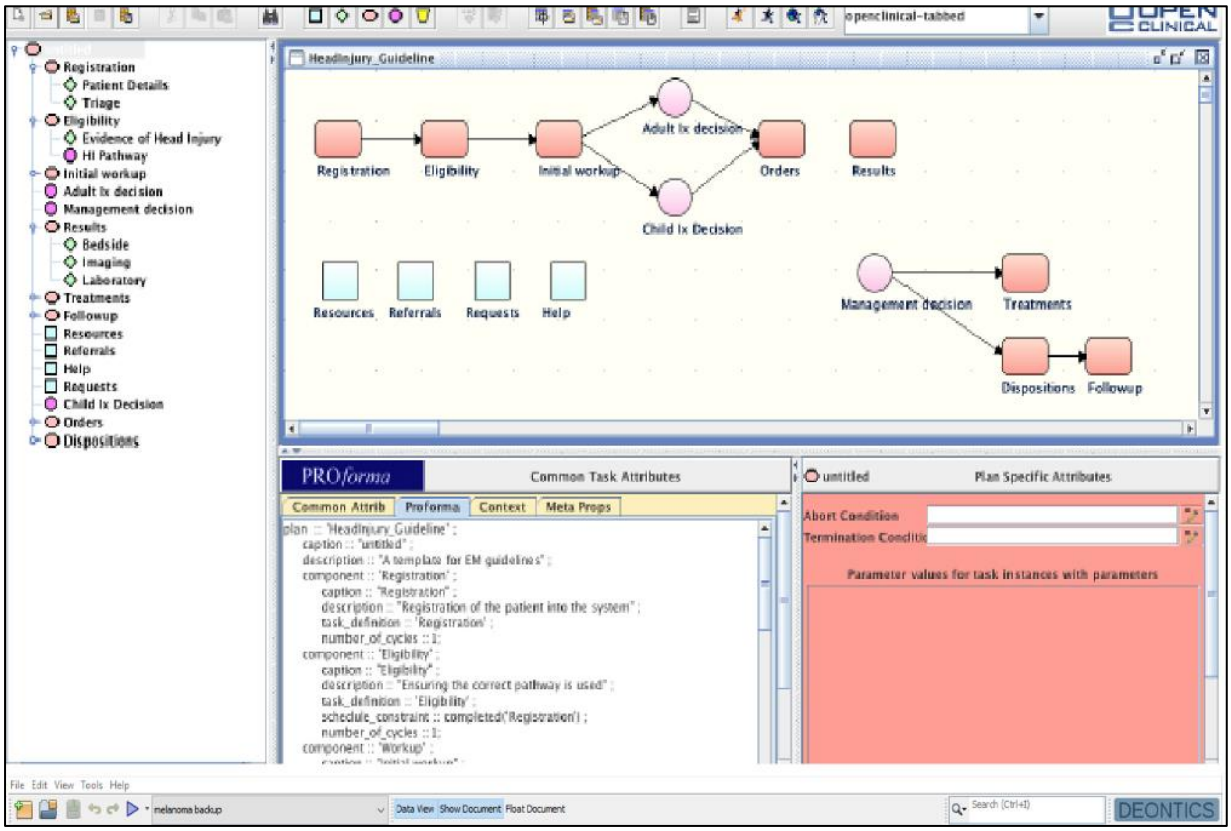

Figure 3.1: illustration taken from (Fox, 2017) the ProForma application development tool (TALLIS)

Similar language specifications and tools have been developed in this area. One of the most comprehensive comparison of these formalisms is (de Clercq *et al.*, 2004), which covers representations such as the Arden Syntax, GLIF, ProForma, Asbru and EON. This survey does not provide a "like-for-like" summary but does discuss the benefits and drawbacks of each representation (e.g. ProForma provides a low-level syntax; Asbru provides a rich set of temporal constructs though lacks an implementation engine; etc). It does note that as of the time of writing (2004), no single implementation had been created to enable these representations in the real world. However, the MobiGuide project – mentioned in section 3.4.2 – has aimed to implement these representations more robustly and has had moderate

success in introducing computer-interpretable guidelines into a clinical setting (Peleg, Shahar, Quaglini, Broens, *et al.*, 2017b).

Another application in this field is the use of ontology technologies. An ontology *"encompasses a representation, formal naming, and definition of the categories, properties, and relations of the concepts, data, and entities that substantiate one, many, or all domains"* (Ontology, 2018). Similar in theme to the descriptive and representative nature of formalisms described above, many groups developing clinical guideline formalisms have attempted to incorporate ontologies into their work, with the goal of introducing comprehensive context into the decision-making process, necessary for following clinical guidelines.

An earlier paper by the same team that had compared clinical guideline formalisms (de Clercq *et al.*, 2001) looked at the combination of ontologies with those formalisms. The primary argument was that the simple action-task descriptions of clinical guideline formalisms does not sufficiently capture enough information to represent and enact a guideline in the real world. Other research groups have approached this method of representation as well: two examples outlining the contemporary state-of-the-art were (Lezcano, Sicilia and Rodríguez-Solano, 2011), which discussed the integration of OWL (Web Ontology Language) and SWRL (Semantic Web Rule Language) frameworks in a clinical environment specifically for guidelines, and (Heymans, McKennirey and Phillips, 2011) which surveys the use of the SNOMED-CT – a computer-interpretable clinical terminology – and how it can be translated into an OWL standard ontology.

After review of these technologies, amongst various criticisms, two stand out as the most relevant to this thesis. First, and most importantly, is the requirement to add information to imbue meaning to the data. Whilst the goal is laudable – to set up a uniform and interchangeable map of meaning between all entities on the Internet – the result often requires work by the end user (dealing with notations and languages, which are not their domain of expertise), or the same work by the informatics scientist (who is not a domain expert in whatever field of research they are requiring the input – e.g. clinical guidelines). This tends to result in an over-specification of functional requirements – e.g. the requirement for a large "enactment engine" in ProForma – and a requirement for extra work from both

parties. Second – in the case of ontologies – is the bias that can often creep into the specification. This is true of all descriptions (e.g. a database schema) but due to its comprehensive nature (the desire to "describe everything"), this bias can have a large negative impact on the data processing, which would be minimised if the descriptive part of the model is more constrained.

An example paper describes the state-of-the-art in this area as of 2017 (Fox, 2017), which summarises the use of the ProForma language, as part of the CREDO "cognitive computing" stack, after 20 years of development. Similar issues as those described above can be seen with this technology - there are still no agreed standards at a high-level (of use to the clinical community), the use of the technology requires tight integration with an over-specified language and implementation (the screenshots, such as figure 3.1, show interfaces that closely resemble technically-oriented development environments), and the reported study applications of the technology did not provide enough evidence of the formalism being reliable enough. Some were small, such as N=144 in (Bury *et al.*, 2005), others had isolated statements of positive results (*"Radiographers… performed better when using advice from the system")* without supporting numbers) (Taylor, Fox and Pokropek, 1999), whilst further had positive results based on general high-level criteria (61% increase in recruitment based on main criteria only) (Patkar *et al.*, 2012).

Although there is substantial work in the field of guideline formalisms, it is felt that for the purposes of building a technology framework to address the particular challenges in this thesis, many of the characteristics that these digital guideline technologies – such as flow-control representation – appeared to be well represented by the more general concept of processes, and their classification as process models.

3.4.4. Process models in a clinical context

Process models are processes of the same nature that are classified together into a model (Process modelling, 2018). As the name suggests, they incorporate a process-based model, which is the most useful characteristic of the guideline formalisms discussed in section 3.4.3. This general feature is something relatable from many aspects of professional life, such as the development and use of flow-

charts (states, actions, transitions between each, and decision points requiring contextual input) in many business and administrative contexts.

The idea of layers of abstraction is pertinent to introduce here – these can be thought of as different layers of representation of data, which have, from the lowest level to the highest, increasing levels of sophistication and complexity. An example would be the raw physiological data from a bedside ICU machine at the lowest level (one number measuring a single parameter at a single time-point), leading to a clinical diagnosis from a clinician at the highest (a complex representation of data, based on inputs and context from many sources).

Figure 3.2: visual abstraction of the layers involved in process modelling (Kless, 1993)

The use of a process model allows a specific level in these layers of abstraction to be chosen, which is an important flexibility that the over-prescription of the guideline formalisms prohibits. (Perimal-lewis et al. 2012) claims that the fundamental element required for the construction of a process model is the historical event log of a process, and this lends itself to the description of actions and reactions that occur in a medical context. This research area is referred to as "process mining" and is usually applied to the higher-level patient care work-flows within a hospital, such as transport of patients and allocation of resources such as bed-spaces. An example study like this would be (Mans, et al. 2009), which investigate the different management processes using various process mining views on flow-control structures, and how these can improve the organisation and performance within a hospital, by identifying redundant clinical pathways.

This area is related to the more general domain of business process management (BPM) not usually realized as medical processes, but critical in the use of event/reaction flow-diagrams to formally describe processes that occur within complex organisations. An example of this is (van der Werf, Verbeek and van der Aalst, 2012), which looks at tools to automate the compliance of a business to

specific guidelines, typically referred to as an “audit”. The idea behind this work is to develop an awareness of the context of a process, which can often impact the perceived compliance to a guideline, without being evident in the audit itself.

At one level, the process mining work referred to above nearly always focuses on the clinician behaviour as part of a corporate body, with a view to improving those corporate processes such as (Perimal-lewis et al. 2012). Pattern extraction science – at a lower level – focuses on mathematical techniques to detect individual events. The connection between these two levels, which is where the work of this thesis is focused, is rare, though does exist. (Huang, Lu and Duan, 2012) looks at the “clinical pathway” area, where a clinical event log is analysed, and common remedial medical behaviours are extracted. The work was validated by clinical experts as a true representation of some of their behaviours, but it did conclude that the general nature of the conclusions meant that more specific work was required, and that some critical behaviours were missed. This is an example of where the focus on a specific condition – in this case TBI – would help in identifying processes more exactly and in a way that is immediately useful to clinicians working in the ICU.

Aims of the thesis

In general, there is much work that attempts to improve adherence to clinical guidelines, or to improve the quality of information that can be extracted from studies, in order to refine guidelines more effectively. Many use non-technological methods, whilst those using technology often either provide intervention at the point of care-delivery, such as apps, or require a large input of knowledge in order to model a clinical process, such as guideline formalisms. What appears to be missing is the ability to provide new information on a clinical management process to a clinician, with a minimal requirement for knowledge (i.e. the information is extracted using only what exists at the bedside already). With the abundance of data available from modern ICUs, and the particular problems facing the output of TBI studies, this would seem to be an avenue that may provide worthwhile results and it is this gap that this thesis intends to address.

Throughout the course of the literature review, several pieces of technological work were encountered that appeared relevant to this goal. The first was work which combined the technologies of process mining and ontological descriptions of a medical environment in order to measure similarities between medical processes (Montani *et al.*, 2014). At the time of publication (2014), this work appeared to share similar issues to those described in section 3.4.3, namely an over-specification of the descriptive entities (in this case a use of "taxonomic distance" with a descriptive ontology, requiring large amounts of knowledge to be input) and the limited access to a practical implementation, which made testing the claims of the work difficult.

However, the broader theory of calculating a distance between two process models, appeared to be a viable avenue of investigation. At the start of 2014, the work of (Perimal-lewis, Vries and Thompson, 2014) had been published describing the application of process models to the administrative maintenance of medical domains, such as hospitals, and was already showing promise as a possible representation to use in this thesis. The shared elements of these two works provide a base theory to use when describing a generalised approach to representing guidelines and understanding differences between two time-varying protocols.

Further investigation into process models to represent guidelines generally, led to the use of Business Process Management Notation (BPMN) as a representation of the BTF guidelines. Authoritative work by a research group in Eindhoven, which had been used for comparison by the (Montani *et al.*, 2014) team, appeared to be highly relevant, though not specifically tailored to the medical domain. Their use of BPMN in process models was documented in (Dijkman, Dumas and Ouyang, 2008) but the work that proved most useful had been published a year later in (Dijkman, Dumas and Garcia-Banuelos, 2009), where they robustly calculated the similarity "distance" between two process models using different algorithmic approaches.

With these component technologies identified, it was possible to collate these approaches into a framework and return to the original clinical problem of representing and comparing two sets of clinical protocols – one from text guidelines, the other from actual clinical data. The final requirement was that there needed to be an understanding of the "level" at which the interpretation of guideline adherence can occur. To illustrate what this means, the following four levels of data interpretation were identified:

1. Raw data from the bedside
2. Clinical management processes
3. Statistical analysis of studies
4. Meta-analyses of grouped trials and studies

From the review conducted, it was noted that the second level in this list was one that typically received less attention than the others - most often raw data would directly contribute to a statistical analysis, with no analysis or interpretation of what the grouped information means in a live ICU context (e.g. a coarse-grained binary classification of say, patients that have received steroids during a patient stay against those that haven't).

Therefore, two broad conclusions were drawn:

1. There is a desire for improved guideline adherence through novel methods in general, and in the area of TBI in particular.

2. No single or group of technologies is sufficiently established to perform an analysis of clinical management processes for the purpose of understanding guideline adherence.

Informed by this final discussion, the original hypotheses of this thesis were formulated:

1. *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*

2. *Having extracted treatment processes, one is able to develop a method to compare those against other treatment processes to establish the degree of similarity between them*

3. *One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence*

And using the technological components identified, a research plan was drawn up.

4. Expression of clinical guidelines as process models

Chapter summary

This chapter describes the process of expressing the BTF guidelines in a format that represents them as process models, and which can be later compared to a similar format derived from the real ICU data.

The method used is a design language known as Business Process Modelling Notation (BPMN). The key features are introduced alongside a brief discussion of alternatives.

The template BPMN diagram is shown with key activities described (such as event start, and clinical management reaction). Temporal event and activity annotations are briefly described (such as time-window size and time to treatment administration).

The three threshold monitoring guidelines are described in text, then shown in BPMN representation. Finally, there is a brief discussion about validation in terms of the relationship of these guidelines to others in the BTF, the location of the implementation data for these models in the database, and the feedback from a domain expert presented with these diagrams (which, in brief, was: "reasonable, but beware of the wider clinical context").

There are 18 TBI guidelines for severe in-hospital treatment, in the 4th edition of the official Brain Trauma Foundation TBI management document. These cover various types of injury and treatment (BrainTrauma, 2018) – three more than in the 3rd edition[8]. Of these, the three guidelines that were specifically investigated were the monitoring thresholds for:

- Systolic blood pressure (SBP)
- Intracranial pressure (ICP)
- Cerebral perfusion pressure (CPP)

Clinically relevant thresholds for SBP, ICP and CPP are routinely debated and feedback from monitoring adherence to these guidelines in clinical environments would be of particular interest to the research community.

The conversion of BTF guidelines to a process model takes the form of expression in a format known as Business Process Model Notation (BPMN) (Camunda, 2018). BPMN is a representation that expresses processes in a graphical format with basic features such as flow objects (e.g. events and activities), connecting objects (e.g. sequence and message flow) and artefacts (e.g. annotations). It shares features with other software design languages, such as activity diagrams in the more commonly used Unified Modelling Language (UML), but has a primary focus on processes rather than objects. The flow-charting features of UML were considered for the expression of the BTF guidelines but were rejected due to the characteristics of BPMN being directly relevant to process models.

Another possibility for capturing the BTF guidelines was to use a domain ontology. This would be similar in implementation to the work that was originally conducted in this thesis for physiological event detection (see Appendix B). Greater context could potentially be captured using an ontology, but BPMN was still considered an better choice due to its ability to model combined temporal and spatial processes more readily.

[8] During the course of this research work (2011 – 2018), the BTF guidelines underwent a major revision from their 3rd to the 4th edition. This highlighted many of the issues raised in chapters 2 and 3 about clinical guidelines, and is detailed in Appendix C. For the purposes of the main body of work, unless explicitly stated, the updated 4th edition guidelines are assumed.

For each guideline detailed in this chapter, the text is shown along with a BPMN diagram. The representation of all three guidelines follows a similar general structure, shown in figure 4.1, and can be considered as a template for all three guidelines. Starting at the top-left of the diagram: a pressure event occurs (with a trigger threshold value), and a treatment is applied ("clinical management reaction"). The level is checked again: if the pressure has returned to a "safe" level then the treatment cycle is stopped; if not, the process continues round the cycle again. Depending on the specific guideline there may be additional factors between the event start and the application of treatment.

It is noted here that this template shows three event and activity "annotations", additional to information derived from the BTF guidelines:

- mandated time window size
- nature of treatment ("repeat/single")
- time from treatment

Though these annotations are not mentioned in the BTF guidelines, they were considered temporal requirements based on both feedback from a domain expert, and the application of common sense during evaluation of real datasets (e.g. specifying a time for a reaction to occur within). For clarity, and because they are constant across all guidelines, these annotations are omitted in figures 4.2-4.5.

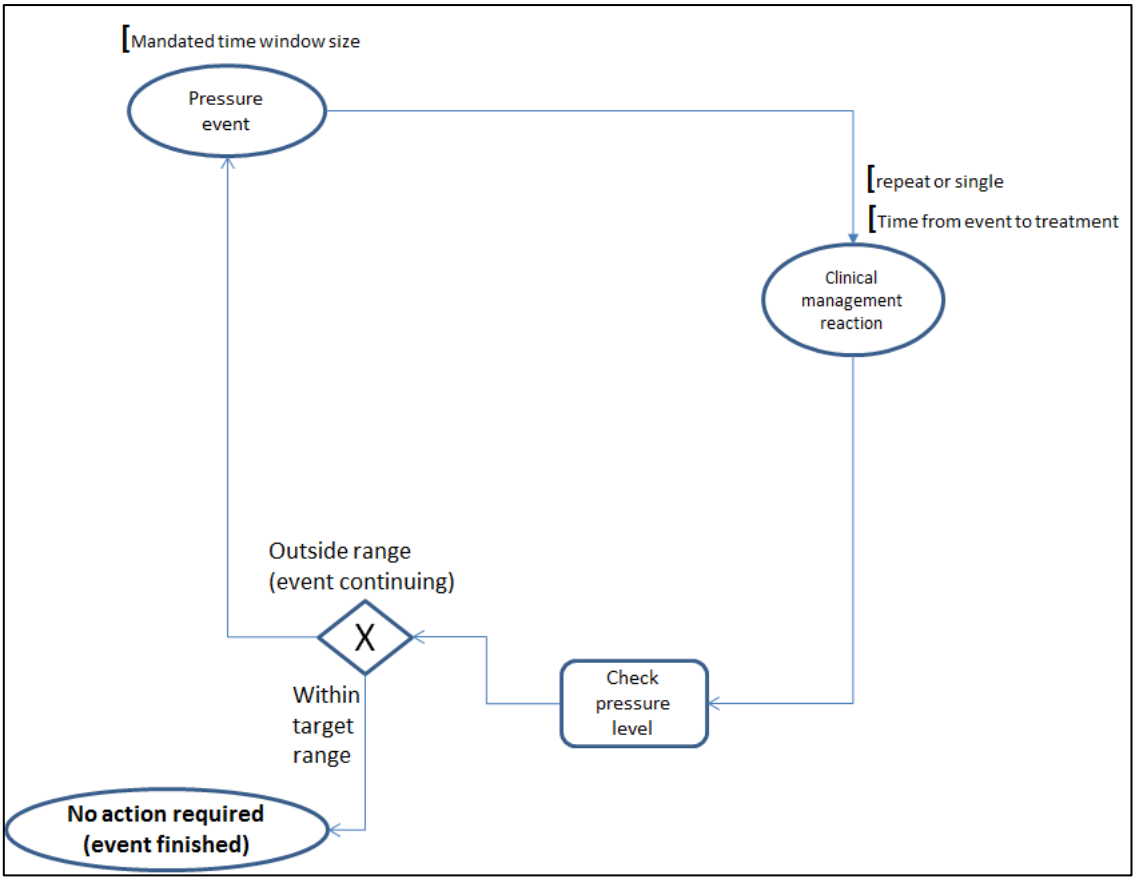

Figure 4.1: a general BPMN "template" representation of the three threshold-monitoring BTF guidelines

4.1. Systolic blood pressure

There is a single, level 3, recommendation in the BTF guidelines for the monitoring of systolic blood pressure:

- **Level 3:** *"Maintaining SBP at >= 100 mm Hg for patients 50 to 69 years old or at >= 110 mm Hg or above for patients 15 to 49 or > 70 years old may be considered to decrease mortality and improve outcomes"*

This translates to the BPMN diagram shown in figure 4.2. An event occurs (systolic < 110 mmHg), the age is checked, and a treatment is applied. The SBP level is checked again – if the SBP has returned to a "safe" level (which varies depending on age – hence the BPMN "message" symbol) then the treatment cycle is stopped; or continues if not.

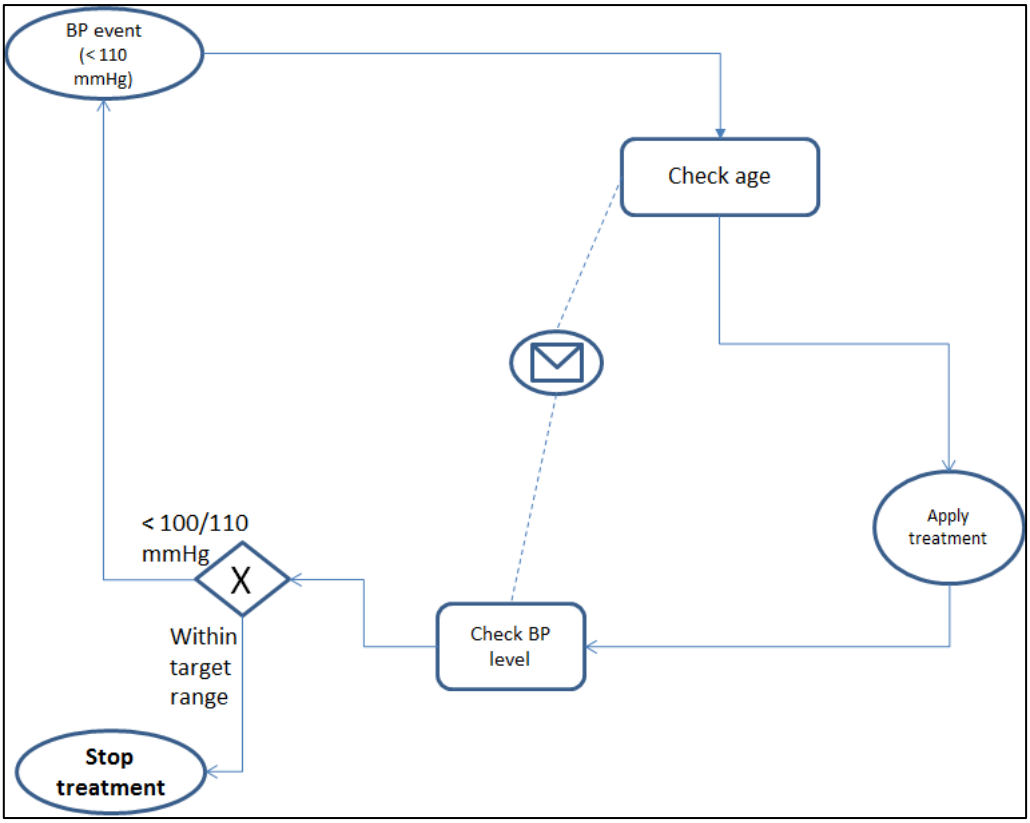

Figure 4.2: BPMN diagram of the systolic blood pressure monitoring process according to the BTF guidelines

4.2. Intracranial pressure

The BTF guidelines contains two recommendations for intracranial pressure monitoring (ICP):

- **Level 2b:** *"Treating ICP > 22 mm Hg is recommended because values above this level are associated with increased mortality"*
- **Level 3:** *"A combination of ICP values and clinical and brain CT findings may be used to make management decisions"*

This translates to the BPMN diagram shown in figure 4.3. An event occurs (> 22 mmHg), other evidence such as brain CT scans (diffuse injury II, III or IV) and other clinical findings (e.g. mass lesion), assigned and input by the treating clinician during the patient stay, are considered, and a treatment is applied. The ICP level is checked again – if the ICP has returned to a "safe" level then the treatment cycle is stopped; or continues if not.

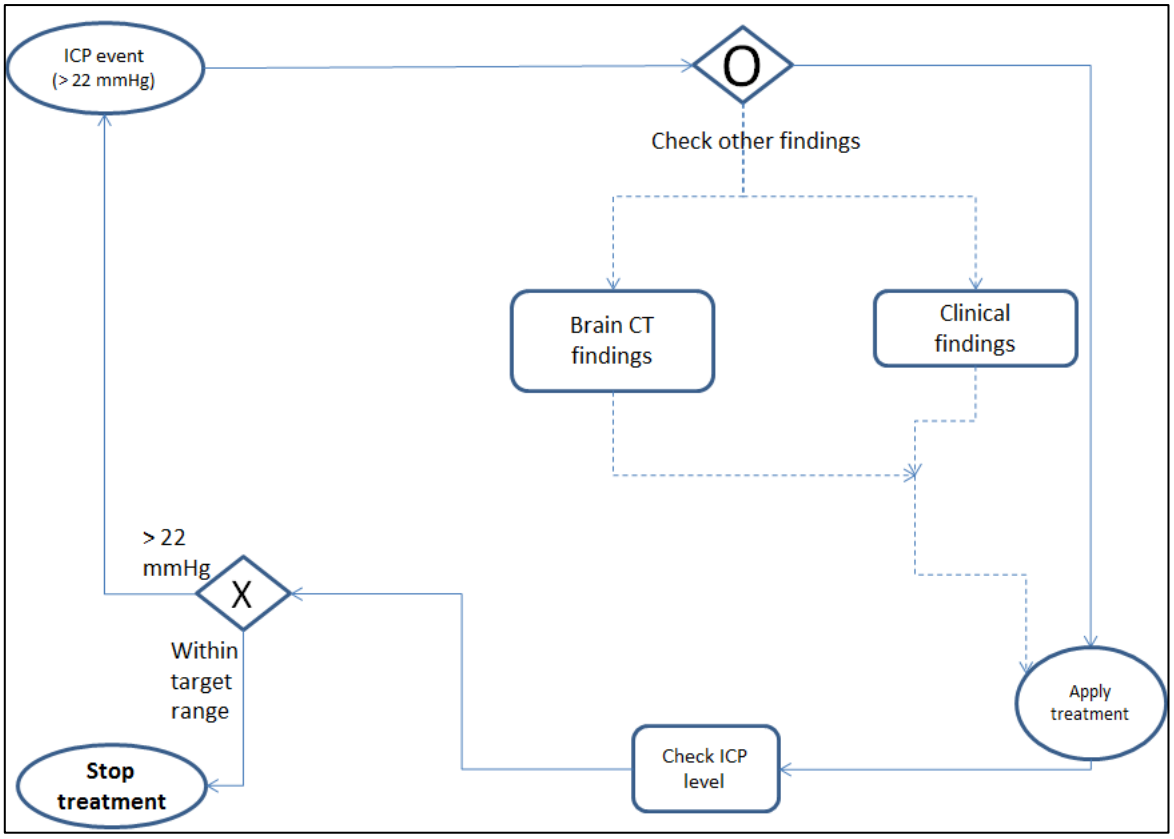

Figure 4.3: BPMN diagram of the ICP monitoring process according to the BTF guidelines

4.3. Cerebral perfusion pressure

Two recommendations for the management of cerebral perfusion pressure (CPP) are contained in the BTF guidelines:

- **Level 2b:** *"The recommended target CPP value for survival and favourable outcomes is between 60 and 70 mm Hg. Whether 60 or 70 mm Hg is the minimum optimal CPP threshold is unclear and may depend upon the auto-regulatory status of the patient"*
- **Level 3:** *"Avoiding aggressive attempts to maintain CPP > 70 mm Hg with fluids and pressors may be considered because of the risk of adult respiratory failure"*

This translates to the BPMN diagram shown in figures 4.4 and 4.5. In figure 4.4, an event occurs (< 60 mmHg), auto-regulation is checked, and a treatment is applied. The nature of the treatment is checked in a feedback loop and the CPP level is checked again, depending on the optimised value, based on auto-regulatory status. If the CPP has returned to a "safe" level, then the treatment cycle is stopped; or

continues if not. Figure 4.5 shows the detail of the feedback loop between pressor and fluid management.

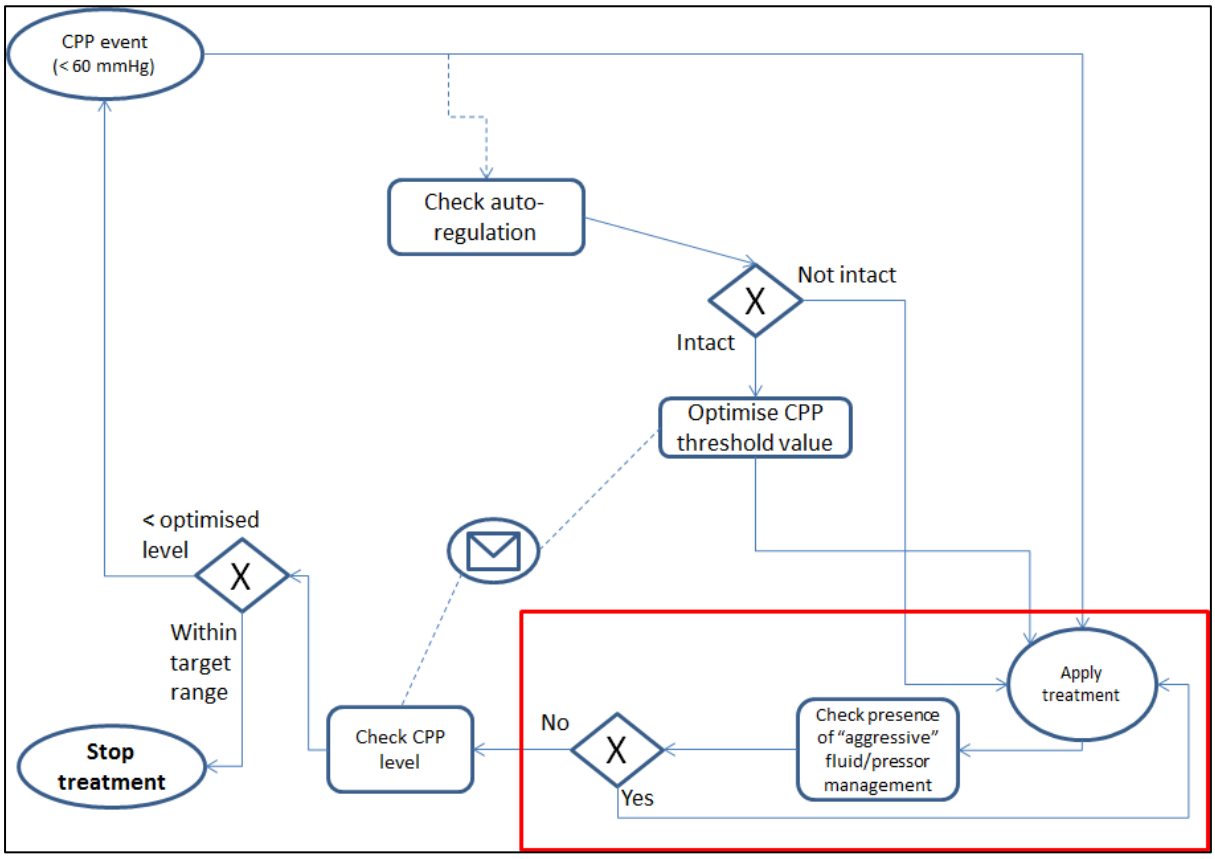

Figure 4.4: BPMN diagram of the CPP monitoring process according to the BTF guidelines

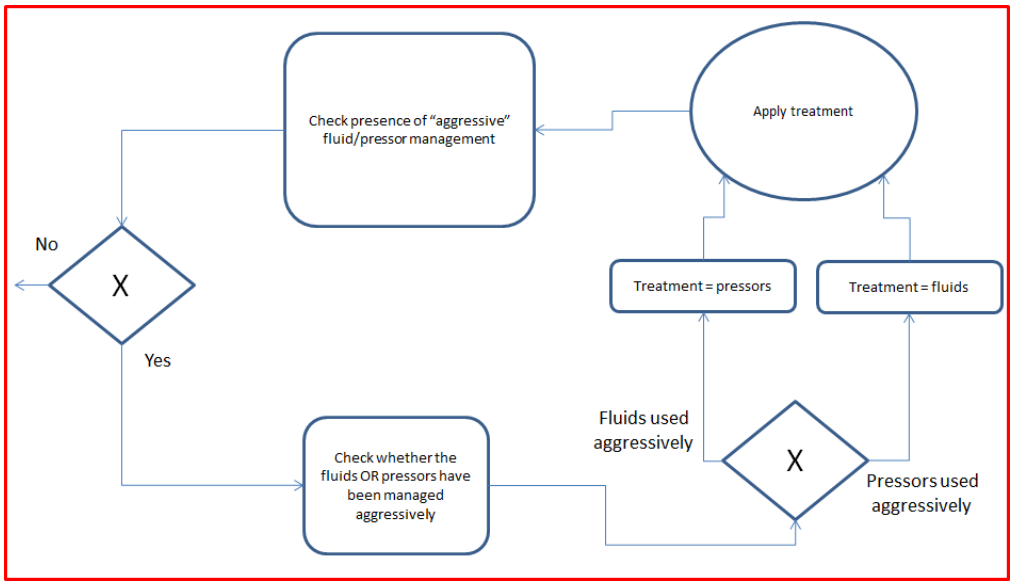

Figure 4.5: BPMN diagram of the detailed nature of the CPP treatment

4.4. Discussion

Figures 4.1-4.5 represent a single instance of a reactive process to a trigger, which in each case, is a pressure-level threshold crossing. Therefore, continuous monitoring can be represented by repeating this individual cycle many times, creating a much larger process model. BPMN diagrams are usually written from left to right, but the choice of a looped display for these guidelines deliberately reflects this cyclical link between the conditional threshold values "guarding" the beginning and end of the process.

Overlap from the other BTF guidelines can only be found in the guidelines referring to monitoring recommendations for ICP and CPP, and even here these only amount to stating that monitoring provides a general benefit to ICP/CPP management[9]. So there is no quantitative or structural impact on these BPMN process representations of the threshold guidelines. There are no contraindications in any of the other BTF guidelines, though the possibility exists that there may be wider clinical considerations that are not part of the BTF focus.

In terms of the implementation of these models – in the datasets used to evaluate the work – SBP, ICP and CPP are available in the physiological data stream. Required corollary information such as age (SBP), CT/clinical findings (ICP), and ancillary monitoring, auto-regulation and pressor/water load (CPP) are available in the captured treatment data and clinical notes where available. All other contributing information is standardised into the treatment profile database (schema details are shown in chapter 5). This corresponds to the BPMN notion of a "rule engine" or the use of business rules to evaluate the steps in the process.

Finally, the most complex of the manually extracted BPMN representations – cerebral perfusion pressure – was given to a domain expert for validation. As the

[9] However, despite the lack of overlap, there were significant differences between the 3rd and 4th editions of the recommendations supporting ICP monitoring initiation. These do not bear directly on the work of this thesis but should be noted here. In particular the following two level 2b recommendations no longer meet the required level of evidence:

"ICP should be monitored in all salvageable patients with a TBI (GCS 3-8 after resuscitation) and an abnormal CT scan. An abnormal CT scan of the head is one that reveals hematomas, contusions, swelling, herniation, or compressed basal cisterns."

"ICP monitoring is indicated in patients with severe TBI with a normal CT scan if >= 2 of the following features are noted at admission: age > 40 years, unilateral or bilateral motor posturing, or SBP < 90 mmHg."

representations were structurally similar the feedback from this was considered applicable to all three. Though requiring an explanation of the notation, the expert confirmed that these were “reasonable” representations of the processes described, and that they were valid to follow in a neurological ICU context. However, it was suggested that as they represented a singular process they may lack surrounding contextual information which can be difficult to express in formal terms. As mentioned in the introduction to this chapter, the possibility of using an ontology to provide this context may be an avenue to explore in future work.

From this work we now have three guidelines expressed in a common and comparable process model representation and can now construct a similar representation for the actual ICU data.

5. Translation of ICU data to process models

Chapter summary

This chapter outlines the conversion of the real ICU data into an event log – a sequential representation of EUSIG pressure events and treatment administration – and subsequently into a process model, comparable to the models generated in the last chapter from the BTF guidelines.

The parameters of the EUSIG pressure events are described, which are a range of values for threshold and event/clear hold-down. The algorithm to extract this pattern from the physiological data is then described, along with the assumptions made in terms of associating treatment administration to events detected in the physiological output, and the association algorithm.

The generation of process models from the event log is then described, along with the standardised database format that the event log is stored in, and also the BPMN template representation of the generated process model. Finally, a discussion of the attempts to generate the process model automatically, using a technique called "process mining", is outlined, along with the advantages and disadvantages of manual versus automatic approaches, and why the manual one was finally settled upon.

5.1. Conversion of ICU data to event log

To convert the ICU data into an event log, the word "event" is used to describe two separate entities and therefore requires clarification:

- A pressure event – as defined by EUSIG (Jones *et al.*, 1994) (see also section 2.4) – is a pattern within time-varying physiological data, well established in clinical literature. Throughout the rest of this work this will be referred to as a "EUSIG-event".
- An event that constitutes the basic unit of an event log used to create a process model, is an object of higher order abstraction which consists of a "EUSIG-event" *and* an associated treatment annotation.

5.1.1. EUSIG-event detection

In order to detect EUSIG-events in time-varying physiological data, the following key structural characteristics are required:

- **Threshold** – indicating when a EUSIG-event has started or finished by the physiological values crossing this value in one direction or the other
- **Hold-down** – the minimum time for which consecutive physiological readings have remained above the threshold, indicating that a EUSIG-event has unambiguously occurred (i.e. confirmation of the EUSIG-event start)
- **Clear hold-down** – the time for which consecutive physiological readings have remained below the threshold, indicating that a EUSIG-event has unambiguously finished (i.e. confirmation of the EUSIG-event end)
- **Duration** – the length of time from the start of the hold-down to the start of the clear hold-down
- **Value range** – the individual physiological readings during the EUSIG-event (sampled for this work at a rate of minute-by-minute)

Therefore, the values input to a EUSIG-event pattern include the threshold crossing value, the direction of crossing, and the time definition of the event/clear hold-down. Table 5.1 shows the values that can be input for threshold, direction and hold-down, which will extract different overall pattern structures from the physiological data.

Parameter	Threshold values (mmHg)	Direction of threshold-crossing	Event/Clear Hold-Down (mins)
ICP	15, 20, 25, 30, 35	Up	5, 10, 15, 20
CPP	50, 60, 70	Down	5, 10, 15, 20
SBP	90, 100, 110	Down	5, 10, 15, 20

Table 5.1: list of physiological EUSIG parameters

Figure 5.1 shows a schematic of a single physiological EUSIG-event, with a time-window for treatment overlaid (the significance and use of time windows for association are discussed in sections 5.1.3 and 5.1.4.). A threshold is crossed and remains high for a specific period (the hold-down) indicating that a EUSIG-event has started. The clear hold-down indicates that the EUSIG-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.

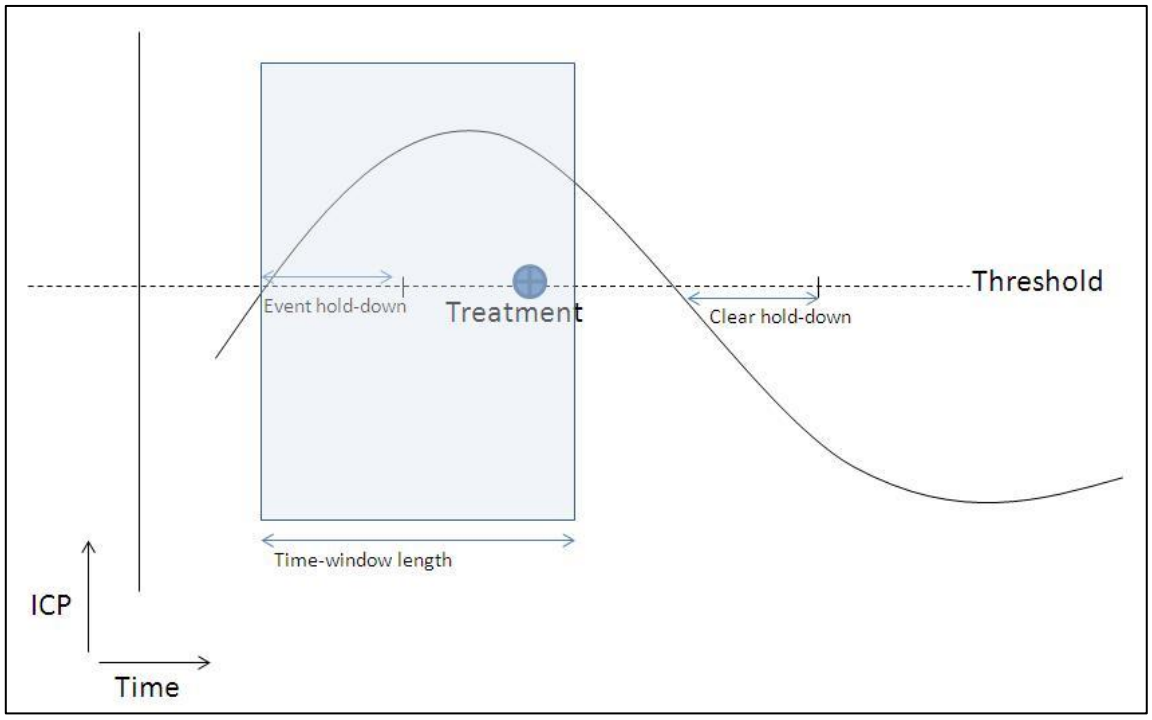

Figure 5.1: EUSIG-event definition for a given time-varying physiological data stream

To borrow language commonly used in data science, this EUSIG-event pattern can be thought of as a complex "object" with various attributes. In the context of TBI, the structural details of this defined "object" are unlikely to change, as this is a generally accepted definition of a pressure event (Jones *et al.*, 1994), (Donald et

al., 2012). This makes it a re-usable pattern, ideal for searching physiological time-series data.

In terms of the variable values – intracranial pressure (ICP), cerebral perfusion pressure (CPP), and systolic blood pressure (SBP) – the optimum values recommended by the BTF guidelines of 22 mmHg (ICP), 60 mmHg (CPP), and 110 mmHg (SBP), are supported by clinical literature. For instance, in studies outside those directly supporting the BTF threshold guidelines (e.g. class 2 studies (Berry *et al.*, 2012), (Sorrentino *et al.*, 2012), (Allen *et al.*, 2014) for SBP, ICP and CPP respectively), the lowest grades of events listed in (Jones *et al.*, 1994) are 20 mmHg (ICP), 60 mmHg (CPP), and 90 mmHg (SBP). Similarly, in (Lazaridis *et al.*, 2014), a study is conducted into individualised ICP levels based on PRx (the pressure-reactivity index) and quotes the traditionally recommended levels of ideal ICP as 20-25 mmHg.

With these key pieces of structural and numerical information about EUSIG-event definition in place, the program can be built that detects this pattern within the data-set and compiles the event log required to generate a process model. The algorithm driving the event detection is now described in section 5.1.2.

5.1.2. Event-detection algorithm

To detect the pattern described in section 5.1.1 from a physiological data stream, the following procedure is used.

Step 1 – Create the list of parameter definitions in program memory ahead of processing the event-detection algorithm. A parameter in this context represents a physiological data stream – or a physical measurement of the patient's brain (e.g. ICP). A representative parameter object is shown in figure 5.2.

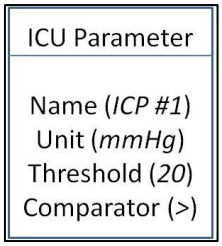

Figure 5.2: ICU parameter object with event definition values

Step 2 – Query the patient database for information[10]

- For each patient, the data is read into an *"n x n vector of vectors"* (i.e. a matrix)
- Each line in the *"n x n vector of vectors"* 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 (see figure 5.3)
- The header line is used to identify the column index for the parameter that is of particular interest (e.g. ICPm, CPP)

Timestamp	ICPm	HRT	TC	BPm	SaO2	BPs	BPd	CPP	RR	SaO2pls
18/07/2004 0:00	8	37	37.1	93	98	169	60	85	-1	-1
18/07/2004 0:01	8	37	37.1	93	98	169	60	85	-1	-1
18/07/2004 0:02	8	37	37.1	93	98	170	60	85	-1	-1
18/07/2004 0:03	8	37	37.1	94	98	171	61	86	-1	-1
18/07/2004 0:04	8	37	37.1	93	98	170	60	85	-1	-1
18/07/2004 0:05	8	37	37.1	93	98	169	60	85	-1	-1
18/07/2004 0:06	8	37	37.1	93	98	170	60	85	-1	-1
18/07/2004 0:07	8	37	37.1	93	98	169	60	85	-1	-1
18/07/2004 0:08	8	37	37.1	94	98	170	60	86	-1	-1

Figure 5.3: example of one-line-per-timestamp structure (minute-by-minute sampling)

Step 3 – For each parameter in the list, the event-detection algorithm detailed in figure 5.4 is executed. The algorithm uses the inputs from the patient parameter feeds, and for each definition of numerical variable (table 5.1) checks for event start (loop 2.4) or event clear (loop 2.3) to build the corresponding structure. Using this algorithm, the event structure can be extracted for all presented EUSIG-event definitions.

[10] Each line is processed in memory individually then written to persistent storage to make efficient use of memory heap space

```
1. Retrieve all of the parameter information for that indexed parameter object (name, unit, threshold, etc).
2. For each hold-down definition:
  2.1. Read in the line, timestamp and value (from step 2) and check the time between this timestamp and the last
  2.2. if (gap > 1min)
    2.2.1. Reset all event metrics and jump to end of the entire checking loop
  2.3. if (event is in progress)
    2.3.1. Is value still above threshold?
    2.3.2. if (no)
      2.3.2.1. Is the clear condition met?
      2.3.2.2. if (yes)
        2.3.2.2.1. if (potentialClear option is false)
          2.3.2.2.1.1. Set the potentialClear variable to true
          2.3.2.2.1.2. Increment the clear hold-down count
        2.3.2.2.2. if (clear hold-down count equals the hold-down definition)
          2.3.2.2.2.1. Note the event end time and add to event object
          2.3.2.2.2.2. Add the event to the list of events and increment the event index
          2.3.2.2.2.3. Add value and timestamp to the event list
          2.3.2.2.2.4. BREAK - GO TO 2.1.
  2.4. if (event is not in progress)
    2.4.1. Is value still below threshold?
    2.4.2. if (no)
      2.4.2.1. Is the event condition met?
      2.4.2.2. if (yes)
        2.4.2.2.1. if (potentialEvent option is false)
          2.4.2.2.1.1. Set potentialEvent to true
          2.4.2.2.1.2. Set event hold-down count to zero
        2.4.2.2.2. else
          2.4.2.2.2.1. Increment the event hold-down count
        2.4.2.2.3. if (event hold-down count equals hold-down definition)
          2.4.2.2.3.1. Note the event start time and add to event object
          2.4.2.2.3.2. BREAK - GO TO 2.1.
      2.4.2.3. if (event condition NOT met)
        2.4.2.3.1. Reset potentialEvent to false
        2.4.2.3.2. Reset event hold-down count
        2.4.2.3.3. BREAK - GO TO 2.1.
  2.5. GO TO 2. UNTIL END OF DEFINITIONS
```

Figure 5.4: the event detection algorithm, rendered in style recommended by (Zobel, 2014)

5.1.3. Association of treatments with events

Once the EUSIG-event pattern has been extracted from the physiological data, the next task is the association of treatment information with the EUSIG-events to create the event log. As described in section 2.4, features common to nearly all modern high-resolution ICU data-sets include the annotation of treatments administered to a patient during their stay in intensive care, for example a nurse administering analgesics to provide pain relief, or a ventilator machine being attached to a patient to allow steady assistance of breathing.

Again, drawing on data science terminology, the structure of a treatment object varies depending on the nature of the treatment and can be simple or complex (see figure 5.5). In its simplest form a treatment would be represented by a timestamp, and a dosage of a certain amount of drug. A more complex representation would be, for instance, 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 (also increasing the complexity of the object structure).

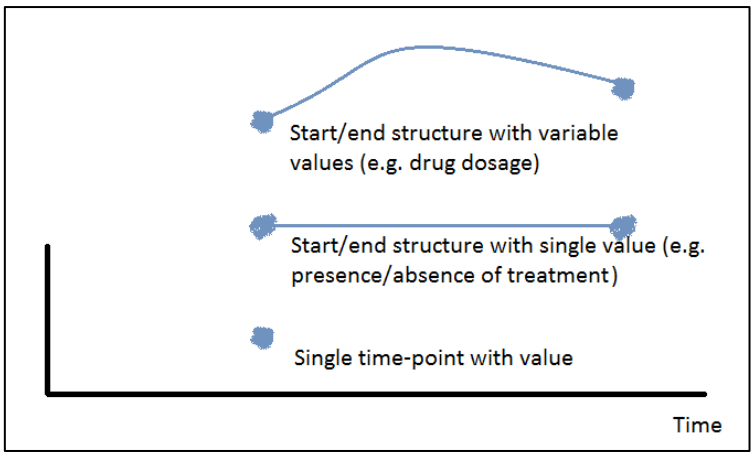

Figure 5.5: schematic example of three treatment types - simple, complex, and time-varying

Association between an event and an action can be calculated in many ways and to unambiguously establish causal association requires a lot of contextual information. Highly-targeted specification of the data-set would be the ideal method (e.g. a clinician directly highlighting for which event they are administering the treatment) but it is often the case that such specification is not available (Enblad et al., 2004). Therefore, any method that tries to establish this association can only do so to a limited degree of certainty.

In the context of this thesis, the association being discussed is indeed causal (a treatment was applied in response to a particular event). If the treatment data is well annotated with contextual information, then a parameter target will have been explicitly noted. However, not all treatments will have annotations like this, and in any case, an independence of the treatment from the physiological events – at least in the mind of the clinician noting the treatment – does help achieve a truly independent representation of clinical management.

Therefore, to account for these issues, the following assumptions have been made when it comes to treatment association with physiological events:

- If multiple treatments fall within an event time window, then – subject to explicit contradictory information in all the multiple annotations – the first treatment is associated.
- If multiple events of the same type occur in a short period, the associated time windows are conflated to one covering the full period (from the start of the first event to end of the time window after the last event)

- The treatment information has been reduced to the simplest point-like structure possible, consisting of: 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 (as this state-change is the most significant part of the treatment). The start point is the information used for noting that an treatment annotation has occurred.

These assumptions do add uncertainty to the process which is addressed in the discussion of this thesis (chapter 11). In the aggregate output of the Brain-IT dataset (section 9.4) the investigation of different values of thresholds sizes also attempts to mitigate this uncertainty by looking at the different range options.

With these considerations in mind, the event detection algorithm is run against the physiological dataset and for each event detected, several time-windows differing in length (30, 60, 90 and 120 minutes) are overlaid, to identify associated treatments.

5.1.4. Event/treatment association algorithm

This association is implemented as follows: for each patient an "association object" is instantiated, shown in figure 5.6. It contains centre and patient identifiers, a list of EUSIG-events and treatment values (e.g. sedation), and a list of associated treatment times and association number counts.

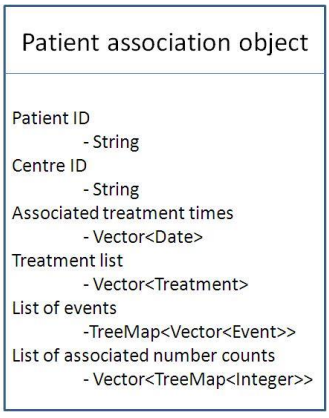

Figure 5.6: a patient association object

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),

each hold-down definition, and each time-window definition, the association algorithm shown in figure 5.7 is run.

```
1. For each event:
    1.1. Get the event start time
    1.2. Define a time-window instance of set duration that begins at the event start
    1.3. For each treatment:
        1.3.1. Get the treatment time
        1.3.2. Get the treatment target, description and value
        1.3.3. Identify all treatment instances that have the tags “cpp”, “icp” or “hypotension” anywhere in the three string values
        1.3.4. Identify all the treatments that are end tags
        1.3.5. If the treatment time is within the time-window bounds:
            1.3.5.1. If the treatment is not an end tag and the event does not already have an associated treatment:
                1.3.5.1.1. Add the treatment to list of treatments associated with this event
                1.3.5.1.2. Set the Boolean flag indicating the event now has an associated treatment
                1.3.5.1.3. Get the time to this treatment
                1.3.5.1.4. BREAK - GO TO 1.3
    1.4. Add the list of associated treatments to the time-window object for this event
    1.5. If the associated treatment list is greater than zero:
        1.5.1. Increment the associated event counter
    1.6. Add all time and treatment data gathered to the patient’s association data object
2. GO TO 1. UNTIL END OF EVENTS
```

Figure 5.7: the treatment association algorithm, rendered in style recommended by (Zobel, 2014)

Using this algorithm, treatment annotations are associated with the detected EUSIG-events, which are used in combination to create the event log, ready for process model generation.

5.2. Process model creation from event log

The output from the creation of the event log is the association of EUSIG-events with treatment information from TBI datasets. This data is then represented in a standardised format, so that future TBI datasets can be processed and compared in a similar way. Currently this standardised format is implemented in a MySQL database (known as the “treatment profile” database), and the entity-relationship diagram (the description of the database schema) is shown in figure 5.8. As noted in section 4.4, this corresponds to the BPMN notion of an “engine” database.

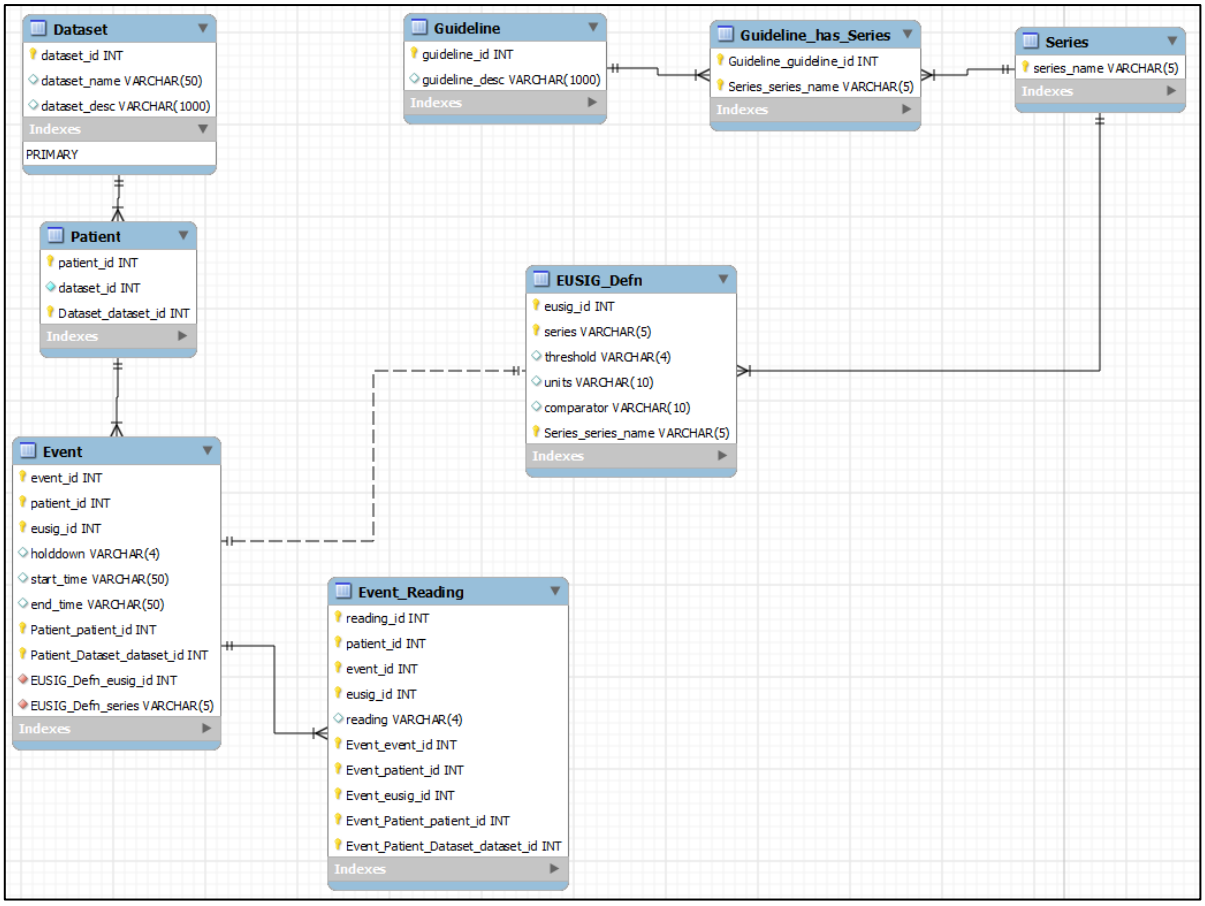

Figure 5.8: E-R diagram of the standardised "treatment profile" database. Strong key relations (where both entities can exist in their own right) between tables are denoted by an unbroken connecting line, with the foreign key in the table with the "tripod" arrow end. The single weak key relation (where one entity depends on the existence of another) is denoted by an broken line.

As discussed in section 5.1, when considering the definition of an "event" whilst developing the process model, the event log actually encompasses both the EUSIG-events and the application of treatments. This is the only abstraction that is made when constructing the standardised "treatment profile" database schema[11].

From this standardised format, it is now possible to create a process model that reflects what has occurred in the ICU. The general BPMN representation of this is shown in figure 5.9. This is the *actual* process model representing what *has occurred* in the ICU (c.f. the *ideal* process model, figures 4.2-4.5 in chapter 4, which represent what *would occur* if the guidelines were strictly adhered to). The main point of comparison between *actual* and *ideal* is the presence and nature of

[11] Again, by making this abstraction, theoretically some information may be lost if the raw data fails to follow this structural definition of an event. However, dramatic differences in this structure are unlikely to be encountered. The most unusual difference found in the clinical literature was (Lazaridis *et al.*, 2014), which used PRx values to calculate the optimum threshold-crossing value for ICP, and this still respected the basic structural definition of a EUSIG event.

the “clinical management reaction” activity (the dashed line representation in figure 5.9, which is added for clarity, and is not strict BPMN), though other attributes such as the time taken or single/repeat nature, also contribute to the comparison. Similar to the BPMN representations in the last chapter, the full process is constructed by concatenating each event represented together (similar to the “continuous monitoring” analogy, also in chapter 4).

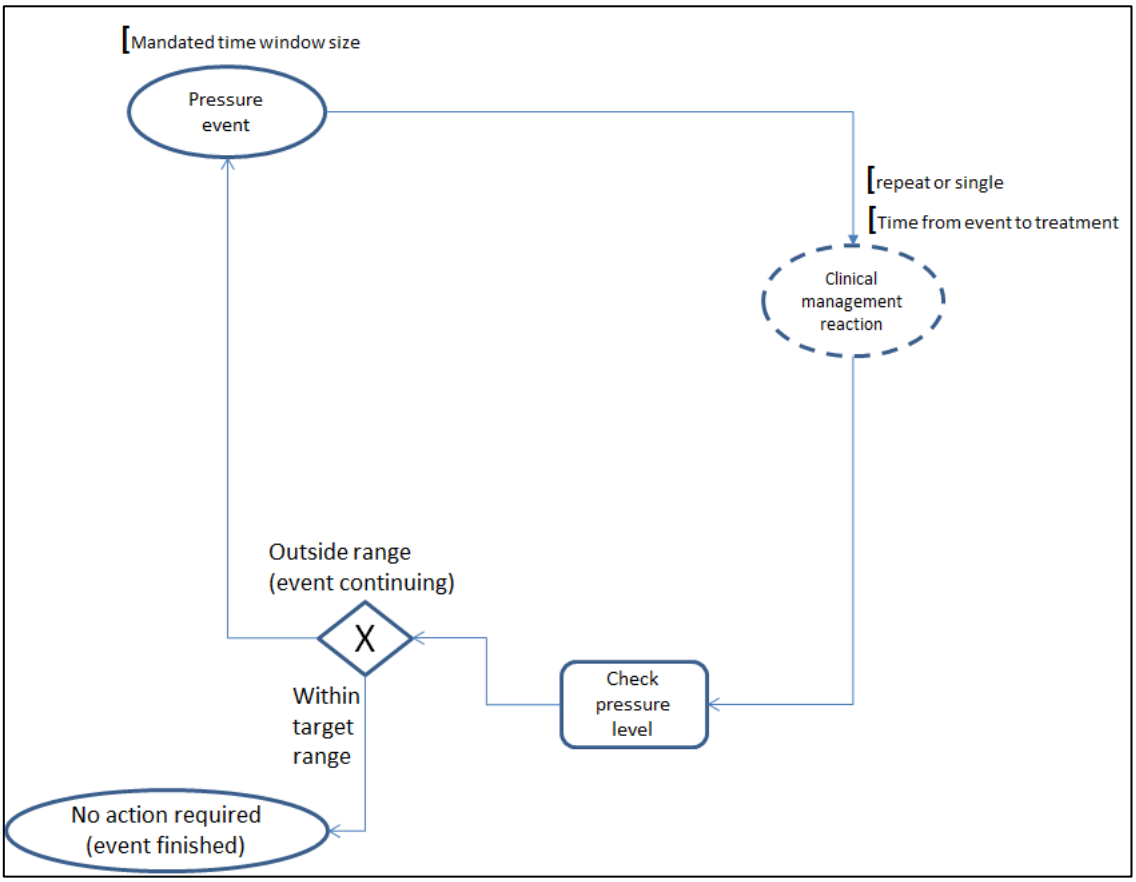

Figure 5.9: generic process model of reaction to a pressure event

Whilst figure 5.9 shows a manually generated process model, another method of generating a process model is to do so automatically ("process mining"). For this thesis, this approach was initially attempted, and the output is detailed in Appendix D.

In brief, repeated attempts to create process models automatically were highly susceptible to "noise" and produced output which bore no resemblance to real-world processes within an ICU. During the development of the automated method, there were several steps of manual arbitration to modify the output which, when pursuing a realistic model, were occurring so frequently that a fully manual

generation was eventually considered acceptable. These arbitration points included the choice of miner program (e.g. the “simple heuristic miner”), the choice of algorithm (“alpha”), and the data range (e.g. selection and deselection of activity tags).

When considering the impact of generating a process model manually versus one generated automatically, the following points for and against each approach were identified:

- A manually generated process model is subject to similar assumptions and biases that would be present in the BTF guideline models (*“ideal”*) generated in chapter 4. Errors surrounding both these generation processes may therefore multiply.
- An automatically generated process model was noisy and had relatively unrealistic output (i.e. diverged from the known real-world processes to an unrepresentative degree). Also, although subjective bias is less in automatically generated models, it is not removed completely, and often enters the calculation through the arbitration points mentioned above.

With these arguments in mind, the construction of manually generated process models was considered a viable option and the work continued along this path.

6. Calculating distance and similarity of process models

Chapter summary

The two process models that have been derived in chapter 4 (the ideal guideline-mandated process) and chapter 5 (the actual process that occurred in the ICU) – have been expressed in a common model notation (BPMN). This chapter now describes the calculation of a distance between those two models, which involves the evaluation of scalar and structural distances.

The process model comparison work that this thesis references (Dijkman, Dumas and García-Bañuelos, 2009) *is briefly discussed with emphasis on the comparison between a "pure" process model and an execution trace.*

The detailed method on how this comparison is applied in this thesis is then outlined, showing how the two process models (actual vs ideal) are evaluated on a minute-by-minute basis. The quantitative distances and weightings are then tabulated along with the qualitative reasons for non-adherence.

Finally, a worked example is shown for illustration along with a brief discussion of the presentation of results: minute-by-minute windowing, interquartile range tables, and a grid providing a summary-measure estimate of the clinical severity as a result of the non-adherence.

As mentioned in chapter 3 (section 3.5), highly relevant research that looks explicitly at a calculation of distance between two process models, can be found in (Dijkman, Dumas and García-Bañuelos, 2009). Dijkman's work proposes a method of comparing process models by converting the BPMN representation to a directed graph – a minor modification to a set of nodes (points on the model) and directional edges (connectors between those points) – then searching through a repository of process models and calculating a distance between each of those. The calculation itself is composed of three conversion steps: node label replacement (referred to as "string-edit similarity"), distance between two nodes ("graph-edit distance"), and a weighted description of that distance ("graph-edit similarity").

Other methods of calculating distances between complex, multi-dimensional objects exist. For the work of this thesis, two methods of comparison were considered: graph-based and document-based. These were chosen due to the process-oriented nature of guidelines and their communication in text documents. The latter of these methods primarily involves comparison of label strings with concepts such as a "bag of words", where a distance between two texts is calculated based on the overall similarity of all the words in a document (similar in function to Google's PageRank algorithm) (Xing and Ghorbani, 2004). Initially, this was considered the most appropriate technique, as the BTF guidelines were themselves text documents. However, a graph comparison approach was eventually settled on as the real ICU data processes had a flow-control structure, which could also be derived from the text of the BTF guidelines. This was considered to capture the processes more accurately that a document-based approach, and of the literature surveyed, the work by Dijkman appeared to be the most relevant.

The final conversion step in (Dijkman, Dumas and García-Bañuelos, 2009) – the "graph-edit distance" – is a complex calculation, known as an "NP complete" problem as there are a vast (potentially infinite) number of distance solutions to find. Several algorithms to do these calculations were assessed within (Dijkman, Dumas and García-Bañuelos, 2009), each with different characteristics that trade-off between completeness and efficiency: these are the "Greedy", "A-star",

”Process heuristic”, and “Exhaustive” algorithms. Focusing on the algorithm’s performance in their paper, the conclusion was that the “Greedy” algorithm (searching for local optima) and “A-star” (a well-known shortest-distance algorithm) were the best performing in terms of speed versus acceptable completeness (“A-star” being slightly slower but more accurate). These algorithms do not play a significant role in the work of this thesis, as a relatively “simple” exhaustive algorithm was sufficient due to the constraints imposed by the problem addressed, such as the clear identification of start/end points in the EUSIG events, and the well-defined structure of the nodes being compared. However, a valid and likely avenue of future work would be to apply these algorithms to the problem space in this thesis for further efficiency gains, or to allow the expansion of the available categories to better reflect the detail of a neurological ICU. Therefore, the details of these algorithms are listed in Appendix E.

To re-state for the work of this thesis: one process model – drawn from the BTF guideline – represents what the *ideal* clinical response would have been given the context of events, the patient situation, etc. The other process model is generated for the *actual* timeline from the treatment profiles database, which is a model representing what actually happened in the ICU for the same patient context. Therefore, an important point that should be emphasized is that the former is a “model”, in the purest sense of the word, whilst the latter is an execution trace – an “instantiation” of the process model using exact numerical input reported from the immediate situation[12]. In process model theory, it is considered acceptable to make interchangeable comparisons between these entities – a short discussion on this is available in (van der Aalst, de Medeiros and Weijters, 2006), describing the different levels of comparison available given local execution circumstances to a particular model.

In summary, the distance between the actual and ideal processes is calculated using the main components of the conversion work established by Dijkman. The rest of this chapter now describes the detailed method for this implementation.

[12] A useful analogy to consider for this concept is that the process model can be thought of as written code, whilst the execution trace can be thought of as the execution of that written code.

6.1. Process models for comparison

The BPMN representations of both the guideline and the output from the ICU data have been constructed in chapter 4 (e.g. figures 4.4 and 4.5 for the CPP guideline) and chapter 5 (figure 5.9). To illustrate this more clearly, a time-varying comparison between the two has been drawn in figure 6.1, showing an "unwrapped" (left-to-right, rather than in a loop) version of the BPMN model and their points of comparison.

When an event is triggered, for each minute of that event, a distance is calculated between the guideline process model (top) and the current state of the ICU output (bottom). This single evaluation is what is shown in figure 6.1.

If a reaction is required by the guideline model (e.g. to administer hypertonic saline) and that reaction is not found in the ICU model at that time-point, then the distance between the two models will be greater than if it was found. The distance would also vary according to differing types of reactions. For instance, if hypertonic saline were required but vasopressors was found, the distance would be non-zero, but less than the maximum possible. If the reaction found is exactly the same as that required by the guideline, then the distance for that component would be zero.

This evaluation occurs each minute that the event continues (figure 6.2) and therefore produces a distance number for each minute of that event. In this model, the greatest distance occurs if there is a mis-match between clinical management reactions. But smaller distances can also occur if there is a difference in the nature of the reaction (e.g. dosage) or the time taken to treatment. The size of these differences relative to each other is controlled by the assigned weightings (discussed in section 6.2).

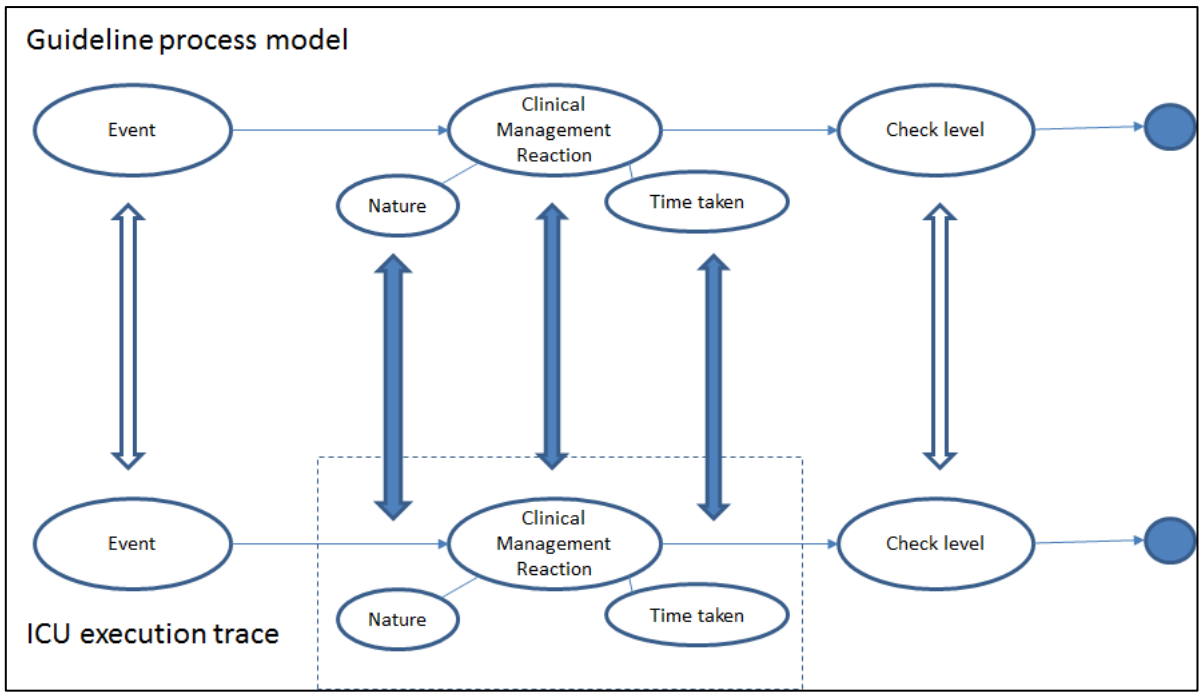

Figure 6.1: Like-for-like comparison of the ideal and actual process models

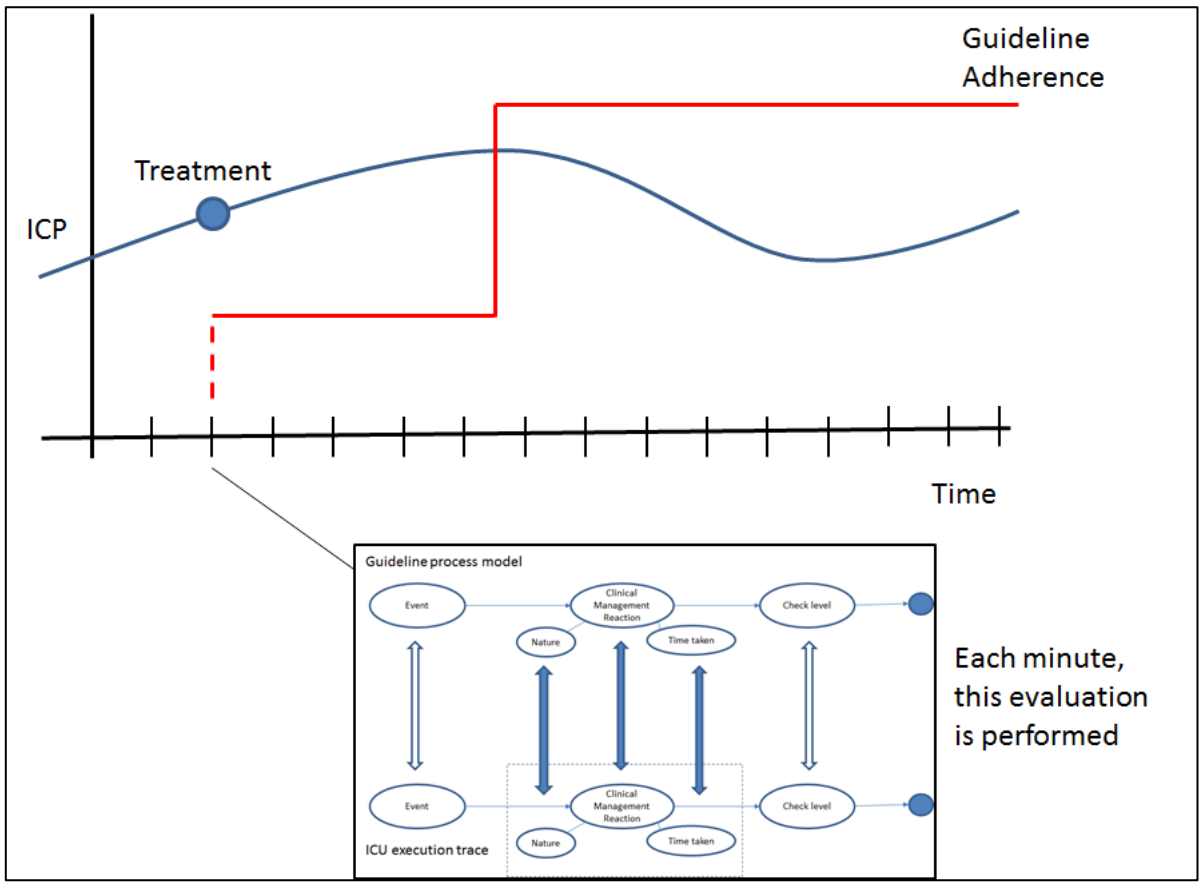

Figure 6.2: Implementation of the process model comparison conducted each minute

6.2. Quantitative and qualitative comparison detail

Returning to the underlying theory, the three steps to calculate a distance between process models, are defined in Dijkman's work as follows:

- **String-edit similarity** – this is the direct similarity between two labelled nodes on the model. The primary attribute of a node is its label, and Dijkman approaches this with a view on the similarity of the literal strings themselves. In this situation however, the labels follow the assigned values relevant to the domain (in this case, an example is the categorisation of TBI treatments). A full list of the accepted node values is shown in table 6.1.

- **Graph distance** – this is the number of steps that must be taken for one process model to become the equivalent of the other (including node substitutions, or string-edit similarity calculations). The steps involved can either be substitutions, insertions or deletions of either nodes or edges of the process model.

- **Graph-edit similarity** – the final similarity calculation is produced by adding weighting "costs" to the steps achieved in the graph distance calculation. These weightings are outlined in table 6.2.

The formal definition of node and edge substitutions in mathematical terms (according to Dijkman) is shown in figure 6.3.

The fraction of inserted or deleted nodes, denoted fskipn*, the fraction of inserted or deleted edges, denoted* fskipe *and the average distance of substituted nodes, denoted* fsubsn*, are defined as follows.*

$$\text{fskipn} = \frac{|\text{skipn}|}{|N_1|+|N_2|} \quad \text{fskipe} = \frac{|\text{skipe}|}{|E_1|+|E_2|} \quad \text{fsubn} = \frac{2.0 \cdot \Sigma_{(n,m)\in M} 1.0 - \text{Sim}(n,m)}{|\text{subn}|}$$

The graph edit similarity induced by the mapping M is:

$$1.0 - \frac{\text{wskipn} \cdot \text{fskipn} + \text{wskipe} \cdot \text{fskipe} + \text{wsubn} \cdot \text{fsubn}}{\text{wskipn} + \text{wskipe} + \text{wsubn}}$$

The graph edit similarity *of two graphs is the maximal possible similarity induced by a mapping between these graphs.*

Figure 6.3: Mathematical expression of the weightings used in the graph-edit similarity calculation (Dijkman, Dumas and García-Bañuelos, 2009)

- *wsubn* – this co-efficient describes the weighting attached to the substitution of a node
- *wskipn* – this co-efficient describes the weighting attached to the insertion/deletion of a node
- *wskipe* – this co-efficient describes the weighting attached to the insertion/deletion of an edge
- *fsubn* – the formula shown in figure 6.3 shows the graph-edit distance (itself made up of the string-edit distance (*1.0 – Sim(n,m)*) divided by the set of all nodes that have undergone substitution

Table 6.1 outlines the possibilities for the labels in the nodes that make up the comparison (figure 6.1), represented as substitutions of the node. The options are categorical apart from the time taken, which is measured as continuous, but assigned to the node as a categorical full-minute number.

Node	Label options
Treatment type	Ventilation, Sedation, Analgesia, Paralysis, Volume expansion, Inotropes, Anti-hypertensives, Anti-pyretics, Hypothermia, Steroids, Cerebral vasoconstriction, Osmotic therapy, CSF drainage, Head elevation, Barbiturates, Other
Nature of treatment	Single, repeat
Time taken	Time between event and treatment

Table 6.1: Node label possibilities for the different nodes in each process model

The most likely structural change in the comparison of actual against ideal processes involves the treatment itself (the node labelled as "clinical management reaction" in figure 6.1). Therefore, the algorithm for calculating the structural difference between models is:

- A node is deleted, which means that a corresponding edge is deleted as well
- If the node is the "clinical management reaction" (i.e. a central node in the process):

    - This deletes the entire group of nodes and associated edges (a major difference between the models)
- Else (the node is peripheral in the process):
    - The node has either the “nature” or “time taken” label
    - The detailed specification of the structure is reduced (which is a relatively minor difference between the models)

Considering the relative importance of the “clinical management reaction” node, and connecting edges, versus the other node types, table 6.2 outlines the weighting costs of each of the steps involved when comparing one model against the other (again with reference to figure 6.1).

These weightings have been assigned as a measure of how important each difference is, relative to each other. For instance, the assignment of 0.99 to the central node being deleted, indicates this is the most important difference in the list (arrived at by adding the base cost of deleting a node, 0.75, and the node being central, which is 0.24). Correspondingly the lowest importance (“nature”) is assigned 0.25 as it is a relatively minor difference. These weightings have been chosen in attempt to model the importance of the different nodes as accurately as possible but require further calibration and consensus from clinical domain experts.

Conversion step	**Normalised weighting (0 – 1)**	**Variable**
Nature label switched	0.25	*wsubn*
Time taken label switched	0.5	*wsubn*
Treatment type label switched	0.5	*wsubn*
Node deleted (base cost)	0.75	*wskipn (base)*
Edge deleted (base cost)	0.6	*wskipe*
Deleted node is central	0.24	*wskipn (additional)*
Deleted node is peripheral	0.08	*wskipn (additional)*

Table 6.2: Weighting values of each step involved, when converting from one model to the other

The calculations in figure 6.3 and the assigned weighting values from table 6.2 provide quantitative information about guideline non-adherence. However, for multi-dimensional and interacting structures such as a guideline process, it is

important that the qualitative information supporting that numerical output is retained as well. For node deletion it is noted that there is a base and additional cost of deletion (so *wskipn* is repeated), the additional cost depending on whether the node deleted is central or peripheral.

The qualitative reasons for non-adherence to a guideline are constructed by recording the individual steps taken to get from one model to the other. As the distance between one model and another is the shortest series of steps to convert one into the other (the "graph edit distance"), it follows that each documented step is being taken for "a reason". The final (compound) list of these reasons provides the trace of qualitative information. In this way maximum information is retained throughout the evaluation of guideline adherence. Table 6.3 shows the one-to-one mapping of the qualitative reason for each step taken, when converting one model into another.

Conversion step	**Qualitative reason for guideline deviation**
Nature label switched	"Nature of treatment is different"
Time taken label switched	"Time taken to administer treatment is outside window"
Treatment type label switched	"Type of treatment is different"
Node deleted	"A component is missing: " [component specified below]
Edge deleted	"A component is missing: " [component specified below]
Deleted node is central	"Treatment"
Deleted node is peripheral	"Nature/Time taken"

Table 6.3: Qualitative reasons for guideline deviation

6.3. Worked example

To give some practical context to the theory discussed in section 6.2, the following worked example demonstrates the calculation of guideline adherence. Figure 6.4 shows the minute-by-minute output for the worked example described in this section.

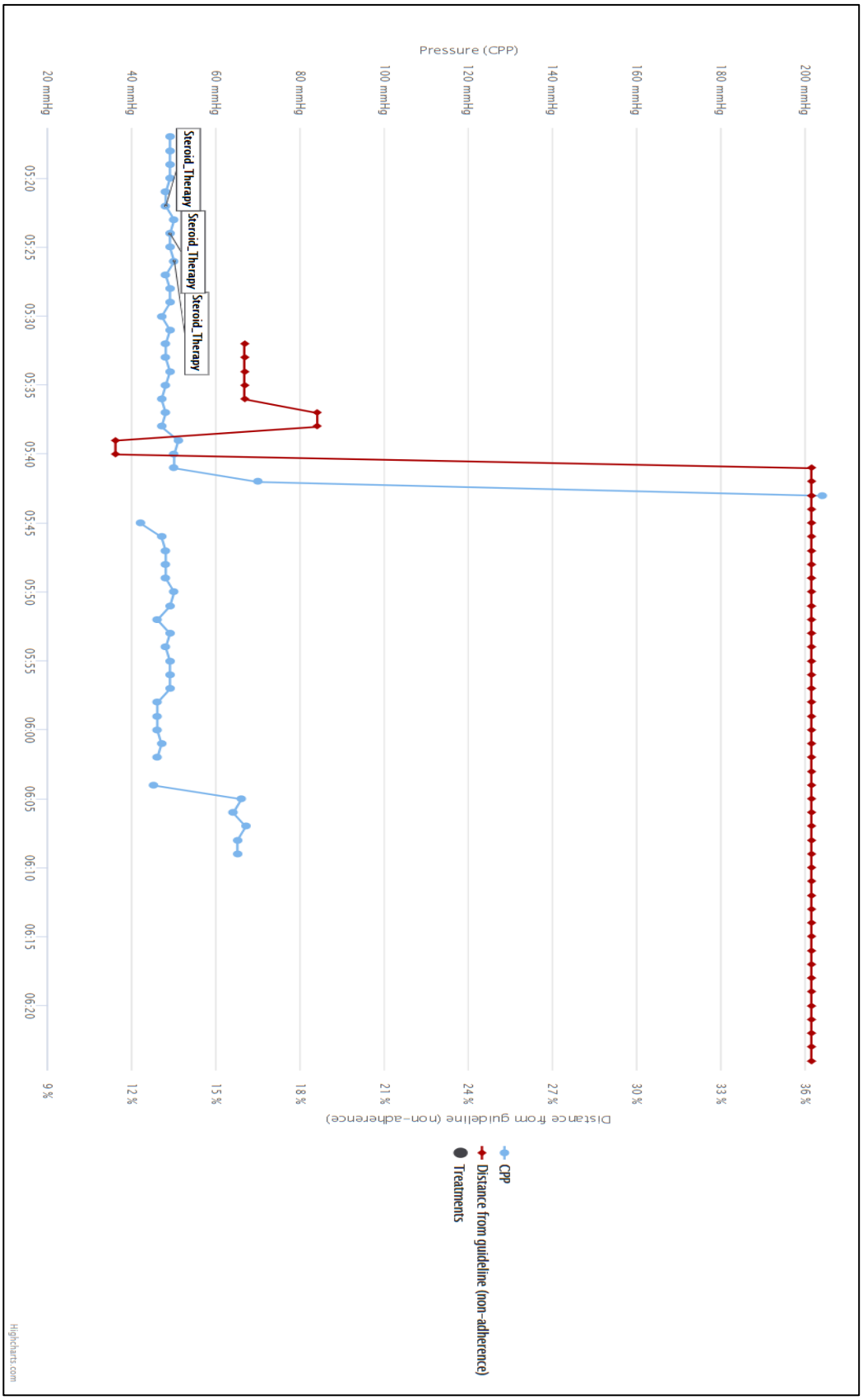

Figure 6.4: Minute-by-minute output of guideline adherence of a single CPP EUSIG-event

6.3.1. Calculation template

The following situation in an ICU provides the context: in response to a CPP event, a clinician has administered three doses of steroids within close proximity to each other and within 15 minutes of the event start. The guideline only recommends two doses within that time-frame.

Therefore, the inputs to this calculation are:

- A single **CPP** EUSIG event (**< 50** mmHg) starting at **2004-05-28 05:17**
- Three **steroid therapy** treatments at **2004-05-28 05:22**, **2004-05-28 05:24** and **2004-05-28 05:26** (all within a **15-minute** time-window since the EUSIG-event start).

The anticipated outputs are:

- The guideline adherence information will be composed of the following factors:
    - The incorrect type of treatment administered after a high-load of that treatment has been established (in this case the third instance of "steroid therapy").
    - Time taken from the EUSIG-event start to treatment administration.
    - Default high non-adherence level once the time-window has expired and a further treatment has not been administered.

For each time-point within the time-window, the calculation includes a measurement of distance for the five points of the model (with reference to figure 6.1). Table 6.4 shows a template of how the distance weightings for this particular case are calculated. The five factors add together to create 100% of the single distance score (non-adherence), so the value of each contributes 20% of the overall score. The practical difference between guideline (ideal) and ICU (actual) is described in parentheses in the second column. Note that Type refers to the difference between actual administration and ideal, whereas Nature refers to any dose of a given drug.

Node	Similarity weightings
Event start	0.0 (no difference)
Type	0.5 (load of treatment type is high)
Time taken	Time to treatment for each dose
Nature	0.25 * number of doses over (+1)
Check level	0.0 (no difference)

Table 6.4: Template of values for each non-adherence instance

6.3.2. Calculation instances

In this calculation, the output shows four instances of non-adherence within the time-window 11.4%, 16%, 18.6% and a large distance of 36.2%[13] outside it (figure 6.4).

The distance for the large "default" instance (36.2%) is calculated using the graph-edit similarity formula as this is a comparison between two structurally different process models. Table 6.5 shows the contribution of each node to the overall guideline non-adherence distance.

Node	Similarity weightings (37.8% instance)
Event start	0.0
Type	0.56
Time taken	0.54
Nature	0.79
Check level	0.0

Table 6.5: Component contributions of each factor for the 37.8% instance

These are arrived at using the weightings listed in table 6.2 and the following additional variable assignments:

- *fsubn = **1.0** (this is the substitution weighting, which in this case is a normalised value of 1.0 as it is at maximum with the missing node)*
- *fskipn = fskipe = **0.6** (for Type - fraction of all nodes substituted = 3/5)*
- *fskipn = fskipe = **0.2** (for Nature/Time taken - fraction of all nodes substituted = 1/5)*

[13] There is an issue with this figure, discovered post-submission, that leads it to be calculated as 37.8% (previously thought to be 36.2%). See Appendix F.3 for an explanation. Wherever the 36.2% figure is encountered with regard to the "default" distance, read 37.8%.

Using the similarity mapping shown in figure 6.3, the distance is calculated as follows:

- *((wskipn * fskipn) + (wskipe * fskipe) + (wsubn * fsubn)) / (wskipn + wskipe + wsubn)*

Therefore, the similarity value of the "Type" component is (rounded to 2 decimal places):

- *((0.6 * 0.99) + (0.6 * 0.6) + (0.5 * 1.0) / (0.99 + 0.6 + 1.0))* = ***0.56***

The similarity values of the "Nature" component is:

- *((0.2 * 0.83) + (0.2 * 0.6) + (0.25 * 1.0)) / (0.83 + 0.6 + 1.0))* = ***0.54***

The similarity values of the "Time taken" component is:

- *((0.2 * 0.83) + (0.2 * 0.6) + (0.5 * 1.0)) / (0.83 + 0.6 + 1.0))* = ***0.79***

Add these five factors together to get the full contribution (normalised as a percentage):

- *((0.0 + 0.56 + 0.54 + 0.79 + 0.0) / 5.0) = 0.378* = ***37.8%***

For the distances within the time-window, the calculation follows the string-edit formula as the calculation between the two process models is structurally similar. Table 6.6 shows the contributing values.

Node	**Similarity weightings (instances)**		
	16%	**18.6%**	**11.4%**
Event start	0.0	0.0	0.0
Type	0.5	0.5	0.0
Time taken	0.3	0.43	0.56
Nature	0.0	0.0	0.0
Check level	0.0	0.0	0.0

Table 6.6: Component contributions of each factor for the instances within the time-window (16%, 18.6%, 11.4%)

With the string-edit distance, the five factors are simply added up and normalised to a percentage, so the calculations and results are:

- *((0.5 + 0.3) / 5.0) * 100 =* ***16%***
- *((0.5 + 0.43) / 5.0) * 100 =* ***18.6%***
- *((0.56) / 5.0) * 100 =* ***11.4%***

6.3.3. Presentation

Using these distance calculations, the final quantitative and qualitative expressions of guideline adherence are generated. The basic units of adherence are two – potentially interdependent – categories: degree of non-adherence (expressed as a percentage) and the duration of these levels of non-adherence (in minutes). The adherence levels detailed at each minute time-point of a patient's stay are compiled into a set of contiguous "instances" of non-adherence. The blue line indicates the CPP level, whilst the red line indicates the corresponding level of guideline adherence.

All non-adherence information for each patient can be viewed in the application referenced in the "Additional Resources" page at the front of this thesis. This particular (test) patient can be viewed by inputting the summary details shown in figure 6.5.

Summary criteria

- Dataset: **Brain-IT**
- Patient ID: **15026161**
- Physiological series: **CPP**
- Threshold (mmHg): **50**
- Hold-down (mins): **5**
- Time window (mins): **15**

Figure 6.5: Summary input information for the worked example

Briefly, a more detailed description of the "default" distance, calculated at the beginning of this section, is merited here. This has been labelled as "default" due to its overwhelming prevalence in the datasets tested, and it being the value that the distance metric "defaults" to when the most common situation occurs: namely, when a treatment should have been administered in the time-window following an event (but hasn't). The value is output at 36.2% in the cases reported, due to the choice of quantitative weightings (table 6.2). An intuitive assumption is that this case should have a distance score of 100%. However due to the fact that

the five components (figure 6.1) are evaluated semi-independently, each contributing 20% to the overall distance, and following the weightings applied, the common output – as worked through in this section – is 36.2%. Possible future work would be to calibrate this figure to have a more meaningful clinical analogue. The concept of this default distance is returned to in chapters 8, 9 and 10, in particular in section 9.3.2.

7. Framework implementation

Chapter summary

This chapter presents the implemented solution to the framework proposed in the thesis. The high-level design is introduced with a step-by-step description of each component, along with references back to the original hypotheses.

The datasets to be used to evaluate the technology are then described including: Brain-IT (specialist neurological ICU data), MIMIC III (non-specialist ICU data), and ICCA (ICU data available from bedside machines, and accessible to a domain expert).

The hardware and software implementations are described, with a brief description of the main code features implementing the key methods from chapters 4 to 6.

The final section describes the presentation of the application. It is a web-enabled system that provides adherence information on individual patients in a dataset, including minute-to-minute guideline adherence for a single patient stay, total duration and distance of non-adherence, and interquartile range spread to understand the significance of the different non-adherence instances. A novel method of presenting clinical severity of the combined output of guideline non-adherence and duration using risk analysis charts is also described.

The overall goal of this framework is to represent both the BTF guidelines and the real ICU data as process models, then compare the distance between the two. Figure 7.1 shows a high-level schematic of the steps required to achieve this, reflecting the methodology outlined in chapters 4 to 6 (repeated from figure 1.1).

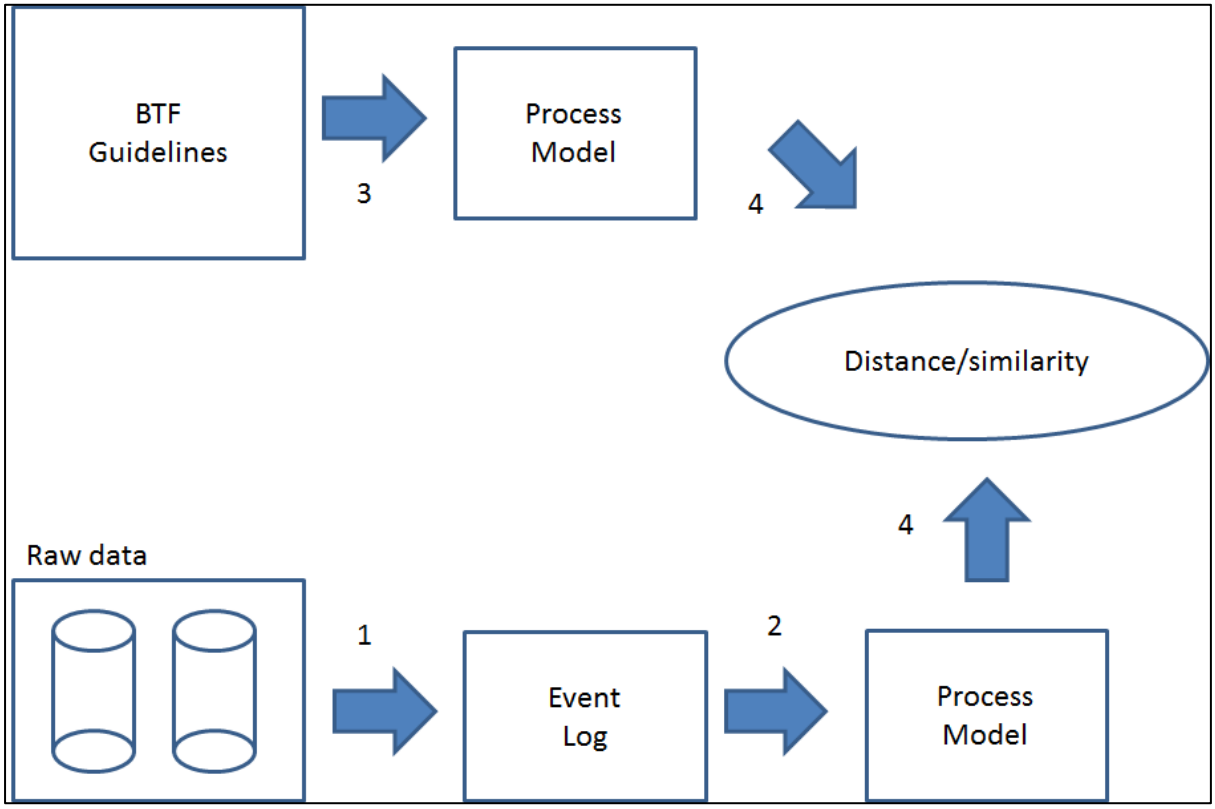

Figure 7.1: Simple schematic of the architectural process underpinning the proposed research

The steps involved in comparing two sets of process models in this context, are numbered in figure 7.1 and represent the following:

1) Create an event log from the raw ICU data, using EUSIG event definitions
2) Create a process model representation of the ICU data from that event log
3) Create a process model representation of the BTF guidelines from their text
4) Compare the two process models and evaluate how similar they are

Recalling the hypotheses formulated for this thesis:

1) *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*

2) *Having extracted treatment processes, one is able to develop a method to compare those against other treatment processes to establish the degree of similarity between them*
3) *One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence*

Steps 1 and 2 in figure 7.1 are directly related to the first hypothesis of this research work: a process model and event log generated from the raw data encapsulates a treatment process for ICP/CPP management.

Step 4 in figure 7.1 relates to the second hypothesis: a comparison between treatment processes is performed between the process models generated from the raw ICU data and the BTF guidelines. Step 3 acts as an intermediate step for the second hypothesis: the process model is generated from the guidelines in order to perform the comparison.

The third hypothesis is realised by the full implementation of the framework as a web-enabled application.

7.1. Application to ICU datasets

To apply the system to real TBI data, the following datasets have been sourced and have guided the development of this research:

- Brain-IT (Piper *et al.*, 2010)
- MIMIC III (Saeed, 2007)
- Philips ICCA (ICCA, 2018)

These three datasets have been chosen as they represent common formats for the storage and processing of ICU data in a clinical and research setting. Brain-IT is a consortium that has developed a data schema specifically for the collection of TBI data, with a particular focus on the management of ICP/CPP. MIMIC III is a large-scale repository of end-hour averaged data, chosen for its general focus on non-specialist ICU data. The Philips ICCA system is a bedside data collection system that is used by many neurological ICUs in the developed world.

7.1.1. Specialist neurological ICU dataset (Brain-IT)

The Brain-IT dataset is a collection of 262 TBI patients with clinical information collected from 22 specialist neurological ICUs across Europe. The data consists of physiological, treatment, lab results, surgeries, and other important clinical events, and was collected using a variety of technologies available at the time (2002 to 2007) such as interfacing with bedside monitoring machines to collect physiological data and using PDA Palm Pilots to collect treatment annotations.

This dataset has been used as the primary one for implementation, due to its comprehensiveness, coverage and availability. The consortium focus on ICP/CPP management facilitates the identification of targeted treatments, but also represents a potential bias in treatment focus that must be considered when evaluating the final output.

7.1.2. Non-specialist ICU dataset (MIMIC III)

To provide a counterpoint to the specialist focus of the Brain-IT dataset, the MIMIC III dataset has been chosen specifically because its primary collection purpose is *not* TBI or ICP/CPP management. This allows the generality of the system to be evaluated and whether it can be transferred from one context to another using similar, but different, data structures.

The MIMIC III dataset is a comprehensive collection of ICU data collected from wards across the United States from 2001 to 2012, with over 38,000 patient records. A sub-set of these patients are TBI and SAH (Sub-Arachnoid Haemorrhage) patients, which make them ideal candidates to standardise the output and run the guideline framework against. The structure is end-hour averaged physiological data, and the treatment data has a large coverage of the possible permutations of annotation labels strings (e.g. “CSF”, “cerebrospinal fluid”, “CS fluid”, etc). The challenge with this end-hour structure is viewing the output at a resolution that will give sufficient information to be useful in a clinical setting.

7.1.3. ICU data collection system (ICCA)

A set of annotated patient records were extracted from the Philips ICCA (IntelliSpace Critical Care and Anaesthesia) system. The intention of using this dataset was to make use of state-of-the-art ICU technology, in common use in

modern neurological ICUs. The physiological data structure is millisecond wave-form, down-sampled to minute-by-minute for this work, whilst the treatment data is primarily ventilation support and drug administration.

An additional feature of this dataset was that it was possible to use it to evaluate the guideline adherence system against patient context information provided by domain experts. To that end, a set of physiological and treatment data was compiled for three TBI patients, by two domain experts at the Queen Elizabeth University Hospital in Glasgow, UK. They were asked to provide notes on patient context and a brief rationale behind the clinical management of those patients. Some of the patients selected had contexts that specifically (and deliberately) provided circumstances where following the BTF guidelines was not the ideal clinical option.

7.2. Implementation

The guideline adherence application has two main modules. The first is a standalone program, written in the Java programming language, that translates the raw ICU patient data into an event log. The two main classes in this program are:

- *"TreatmentAnalysis"* – this interfaces with the ICU patient data format (e.g. MS Access, MS SQL Server) and translates that format into a standardised representation of physiological and treatment information
- *"EventDetection"* – which, based on input EUSIG definitions, converts the physiological data into higher level abstraction events and stores them in the treatment profile database (algorithm and database details are described in chapter 5).

The second module is the guideline adherence calculation program, which is also coded in Java and in the web-enabled version, Java Server Page (JSP). It is an n-tier web application, connected to the standardised treatment profile database and indexed on the unique patient IDs of each dataset.

There are four components of the code that merit a brief description:

- The process model is represented as a complex object with component objects of nodes and edges. These nodes and edges are collection variables for each process model instance, with **n-1** edges where there are **n** nodes.
- The evaluation of the distance between process models is represented by composing the ideal process model, composing the corresponding actual process model for that time-point, and calculating the distance between the two.
- The string-edit and graph-edit calculations include the specific, non-abstracted information relevant to each guideline, with business logic rules to apply that information.
- Contextual input from the guidelines (e.g. check age for SBP, check mass lesion presence for ICP, check fluid/pressor load for CPP) is present in two places: one on the original event detection algorithm, and one within the individual time-point distance calculation.

The programs are run on a virtual machine hosted by the NeCTAR (*"National eresearch Collaboration Tools And Resources"*) cloud platform at the University of Melbourne, running the Ubuntu 16.04 (Xenial) amd64 operating system, on a machine with 1 VCPU, 4Gb of memory and 30Gb of storage, and the Java Virtual Machine (JVM) modified to make use of all of the available memory[14].

It is noted that the event information (EUSIG-event plus treatment annotation) is stored as formatted (re-usable) information in the treatment profiles database. Higher order information used in the process model creation (e.g. auto-regulatory status, brain/clinical CT findings, etc) are taken directly from the individual data repository (e.g. Brain-IT, MIMIC III, etc).

A web-application was chosen for ease of communication of results and to provide general (but gated) access to the research community. If converted to a bedside application, to evaluate adherence in real-time, it would be better developed as a standalone program, for reasons of security (less exposed threat surface to general Internet) and speed (programmatic access to dedicated memory with local

[14] The "JAVA_OPTS" runtime variable is set with the flag "-Xmx4096m"

response speeds, rather than communication across wider networks). However, for a non real-time audit application, a web application would still be appropriate.

The code is available for download and execution at the repository referenced in the "Additional Resources" page at the front of this thesis.

7.3. Presentation

The guideline adherence output is obtained through the presentation pages of the web-enabled application. When selecting information, several choices are available, such as dataset, individual patient ID, and various options relating to event definition (e.g. threshold level, event hold-down size, etc) and options relating to display (e.g. "show default instance", "select event ID").

7.3.1. Worked example

Repeating the core concepts from section 6.3, the basic units of adherence are two – potentially interdependent – categories: degree of non-adherence (expressed as a percentage) and the duration of these levels of non-adherence (in minutes). The adherence levels detailed at each minute time-point of a patient's stay are compiled into a set of contiguous "instances" of non-adherence. The final overall adherence information for a single patient stay is compiled from these instances of non-adherence and is presented in five sections.

7.3.1.1. Minute-by-minute output

Figure 7.2 (repeated from figure 6.4) shows the physiological read-out of a particular event in a patient's stay. The guideline adherence level is overlaid on a separate series, detailing the quantitative level of non-adherence and listing the reasons contributing to that level at each time-point. The qualitative components of that non-adherence level can be viewed by clicking upon the individual time-point on the non-adherence series.

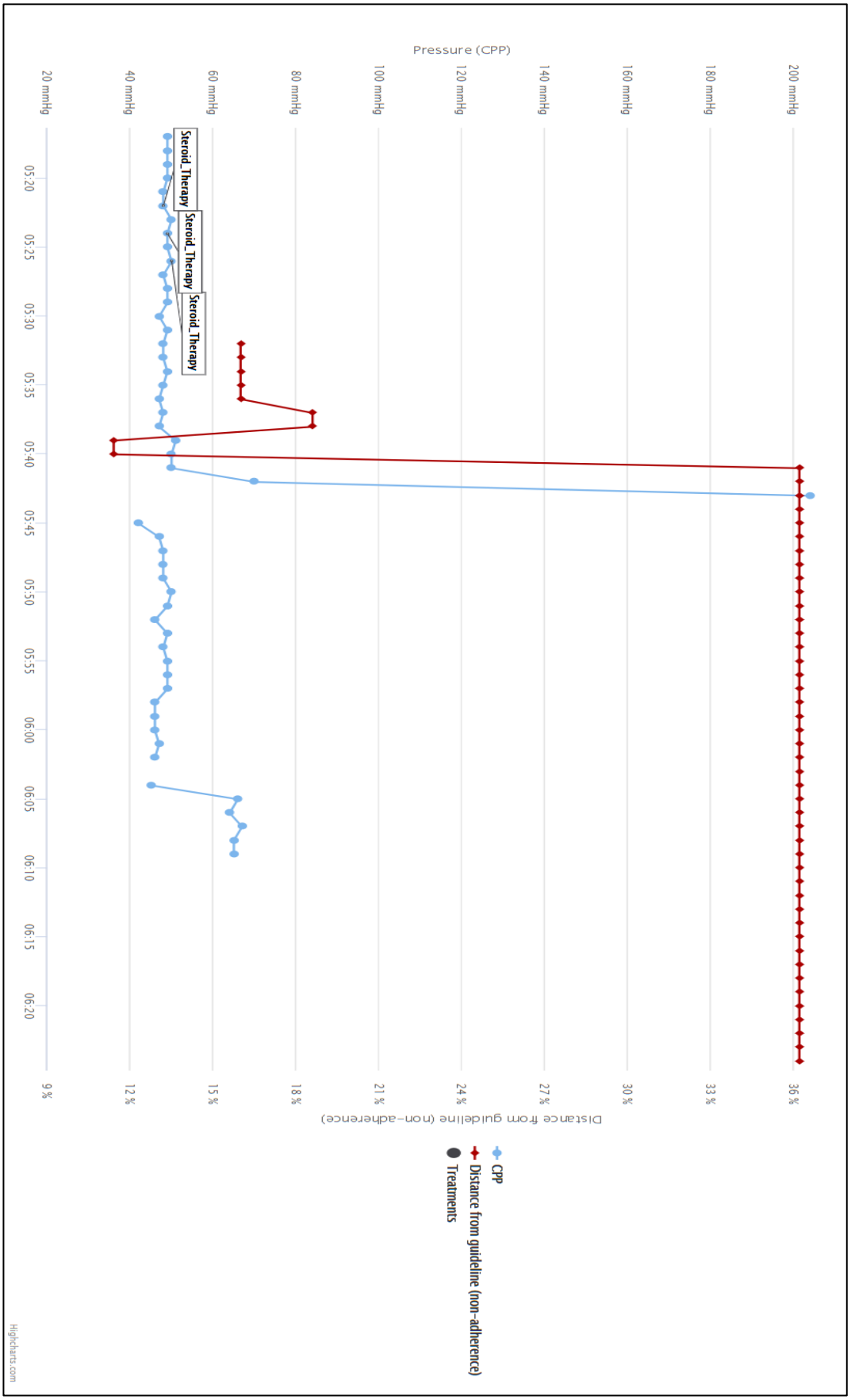

Figure 7.2: Minute-by-minute output of guideline adherence of a single CPP EUSIG-event

7.3.1.2. Instances of non-adherence

Figure 7.3 shows the total time spent at a particular (unique) level of non-adherence in a patient's stay, and the reasons contributing to that level.

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
2	11.4	Time to treatment (11.4)
5	16	Type (10.0), Time to treatment (6.0)
2	18.6	Type (10.0), Time to treatment (8.6)
68	36.2	Treatment missing (36.2)

Figure 7.3: Aggregated instances of non-adherence (for worked example in section 6.3)

7.3.1.3. Contributing reasons of non-adherence

The reasons listed in the previous two sections are shown in a stacked bar-chart (figure 7.4) to visualise the contributing weight of each.

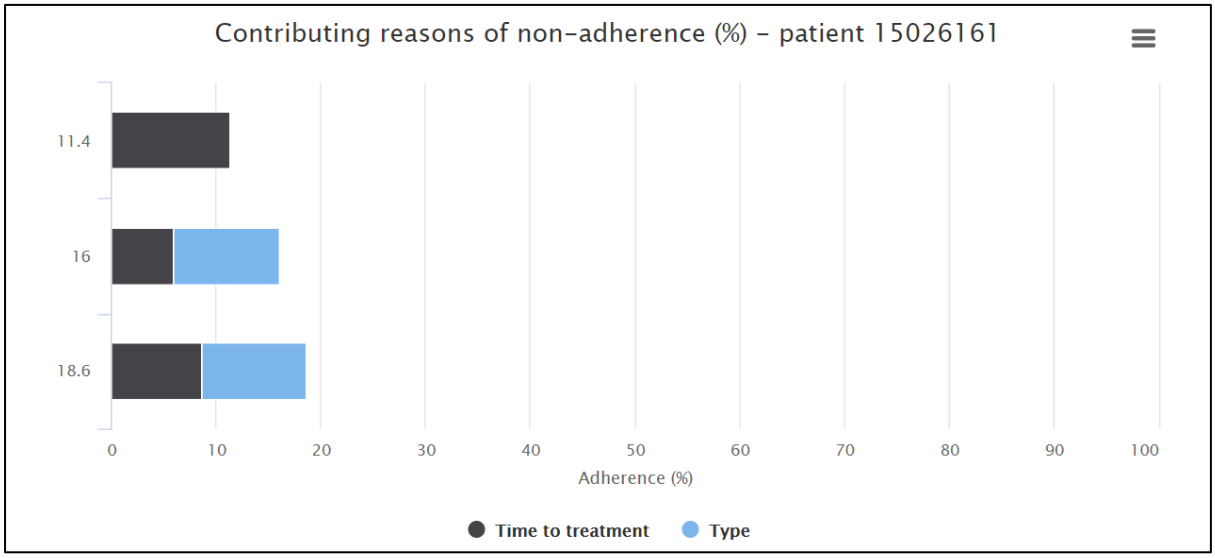

Figure 7.4: Stacked bar chart visualising the relative contributions of each reason to each instance level (for worked example in section 6.3)

7.3.1.4. Representative distribution of non-adherence instances

The Interquartile range of each adherence aspect is measured and shown in figure 7.5. The minimum, q1, mean/median, q3 and maximum quartiles are displayed (along with visualised box-plot in figure 7.6) of the levels of non-adherence, the durations of non-adherence, and two combinatorial metrics of the level and duration: **"A"** which equals **(duration / level)**, and **"B"** which equals **(duration ***

level). Note that the large "default" instance has been removed from figures 7.5 and 7.6 to show the variation in more detail.

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	11.4	16	15.56	16	17.3	18.6
Duration (mins)	2.0	2.0	3	2.0	3.5	5.0
Duration / Non-adherence	0.11	0.14	0.2	0.18	0.24	0.31
Duration * Non-adherence	22.8	30	46.67	37.2	58.6	80

Figure 7.5: An interquartile range of non-adherence for worked example in section 6.3, with default instance removed

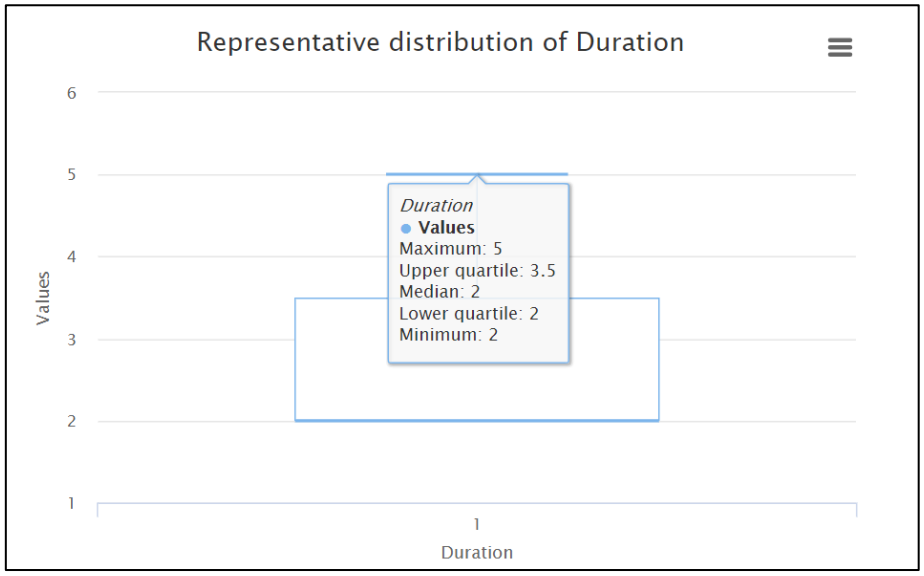

Figure 7.6: Box-plot representation of the interquartile range of the "Duration" aspect shown in figure 7.5

7.3.1.5. Severity chart

In an attempt to apply real clinical interpretation to the guideline non-adherence output, the factors of non-adherence level and duration must be considered in combination. Therefore, as mentioned in section 7.3.1.4, two metrics are defined to indicate this relationship:

- Duration / Non-adherence (A)
- Duration * Non-adherence (B)

To represent this relationship visually, a severity chart is presented that plots both metrics (the mean values of the metrics for all the non-adherence instances) against each other (figure 7.7 for the worked example in section 6.3).

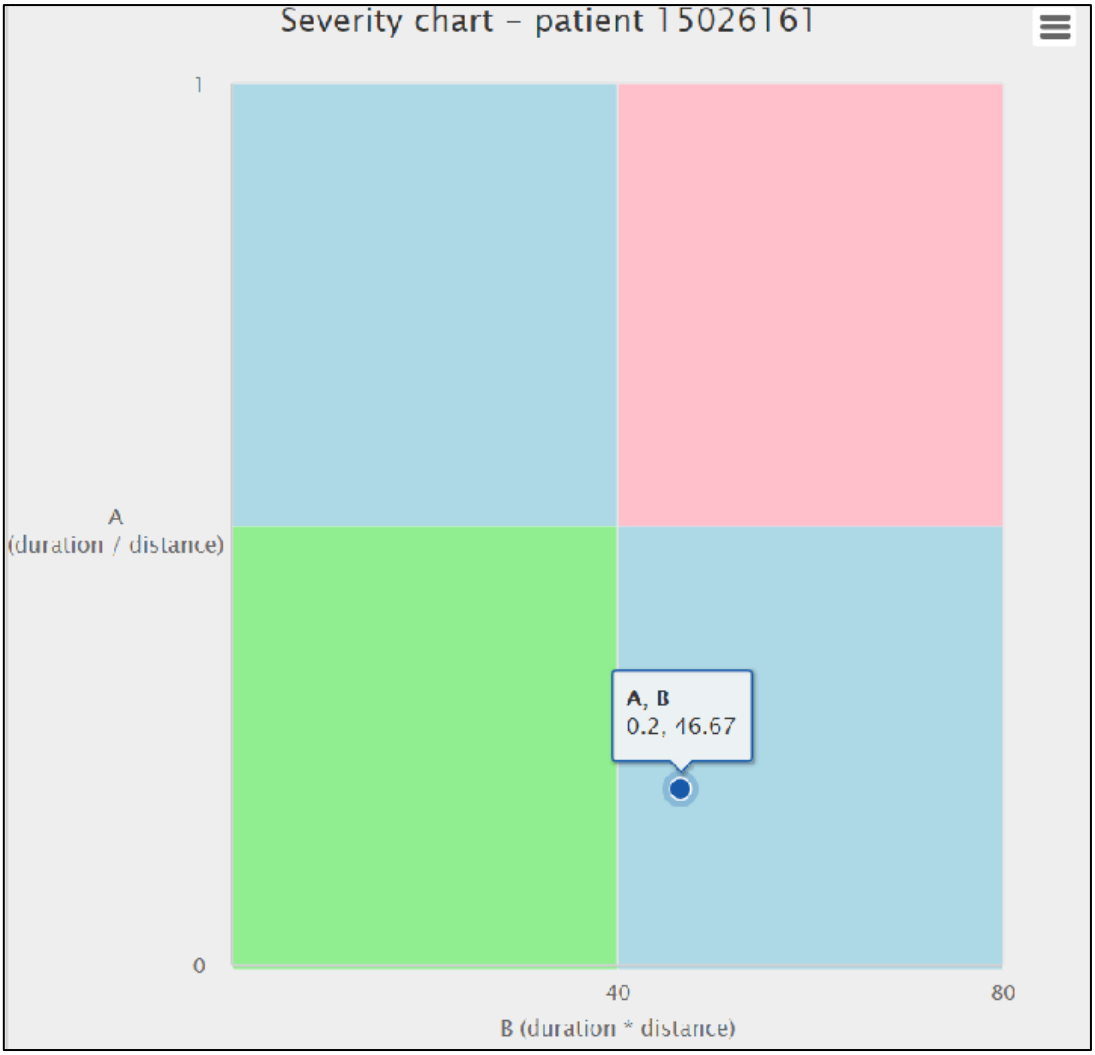

Figure 7.7: Mean values of the two metrics (A and B) plotted against each other in the severity chart

The clinical analogue of these combinations is that if A is very high or very low, the severity occupies either of the two mid-range quadrants. If A tends to 1, then it is either in the least or most significant quadrants. To ascertain which of these latter quadrants the output occupies, B indicates either high (most significant) or low (least significant). The axes are calculated by taking the maximum and minimum values from the set of non-adherence instances output by this patient's stay (the A metric normalised to lie between 0 and 1). The thresholds are calculated by taking the mid-point of the maximum absolute range of the two contributing factors – duration and non-adherence distance – for that patient. Refining where these thresholds should be placed would be follow-up work (discussed further in chapter 11). This method of presentation has been drawn from the domains of business process modelling and risk analysis (Magic Quadrant, 2018) and is hoped to provide

a single point of useful, clinical interpretation (i.e. in one representation, how severe has the impact of guideline non-adherence been – high, low or mid-range) over the distributed output.[15]

[15] Since the viva examination of this thesis, the rationale for the severity chart has changed to supporting a linear scale (measuring metric B only). See Appendix F.2 for further details.

8. Evaluation

Chapter summary

This chapter details the methods and results of four evaluations, which were conducted to assess the performance of the framework against key clinical aspects. These included:

- *A performance test of the framework's primary functions*
- *An evaluation of annotation timings on typical ICU datasets*
- *An evaluation of the framework with contextual domain expert information*
- *An investigation into the possible relationship between the adherence output and patient outcome*

***Framework function performance** – the individual components of the BTF guidelines that contribute to the overall adherence output were tested in this section. Two tests were run to confirm the contribution of two general aspects of the system:*

1) *Whether the association of a treatment annotation with a EUSIG-event (within or outside a set time-window) can provide adherence information.*
2) *Whether a breach of a limit on the number of drug administrations can provide adherence information.*

A further three tests were run to confirm the individual aspects of each separate monitoring-threshold guideline:

1) *Whether the effect of age on the BPs guideline can be detected.*
2) *Whether the effect of mass-lesion/diffuse injury presence on the ICP guideline can be detected.*
3) *Whether the effect of pressor/fluid load balance on the CPP guideline can be detected.*

The results of all the unit tests confirmed expected guideline adherence information.

***Evaluation of annotation timing** - a study was conducted to investigate the accuracy of treatment annotation timing. A set of three Traumatic Brain Injury*

(TBI) and five Sub-Arachnoid Haemhorrage (SAH) patients with annotation information recorded by a live observer were compared to the corresponding physiological information produced for them from the local bedside ICCA system. The distances from the "live" annotations to the corresponding marker in the physiological output were recorded.

This was repeated for a second set of patients (x3 TBI), which had treatment annotations produced by regular "non-live" annotations (e.g. at the end of an ICU ward shift). Similarly, the timing distances from the recorded annotations to the corresponding marker in the physiological output were recorded.

The live annotations provided highly accurate timing information: in four patients, 24 events out of 32 were closely matched, with a mean distance of 3 minutes between events and a median of 1 minute. The non-live timings had no events matched within the asserted time limit (15 minutes), and timing distances of hours when that time limit was removed.

***Evaluation with domain expert information** – guideline adherence was calculated for three TBI patient datasets extracted from the ICCA bedside system. These cases were presented alongside clinical notes that outlined the different patient contexts. The results from the adherence information showed that Patient #1 indicated non-adherence from the type of drug administered, with a spike in non-adherence due to high dosage; patient #2 indicated high severity of guideline non-adherence; patient #3 indicated a low number of associated events and treatment annotations, resulting in a mid-range severity output. When cross-referenced with patient context, all patients indicate required refinements of the framework.*

***Evaluation of relationship to patient outcome** – a unique feature of the Brain-IT dataset is the capture of 6-month GOSe score indicating patient outcome post-injury and discharge. Using this information a logistic regression was performed between the guideline adherence information and the patient 6-month GOSe. No statistically significant relationship was found.*

8.1. Unit testing

This section outlines the steps taken to make sure that each aspect of the guideline adherence calculation performed correctly given the appropriate situation. Two tests were performed which evaluated two general aspects common to all three guidelines (treatment presence and dosage level), and a further three were performed to test the unique aspects of the individual guidelines (age effect on SBP, mass lesion/diffuse injury effect on ICP and pressor/fluid load on CPP). As the aim of these tests was to identify issues with the performance of the adherence algorithm, test patient data had been created. Section 8.3 details performance of the system on real-world (unedited) datasets.

8.1.1. General tests

Two tests were run to confirm two general features of the adherence framework. The two cases are the presence of a treatment annotation within the specified time-window, and a treatment annotation either not present or outside the time-window. The anticipated outcome was variation due to the presence of the associated treatment (lower than default output) and variation due to a dosage considered too high (higher than surrounding context). Table 8.1 outlines the short-hand notation used for reporting the results: there are four possible states of two features (output #1 – the dosage – can either be H (high) or L (low), and output #2 – the presence – can either be I (in) or O (out)). Therefore, to test all use-cases, contexts of HI, HO, LI or LO are constructed and tested.

Feature	Notation
Treatment dosage too high (over x2 doses)	H (high)
Treatment dosage within range	L (low)
Treatment inside time-window	I (in)
Treatment outside time-window	O (out)

Table 8.1: Notation indicating the use-case being tested

8.1.1.1. Testing presence of annotation within and outside time-window

For this particular evaluation, treatments were artificially added to the dataset. Table 8.2 shows the base-line parameters used for most of these tests (general and

ICP). Different events were required for the BPs and CPP unit tests, which are shown in tables 8.3-8.5.

Feature	Value
Patient	**15026161**
EUSIG definition	**ICPm** with threshold of **15 mmHg**
Hold-down	**5 mins**
Time-window	**15 mins**
Event index	**0**
Event start	**01:58** on **2004-05-28**
Event end	**02:55** on **2004-05-28**

Table 8.2: Base-line features used for unit tests

Using these features, the **LI** case (dosage within range, treatment within time-window) was tested as follows:

- Add a treatment (any type) inside the time-window at **02:08** on **2004-05-28**

The result – shown in figure 8.1 – indicated a guideline adherence value of **12.6%** until the end of the time window was reached, when the adherence level goes back to a default value of **36.2%**. The 12.6% value came from the time taken to treatment, with type and nature being fully adherent in this case (and therefore contributing zero to the distance value).

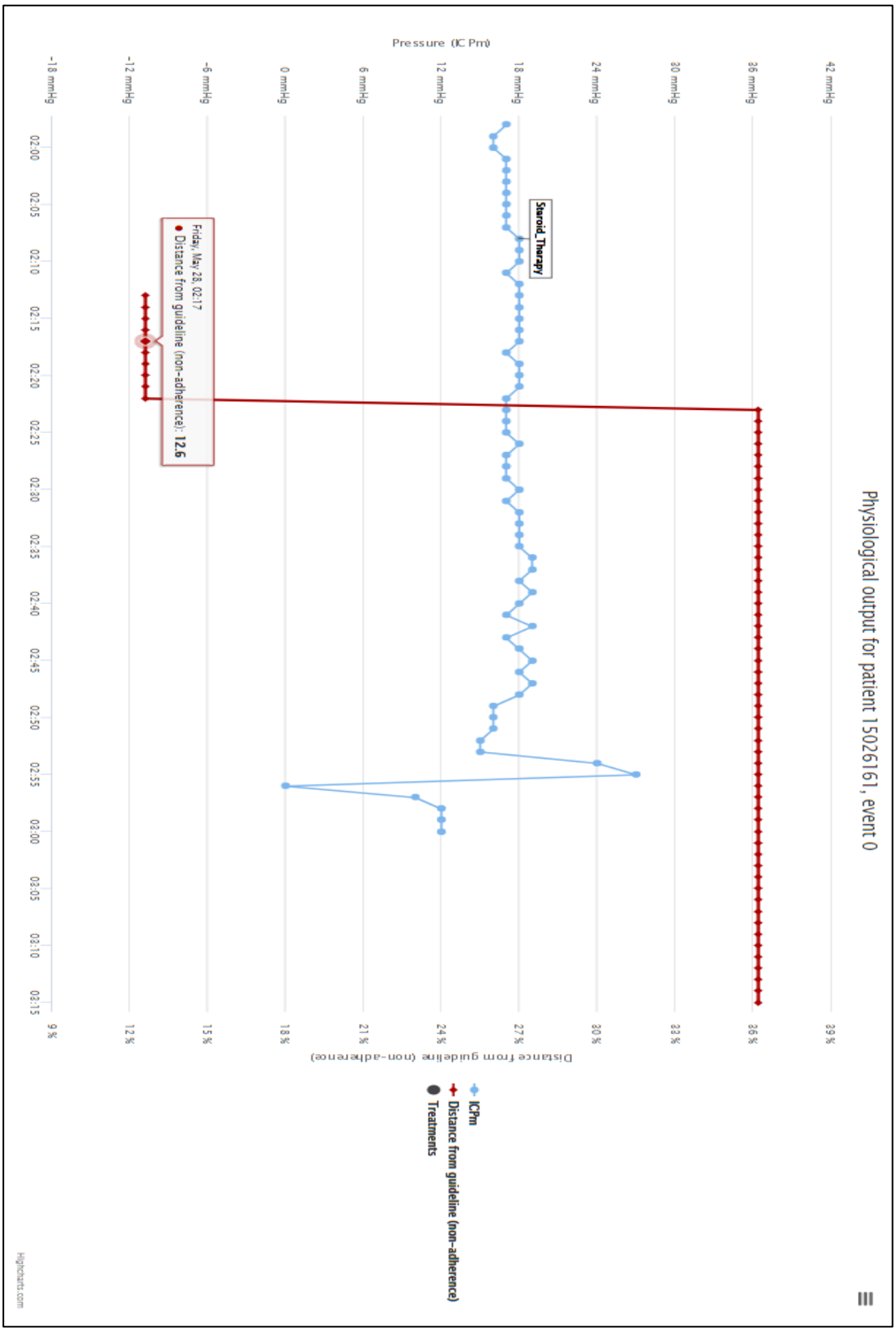

Figure 8.1: Unit test #1 - a treatment annotation occurs within the 15-minute time window since event start

The **LO** case (dosage within range, treatment outside time-window) was tested as follows:

- Add a treatment (any type) outside the time-window at **02:18** on **2004-05-28**

The result – shown in figure 8.2 – indicated a guideline adherence value of **36.2%**, which dropped to **20%** when the treatment was encountered and went back to **36.2%** once the time-window had expired (i.e. 15 minutes later). Note: the flag doesn't show for this annotation, which is a deliberate user-interface choice.

The initial high output was due to no treatment being found within the time-window since the EUSIG-event start. The drop to 20% indicated a treatment now encountered, but outside this initial time-window, which resulted in a higher adherence level than before. This reverted back to the default value as the time-window expired and no further treatment was found.

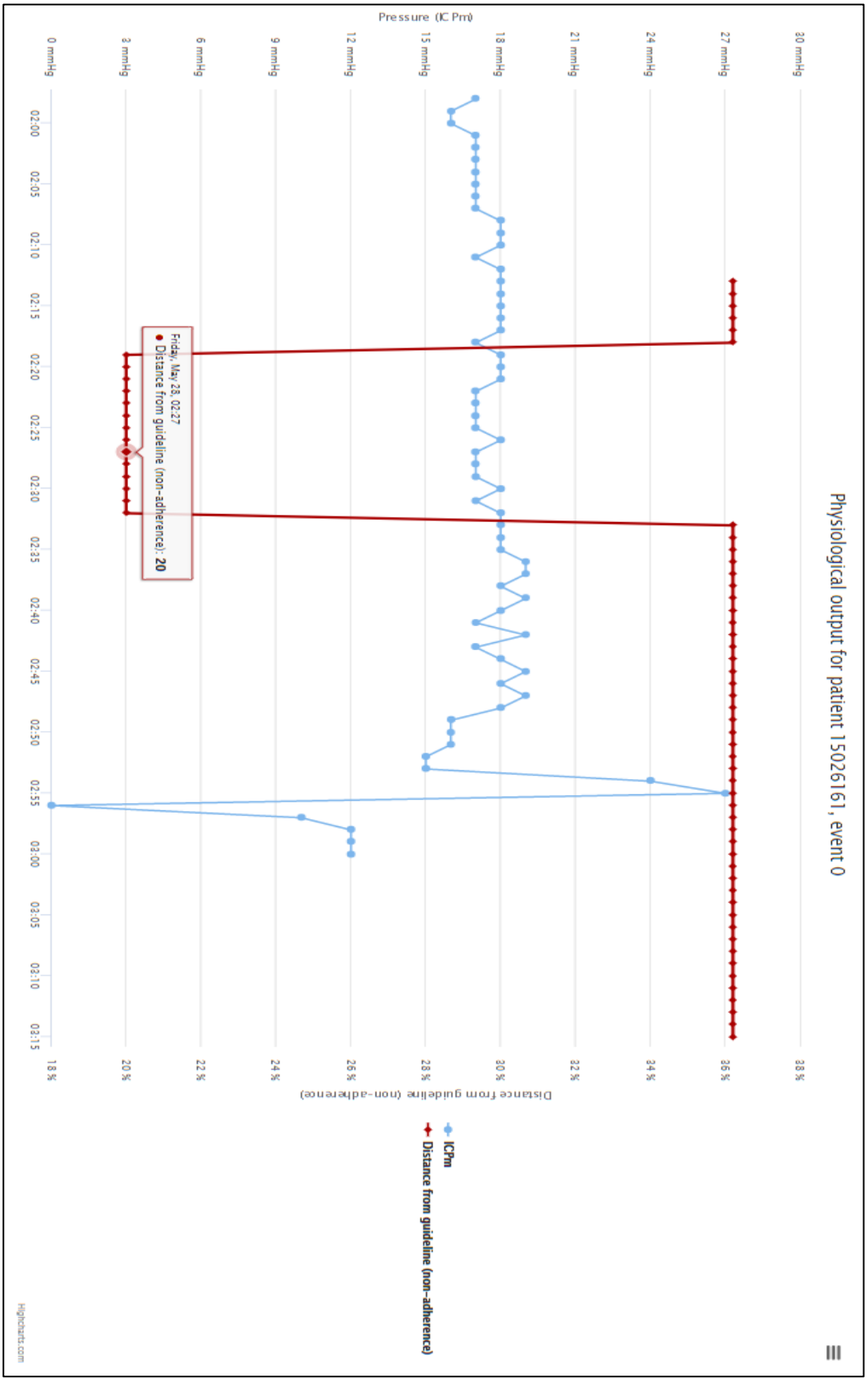

Figure 8.2: Unit test #2 – a treatment annotation (not shown) occurs outside the time window since the event start

8.1.1.2. Testing dosage administered

The **LI** case (dosage within range, treatment within time-window) had been tested in section 8.1.1.1. with the result shown in figure 8.1. This served as a benchmark against which to test the dosage/nature component of the output.

The **HI** case (dosage too high, treatment(s) within time-window) is tested as follows:

- Three treatment annotations (same type – in this case steroid therapy) are added within the time-window at **02:04** on **2004-05-28**, **02:06** on **2004-05-28**, and **02:08** on **2004-05-28**

The result – shown in figure 8.3 – showed a stepped output of adherence with contiguous instances of **7.4%**, **10%**, **12.6%** and finally the default value of **36.2%**.

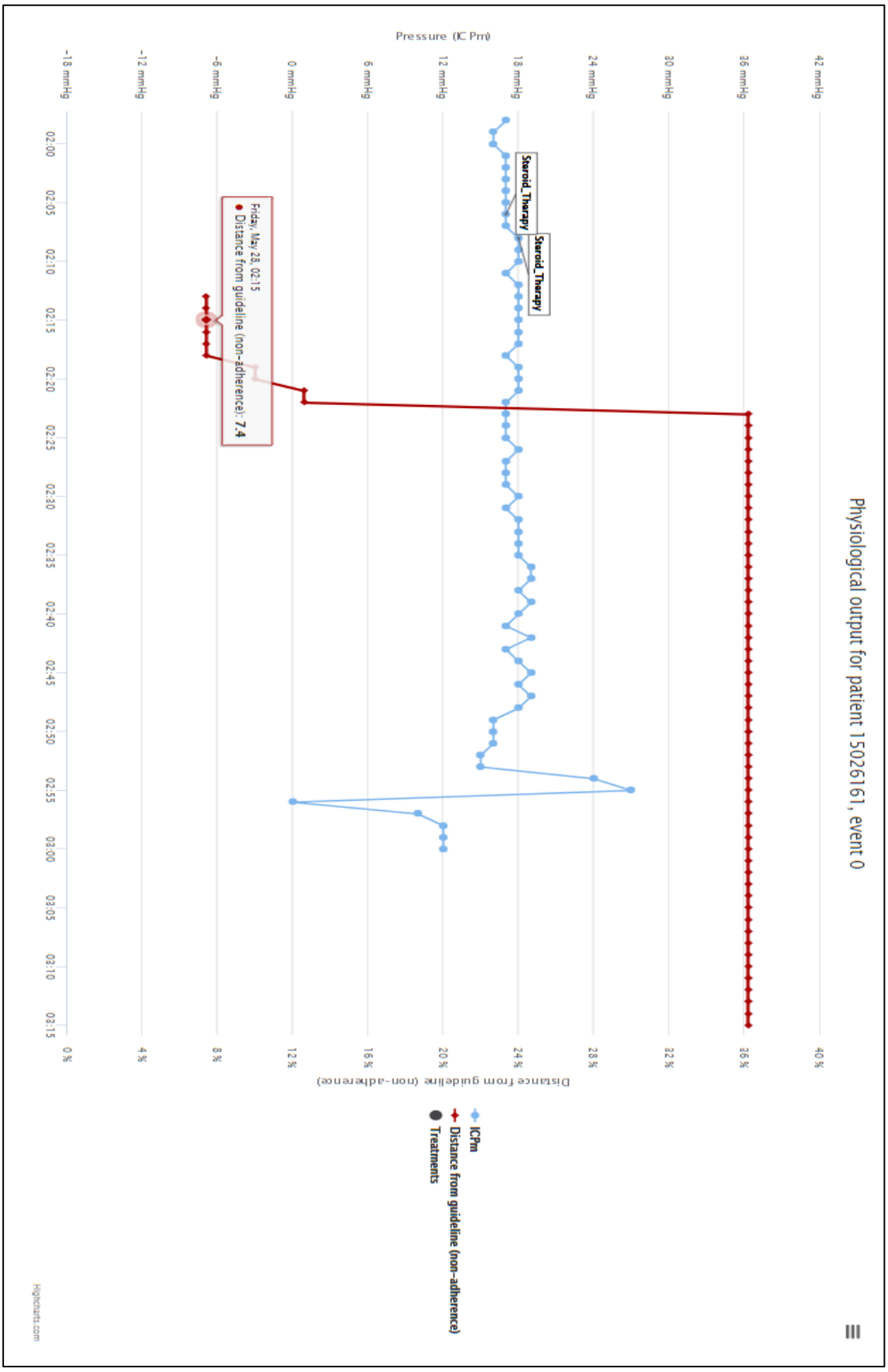

Figure 8.3: Unit test #3 - three treatment annotations within a time-window since the event start

The **HO** case (dosage too high, treatment(s) outside time-window) was tested as follows:

- Three treatment annotations (same type – in this case steroid therapy) are added, two within the time-window, one outside, at **02:08** on **2004-05-28**, **02:13** on **2004-05-28**, and **02:18** on **2004-05-28**

The result – shown in figure 8.4 – showed a stepped output of adherence with contiguous instances of **12.6%**, **19.4%**, **20%** and finally the default value of **36.2%**. Note that the annotation flag at 02:18 does not show.

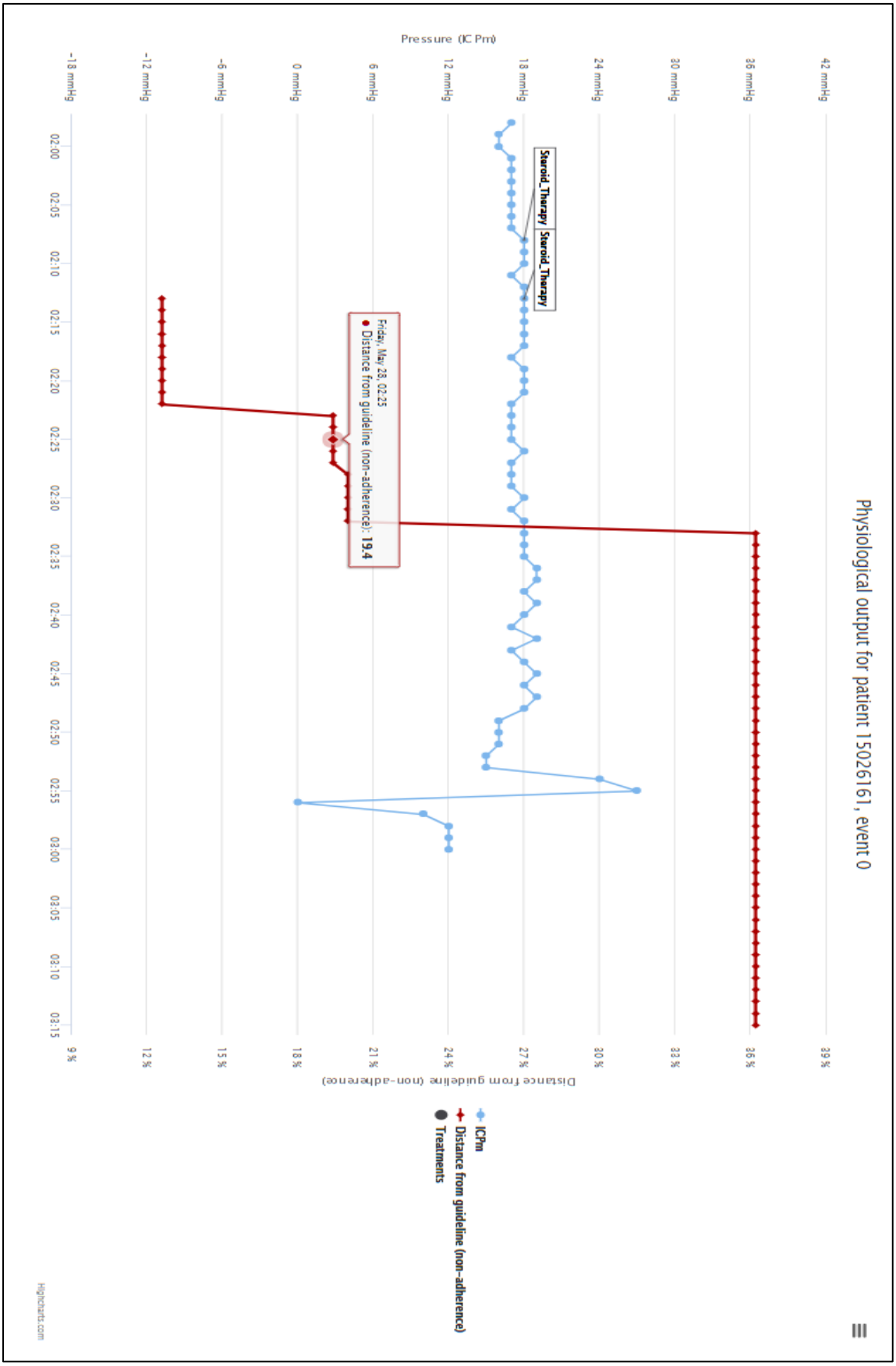

Figure 8.4: Unit test #4 - three treatment annotations, two within the time window, the third (not shown) outside the time window

The first two unit tests illustrated the difference in impact of a treatment annotation which occurred within the prescribed time window and one that did not. The latter reduced the adherence level from the default value of 36.2%, but only to 20%, whereas the former is considered more “adherent” and therefore had a level of 12%.

A similar difference in pattern occurred between tests #3 and #4: several stepped levels of adherence occurred after the first measurement (15 minutes after the event start), which were separated by the time taken from the event start to the administration of the treatment. This same stepped output was present in test #4, but the distances were greater due to the fact of the final treatment annotation being outside the time window.

8.1.2. Unit tests for individual guidelines

Each individual BTF guideline has a component unique to that guideline only. These can manifest in two ways: either in the choice of EUSIG definition to render the physiological output (e.g. depending on the patient age, a BPs EUSIG definition of 100 or 110 mmHg is selected) or in the individual distance evaluation at a specific time-point (e.g. the presence of a mass lesion in an ICP adherence reading, causing the adherence value to be different to when the lesion is not present). The following three tests (with two expected outcomes each) verify these unique components.

8.1.2.1. Systolic Blood Pressure (BPs)

The recommendations from the BTF guidelines on BPs manifest in the two physiological patterns, which result from the two EUSIG threshold definitions (100 and 110 mmHg respectively). This was tested as follows:

- Add a treatment (any type) inside the time-window after a BPs event occurring **(>110 mmHg)** at **02:22** on **2004-05-29** (event **1**). Event detail shown in table 8.3.
- Add a treatment (any type) inside the time-window after a BPs event occurring **(>100 mmHg)** at **09:02** on **2004-05-29** (event **0**). Event detail shown in table 8.4.

Feature	Value
Patient	**15026161**
EUSIG definition	**BPs** with threshold of **110 mmHg**
Hold-down	**5 mins**
Time-window	**15 mins**
Event index	**1**
Event start	**02:13** on **2004-05-29**
Event end	**03:17** on **2004-05-29**

Table 8.3: BPs definitions used for first BPs unit test

Feature	Value
Patient	**15026161**
EUSIG definition	**BPs** with threshold of **100 mmHg**
Hold-down	**5 mins**
Time-window	**15 mins**
Event index	**0**
Event start	**08:54** on **2004-05-29**
Event end	**09:22** on **2004-05-29**

Table 8.4: BPs definitions used for the second BPs unit test

The results of these tests – shown in figures 8.5 and 8.6 – indicated adherence relating to the time taken to treatment and the default values resulting outside the time window. The adherence values thus showed the impact that age has upon the guideline adherence output (albeit in temporal output only).

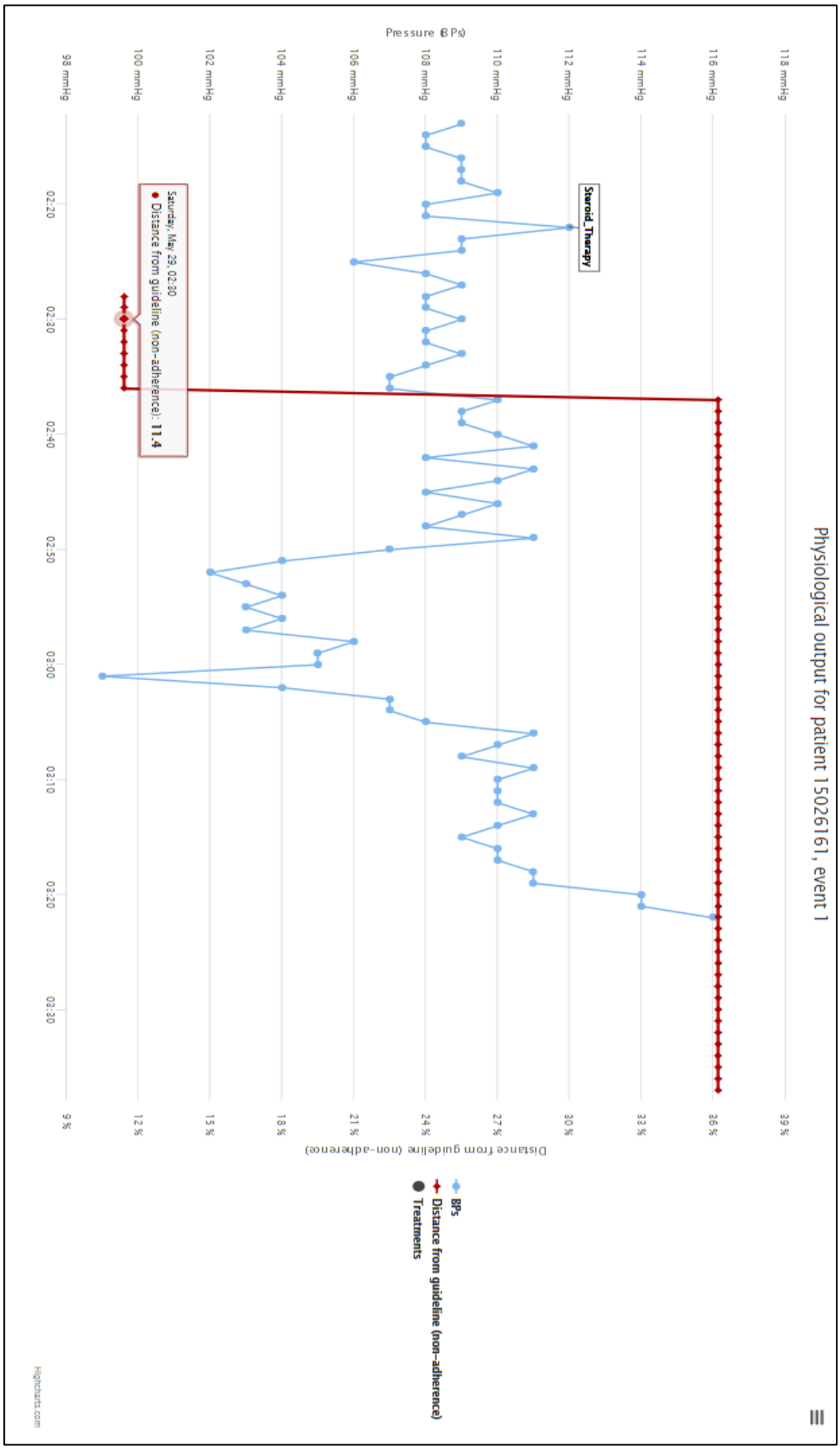

Figure 8.5: Unit test of BPs (110 mmHg threshold) guideline

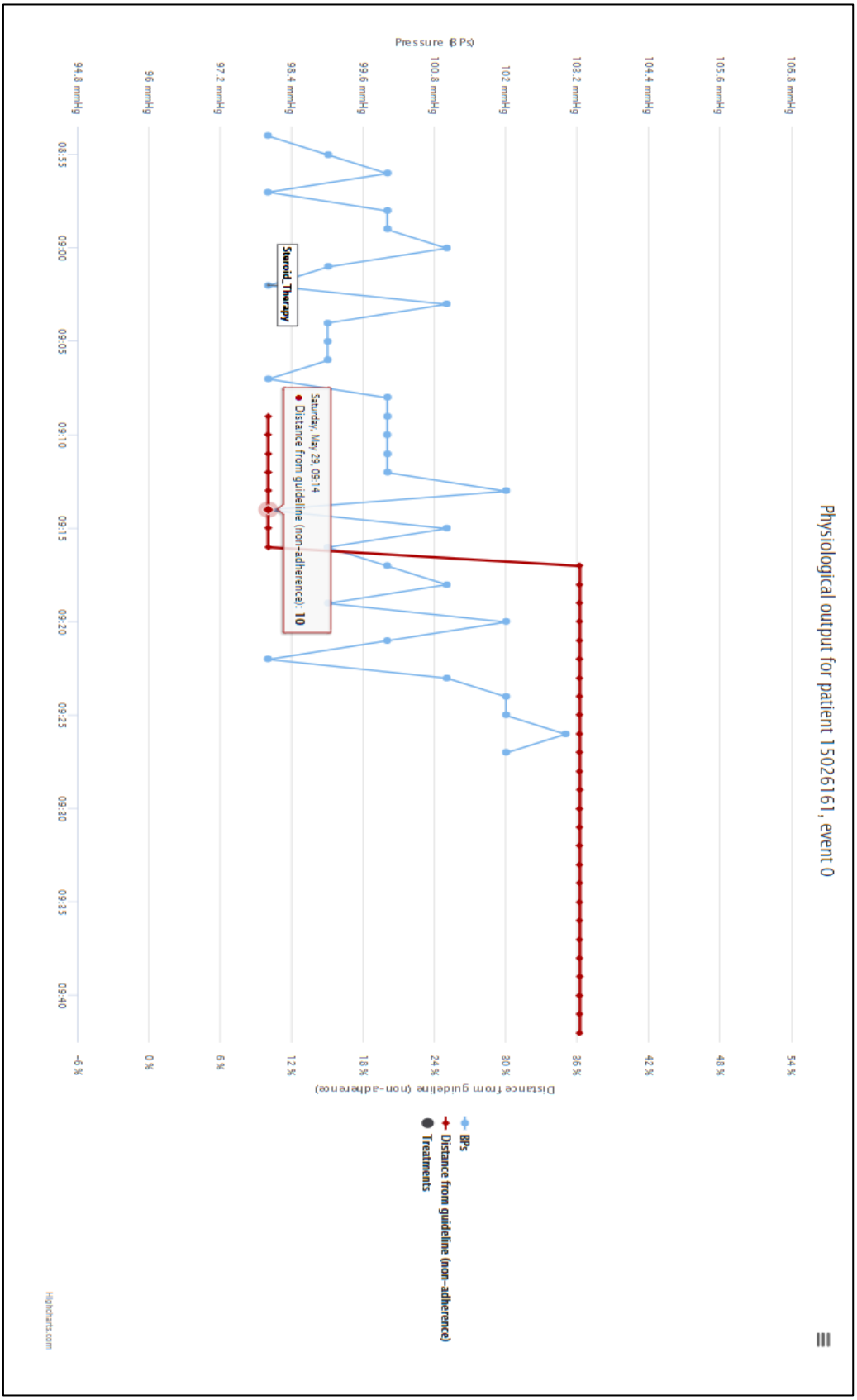

Figure 8.6: Unit test of BPs (100 mmHg threshold) guideline

8.1.2.2. Intracranial Pressure (ICP)

The unique feature of the ICP guideline from the BTF is the presence or not of a mass lesion or diffuse injury. This was tested as follows:

- Add a treatment (any type) inside the time-window at **02:08** on **2004-05-28** (event **0**) on patient **15026161**, which has a mass lesion/diffuse injury.
- Add a treatment (any type) inside the time-window at **02:08** on **2004-05-28** (event **0**) on patient **15026161** (inputs artificially modified) which has no mass lesion/diffuse injury (therefore the treatment was not necessarily merited).

The results of these tests – shown in figures 8.7 and 8.8 – indicated the standard **12.6%** distance defaulting to **36.2%** outside the time window, when the mass lesion/diffuse injury was present, as the treatment was merited in this case (figure 8.8). When the mass lesion/diffuse was not present, the distance within the time window was **22.6%**, larger as a warning that the treatment was not necessarily merited.

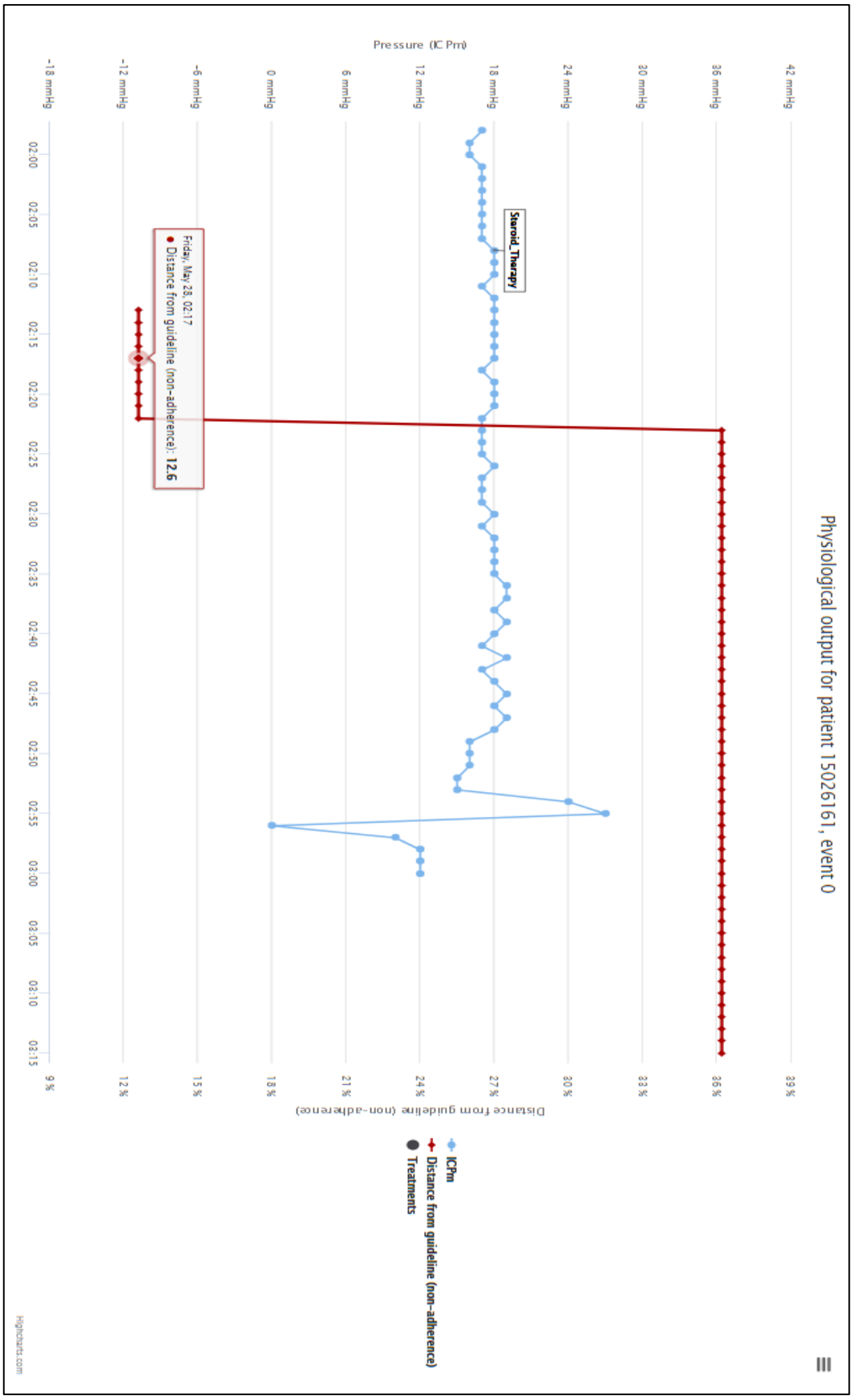

Figure 8.7: Repeat of (general) unit test #1, which shows ICPm distance within time-window in the presence of a diffuse injury

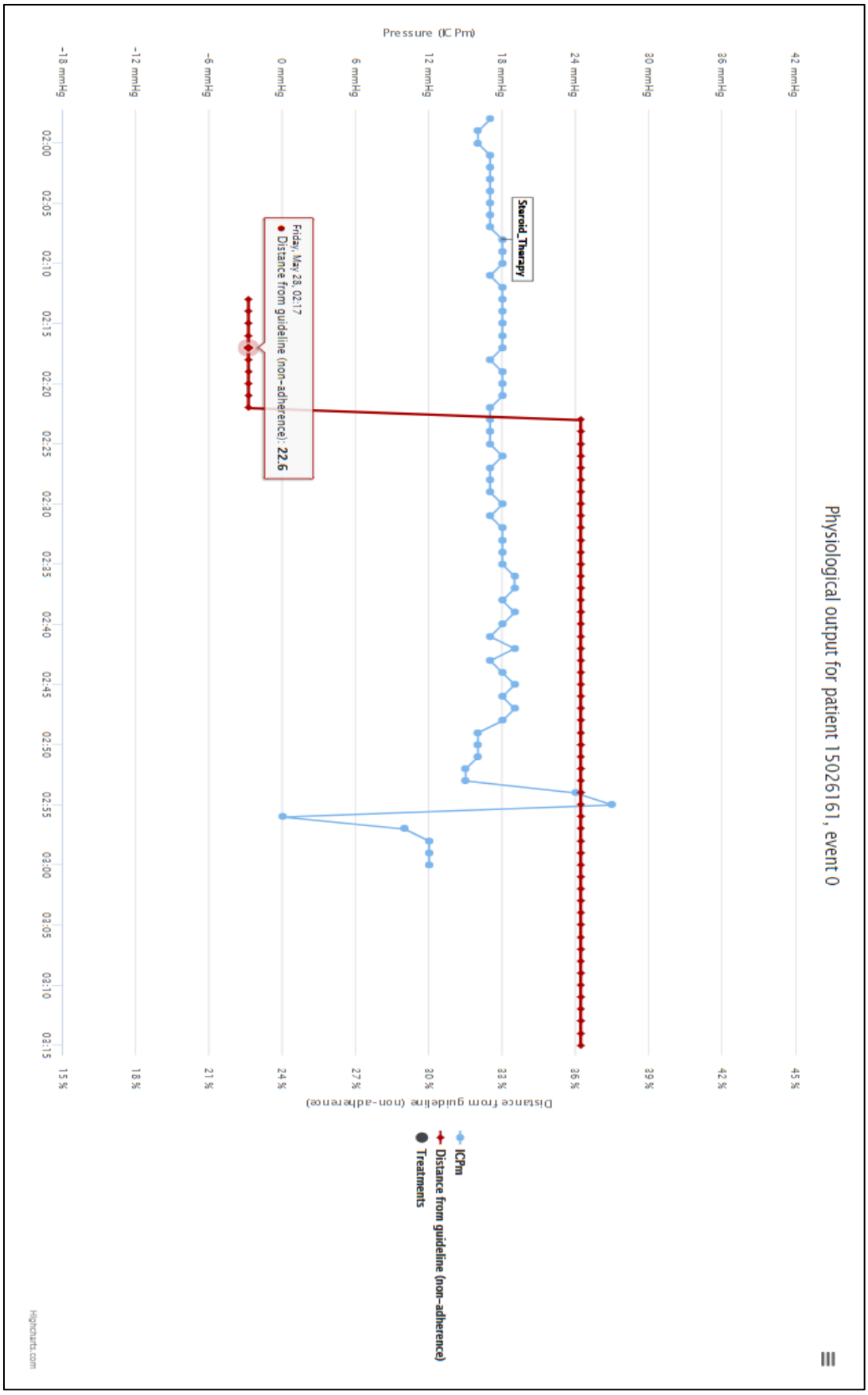

Figure 8.8: Repeat of (general) unit test #1, which shows ICPm distance within time-window without the presence of a diffuse injury

8.1.2.3. Cerebral Perfusion Pressure (CPP)

The unique feature of the CPP guideline in the BTF is the choice between administration of pressors and fluids depending on the context of patient history (the relationship is inverse, so if the pressor-load is high, fluid is recommended and vice versa). This was tested as follows:

- Add three treatments inside the time-window after CPP event 1 on patient 15026161, x2 inotropes at **05:22** and **05:24** on **2004-05-28**, and x1 of osmotic therapy at **05:26** on **2004-05-28**.
- Add three treatments inside the time-window after CPP event 1 on patient 15026161, all inotropes at **05:22**, **05:24** and **05:26** on **2004-05-28**.

Feature	Value
Patient	**15026161**
EUSIG definition	**CPP** with threshold of **50 mmHg**
Hold-down	**5 mins**
Time-window	**15 mins**
Event index	**1**
Event start	**05:17** on **2004-05-28**
Event end	**06:04** on **2004-05-28**

Table 8.5: CPP definitions used for all CPP unit tests

Note that because the treatments were all within the time-window of this event and that the maximum dosage was three, then the other factors (time-taken and nature) were controlled and only the type difference was being tested.

The results of these tests – shown in figures 8.9 and 8.10 – were contiguous instances of adherence at **16%**, **8.6%**, **11.4%** and finally reverting to the default value of **36.2%** in figure 8.9.

In figure 8.10, where pressors were administered despite the pressor-load already being high, gave contiguous instances of **16%**, **18.6%**, **11.4%** and the default value of **36.2%**. The middle level was higher due to this extra (non-recommended) administration of pressors.

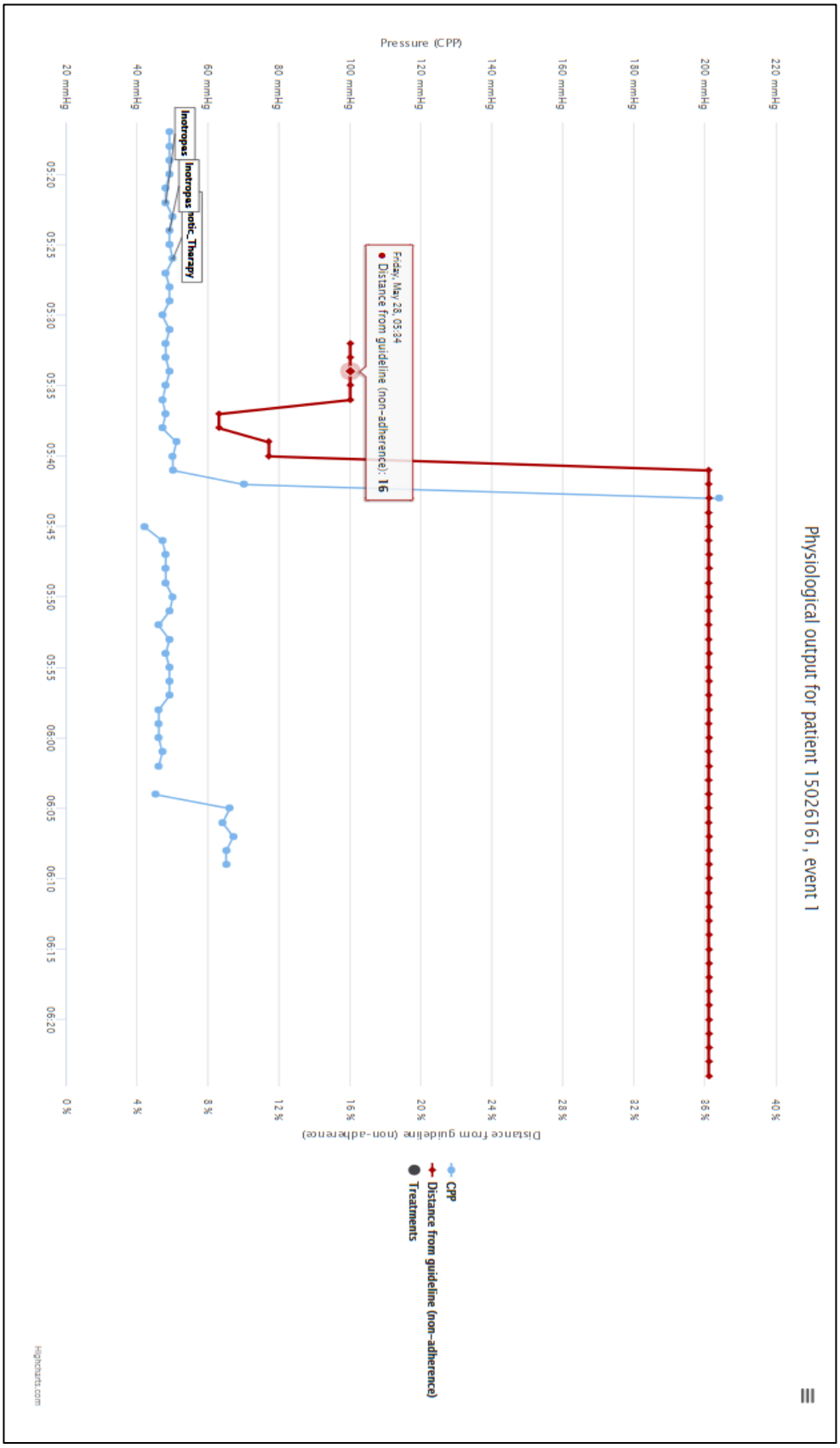

Figure 8.9: Unit test #6 – testing the pressor/fluid balance in the CPP guideline (x2 inotropes and x1 osmotic)

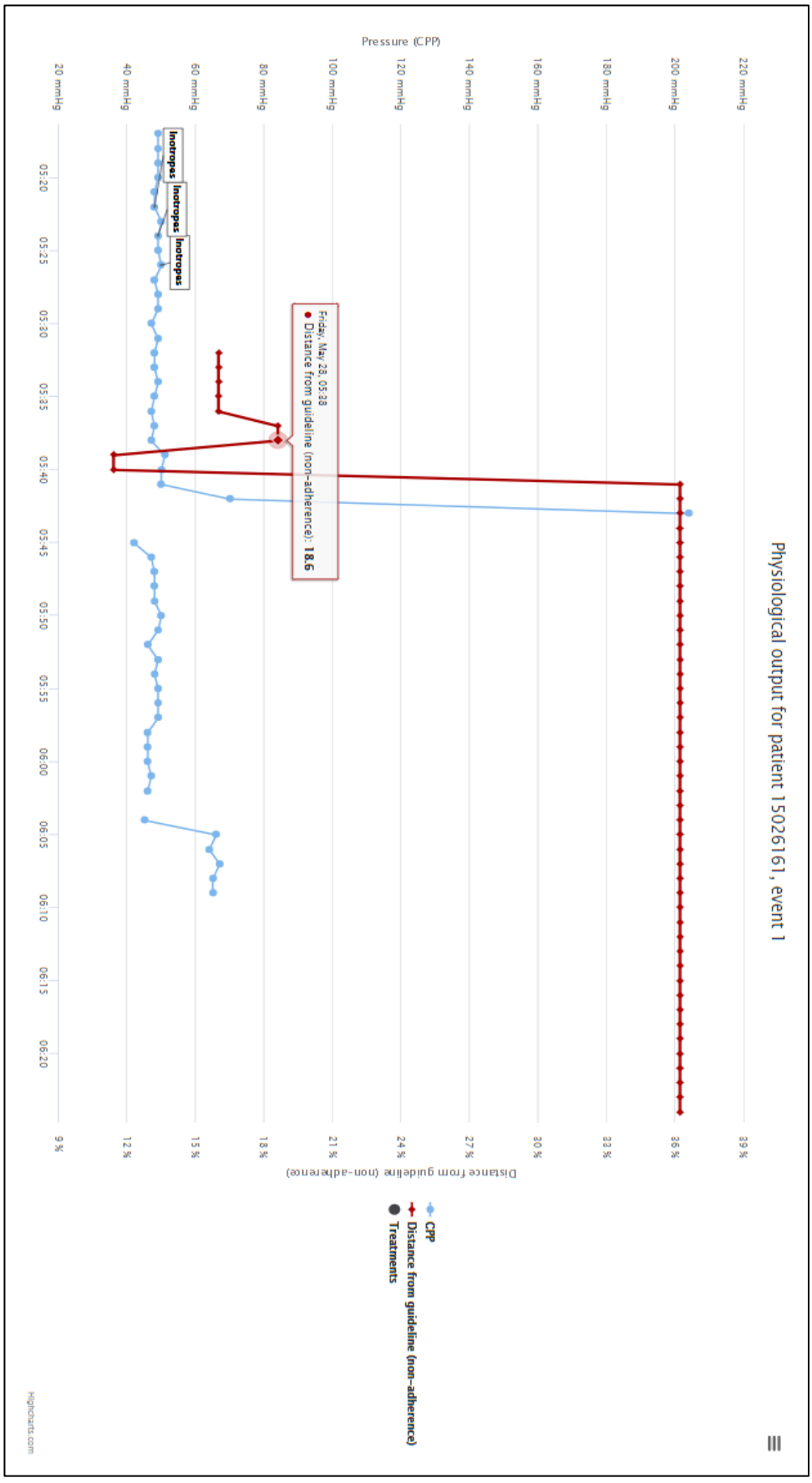

Figure 8.10: Unit test #7 – testing the pressor/fluid balance in the CPP guideline (x3 inotropes)

There is another component of the CPP guideline, not tested here, which is the value of CPPopt. This is dependent on the status of the patient's cerebral autoregulation and has an impact on the EUSIG definition of the CPP physiological reading (similar to the difference in BPs threshold definition dependent on age). This will be future work for the refinement of this system[16].

8.1.3. Discussion

These unit tests have shown the output that results when individual adherence circumstances are input to the system. Two output patterns predominate: small (but significant) variation when a treatment is encountered within the time-window since the start of a EUSIG-event, and much larger periods of "default" adherence levels when a treatment is not encountered. This binary pattern is expected but underlines the importance of accurate and plentiful treatment annotations, both for the best performance of this system, and for accurate data capture in general.

The variation that occurs within the time-window output should be the initial main focus of adherence measurement or guideline improvement. The adherence levels presented above indicate a "first cut" of the output given the assigned weightings attached to the reasons for non-adherence, and future work would include the refinement of these weightings in order to better understand the clinical priorities in a given moment (additional insight into this can be found in section 8.3 using the ICCA dataset).

[16] The methods and API for this was implemented in the code but relied upon a moving calculation of the Pearson correlation coefficient, which was not providing suitable output by the time of thesis submission.

8.2. Evaluation of treatment annotations timings

An early finding of this thesis was that the accuracy of treatment annotations, in particular the timing, was one of the most important aspects in obtaining reliable adherence output. This section describes a validation study that was conducted to establish how accurate clinical annotations are, in a real and representative clinical environment.

8.2.1. Study description and method

This validation study was conducted using a neuro-intensive ICU dataset that was compiled for a previous project, attempting to improve arterial hypotension in ICU patients by detecting and analysing artefacts in physiological data streams (Lal *et al.*, 2015). In order to identify the artefacts, a live observer had been required to monitor the management of patients in the ICU and make annotations when a clinically relevant action or event occurred. This generated a unique neurological ICU dataset which included "live" treatment annotations.

This set of "live" annotations was used to measure the difference in timing from the annotation (manually recorded) to the corresponding marker in the physiological data readouts for a given patient (e.g. ABP drops to near zero, or registers volatile output, when a blood sample is taken). The same process was repeated for a dataset that had been annotated in a "normal" ICU environment ("non-live"), to attempt to find an estimate of the difference between annotation timings in that ideal situation, and those in regular, resource-limited ones.

Therefore, the two specific research questions that were considered for this study were:

1. Given ideal conditions (a dedicated research nurse standing at the bed-space annotating events as they occur – "live") - what is the average/range of timing differences between the event annotation time and the actual event as measured from the physiological data?
2. Given "real-world" conditions where the normal bed-side nurse is annotating events when they can ("non-live") - how does the difference in timing (average/range) compare against the "live" ideal conditions?

To investigate these questions, a list of annotations consisting of treatments that could be clearly identified on a physiological output (e.g. BP sampling, BP transducer goes to zero, patient is turned/moved, etc) was drawn up. The term used to describe these events is a 'zero-drop event'. Both datasets (live and non-live annotations) were then examined and occurrences of these treatment annotations in the physiological output identified and compared against the manually observed and recorded events. The average timings between both sets of events (annotation to physiological event) was noted.

8.2.2. Results

In the "live" dataset, there were **4** patients with a total of **32** "zero-drop" events. In the "non-live" dataset, there were **3** patients with a total of **27** "zero-drop" events.

The time distance between the zero-drop in the physiological data and the treatment annotation was limited to 15 minutes (considered a reasonable time to assert that the two recordings are the same physical event). Table 8.6 shows the live dataset events in this group (**24** out of **32**). No matches were found in the non-live dataset (**0** out of **27**).

Patient ID	Timestamp	Annotation	Zero-drop start	Distance (mins)
CSO_0083	30/10/2014 21:25	Blood sample	30/10/2014 21:26	1
CSO_0083	31/10/2014 00:08	Blood sample	31/10/2014 00:09	1
CSO_0083	31/10/2014 06:04	Blood sample	31/10/2014 06:07	3
CSO_0083	31/10/2014 06:37	Zero ABP reading	31/10/2014 06:51	11
CSO_0083	31/10/2014 18:27	Zero ABP reading	31/10/2014 18:39	12
CSO_0083	31/10/2014 18:37	Blood sample	31/10/2014 18:39	2

CSO_0083	31/10/2014 20:49	Zero ABP reading	31/10/2014 20:50	1
CSO_0083	01/11/2014 06:25	Blood sample	01/11/2014 06:37	12
CSO_0083	01/11/2014 06:37	Zero ABP reading	01/11/2014 06:38	1
CSO_0083	01/11/2014 07:00	Zero ABP reading	01/11/2014 07:01	1
CSO_0086	04/11/2014 07:46	Zero ABP reading	04/11/2014 07:47	1
CSO_0086	04/11/2014 14:35	Blood sample	04/11/2014 14:36	1
CSO_0086	03/11/2014 08:50	Blood sample	03/11/2014 08:51	1
CSO_0086	03/11/2014 16:42	Blood sample	03/11/2014 16:43	1
CSO_0086	04/11/2014 06:21	Blood sample	04/11/2014 06:22	1
CSO_0112	27/11/2014 14:45	Zero ABP reading	27/11/2014 14:59	14
CSO_0112	27/11/2014 14:57	Blood sample	27/11/2014 14:59	2
CSO_0115	01/12/2014 19:01	Blood sample	01/12/2014 19:02	1
CSO_0115	02/12/2014 02:09	Blood sample	02/12/2014 02:10	1
CSO_0115	02/12/2014 06:22	Blood sample	02/12/2014 06:23	1
CSO_0115	02/12/2014 20:53	Blood sample	02/12/2014 20:54	1
CSO_0115	03/12/2014 00:02	Blood sample	03/12/2014 00:03	1

CSO_0115	03/12/2014 04:18	Blood sample	03/12/2014 04:19	1
CSO_0115	03/12/2014 06:41	Blood sample	03/12/2014 06:43	1

Table 8.6: Live dataset events

Total zero-drop events with no corresponding annotation found were: **27** (non-live), **8** (live).

Table 8.7 shows the average time distance (mean and median) for recorded annotation time to the nearest event detected in the live physiological dataset.

Dataset	Annotation number	Mean distance (mins)	Median distance (mins)
Live	24	3	1

Table 8.7: Live average time distance for all recorded zero-drop events to first-encountered annotation

8.2.3. Discussion

8.2.3.1. Data processing issues

As the study concerned real-world data, there were several issues encountered during processing:

- The "zero drops" were not always easy to identify. For instance, there had to be a robust definition of "zero", as many machine outputs give readings of high volatility, rather than a simple disconnected value (e.g. high negative values, or "close to" zero).
- Similar to the problem of association of events with treatments, it is not always clear which zero drop relates to which annotation. The correspondence in the live annotated dataset was so high that this problem was minimal in that dataset but became much more of an issue in the non-live dataset (to the point where clinical sense of the output was effectively lost). A mitigation for this was to set the time-window cap mentioned previously (in this case 15 minutes).
- The physiological data is provided in waveform format, thus gives readings per milliseconds. This was originally condensed to minute-by-minute (similar

to the resolution of the Brain-IT dataset) but it became obvious that zero-drop events often have a duration lower than a minute, and the resolution was raised to second-by-second. This sampling frequency was deemed to be the most relevant to the problem space (e.g. blood sampling would never take less than 1 second), but the choice of resolution may still require future consideration.

8.2.3.2. Interpretation

The goal of this study was to use a unique feature of the CSO neuro-intensive dataset, to provide some additional insight into the timing accuracy of treatment annotations in a neurological ICU.

The results show that when observed in real-time, the accuracy of timing does correspond well to the physical output of the physiological data at the bedside. However, the comparison against the non-live dataset was inconclusive. With no matching annotations at all, it is possible that the discrepancy between manual recording of annotations and the physiological output was so great, that none showed in this small analysis. To clarify, it should be noted that even though these annotations are performed at the end of a ward shift, the distance is from the time recorded in the annotation (at some point during the ward shift), and not the actual time of recording. Therefore, the discrepancies occur due to the mis-remembering of times, rather than a systematic error. As this analysis should theoretically match between hours, it is unrealistic to think that these discrepancies are all due to this issue (or that shift nurses would get the timings so wrong), therefore the analysis should be performed again to verify this.

Another possibility in this result is the approach to annotations in this particular sample set: it may be the case the events selected are recorded once per ward shift by agreement between staff.

The high accuracy correspondence of the live observer is obviously very useful but also inefficient when considering the resources required (24 - high-skill - man-hours per day). The more immediate benefits would be to pursue similar accuracies through automated means, such as the automated tracking of drug

infusions, ventilation application or thermal cooling through “smart” blankets and beds.

A comparative timing measure was unable to be obtained in this study, so an understanding of the “average” timing errors was therefore not possible. However, if the study were repeated and a value obtained for the non-live dataset, this could be incorporated into the confidence with which an adherence pattern obtained from the framework can be interpreted.

8.3. Evaluation of framework with domain experts

The purpose of this section was to verify the framework against contextual information from clinical domain experts, who work in neurological ICUs and would be the target users of the system in its production form. A dataset collected from the Philips ICCA (IntelliSpace Critical Care and Anaesthesia) system was chosen for this evaluation as it is a popular bedside patient management software commonly used in intensive care units in the UK and is familiar to the domain experts.

8.3.1. Method

Three TBI patients were selected for analysis by the domain experts, with the following data characteristics:

- A prevalence of EUSIG events in the ICP output
- Active management of ICP required
- Two patients (#1 and #3) required non-intervention due to the nature of the ICP events
- One patient (#2) had one large refractory event throughout the course of their stay

Clinical management of the patients was carried out in accordance with a number of relevant treatment protocols and guidelines (not just BTF) which deliberately tested the ability of this system to provide useful guidance despite competing clinical priorities and possible co-morbidities.

As already mentioned in section 7.1.3, the physiological data structure of the ICP data in the ICCA dataset was millisecond wave-form, down-sampled to minute-by-minute for this evaluation. A code modification of the system was required in the *"TreatmentAnalysis"* module to perform this conversion. This was achieved by averaging the contributing millisecond readings across the corresponding minute, which would incur some loss of precision in the recorded variation across that minute (considered acceptable at this resolution). The treatment data was primarily ventilation support and drug administration, obtained by manual inputs to, and then drawn from, the integrated ICCA system.

The results of guideline adherence output from these patients are presented in three sections:

1) An audit of overall counts of EUSIG-events and treatment annotations
2) Overall adherence measures for all three patients, according to the presentation features described in section 7.3.1, including non-adherence instance tables; a chart of contributing reasons; an interquartile range of instances; a severity chart
3) Individual instances of non-adherence shown in a time-varying chart (section 7.3.1.1).

The detailed clinical notes for the individual patients were as follows[17]:

- ***Patient 1*** *- "Infusions of propofol, morphine, midazolam, nor-adrenaline; Repeat CT scan on 12/07/2017 - decision to stop sedation, disconnect ICP and assess; Repeat CT scan performed - no surgical options; Decision that as ICP not controlled by medical management - remove ICP monitor, stop sedation and assess"*
- ***Patient 2*** *- "ICP consistently >20, overall upward trajectory of ICP (despite infusions of Propofol 2% 400 mg/hr; Morphine 3 mg/hr; Midazolam 13 mg/hr 11:00; Cisatracurium 30 mg/hr; Thiopentone 125 mg/hr 14:00; Nor-adrenaline 0.1 mg/hr 13:34 - increased to 0.2 mg/hr at 14:00"*
- ***Patient 3*** *- "ICP > 20; Associated with rise in ETCO2; Optimisation of ventilation by increasing pressure support (documented at 0900 15/12/2016); Decrease in CO2 leads to decrease in ICP"* [Therefore non-intervention was recommended as the ICP increase was expected to be transient]

8.3.2. Audit: event and treatment counts

The summary numbers in table 8.8 were compiled from the database once the ICCA sample dataset had been processed into the *"treatment_profiles"* database.

Total EUSIG event number: **21**

[17] These are reproduced verbatim except the square brackets in patient 3, which were added as a separate note by one of the clinicians to provide added clarity.

Threshold value (mmHg)	Hold-down value	Count
10	20	3
15	20	10
20	20	2
25	20	3
30	20	3

Table 8.8: Count of individual ICPm EUSIG events from the ICCA sample dataset

Total treatment annotation number: **1721**

Treatment	Count
Ventilation	406
Propofol	119
Noradrenaline	234
Co-Amoxiclav	12
Morphine	388
Potassium Chloride	37
Carbomer Ointment	13
Hydrocortisone	1
Salbutamol	28
Magnesium Sulphate	11
Calcium Gluconate	2
Ranitidine	13
Omeprazole	13
Midazolam	55
Mannitol	1
Glycophos	75
Metoclopramide	14
Benzylpenicillin	25
Temocillin	8
Amoxicillin	12
Paracetamol	54
Lactulose	19
Senna	19
Phosphate-Sandoz	21
Carbocisteine	27

Furosemide	1
Metronidazole	27
Clindamycin	27
Laxido PO/NG	7
Enoxaparin	3
Ceftriaxone	5
Sando-K	17
Meropenem	18
Clotrimazole	9

Table 8.9: ICCA treatment counts

For each time-window value, **8** EUSIG events were associated with treatments, which was **80%** of the most numerous definition (an ICP threshold value of 15 mmHg for a hold-down time of 20 minutes, which was the only hold-down definition available in this sample set – see discussion below). This suggests that the count of annotations was high relative to the physiological output (c.f. to the numbers presented for the older Brain-IT dataset in chapter 9).

8.3.3. Overall adherence measures

For the three patients the results are presented as:

- Table of non-adherence instances
- Charts of contributing reasons
- Interquartile range table and box-plots ("duration with default state removed" as example)
- Severity charts

8.3.3.1. Patient #1

Most instances for patient #1 are variation due to the time to treatment and a contributing factor of incorrect type. However, for two instances, the dosage/nature is a contributing factor as well. The majority of the non-adherence is spent in the default state (36.2%) and a lower, but similar value of 30%. The distribution appears to be spread evenly through all four factors

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
2	12	Type (10.0), Time to treatment (2.0)
2	14.6	Type (10.0), Time to treatment (4.6)
1	16	Type (10.0), Time to treatment (6.0)
6	17.4	Type (10.0), Time to treatment (7.4)
4	21.4	Type (10.0), Time to treatment (11.4)
5	28	Type (10.0), Time to treatment (18.0)
400	30	Type (10.0), Time to treatment (20.0)
498	36.2	Treatment missing (36.2)
11	48.8	Type (10.0), Time to treatment (20.0), Dose (18.8)
12	49	Type (10.0), Time to treatment (20.0), Dose (19.0)

Figure 8.11: Instances of non-adherence for Patient #1

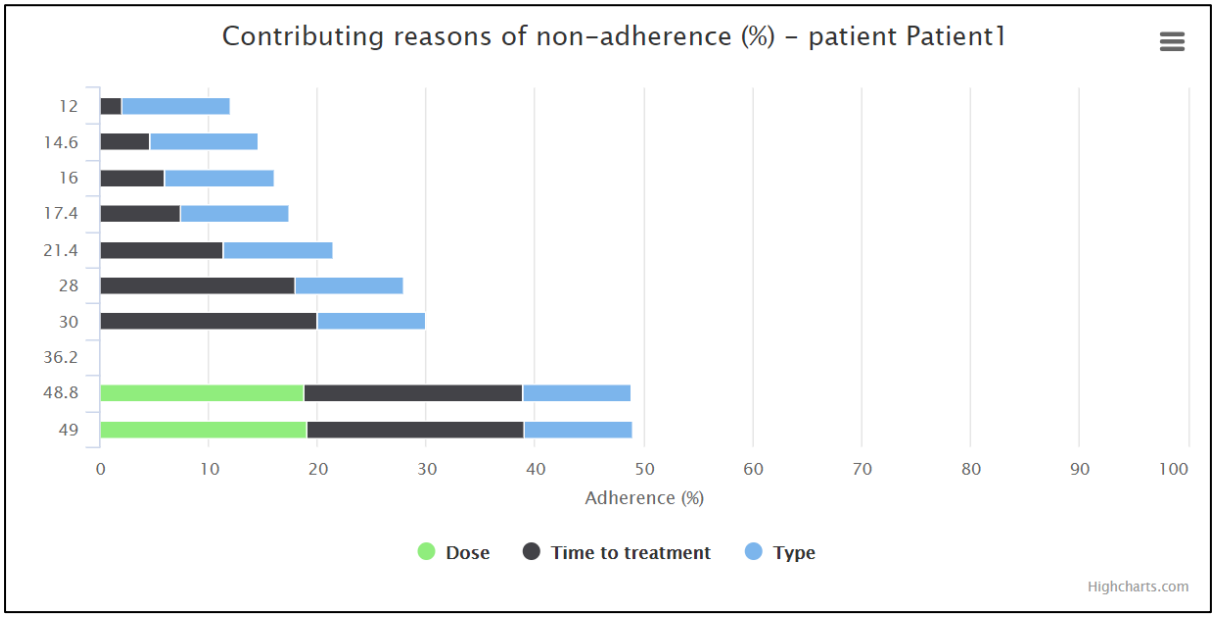

Figure 8.12: Reasons contributing to the different instances of non-adherence for Patient #1

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	12	30	33.53	36.2	36.2	49
Duration (mins)	1.0	5.0	15.95	14.0	20.0	47.0
Duration / Non-adherence	0.02	0.15	0.49	0.47	0.63	1.33
Duration * Non-adherence	16	181	534.77	420	660	1701.4

Figure 8.13: Interquartile range table for instances of non-adherence for Patient #1 with the default state (36.2%) removed

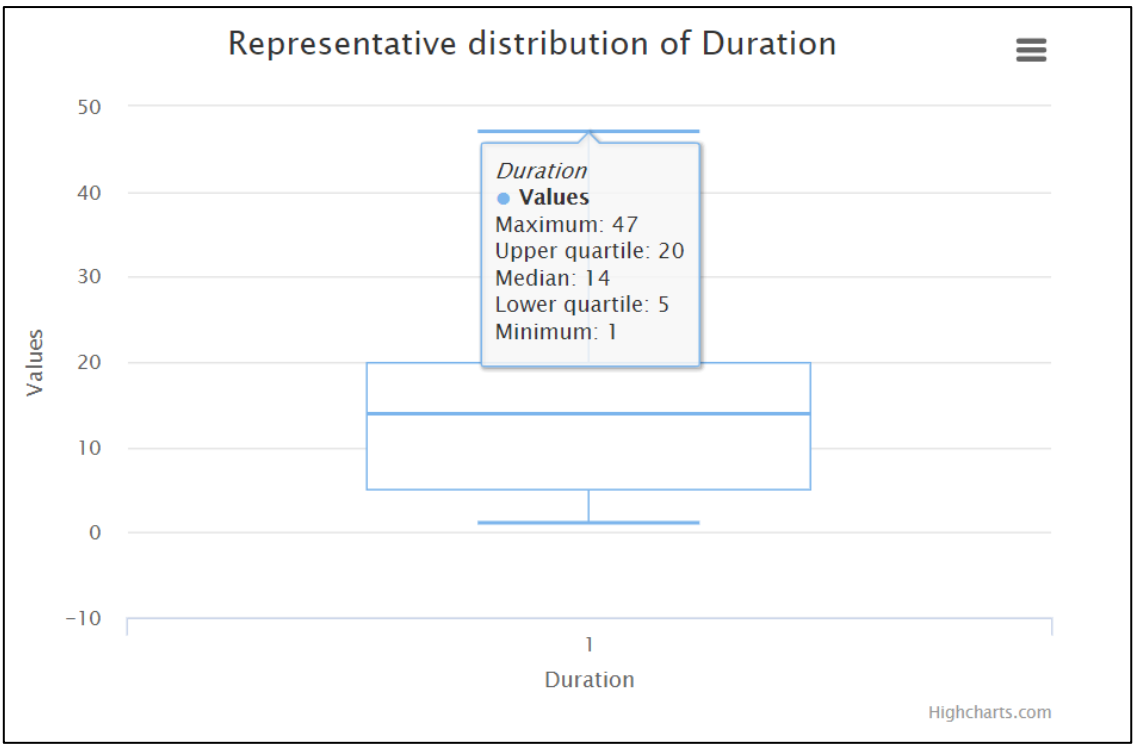

Figure 8.14: Box-plot of the duration aspect of non-adherence

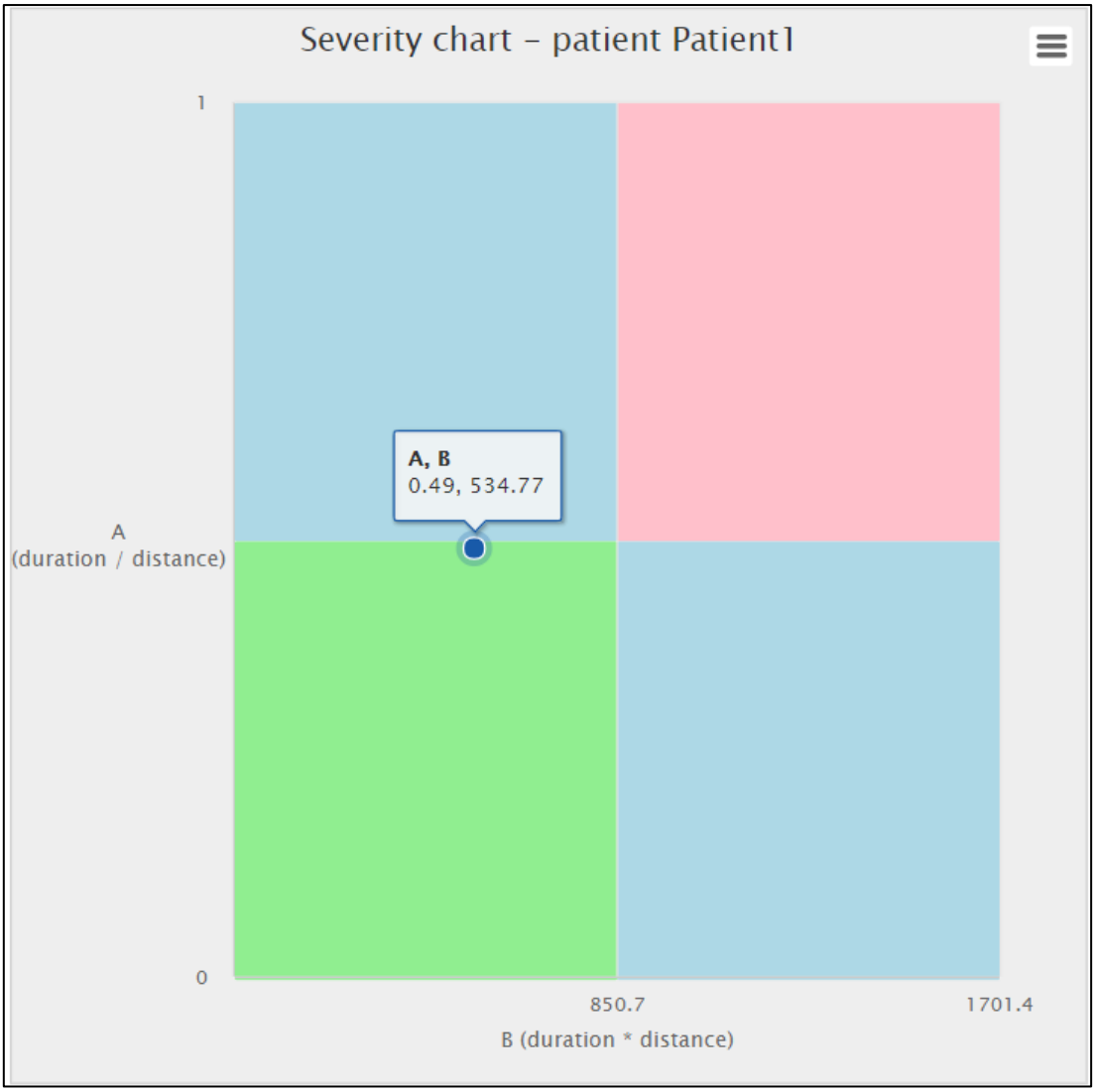

Figure 8.15: Severity chart of the non-adherence instances for Patient #1

8.3.3.2. Patient #2

In this case there is only one instance for the whole patient stay, which is the default instance (36.2%).

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
1447	36.2	Treatment missing (36.2)

Figure 8.16: Table of instances for patient #2

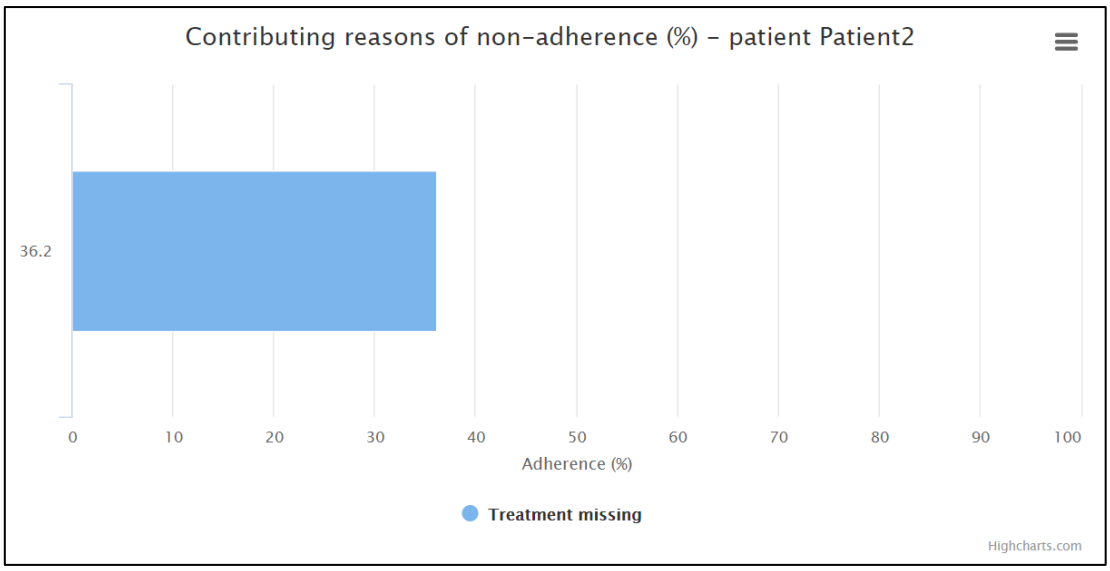

Figure 8.17: Contributing reason to non-adherence for Patient #2

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	36.2	36.2	36.2	36.2	36.2	36.2
Duration (mins)	100.0	1447.0	1447	1447.0	1447.0	1447.0
Duration / Non-adherence	39.97	39.97	39.97	39.97	39.97	39.97
Duration * Non-adherence	100	52381.4	52381.4	52381.4	52381.4	52381.4

Figure 8.18: Interquartile range table for Patient #2

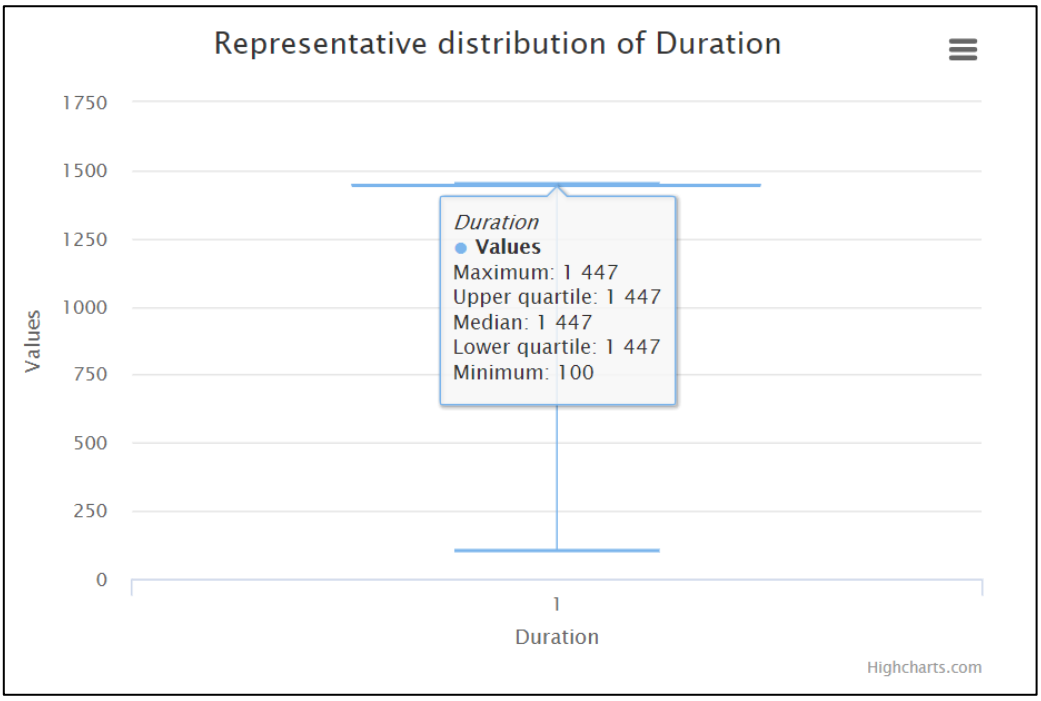

Figure 8.19: Boxplot of interquartile range of Duration of patient #2

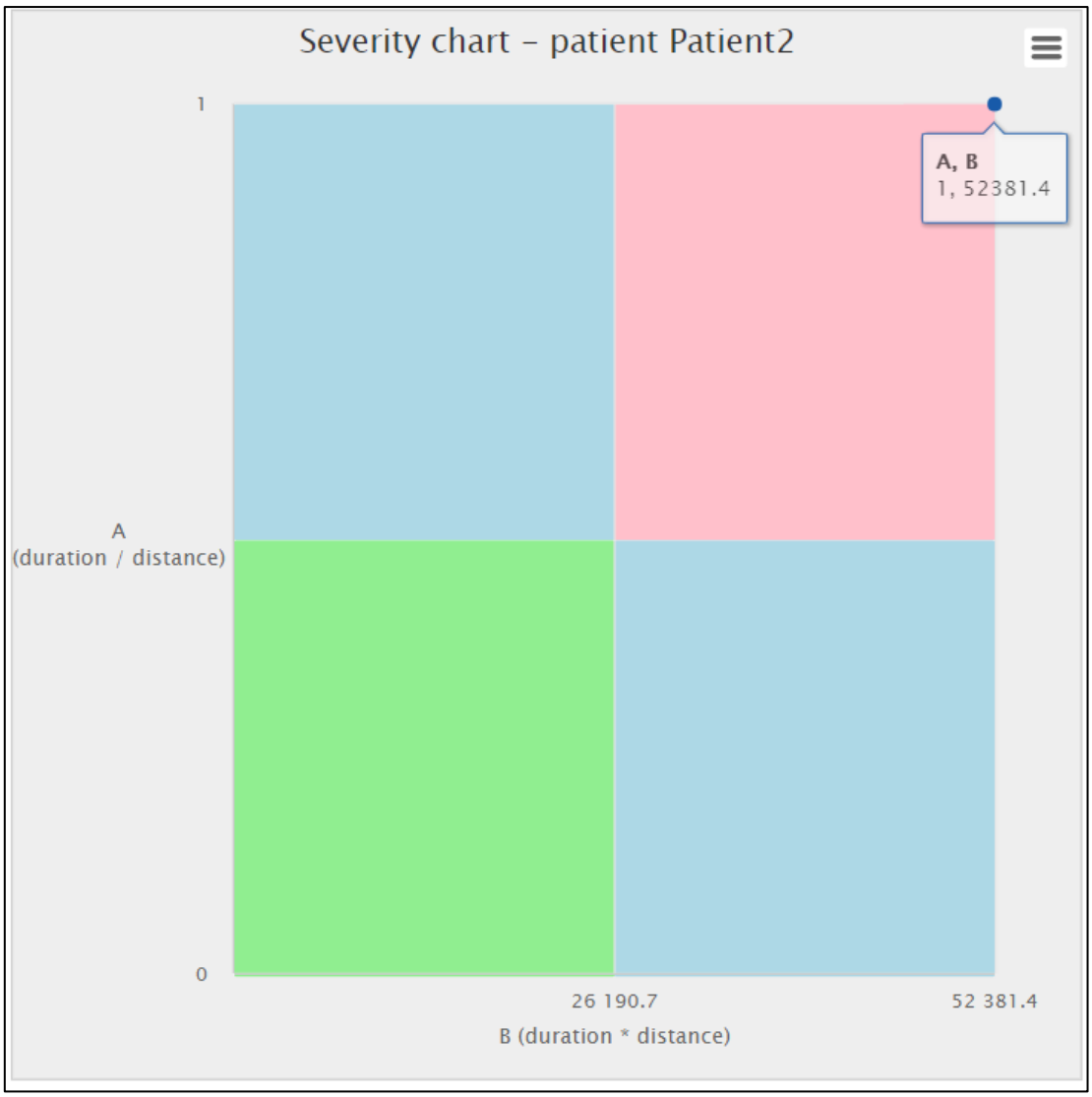

Figure 8.20: Severity chart for Patient #2

8.3.3.3. Patient #3

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
3	13.4	Type (10.0), Time to treatment (3.4)
4	18.6	Type (10.0), Time to treatment (8.6)
170	30	Type (10.0), Time to treatment (20.0)

Figure 8.21: Table of instances for Patient #3, with the default instance (36.2%) removed

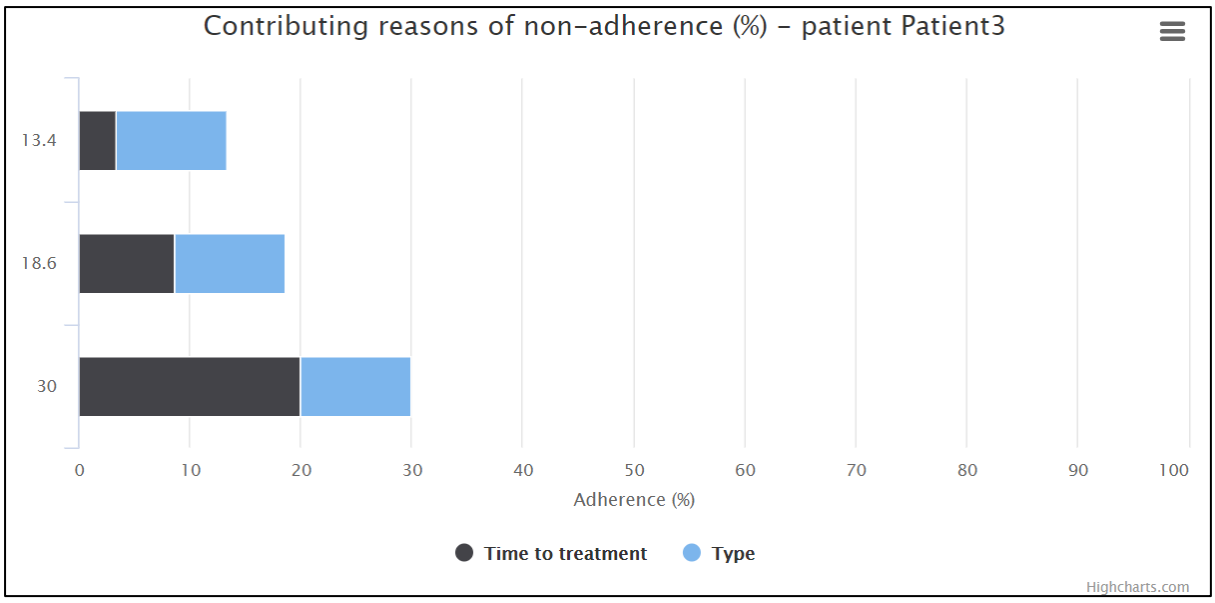

Figure 8.22: Contributing reasons for Patient #3

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	13.4	30	29.46	30	30	30
Duration (mins)	3.0	14.0	13.62	14.0	17.0	26.0
Duration / Non-adherence	0.13	0.47	0.47	0.47	0.57	0.87
Duration * Non-adherence	40.2	420	401.12	420	510	780

Figure 8.23: Interquartile ranges for Patient #3

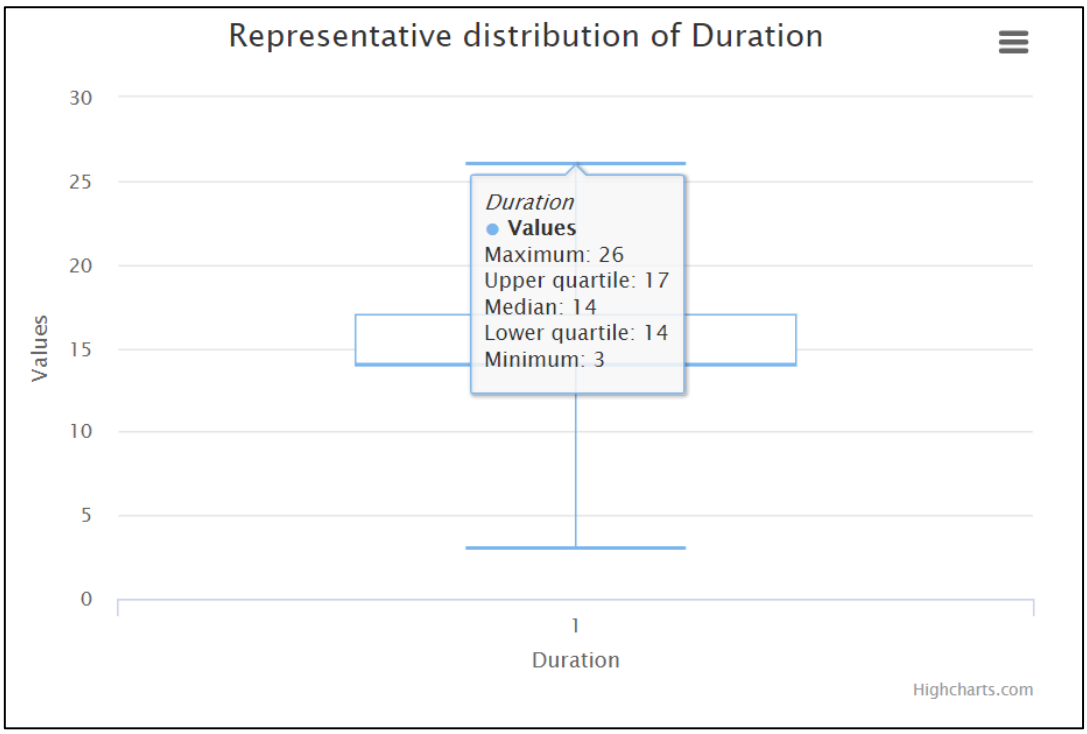

Figure 8.24: Boxplot visualisation of the Duration aspect of Patient #3

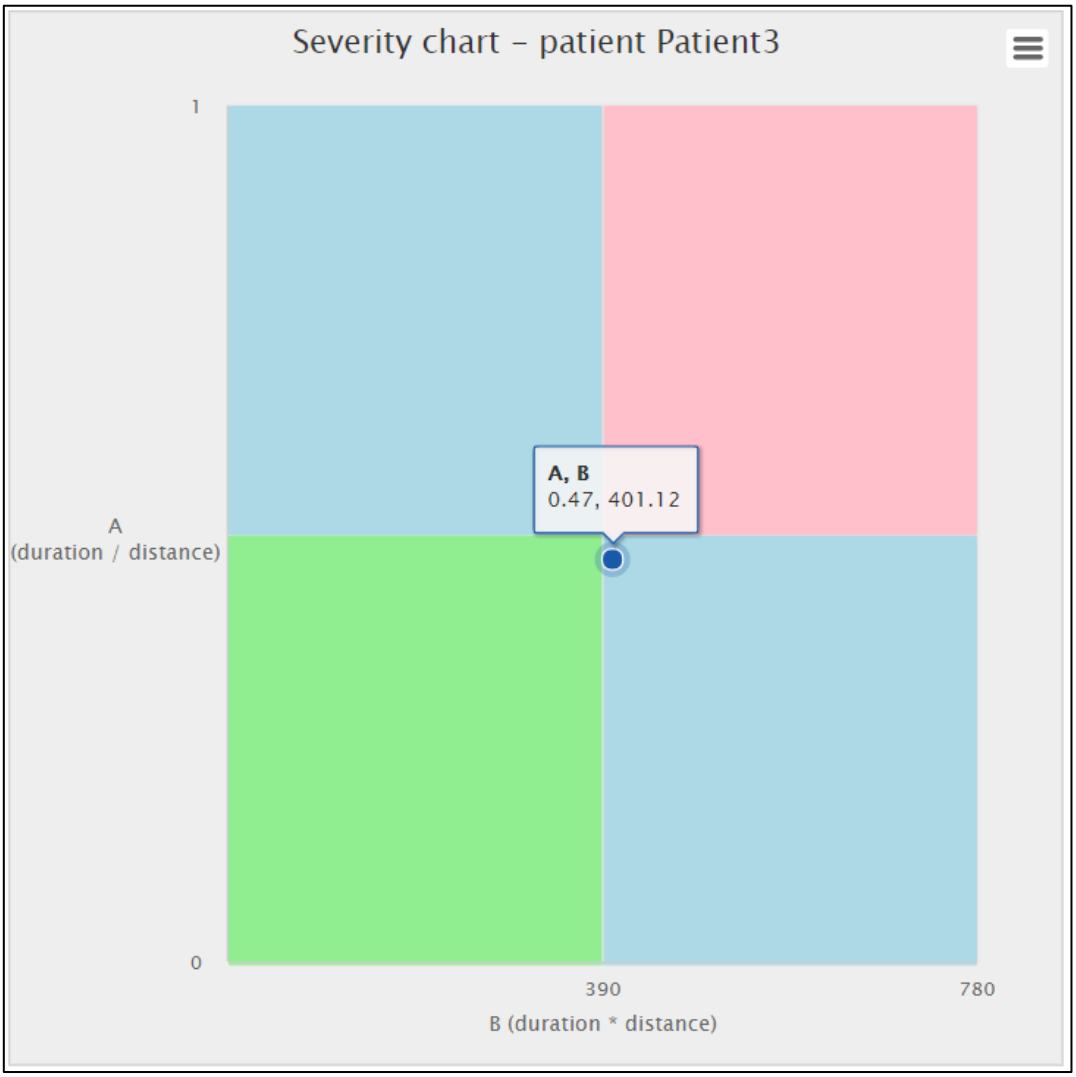

Figure 8.25: Severity chart of Patient #3

8.3.4. Individual charts of adherence

Summary criteria

- Dataset: **ICCA**
- Patient ID: **Patient1**
- Physiological series: **ICPm**
- Threshold (mmHg): **10**
- Hold-down (mins): **20**
- Time window (mins): **15**

Figure 8.26: Summary criteria for event 0 with patient #1

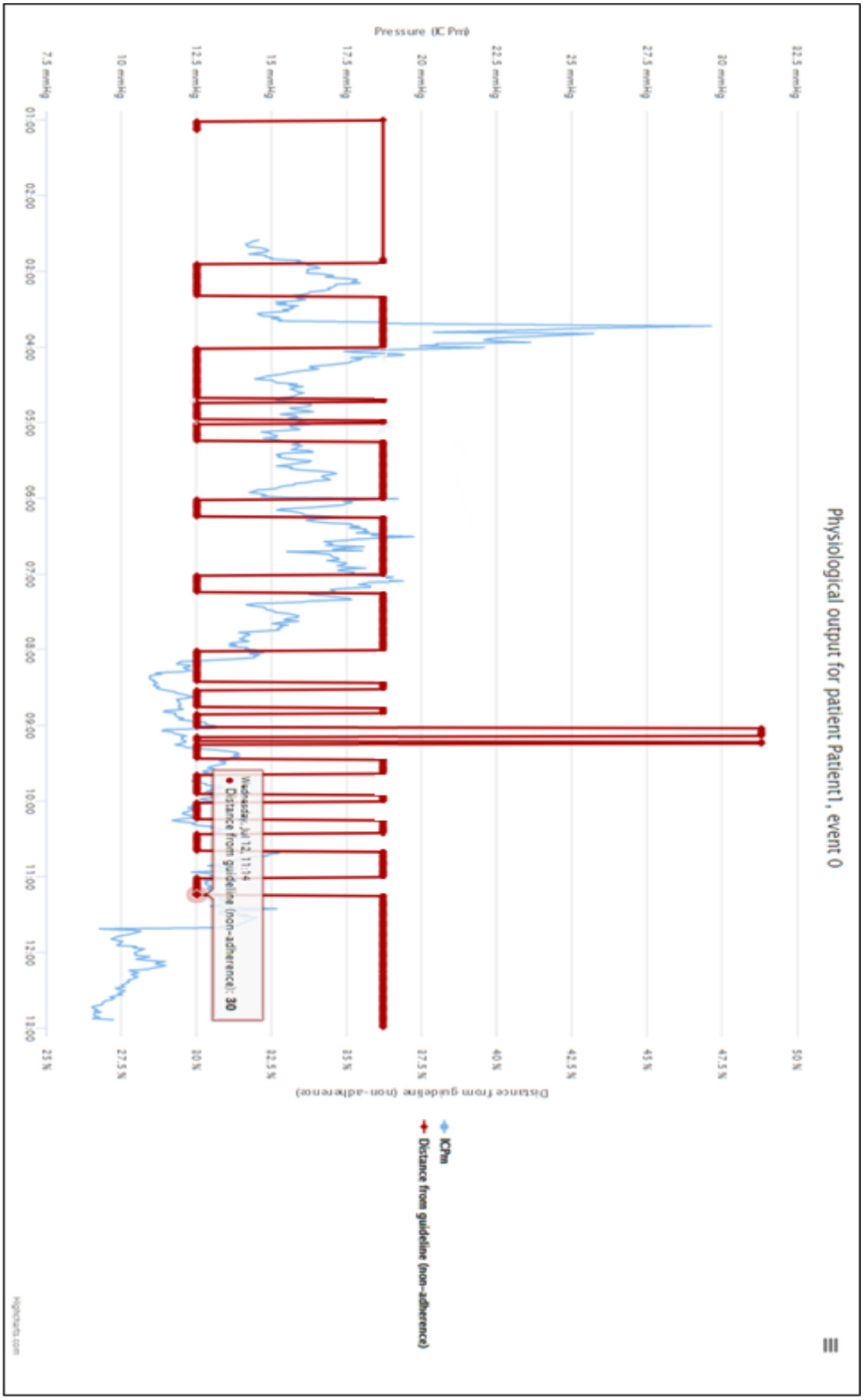

Figure 8.27: Timeline of event 0 ICPm

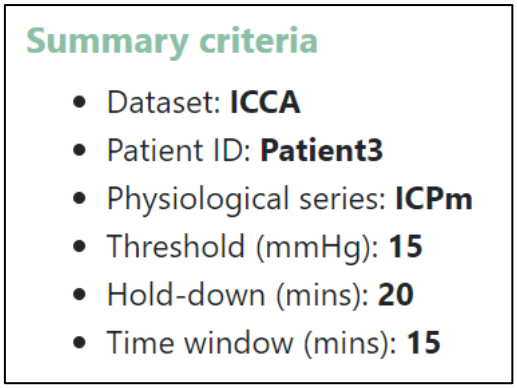

Figure 8.28: Summary criteria for event 0 in patient #3

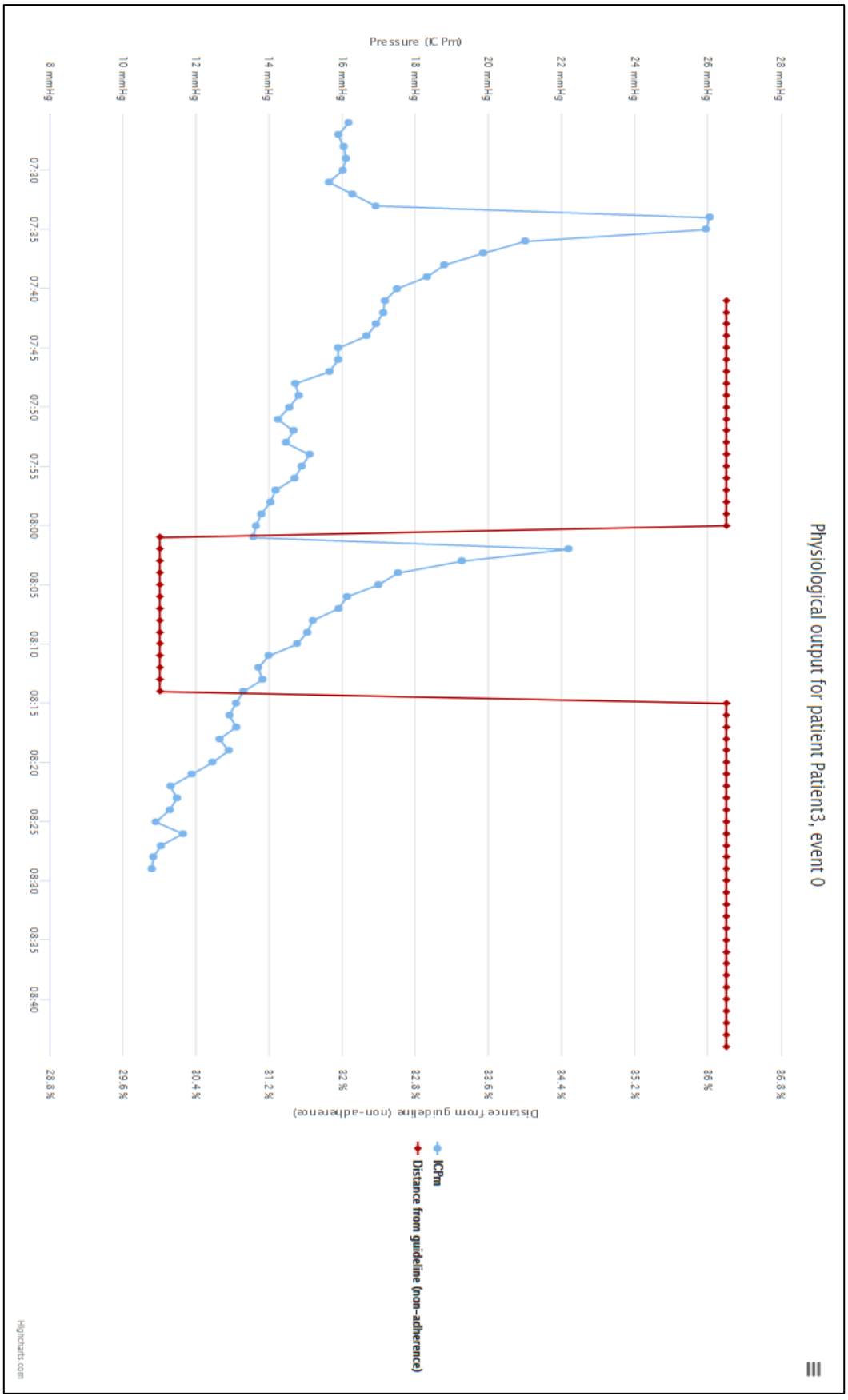

Figure 8.29: Timeline of event 0, patient #3

8.3.5. Discussion

Some general issues identified in this evaluation were the presence of EUSIG-events that only had hold-down definitions of 20 minutes (i.e. no events were detected with hold-down values lower than 20 mins). It is uncertain whether this was an issue of data processing or if this was a clinician choice (no action unless event held-down for 20 minutes) that manifested in the data. It would be anticipated that at least some events would be detected that were sub-sets of that maximum value (e.g. 5, 10 and 15 minutes), therefore this needs to be further investigated and confirmed that these are in fact the only physiological events detected.

Another issue was the categorisation of treatments. The direct drug name was listed in the treatment tables of the framework database, which can either be categorised according to the Brain-IT listing or can be individually incorporated into the framework (e.g. the drug name specifically listed in the code where pressors/fluids are captured to evaluate the type comparison in the CPP guideline). This categorisation would specify the measurement of adherence output more closely.

Patient 1 – The guideline adherence value on patient 1, event 0 (figure 8.27), largely flips between 30% and 36.2% based on whether it has recently encountered a treatment or not. From the overall adherence information for patient 1 (figure 8.11), it can be seen that the drug type contributes consistently to all non-adherence instances with this patient – this is believed to be because the treatment data has been categorised under the individual drug names rather than into drug family types, resulting in the fact that the type is always returned as "not recommended". Therefore, when a treatment is encountered the base level of adherence is generally 30% (composed of a type issue, plus the treatment being outside a given time window). This could be a reasonable representation of the contextual statement *"ICP not controlled by medical management"* for this patient (in the associated clinical notes).

Patient 2 – this patient has no treatment associations. The output is believed to be due to the refractory nature of the ICP increase (mentioned in the patient notes). This failure to associate annotations with a single, large EUSIG-event comes back

to the issue of time-window definition (mentioned in section 8.1). In this case, as the physiological output never returns to a "sub-threshold" position, the pattern matching algorithm is never reset and no treatment is associated even though there are many recorded. This should obviously be refined as an edge-case of the system.

The overall adherence output also reflects the lack of variation: one instance only (figure 8.16), one contributing reason (figure 8.17), and a highly skewed interquartile range output (figures 8.18 and 8.19). The severity chart (figure 8.20) shows the number being as severe as possible. This is indeed a severe case – however the severity is flagged by the (incorrect) reason of lack of treatments, rather than the escalating patient context (refractory ICP event) which is not addressed by the BTF guidelines therefore not captured in this system.

Patient 3 – The number of treatments for patient 3 is much lower than the other two, and therefore the default instance of non-adherence is larger (figures 8.21 and 8.22 from overall adherence information for patient 3). This leads to a severity value that occurs around the middle of the chart, an outcome which suggests that a small perturbation could have large effects on the considered severity (i.e. it would move it across the border between quadrants, though this measure is somewhat subjective). The main issue with the outputs from patient 3 is that when it is cross-referenced back to the original patient notes – which specified that the ICP EUSIG-events in this patient were occurring transiently and were apparently linked to the rise and fall of the patient's ETCO2 – the consistent management position was to take no action. The guideline adherence framework fails to capture this nuance when presenting output, which again leads to an over-statement of the lack of annotations. However, the variation when a ventilation treatment is encountered captures at least part of this clinical management process.

8.4. Evaluation of framework against patient outcome (GOS)

The Glasgow Outcome Score (GOS) provides an insight into the status of a patient six months after the initial brain injury and is on an 8-point scale (GOS-e or "extended" from the original 5-point GOS) as shown in table 8.10.

Score / Label	Description
1 – Death	Severe injury or death without recovery of consciousness
2 – Persistent vegetative state	Severe damage with prolonged state of unresponsiveness and a lack of higher mental functions
3 – Lower severe disability 4 – Upper severe disability	Severe injury with permanent need for help with daily living
5 – Lower moderate disability 6 – Upper moderate disability	No need for assistance in everyday life, employment is possible but may require special equipment
7 – Lower good recovery 8 – Upper good recovery	Light damage with minor neurological and psychological deficits

Table 8.10: Extended Glasgow Outcome Score (GOS-e) and label, with the associated description

The ordinality of this score can be used in different ways, but for the purposes of understanding the relationship to guideline distance/duration, scores 1-4 are classified as "poor" (or value "1"), and scores 5-8 are classified as "good" (or value "0"). Therefore, given a set of guideline non-adherence instances for different time durations, we can assess whether there is any correlation with the outcome of a patient. Examples of similar "adjusted" assessments of other variables' influence on patient outcome from traumatic brain injury can be found in (Edwards *et al.*, 2005) and (Güiza *et al.*, 2013).

8.4.1. Materials and Method

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 post-hoc analyses (Shaw *et al.*, 2009). A comprehensive collection of TBI data with

physiological, treatment, laboratory, surgery and other clinical events, with a particular focus on the management of ICP and CPP, it forms a detailed retrospective view of physiological and treatment data that is well suited to analyses such as the research work in this thesis. This is used for this statistical evaluation due to the presence of 6-month GOSe information for each patient in this cohort.

The steps towards understanding the influence of guideline deviation and correcting for the influence of known TBI factors, are:

- Create a "null" model – this is an "average" model, effectively a 50:50 guess on what the GOS will be in 6 months. It is hoped that information about guideline deviations will be at least better than this.
- Create a set of "unadjusted" models – these are univariate models that show the relationship between the GOS and the covariates (in this case guideline non-adherence distance and duration) and with the known factors influencing TBI outcomes (Edwards *et al.*, 2005). These factors are: age, GCSm, pupil reactivity, major extra-cranial injury, and CT scan availability.
- Create an "adjusted" model – this is a model including all the unadjusted co-variates that may have significance in explaining the model error (i.e. have a p-value less than 0.1).
- Check the variance between the models – this is a method to test if a change has had a positive effect on the overall nature of the model, by comparing the change in variance between the models.

To implement this, the output of all instances of guideline deviation (each instance containing a duration and a distance of deviation from the ideal guideline), are written to a spreadsheet. The first three models use a binomial logistic regression and the final comparison uses the "analysis of variance" (ANOVA) methodology.

A further consideration is the possibility that the two covariates – distance and duration of guideline adherence – are related. Therefore, an analysis of variance between the two unadjusted models is run as well.

8.4.2. Results

To run the logistic regression, the instance data is loaded, split between training/test (80/20 to optimise the variance between both performance and prediction), and fitted using the glm (generalized linear model) package in R, run on Ubuntu Xenial (16.04).

The duration and distance measures were repeat measures so an average of both was used for each patient line. The following column headers were used, with 249 instances (one line per patient, 13 removed due to lack of output, and Patient ID removed as it was an index only):

- GOSe
- Average duration of non-adherence
- Average non-adherence level
- Age
- GCS motor
- Left pupil reaction
- Right pupil reaction
- Facial injury

The availability of CT scan was not individually identified in the Brain-IT dataset so was not included on the list. Facial injury was the closest single data point to capture “extra-cranial injury”, which is the actual indicator from the CRASH/IMPACT studies.

8.4.2.1. Null model

The null model was constructed by running a regression against the mean of the GOS alone.

Model: GOS ~ 1

Call: glm(formula = GOS ~ 1, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.238	-1.238	1.118	1.118	1.118

Table 8.11: Deviance residuals of the null model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.1402	0.1418	0.989	0.323

Table 8.12: Coefficients of the null model

8.4.2.2. Unadjusted models – guideline adherence

The unadjusted model for adherence distance was created by running a regression against the GOS using the average of all instances for each patient. The intercept had a p-value < 0.1.

Model: GOS ~ Avg.distance

Call: glm(formula = GOS ~ Avg.distance, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.5656	-1.2091	0.8337	1.1488	1.1820

Table 8.13: Deviance residuals of the unadjusted distance model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.87802	0.49460	1.775	0.0759
Avg.distance	-0.02455	0.01561	-1.572	0.1158

Table 8.14: Coefficients of the unadjusted distance model

The unadjusted model for adherence duration was created by running a regression against the GOS using the average of all instances for each patient.

Model: GOS ~ Avg.duration

Call: glm(formula = GOS ~ Avg.duration, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.4933	-1.2180	0.9905	1.1393	1.2020

Table 8.15: Deviance residuals of the unadjusted duration model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.057743	0.219086	-0.264	0.792
Avg.duration	0.003917	0.003349	1.170	0.242

Table 8.16: Coefficients of the unadjusted duration model

8.4.2.3. Unadjusted models – known TBI predictors

The unadjusted model for age was created by running a regression against the GOS of the patients. The intercept had a p-value < 0.05 and Age had a p-value < 0.01.

Model: GOS ~ Age

Call: glm(formula = GOS ~ Age, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.6202	-1.1608	0.8055	1.1335	1.4811

Table 8.17: Deviance residuals of the unadjusted age model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.667656	0.315721	-2.115	0.03446
Age	0.022767	0.008023	2.838	0.00454

Table 8.18: Coefficients of the unadjusted age model

The unadjusted model for GCSm was created by running a regression against the GOS of the patients. The intercept had a p-value < 0.05 and GCSm had a p-value < 0.1.

Model: GOS ~ GCSm

Call: glm(formula = GOS ~ NSH_Adm_GCS_Motor, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.3974	-1.1559	0.9723	1.1094	1.3292

Table 8.19: Deviance residuals of the unadjusted GCSm model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.67435	0.32989	2.044	0.0409
NSH_Adm_GCS_Motor	-0.17072	0.08922	-1.914	0.0557

Table 8.20: Coefficients of the unadjusted GCSm model

The unadjusted model for left pupil reaction was created by running a regression against the GOS of the patients. Left pupil reactivity had a p-value < 0.01.

Model: GOS ~ Left pupil reactivity

Call: glm(formula = GOS ~ NSH_Adm_Left_Pupil_Reaction, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.6651	-1.1144	0.7585	1.2417	1.2417

Table 8.21: Deviance residuals of the unadjusted left pupil reaction model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.1499	0.1654	-0.906	0.36472
NSH_Adm_Left_Pupil_Reaction	1.2486	0.4189	2.980	0.00288

Table 8.22: Coefficients of the unadjusted left pupil reaction model

The unadjusted model for right pupil reaction was created by running a regression against the GOS of the patients. Right pupil reactivity had a p-value < 0.01.

Model: GOS ~ Right pupil reactivity

Call: glm(formula = GOS ~ NSH_Adm_Right_Pupil_Reaction, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.6459	-1.1272	0.7726	1.2285	1.2285

Table 8.23: Deviance residuals of the unadjusted right pupil reaction model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.1193	0.1630	-0.732	0.46418
NSH_Adm_Right_Pupil_Reaction	1.1754	0.4417	2.661	0.00778

Table 8.24: Coefficients of the unadjusted right pupil reaction model

The unadjusted model for facial injury was created by running a regression against the GOS of the patients. The intercept had a p-value < 0.1.

Model: GOS ~ Facial injury

Call: glm(formula = GOS ~ Injury_Facial, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.6006	-1.2055	0.8067	1.1495	1.1495

Table 8.25: Deviance residuals of the unadjusted facial injury model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.95551	0.52623	1.816	0.0694
Injury_Facial	-0.01779	0.01093	-1.627	0.1037

Table 8.26: Coefficients of the unadjusted facial injury model

8.4.2.4. Adjusted models

The model for adherence duration and distance (combined) was created by running a regression against the GOS using the average of all instances for each patient and testing for the relationship between the two parameters. The intercept had a p-value < 0.1 and the Avg.distance contribution had a p-value < 0.05.

Model: GOS ~ Avg.duration * Avg.distance

Call: glm(formula = GOS ~ Avg.duration * Avg.distance, family = binomial(link = "logit"),data = train)

Min	1Q	Median	3Q	Max
-1.7507	-1.1708	0.7915	1.1707	1.3059

Table 8.27: Deviance residuals of the distance/duration model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	1.0001521	0.5224688	1.914	0.0556
Avg.distance	-0.0397791	0.0176229	-2.257	0.0240
Avg.duration	-0.0178465	0.0311498	-0.573	0.5667
Avg.distance:Avg.duration	0.0007396	0.0009062	0.816	0.4144

Table 8.28: Coefficients of the distance/duration model

The model for all known TBI indicators (referred to here as the CRASH/IMPACT model) was created by running a regression against the GOS using all five covariates listed above. GCSm and left pupil reaction had a p-value < 0.05 and Age had a p-value < 0.01.

Model: GOS ~ CRASH/IMPACT model (known TBI indicators)

Call: glm(formula = GOS ~ Age + NSH_Adm_GCS_Motor + NSH_Adm_Left_Pupil_Reaction + NSH_Adm_Right_Pupil_Reaction + Injury_Facial, family = binomial(link = "logit"), data = train)

in	1Q	Median	3Q	Max
-1.98821	-1.02000	0.00016	1.05666	1.75047

Table 8.29: Deviance residuals of the CRASH/IMPACT model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	16.65411	1400.33243	0.012	0.99051
Age	0.03226	0.01113	2.898	0.00376
NSH_Adm_GCS_Motor	-0.25987	0.10644	-2.441	0.01463
NSH_Adm_Left_Pupil_Reaction	1.47781	0.75293	1.963	0.04968
NSH_Adm_Right_Pupil_Reaction	-0.41107	0.80073	-0.513	0.60770
Injury_Facial	-0.34513	28.00665	-0.012	0.99017

Table 8.30: Coefficients of the CRASH/IMPACT model

The model for the CRASH/IMPACT model with guideline adherence information added was created by running a regression against the GOS using all five covariates listed above and the duration/distance covariates. Age and GCSm had p-values < 0.01, distance/duration had p-values < 0.05 and left pupil reaction and distance * duration had p-values < 0.1.

Model: GOS ~ CRASH/IMPACT model with guideline adherence

Call: glm(formula = GOS ~ Age + NSH_Adm_GCS_Motor + NSH_Adm_Left_Pupil_Reaction + NSH_Adm_Right_Pupil_Reaction + Injury_Facial + Avg.distance * Avg.duration, family = binomial(link = "logit"), data = train)

Min	1Q	Median	3Q	Max
-1.72050	-0.93226	0.00015	0.93286	1.98057

Table 8.31: Deviance residuals of the CRASH/IMPACT model

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	1.641e+01	1.397e+03	0.012	0.99063
Age	4.017e-02	1.225e-02	3.280	0.00104
NSH_Adm_GCS_Motor	-3.090e-01	1.129e-01	-2.737	0.00620
NSH_Adm_Left_Pupil_Reaction	1.384e+00	8.039e-01	1.722	0.08506
NSH_Adm_Right_Pupil_Reaction	-1.441e-01	8.473e-01	-0.170	0.86495
Injury_Facial	-5.748e-01	2.795e+01	-0.021	0.98359
Avg.distance	3.195e-01	1.336e-01	2.390	0.01683
Avg.duration	1.691e-01	8.296e-02	2.038	0.04150
Avg.distance:Avg.duration	-4.475e-03	2.389e-03	-1.873	0.06111

Table 8.32: Coefficients of the CRASH/IMPACT model

8.4.2.5. Analysis of variance

The deviance between the duration and null models was not high and without statistical significance.

Model 1: GOS ~ Avg.duration

Model 2: GOS ~ 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr (>Chi)
1	198	274.83			
2	199	276.28	-1	-1.451	0.2284

Table 8.33: Analysis of variance between duration and null models

Similarly, the deviance between the distance and null models was not high and without statistical significance.

Model 1: GOS ~ Avg.distance

Model 2: GOS ~ 1

	Resid. Df	Resid. Dev	Df	Deviance	Pr (>Chi)
1	198	273.63			
2	199	276.28	-1	-2.6503	0.1035

Table 8.34: Analysis of variance between distance and null models

Between each other, the relationship did not appear to correlate in a statistically significant way (low deviance and no significant p-value).

Model: Avg.distance ~ Avg.duration

	Df	Deviance	Resid. Df	Resid. Dev	Pr (>Chi)
NULL			199	276.28	
Avg.distance	1	2.6503	198	273.63	0.10353
Avg.duration	1	4.1174	197	269.51	0.04244
Avg.distance:Avg.duration	1	0.6706	196	268.84	0.41284

Table 8.35: Analysis of variance between distance and duration models

When the adjusted models were compared, the explanatory power of the model to patient GOS was largely explained by the CRASH/IMPACT model (large deviance with significant p-value).

Model 1: GOS ~ CRASH/IMPACT model * Duration/Distance

Model 2: GOS ~ CRASH/IMPACT model

	Resid. Df	Resid. Dev	Df	Deviance	Pr (>Chi)
1	135	153.25			
2	138	168.22	-3	-14.969	0.001843

Table 8.36: Analysis of variance between CRASH/IMPACT model with adherence and CRASH/IMPACT model without

8.4.3. Discussion

Through the creation of the regression models in this section, the possible relationship between the guideline adherence output and patient outcome was investigated. When combined with the known indicators in the CRASH/IMPACT studies, the instances of guideline adherence failed to have a statistically significant bearing on the explanatory power of the models.

This is not unexpected, partly due to the weightings used in numerically evaluating the adherence instances, and the low treatment associations in the Brain-IT dataset (see chapter 9). It is possible that with calibration of the weightings and applied to a dataset with a higher density of treatment annotations it may have some relationship.

8.5. Evaluation summary

The evaluations performed in this chapter indicated satisfactory performance in capturing guideline adherence information at a clinical management level (section

8.1) and successfully provided representation against contextual information from real patient notes in an ICU setting (section 8.3). Accurate timing of treatment annotations was identified as a key requirement in section 8.1, and an attempt was made to quantify this timing difference in section 8.2. Unfortunately, the results of section 8.2 only provided information in the ideal “live” situation, but this in itself was an indicator of the required improvement in annotation timings (ideally by automated means). Section 8.4 failed to find a statistical relationship between the adherence output and patient outcome though it is hoped with more refinement of the weightings involved, this may improve.

9. Application of framework to neurological ICU dataset: Brain-IT

Chapter summary

In this chapter, the process model guideline adherence framework is applied to the Brain-IT dataset to evaluate its performance on real-world data.

The information extracted from the Brain-IT dataset is presented in five sections:

1. *The number of identified events based on the EUSIG definitions*
2. *Generation of the event log showing the association of treatment annotations with EUSIG-events and the distribution of treatment categories*
3. *The individual patient output of the comparison of the resulting process models for selected patients:*
    a. *Minute-by-minute adherence*
    b. *Total duration and levels of non-adherence*
    c. *Interquartile ranges showing statistical spread (impact and relevance) of non-adherence*
    d. *"Default" vs "non-default" non-adherence instances*
    e. *Interquartile ranges with default instances removed (highlighting different sources of non-adherence variety)*
    f. *Contribution of non-adherence reasons*
    g. *Clinical severity charts*
4. *Additional insights from applying the guideline adherence framework on the Brain-IT dataset, including the impact of overlaying a clinical response time-window, and the treatment category distribution*

Overall, the physiological data is comprehensive and the EUSIG event pattern easily extracted. However, the annotation frequency and density in the dataset overall is low, leading to low association numbers with EUSIG. Despite this low resolution, clear indications of guideline adherence were found within the dataset, suggesting that the framework does provide viable output for measuring guideline adherence.

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 post-hoc analyses (Shaw *et al.*, 2009). A comprehensive collection of TBI data with physiological, treatment, laboratory, surgery and other clinical events, with a particular focus on the management of ICP and CPP, it forms a detailed retrospective view of physiological and treatment data that is well suited to analyses such as the research work in this thesis.

9.1. EUSIG-event detection

The first stage of evaluation of this dataset is to detect physiological EUSIG (Edinburgh University Secondary Insult Grade) events.

9.1.1. Coverage

The coverage of physiological parameters in the database is summarised in table 9.1. "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 to the analysis.

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 9.1: Physiological parameter coverage in Brain-IT

By inspecting the coverage for the data points used, a level of initial confidence can be gained to see how well represented the parameters in the dataset are. If the parameter is well covered, it is a reasonable expectation that the event detection algorithm will produce useful information. From the results shown in table 9.1, we can see that the pressure measures are all at least above 80%, including those of particular interest: mean intra-cranial pressure (ICPm), cerebral perfusion pressure (CPP) and systolic blood pressure (BPs).

9.1.2. EUSIG-event definitions

Following from the description of EUSIG definitions of ICP, CPP and SBP events (table 9.2), ten threshold definitions are used to cover the range of clinically relevant definitions.

Parameter	Threshold values (mmHg)	Direction	Event Hold-Down (mins)
ICP	10, 15, 20, 25, 30	Up	5, 10, 15, 20
CPP	50, 60, 70	Down	5, 10, 15, 20
SBP	100, 110	Down	5, 10, 15, 20

Table 9.2: Definitions of raised ICP, lowered CPP and lowered SBP events

Similarly, four values are applied representing the differences in hold-down and clear hold-down times: 5, 10, 15 and 20 mins. Therefore, there are a total of 40 (= 10 * 4) ways that a physiological monitoring event can be detected in a dataset.

9.1.3. EUSIG-event counts

The event detection algorithm has been applied to the 40 definitions of EUSIG events. Tables 9.3-9.5 and figures 9.1-9.3 show the count of individual events for each EUSIG definition across all 262 patients (with all profiles now stored in the treatment profile database after processing).

Threshold value (mmHg)	Hold-down value	Count
10	20	2585 (abs)
15	20	2797 (abs)
20	20	1631 (abs)
25	20	643 (abs)
30	20	280 (abs)
10	15	+732
15	15	+991
20	15	+686
25	15	+318
30	15	+132
10	10	+1408
15	10	+1799
20	10	+1452
25	10	+714
30	10	+316
10	5	+3697
15	5	+5130
20	5	+4213
25	5	+2266
30	5	+1088

Table 9.3: Event count for each ICP EUSIG definition (x20). The first five definitions are an absolute count, whilst the following fifteen give the additional count, as the hold-down values are subsets of each other.

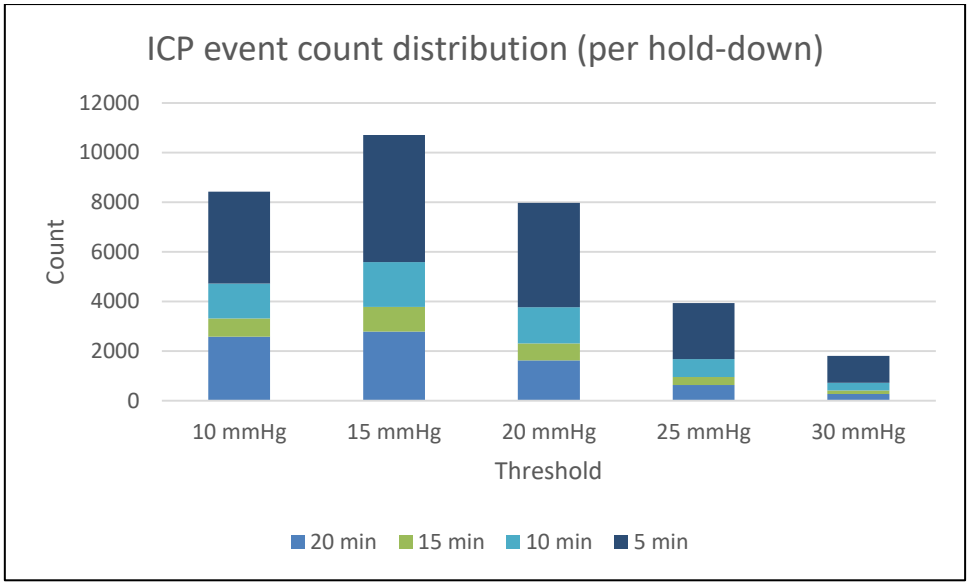

Figure 9.1: Event count distribution for ICP

Threshold value (mmHg)	Hold-down value	Count
50	20	537 (abs)
60	20	1874 (abs)
70	20	3048 (abs)
50	15	+252
60	15	+670
70	15	+1080
50	10	+534
60	10	+1491
70	10	+2085
50	5	+1759
60	5	+4261
70	5	+6278

Table 9.4: Event count for each CPP EUSIG definition (x12). The first three definitions are an absolute count, whilst the following nine give the additional count, as the hold-down values are subsets of each other.

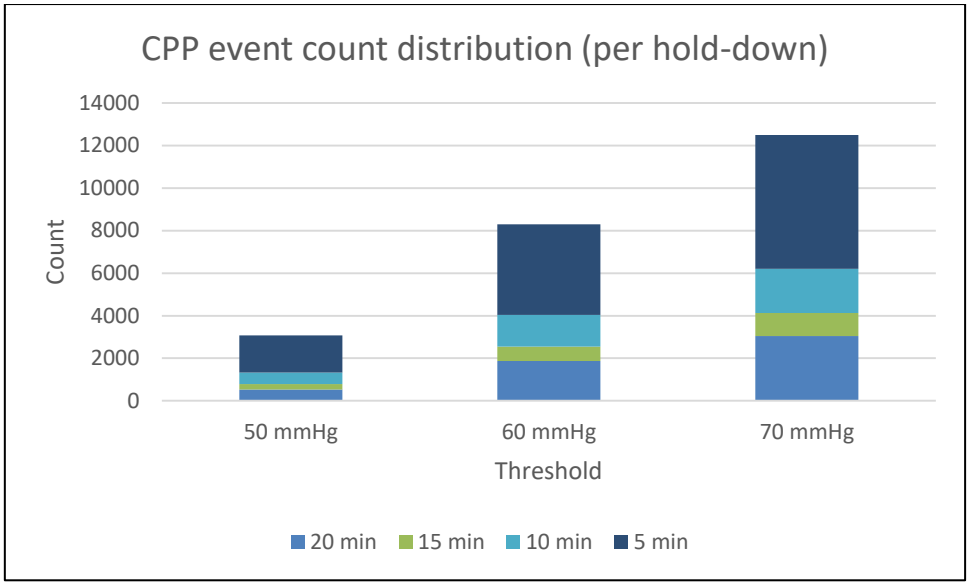

Figure 9.2: Event count distribution for CPP

Threshold value (mmHg)	Hold-down value	Count
100	20	512 (abs)
110	20	1371 (abs)
100	15	+193
110	15	+450
100	10	+384
110	10	+935
100	5	+1334
110	5	+2677

Table 9.5: Event count for each SBP EUSIG definition (x12). The first two definitions are an absolute count, whilst the following six give the additional count, as the hold-down values are subsets of each other.

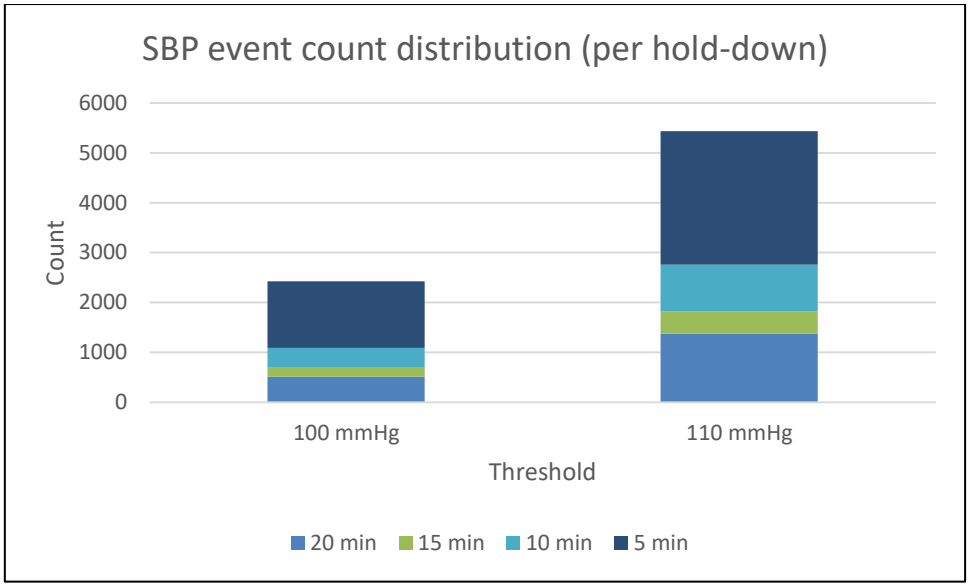

Figure 9.3: Event count distribution for SBP

Figures 9.1 to 9.3 show these event count results for ICP, CPP and SBP as bar charts. The column number in each chart is the total for that parameter (e.g. ICP has 20 corresponding to the five threshold definitions multiplied by the four hold-down values). The interpretation of the cyclical shape for every set of hold-down definitions is as follows: as the hold-down value for that particular definition increases, the number of EUSIG-events captured goes down (e.g. a EUSIG-event with a hold-down value of 20 minutes will be less common than one with a 5-minute hold-down). However, slightly less intuitively, the individual numbers vary according to definition: e.g. the most populous number of events in ICP, in each hold-down definition cycle, come from having a monitoring threshold of > 15 mmHg. This represents a minima inflection point, discussed further in section 9.5 (additional information from treatment associations).

From this initial evaluation of event count numbers, the EUSIG-event pattern has clear representation in the Brain-IT dataset and can be used as a basis for assessing a wider picture of clinical management. The next step is to attempt to apply the association of treatment annotations to these event counts to generate an event log.

9.2. Generation of an event log

To generate the full event log that can then be turned into a process model, requires the association of treatment annotations with physiological events.

Referring back to section 5.1 and the relative definitions of the word "event": the convention in this thesis is to use "EUSIG-event" to refer to the pattern in the physiological output described by a threshold, a hold-down and a clear hold-down; whilst "event" (singular) refers to a EUSIG-event *and* an associated treatment. Multiple instances of this latter definition are chained together to form the full patient event log described in this sub-section.

The treatment labels in the Brain-IT dataset follow the well-defined categories in that data schema and can be associated with the EUSIG-events subject to the considerations listed in section 5.1.3 (i.e. multiple treatments and multiple overlapping event time-windows being treated as one). The event log generated from this process is then available to be converted into a process model.

9.2.1. Treatment categories

Table 9.6 shows the treatment categories, and overall count of individual annotations, that have been identified by the Brain-IT consortium as being critical to identify information relevant to the management of ICP, CPP and SBP.

Treatment category	Count
Ventilation	474
Sedation	1796
Analgesia	1732
Paralysis	1790
Volume expansion	3063
Inotropes	860
Anti-hypertensives	86
Anti-pyretics	1294
Hypothermia	99
Steroids	68
Cerebral vasoconstriction	24
Osmotic therapy	1773
Cerebrospinal fluid (CSF) drainage	661
Head elevation	676
Barbiturates	136
Other	4603

Table 9.6: Overall treatment category count (total = 19135)

9.2.2. Treatment association and event log

To show the representation of instances of treatment/event associations, table 9.7 shows an example of the count of events with associated treatments expressed as an absolute number count and a percentage of overall events for that definition (see table 9.8 for the absolute ICP event number). The definition selected is the ICP threshold value of 15 mmHg, for a hold-down time of 5 mins.

Time-window (mins)	Count (absolute)	Count (%)
15	1822	17.2
30	1982	18.7
45	2197	20.8
60	2246	21.3

Table 9.7: Number of events with associated treatments depending on the time-window definition chosen

Figure 9.4 shows this variation across the five definitions of ICP threshold monitoring, with a constant hold-down time of 5 minutes.

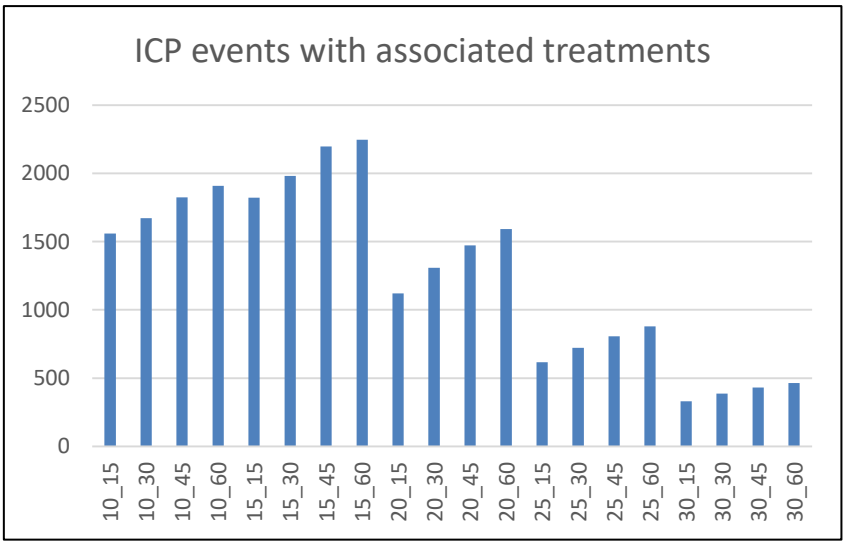

Figure 9.4: ICP associated treatment count, varied by time-window

Overall there are 19135 annotated treatments in the dataset. Table 9.8 shows the absolute number of patients and events that have associated treatments, and the percentage relative to the total number of both (patients and events) for the same definition as above (ICP > 15 mmHg for hold-down of 5 minutes). This validation check shows that whilst the pattern is a good representation of EUSIG-events, it does not necessarily cover the entire patient cohort, and could be improved to

capture greater accuracy (see the discussion later in this chapter and in chapter 11 on the quality and frequency of treatment annotations).

Patients (absolute)	Events (absolute)	Patients (% of total)	Events (% of total)
187	1822	71.3	17.2

Table 9.8: Number of patients and EUSIG-events with associated treatments expressed as absolute and percentage of total of each for a definition of ICP > 15 mmHg for a hold-down value of 5 minutes and a time-window of 15 minutes

This figure of 17.2% is low, though it should be highlighted here that there is believed to be a bug in the data processing, the impact of which is to halve the actual number of treatment associations, meaning the true value is closer to 34%. (See Appendix F for details.)

To convert this association information into an event log, the EUSIG-event and associated treatment instances are collated per-patient and represented as shown in figure 9.5 as an example for patient 15138374.

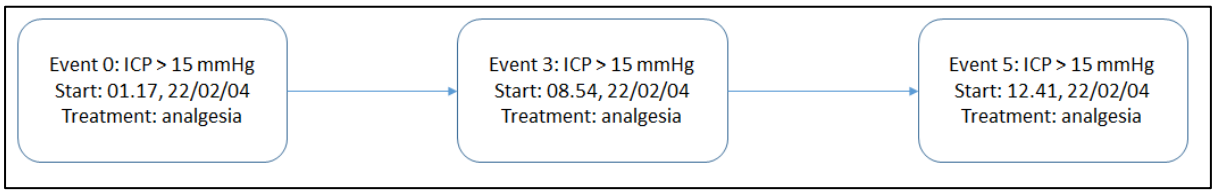

Figure 9.5: Snippet of the event log for patient 15138374

9.3. Process model comparison – individual output

As described in chapter 6, the event log can now be converted to a process model and evaluated for guideline adherence.

9.3.1. Minute-by-minute, aggregate and statistical spread of non-adherence

The following five patients have been selected to indicate information relating to different classes of non-adherence that have appeared during the dataset assessment. They have been selected from the patients that had treatment associations (and hence varied adherence information) and display different characteristics that exemplify the possible outcomes from this framework. Within those parameters and that patient subset, they are structurally representative of the variation shown by the framework over the whole dataset.

The timeline captions indicate the reasons for the pattern of non-adherence for that event (e.g. “treatment missing”, “dosage too high”, etc). As before, each

patient has an accompanying table showing the aggregate (total) non-adherence duration and level, as well as an interquartile range table showing the statistical spread of the various instances of non-adherence (one feature repeated visually as an example box-plot diagram). Table 9.9 shows the EUSIG-event definitions used for these examples.

Feature	Value
EUSIG definition	**ICPm** with threshold of **15 mmHg**
Hold-down	**5 mins**
Time-window	**15 mins**

Table 9.9: EUSIG definition details that the following five examples are sourced from

9.3.1.1. Patient 4026626

For this patient, when the Analgesia has been applied the non-adherence distance shows 2%, which is very adherent to the guideline (treatment present, recommended for this context and not over dosage). Once the time-window period has passed, the event continues but a recent treatment is now missing, therefore the non-adherence jumps back to the "default" level of 36.2%.

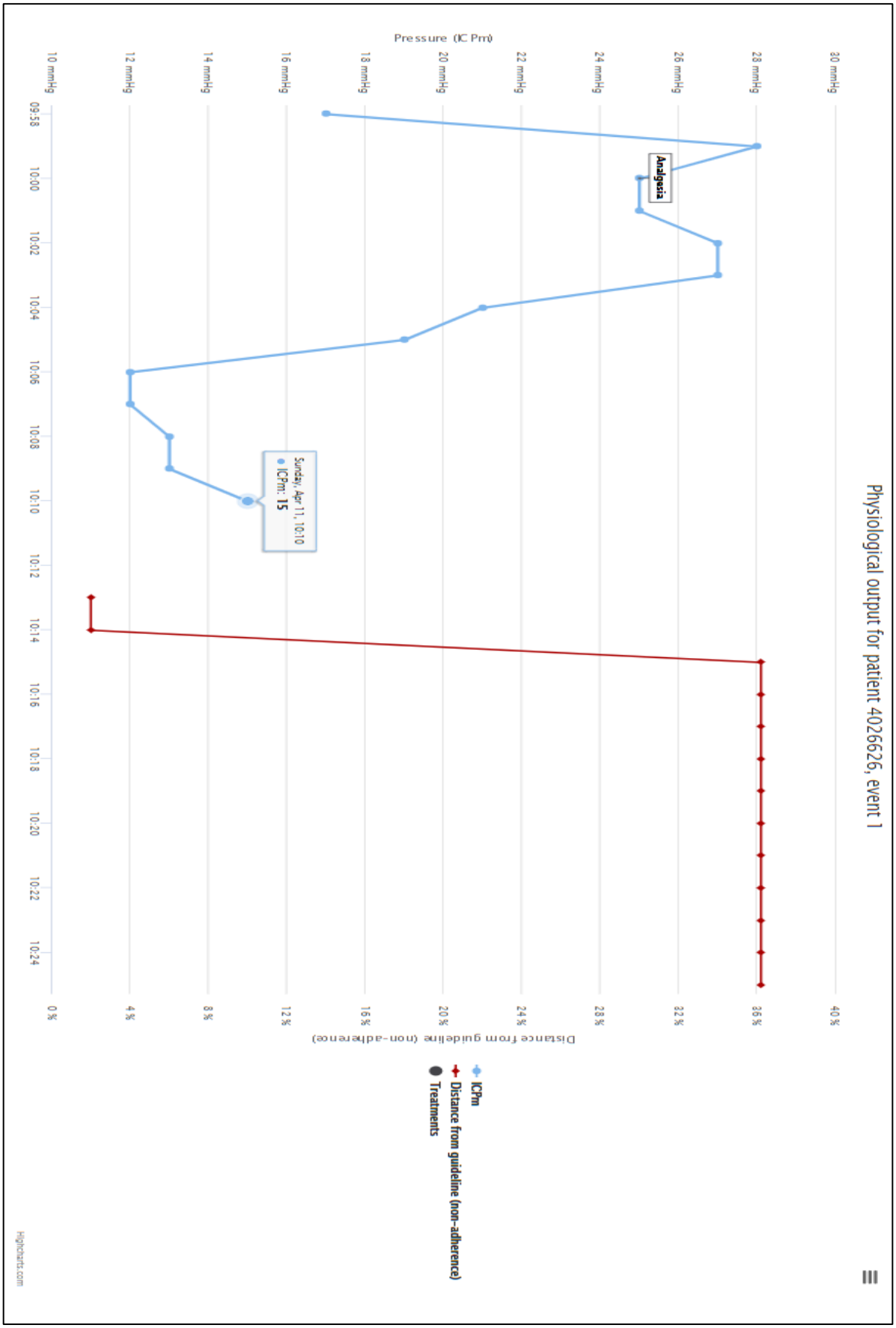

Figure 9.6: Patient 4026626, event 1

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
2	2	Time to treatment (2.0)
2069	36.2	Treatment missing (36.2)

The following items also show overall durations for this patient:

Total time of patient stay (mins): 55125
Total time of patient requiring clinical action after an event (mins): 2071

Total duration of "default" non-adherence state (mins): 2069.0
Total time spent in non-adherence, which is not in the "default" state (mins): 2.0
Expressed as a percentage of all the time spent requiring clinical action: 0.1 %

Figure 9.7: Total output for patient 4026626

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	2	36.2	36.17	36.2	36.2	36.2
Duration (mins)	2.0	13.5	47.07	22.5	42.0	349.0
Duration / Non-adherence	0.3	0.43	1.32	0.66	1.16	9.64
Duration * Non-adherence	4	488.7	1702.31	814.5	1520.4	12633.8

Figure 9.8: Interquartile range table for patient 4026626

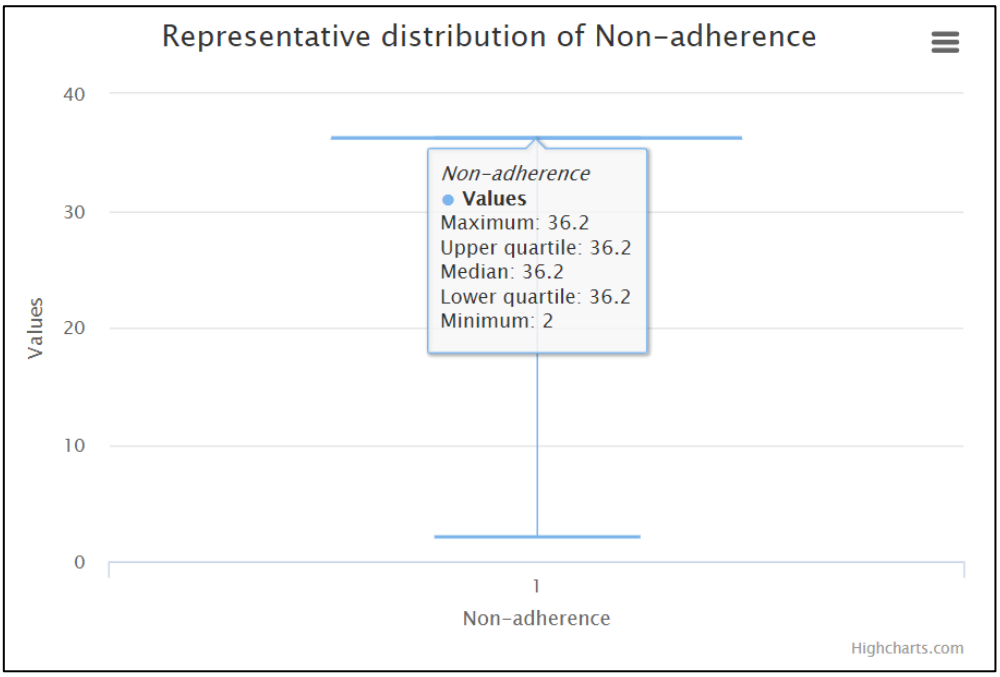

Figure 9.9: Box-plot visualisation of interquartile ranges of Non-adherence aspect

9.3.1.2. Patient 15138374

In this case the adherence level begins at 8.6% due to earlier annotation still being recorded, the window goes back to default for one minute (36.2%) then drops again due to the Volume Expansion treatment.

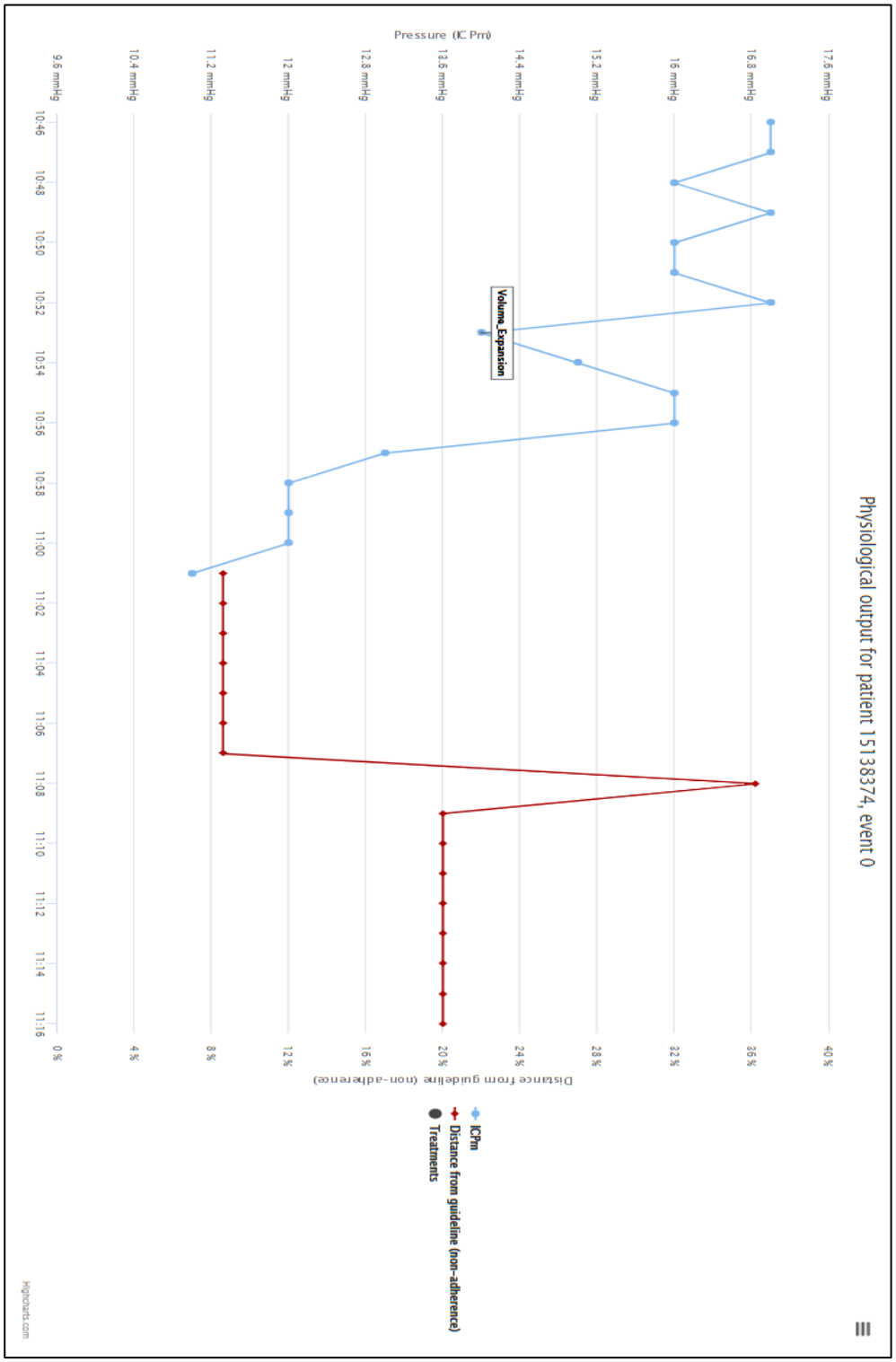

Figure 9.10: Patient 15138374, event 0

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
6	7.4	Time to treatment (7.4)
8	8.6	Time to treatment (8.6)
8	10	Time to treatment (10.0)
10	11.4	Time to treatment (11.4)
1	12.6	Time to treatment (12.6)
11	14	Time to treatment (14.0)
1	15.4	Time to treatment (15.4)
13	16.6	Time to treatment (16.6)
15	18	Time to treatment (18.0)
1	19.4	Time to treatment (19.4)
155	20	Time to treatment (20.0)
1843	36.2	Treatment missing (36.2)

The following items also show overall durations for this patient:

Total time of patient stay (mins): 55125
Total time of patient requiring clinical action after an event (mins): 2072
Total duration of "default" non-adherence state (mins): 1843.0
Total time spent in non-adherence, which is not in the "default" state (mins): 229.0
Expressed as a percentage of all the time spent requiring clinical action: 11.05 %

Figure 9.11: Total table for patient 15138374

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	7.4	36.2	34.18	36.2	36.2	36.2
Duration (mins)	1.0	8.0	28	14.0	21.0	491.0
Duration / Non-adherence	0.03	0.33	0.87	0.58	0.79	13.56
Duration * Non-adherence	8.6	187.9	956.91	434.4	760.2	17774.2

Figure 9.12: Interquartile range table for patient 15138374

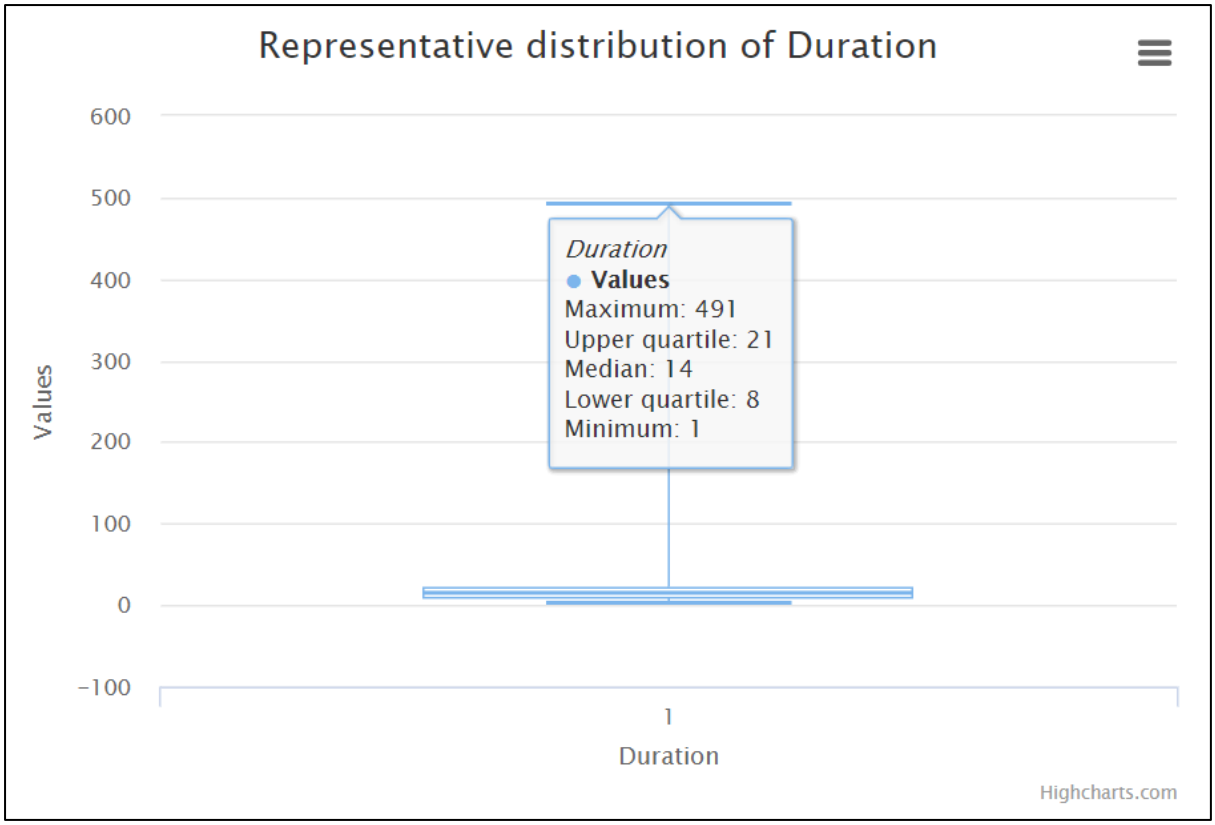

Figure 9.13: Interquartile range for Duration for patient 15138374

9.3.1.3. Patient 26138262

This patient shows a recurrent administration of sedation/analgesia in one patient. The stepped non-adherence score is due to the combined issues of going outside a time-window and reaching a "too high" dose.

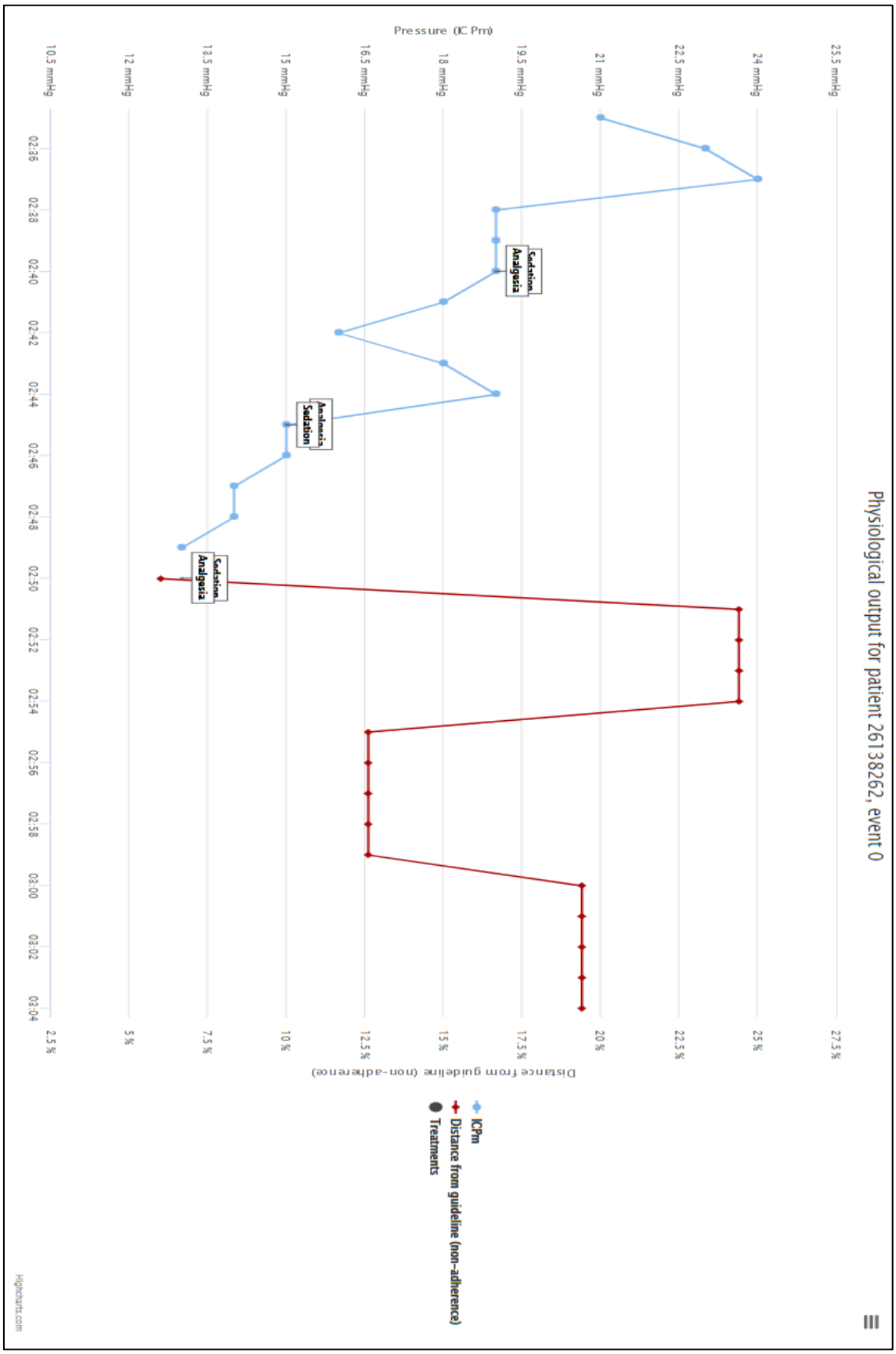

Figure 9.14: Patient 26138262 – shows a recurrent administration of sedation/analgesia in one patient

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
1	6	Time to treatment (6.0)
5	12.6	Time to treatment (12.6)
5	19.4	Time to treatment (19.4)
58	20	Time to treatment (20.0)
4	24.4	Time to treatment (6.0), Dose (18.4)
111	36.2	Treatment missing (36.2)

The following items also show overall durations for this patient:

Total time of patient stay (mins): 5393
Total time of patient requiring clinical action after an event (mins): 184
Total duration of "default" non-adherence state (mins): 111.0
Total time spent in non-adherence, which is not in the "default" state (mins): 73.0
Expressed as a percentage of all the time spent requiring clinical action: 39.67 %

Figure 9.15: Total table for patient 26138262

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	6	20	29.17	36.2	36.2	36.2
Duration (mins)	5.0	5.0	20.44	24.0	37.5	41.0
Duration / Non-adherence	0.25	0.6	0.83	0.89	1.13	1.7
Duration * Non-adherence	30	98.5	596.47	530	1267	1484.2

Figure 9.16: Interquartile range table for patient 26138262

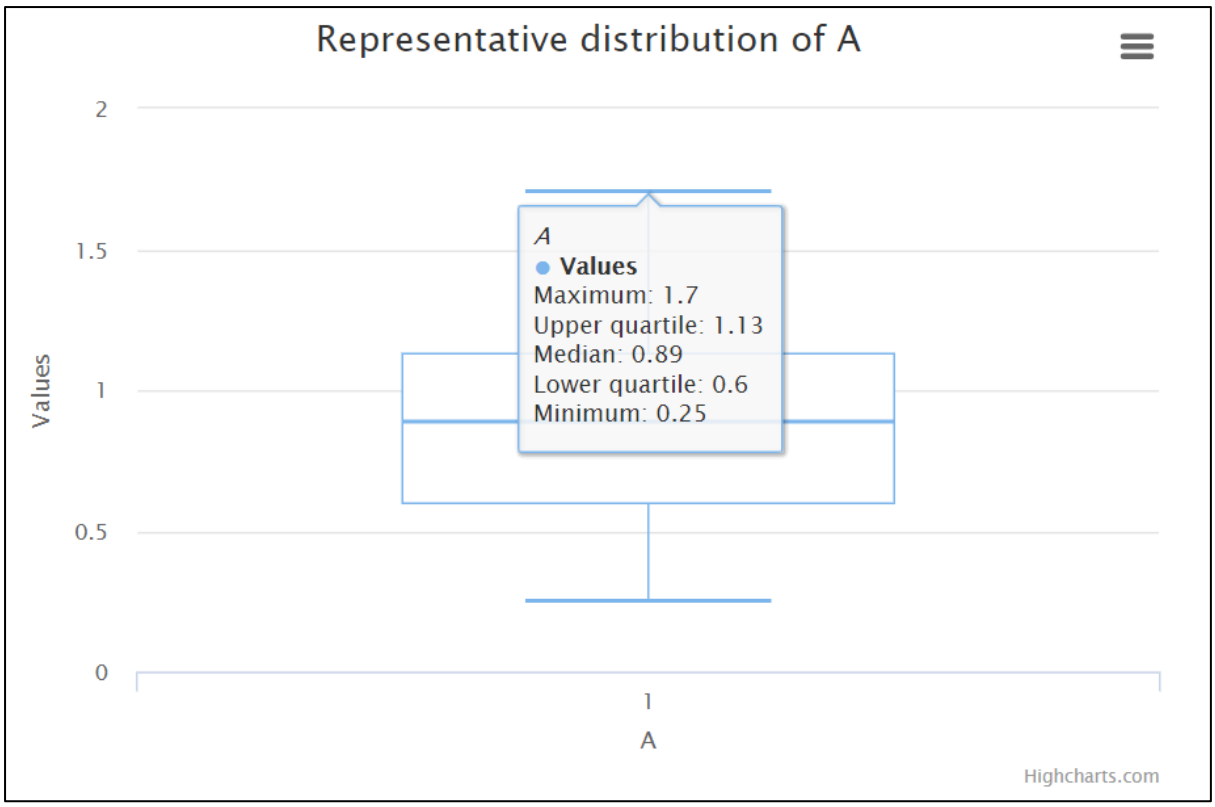

Figure 9.17: Box-plot visualisation of the A metric (Duration / Non-adherence) for patient 26138262

9.3.1.4. Patient 4026152

This patient has a type aspect featuring as the ICP guideline is being assessed and the patient does not have a mass lesion or diffuse injury recorded (all others do).

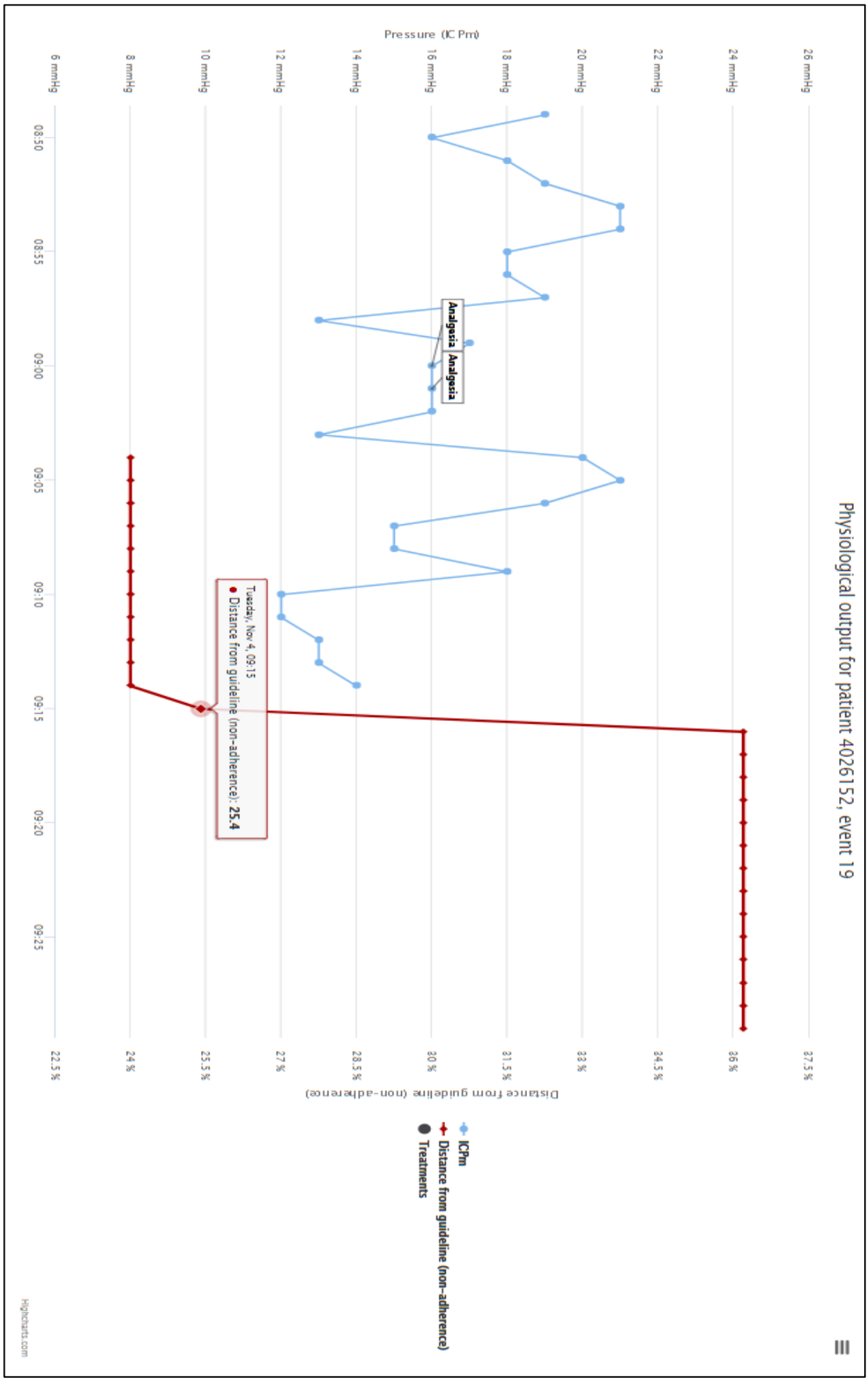

Figure 9.18: Patient 4026152, event 19

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
11	24	Type (10.0), Time to treatment (14.0)
1	25.4	Type (10.0), Time to treatment (15.4)
13	28	Type (10.0), Time to treatment (18.0)
552	36.2	Treatment missing (36.2)

The following items also show overall durations for this patient:

Total time of patient stay (mins): 12630
Total time of patient requiring clinical action after an event (mins): 577
Total duration of "default" non-adherence state (mins): 552.0
Total time spent in non-adherence, which is not in the "default" state (mins): 25.0
Expressed as a percentage of all the time spent requiring clinical action: 4.33 %

Figure 9.19: Total table for patient 4026152

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	24	36.2	35.76	36.2	36.2	36.2
Duration (mins)	1.0	15.0	26.23	17.5	33.0	115.0
Duration / Non-adherence	0.04	0.43	0.74	0.48	0.91	3.18
Duration * Non-adherence	25.4	543	937.99	633.5	1194.6	4163

Figure 9.20: Interquartile range table for patient 4026152

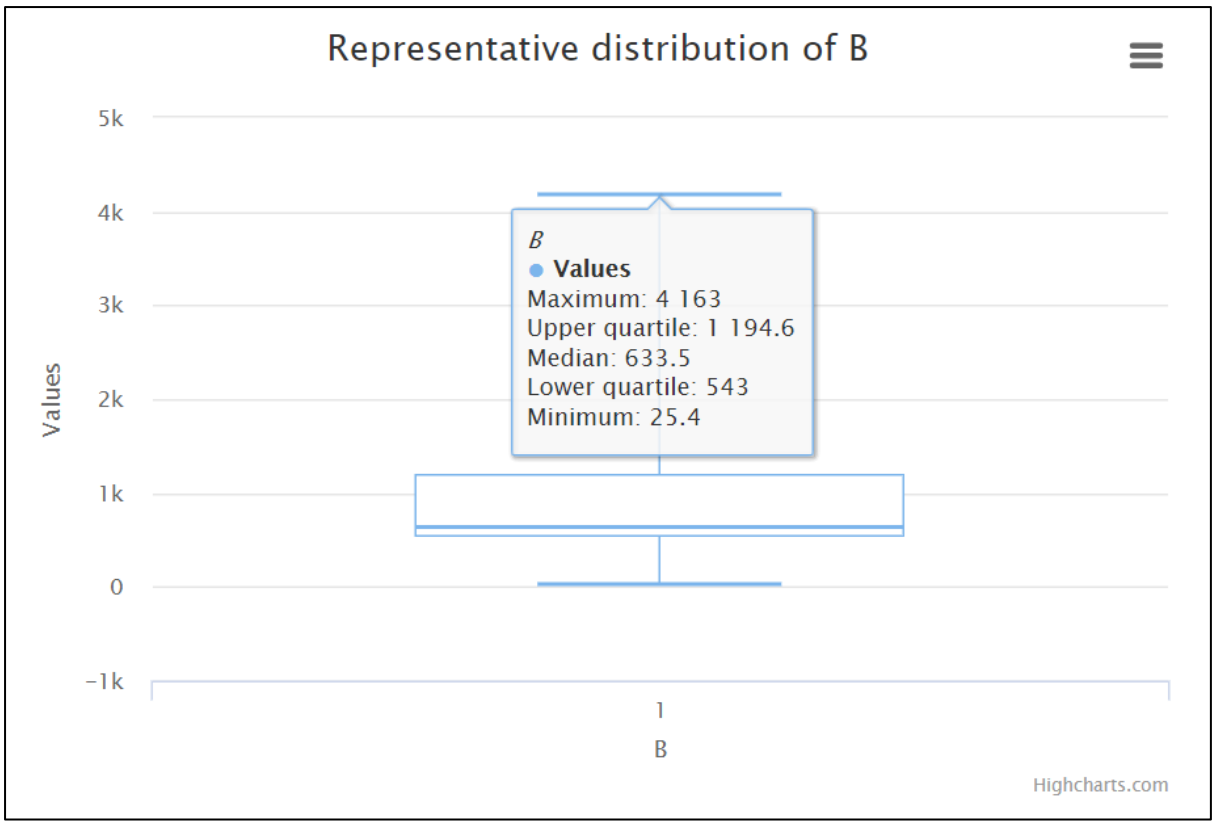

*Figure 9.21: Box-plot visualisation of the B metric (Duration * Non-adherence) for patient 4026152*

9.3.1.5. Patient 64816161

This patient was chosen as they had verifiable "too high" administrations of a particular treatment category (Analgesia). However, these did not register on the framework, possibly because the administration occurred outside a reaction time window.

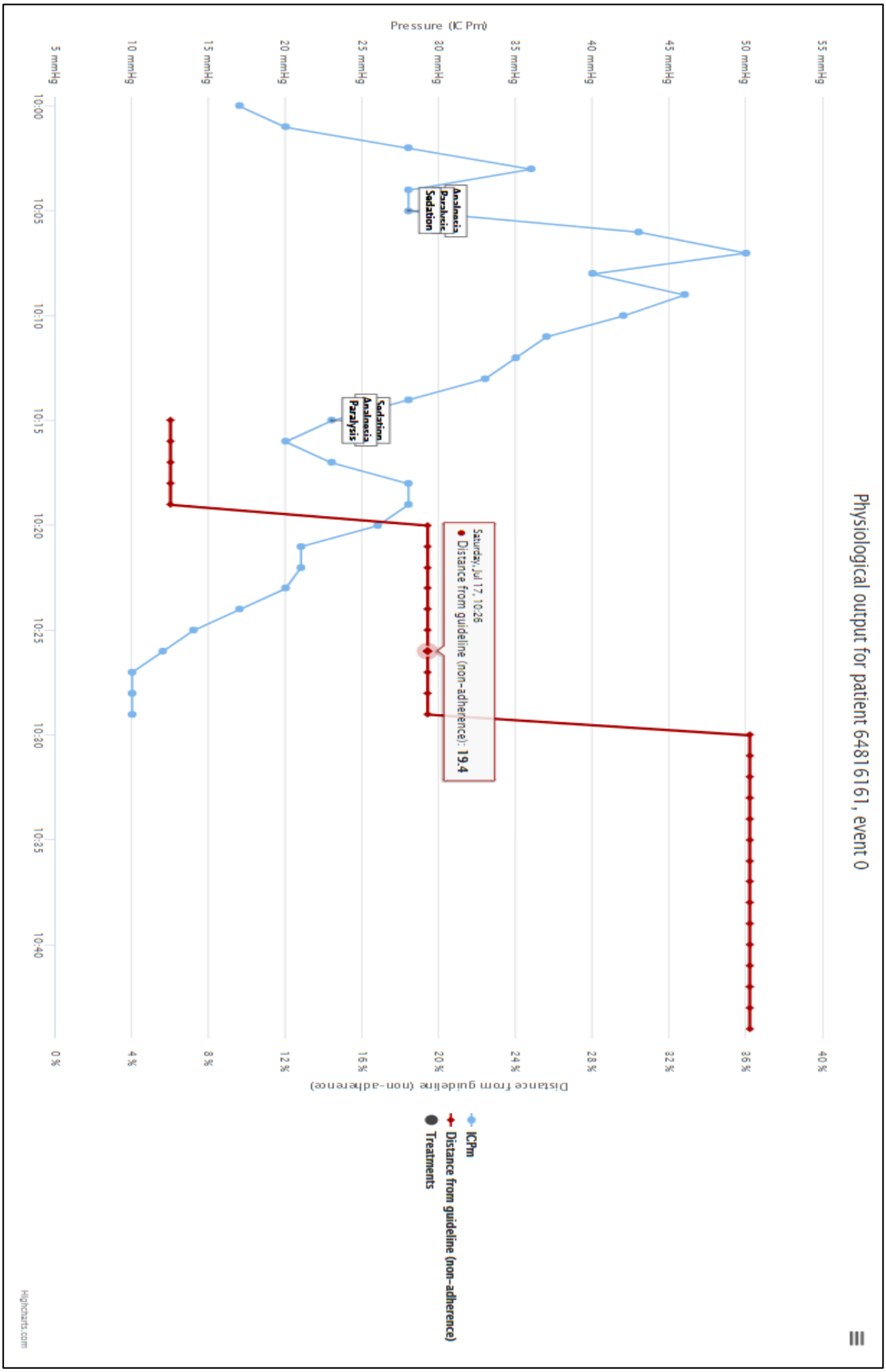

Figure 9.22: Patient 64816161, event 0

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
1	0.6	Time to treatment (0.6)
5	6	Time to treatment (6.0)
6	7.4	Time to treatment (7.4)
16	10	Time to treatment (10.0)
5	16.6	Time to treatment (16.6)
10	19.4	Time to treatment (19.4)
1	20	Time to treatment (20.0)
177	36.2	Treatment missing (36.2)

The following items also show overall durations for this patient:

Total time of patient stay (mins): 55125
Total time of patient requiring clinical action after an event (mins): 221
Total duration of "default" non-adherence state (mins): 177.0
Total time spent in non-adherence, which is not in the "default" state (mins): 44.0
Expressed as a percentage of all the time spent requiring clinical action: 19.91 %

Figure 9.23: Total table for patient 64816161

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	0.6	36.2	31.4	36.2	36.2	36.2
Duration (mins)	1.0	5.5	13.81	11.5	16.5	52.0
Duration / Non-adherence	0.05	0.4	0.67	0.55	0.81	1.67
Duration * Non-adherence	0.6	62.2	433.71	332.3	597.3	1882.4

Figure 9.24: Interquartile range table for patient 64816161

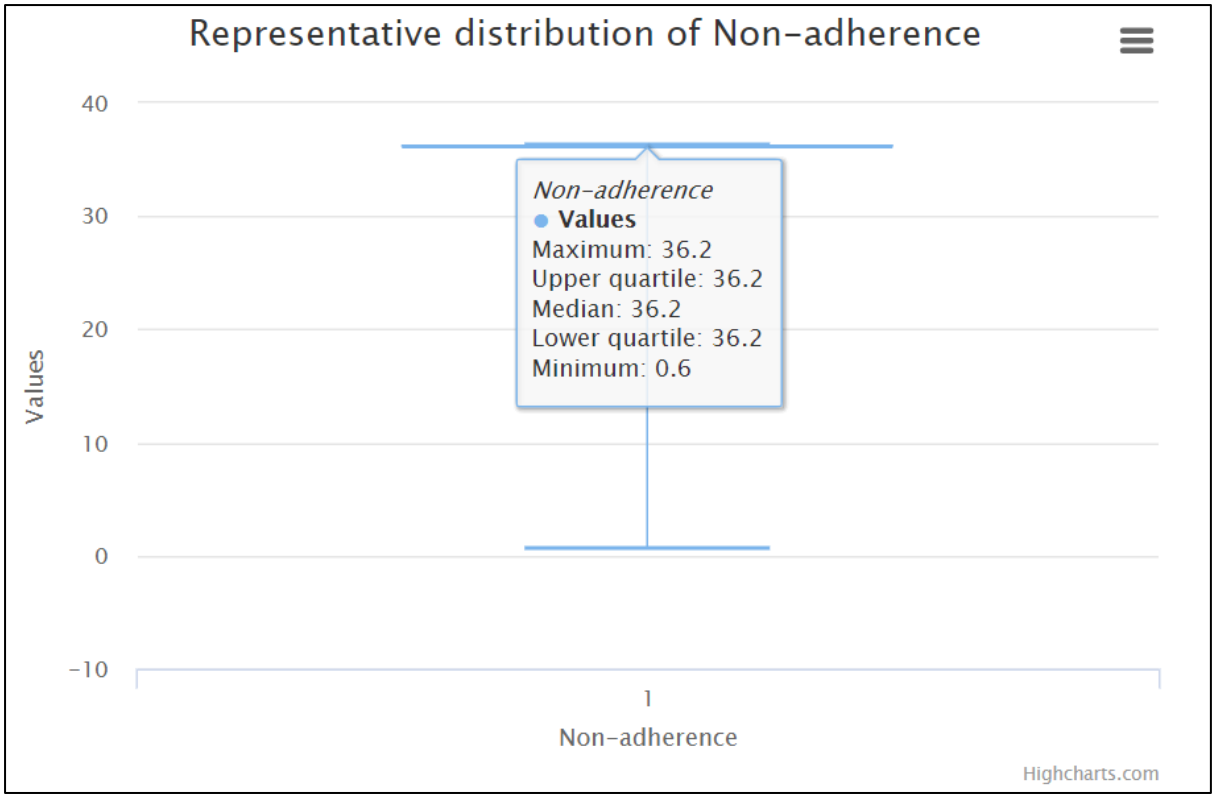

Figure 9.25: Box-plot visualisation of the Non-adherence aspect for patient 64816161

The five patients show various classes of non-adherence. Most have a large time period where treatment annotations haven't been associated with corresponding EUSIG-events (see section 9.3.2). This large presence of default time periods can be seen in the highly skewed interquartile ranges (see the box-plots in figures 71, 74 and 87). However, the multiplicative nature of the A and B metrics causes the distribution of values to be much more evenly spread.

Some patient records are very adherent in places, for instance patient 4026626 shows a distance of only 2% within a specified time window as the treatment annotation is present, recommended and not (yet) over a considered dosage limit. Patient 26138262 shows non-adherence due to dosage limits being breached. Patient 64816161 should show similar dosage output but doesn't due to the (suspected) reason that this has occurred outside the time window. Patient 4026152 shows a contribution of type, as they do not have a mass lesion or diffuse injury, which is an assessed component of the ICP guideline. The instance shown in patient 15138374 shows a spike due to the presence of a treatment, the time-window expiring, then another treatment immediately bringing the adherence distance down again.

9.3.2. "Default" instances of non-adherence

Three classes of time period can be applied to this analysis of non-adherence:

- Time period where a clinical reaction is not required
- Time period where a clinical reaction is required but not provided
- Time period where a clinical reaction is required and is provided

A result that has emerged from the listing of non-adherence instances for a large dataset such as Brain-IT, was that for every patient, there would be a long period of a single level of non-adherence. This has been referred to as the "default" instance for a patient and is likely represented by the second category (clinical reaction required, but not provided). A possible clinical interpretation of this is that there are often periods where a clinician must gauge a patient's status during a EUSIG-event, before administering a treatment (taking time to make a clinical judgement rather than immediately following the guideline). It is also likely to be due to instances where a clinical prognosis is so poor that all interventions or none (palliative care) are enacted (Dr Chris Hawthorne, University of Glasgow, 31st July 2018, pers.comm). A third – and very likely – reason is that the annotations have not been recorded, an estimation of the error that this phenomenon introduces was the purpose of the validation step taken in section 8.2.

An example of this can be seen in figures 88 and 89 (for patient 15138374).

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
6	27.4	Time to treatment (7.4), Dose (20.0)
8	28.6	Time to treatment (8.6), Dose (20.0)
8	30	Time to treatment (10.0), Dose (20.0)
10	31.4	Time to treatment (11.4), Dose (20.0)
1	32.6	Time to treatment (12.6), Dose (20.0)
11	34	Time to treatment (14.0), Dose (20.0)
1	35.4	Time to treatment (15.4), Dose (20.0)
13	36.6	Time to treatment (16.6), Dose (20.0)
15	38	Time to treatment (18.0), Dose (20.0)
1	39.4	Time to treatment (19.4), Dose (20.0)
1998	40	Time to treatment (20.0), Dose (20.0)

Total stay time (mins): 55125
Total time within events (mins): 2072
Duration of "default" state: 1998.0
Duration of "non-default" state: 74.0 (3.7 %)

Figure 9.26: The total table for patient 15138374 with the "default" state highlighted

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
6	27.4	Time to treatment (7.4), Dose (20.0)
8	28.6	Time to treatment (8.6), Dose (20.0)
8	30	Time to treatment (10.0), Dose (20.0)
10	31.4	Time to treatment (11.4), Dose (20.0)
1	32.6	Time to treatment (12.6), Dose (20.0)
11	34	Time to treatment (14.0), Dose (20.0)
1	35.4	Time to treatment (15.4), Dose (20.0)
13	36.6	Time to treatment (16.6), Dose (20.0)
15	38	Time to treatment (18.0), Dose (20.0)
1	39.4	Time to treatment (19.4), Dose (20.0)

Total stay time (mins): 55125
Total time within events (mins): 2072

Figure 9.27: The total table for patient 15138374 with the "default" state removed

This information can be useful in itself to understand the nature of guideline adherence (e.g. do different management strategies have different "wait and see" times, and if so, does this impact on the patient outcome). However, a secondary step to this is to remove this "default" instance, and thereby understand what variation contributes to a non-adherence score, when in the more "active" time periods of clinical management.

Figures 9.26 and 9.27, show one patient with this default state retained (figure 9.26) and the state removed (figure 9.27). The variation of non-adherence reasons is evidently different in these two cases, which gives potentially more detailed information about the reasons/levels for non-adherence during the patient stay (figure 9.28). This feature of being able to view the spread of non-adherence instances with and without the "default" instance showing is built into the web application as a "view toggle".

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	27.4	30	33.44	34	36.6	39.4
Duration (mins)	1.0	1.0	5.69	6.5	10.0	14.0
Duration / Non-adherence	0.03	0.03	0.17	0.23	0.31	0.37
Duration * Non-adherence	28.6	36.7	190.34	182.3	328.3	532

Figure 9.28: Interquartile range table for patient 15138374, but with default state removed

Figures 9.29 to 9.36 show the difference between all four aspects for this patient (15138374) when the default state is retained (figures 9.29 to 9.32) or removed (figures 9.33 to 9.36).

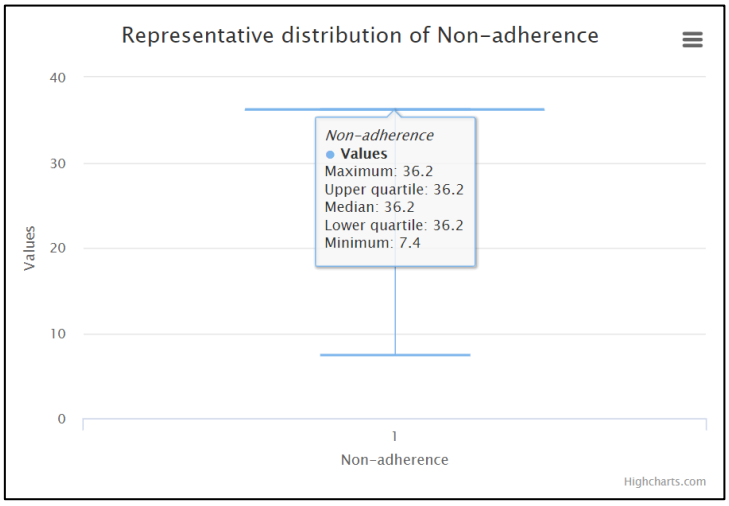

Figure 9.29: Interquartile range for Non-adherence aspect for patient 15138374 (default state retained)

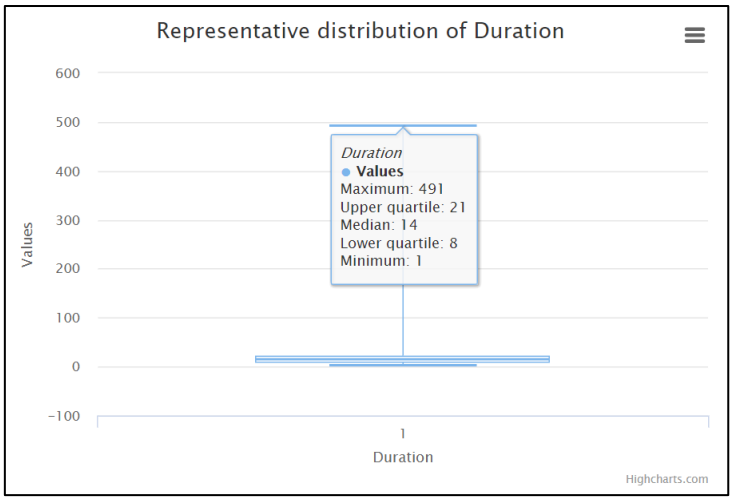

Figure 9.30: Interquartile range for Duration aspect for patient 15138374 (default state retained)

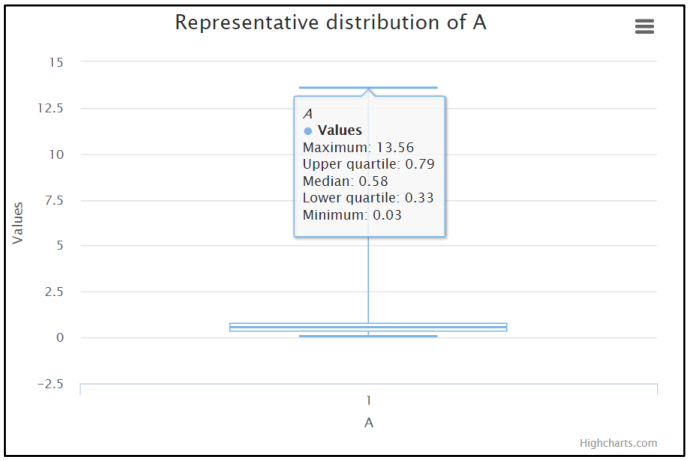

Figure 9.31: Interquartile range for (Duration / Non-adherence) aspect for patient 15138374 (default state retained)

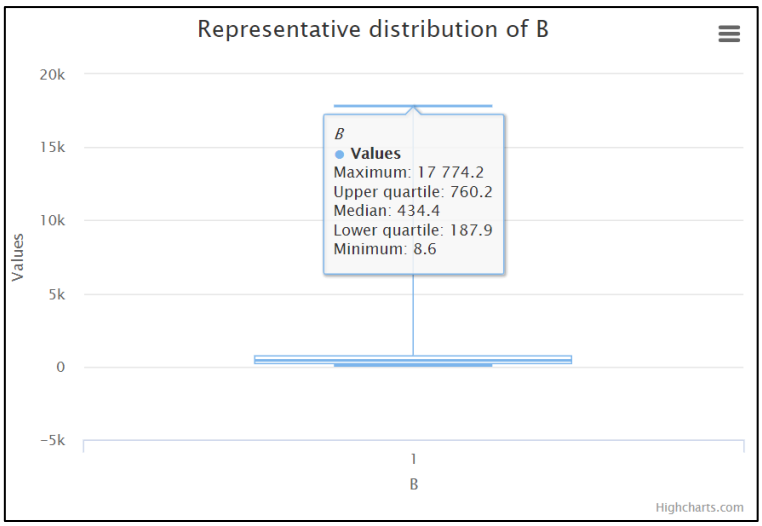

*Figure 9.32: Interquartile range for (Duration * Non-adherence) aspect for patient 15138374 (default state retained)*

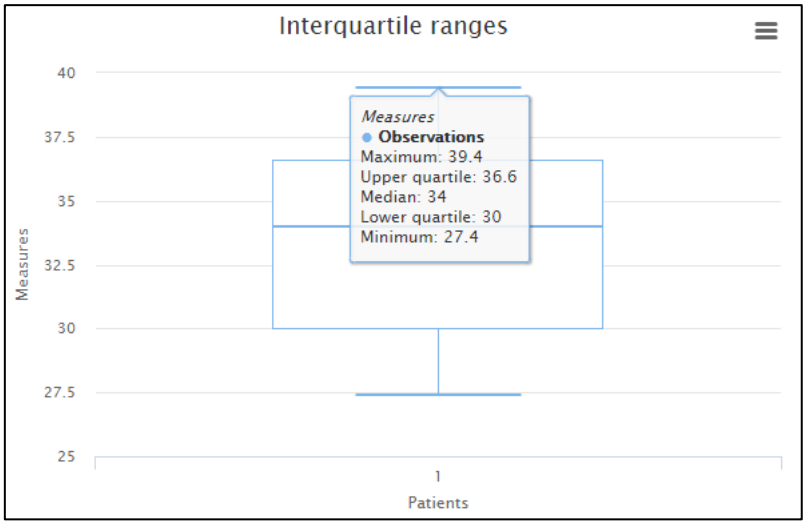

Figure 9.33: Interquartile range for Non-adherence for patient 15138374 with default state removed

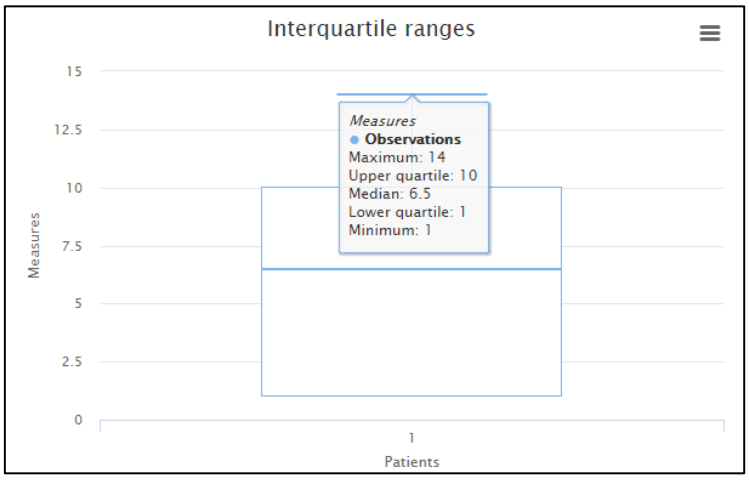

Figure 9.34: Interquartile range for Duration for patient 15138374 with default state removed

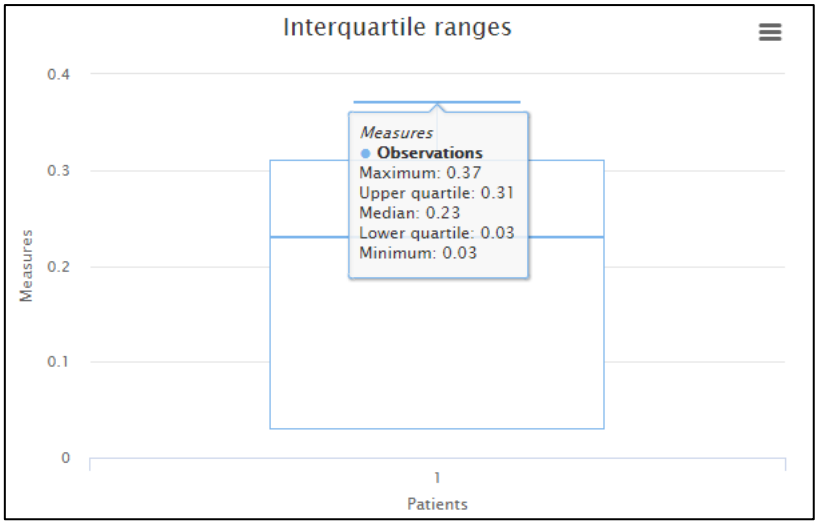

Figure 9.35: Interquartile range for (Duration / Non-adherence) for patient 15138374 with default state removed

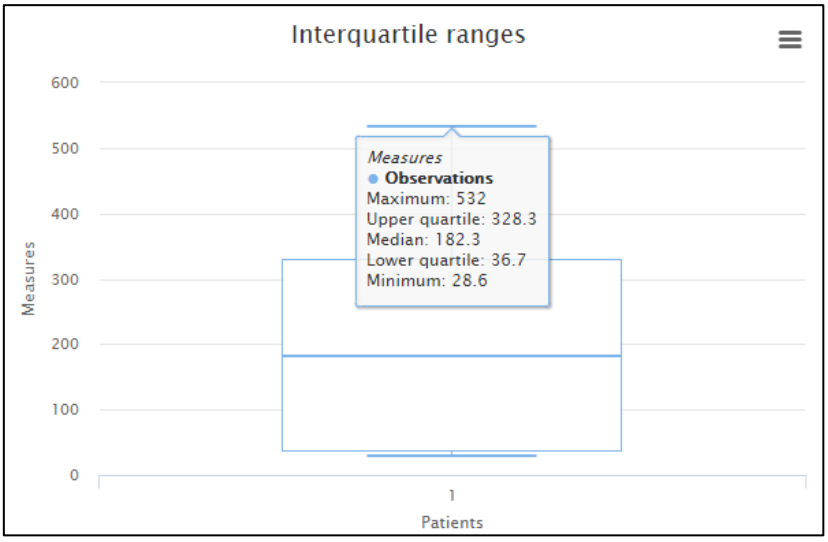

*Figure 9.36: Interquartile range for (Duration * Non-adherence) for patient 15138374 with default state removed*

9.3.3. Variation in reasons for non-adherence

The variation in reasons for non-adherence, is shown for all five patients in figures 9.37 to 9.41. This lists all the instances occurring in a single patient stay and visualises the contribution of the reasons to the overall non-adherence level.

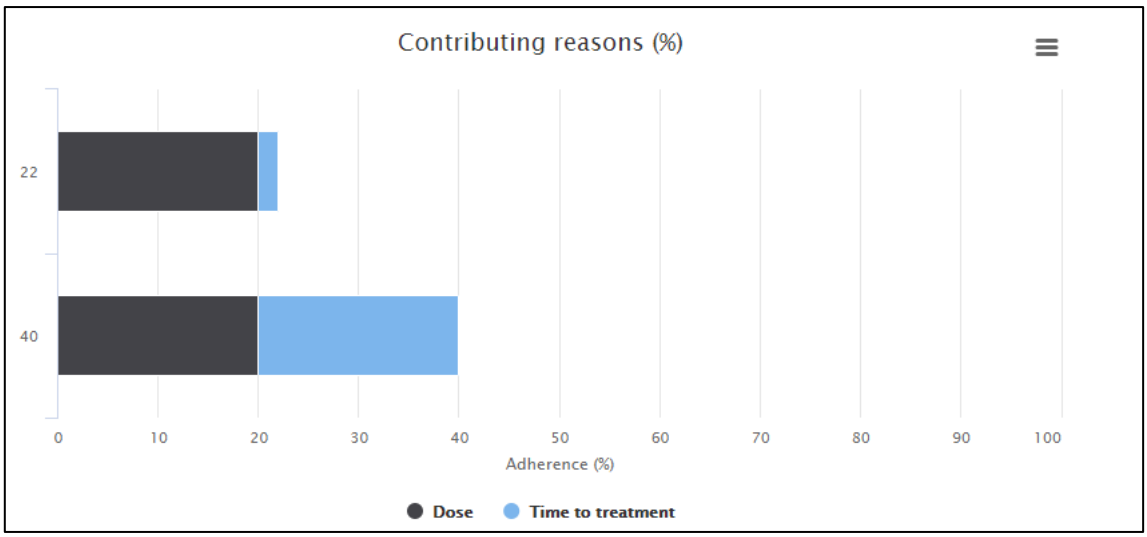

Figure 9.37: Contributing reasons to the variation in 4026626

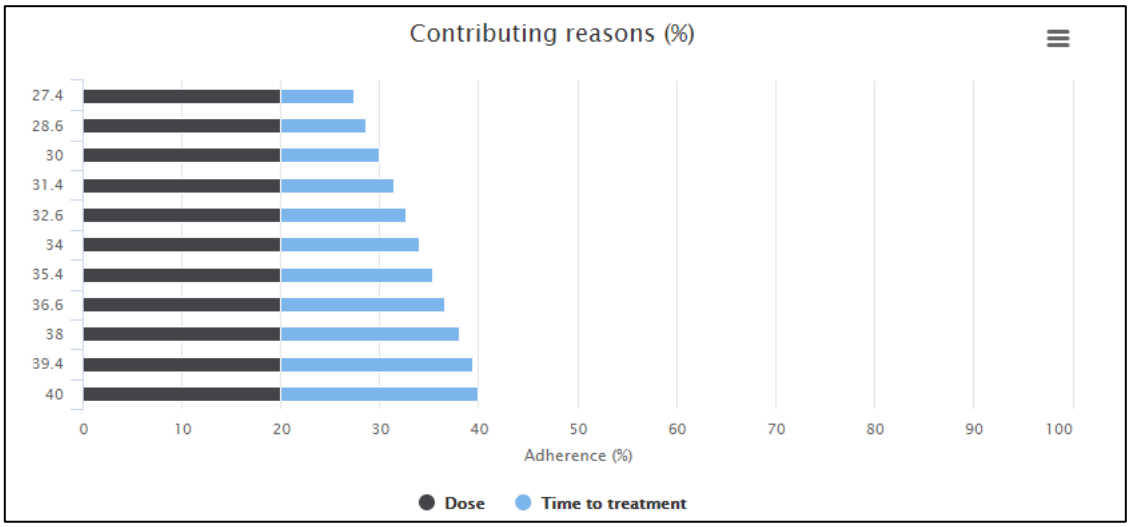

Figure 9.38: Contributing reasons to the variation in 15138374

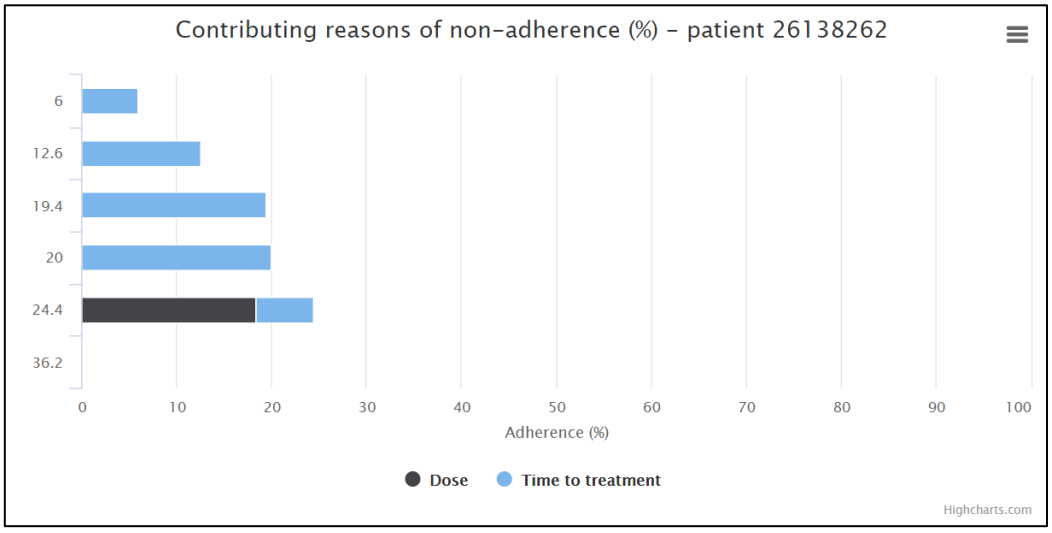

Figure 9.39: Contributing reasons to the variation in 26138262

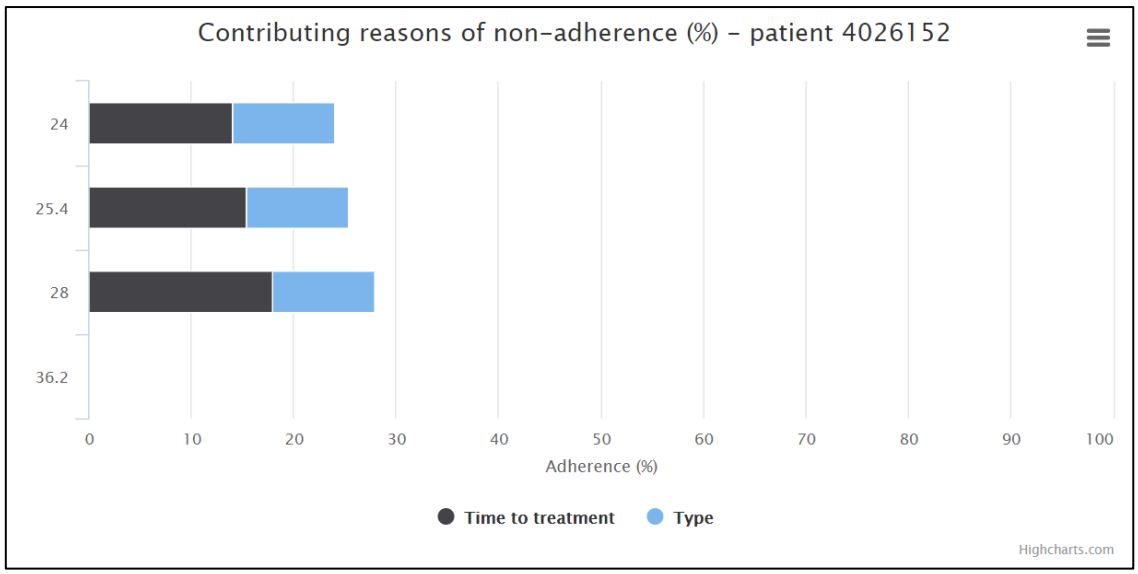

Figure 9.40: Contributing reasons to the variation in 402615

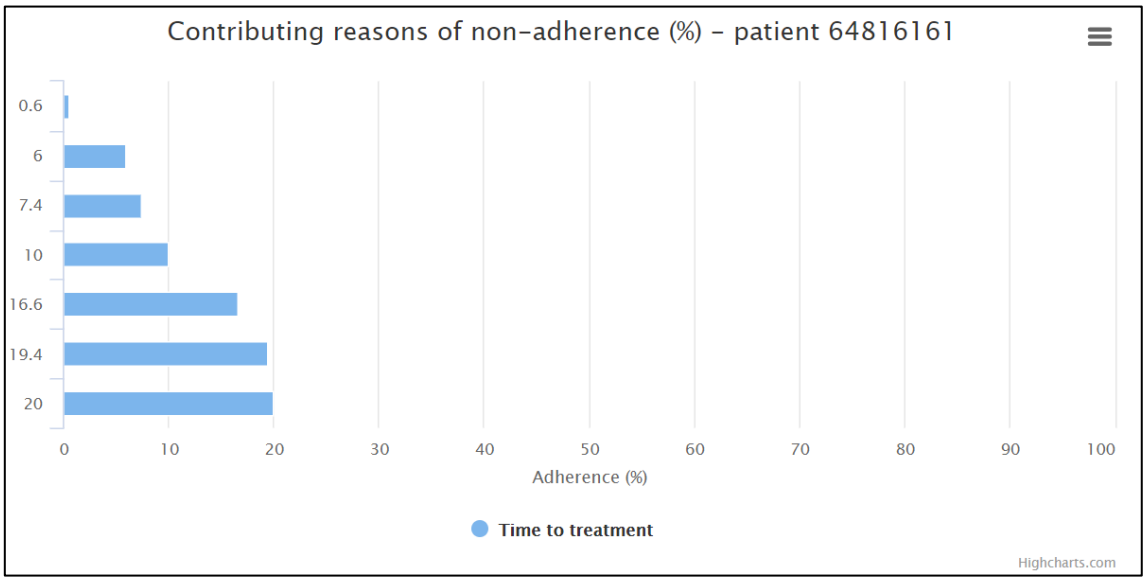

Figure 9.41: Contributing reasons to the variation in 64816161

This feature allows the full qualitative composition of the overall quantitative score to be shown (or expressed another way: the trail of reasons for this non-adherence).

9.3.4. Overall clinical severity of non-adherence

Finally, the clinical severity of the non-adherence instances, by combining duration and distance, is evaluated and presented using the risk-analysis grids (figures 9.42 to 9.46), which are the mean of the combinatorial metrics A (duration / distance) and B (duration * distance). The resulting quadrants are: bottom-left

(green) is least severe, top-right (red) most severe, and the opposing two (blue) are mid-range severity.

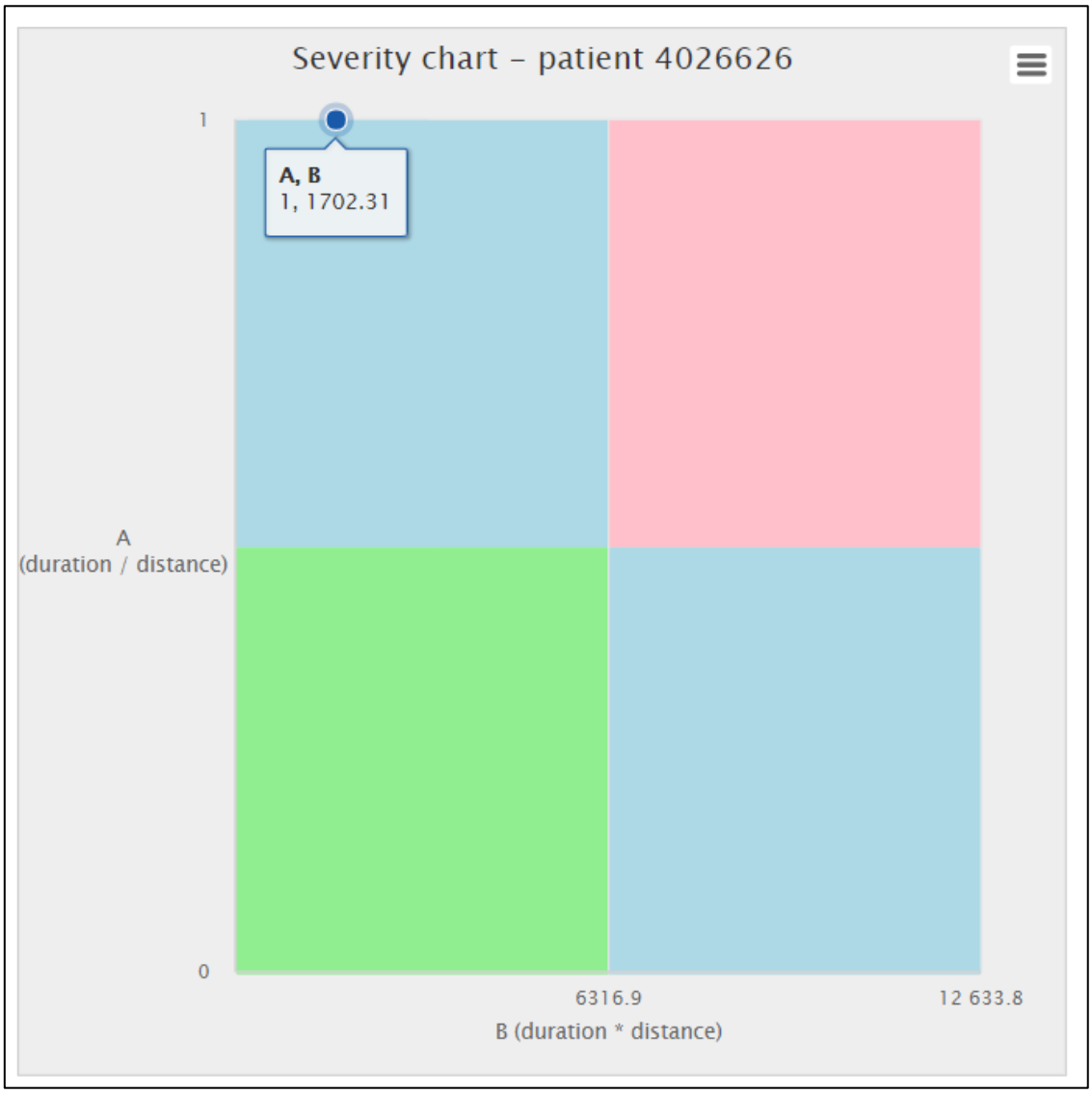

Figure 9.42: Severity chart for patient 4026626

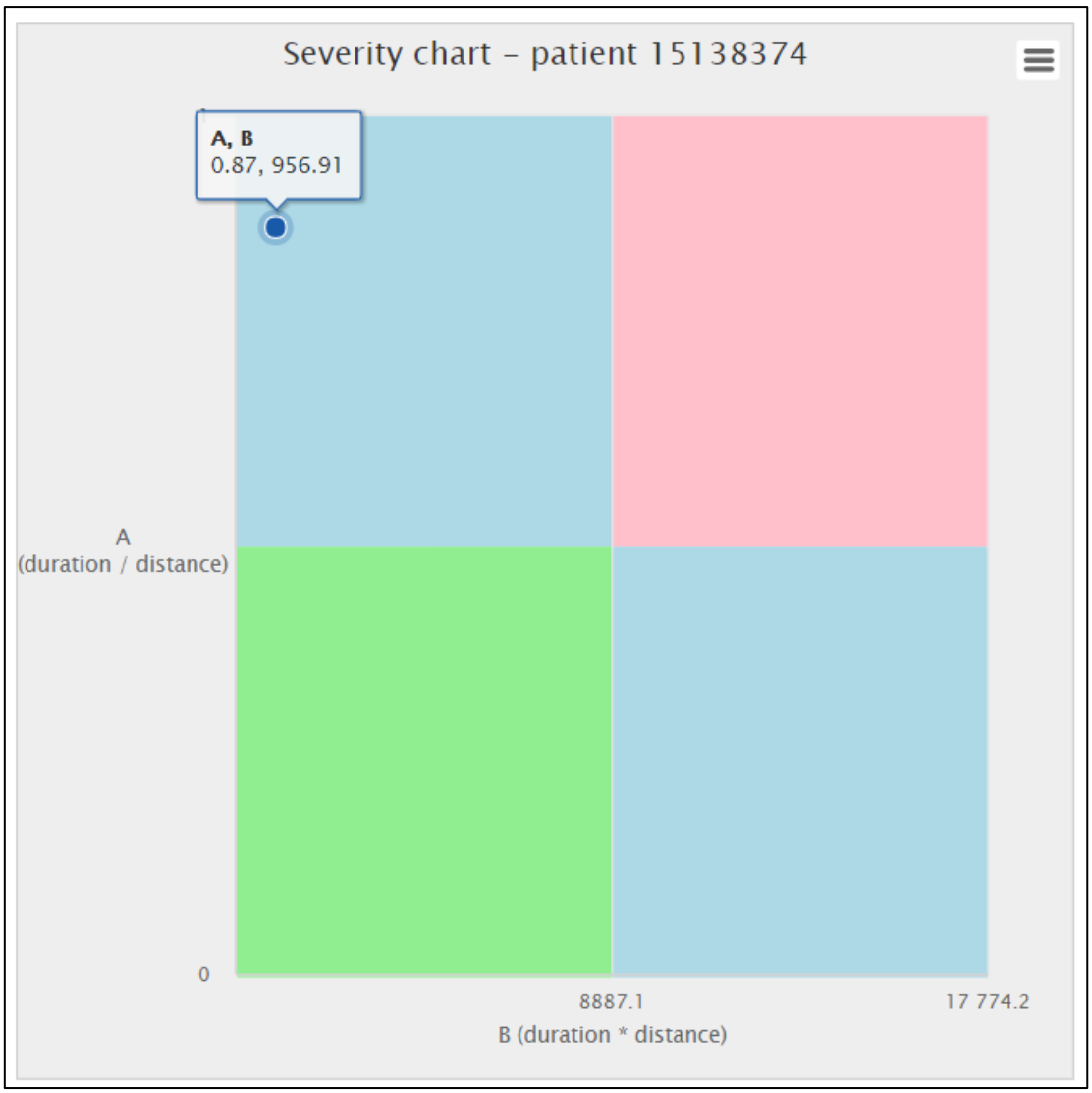

Figure 9.43: Severity chart for patient 15138374

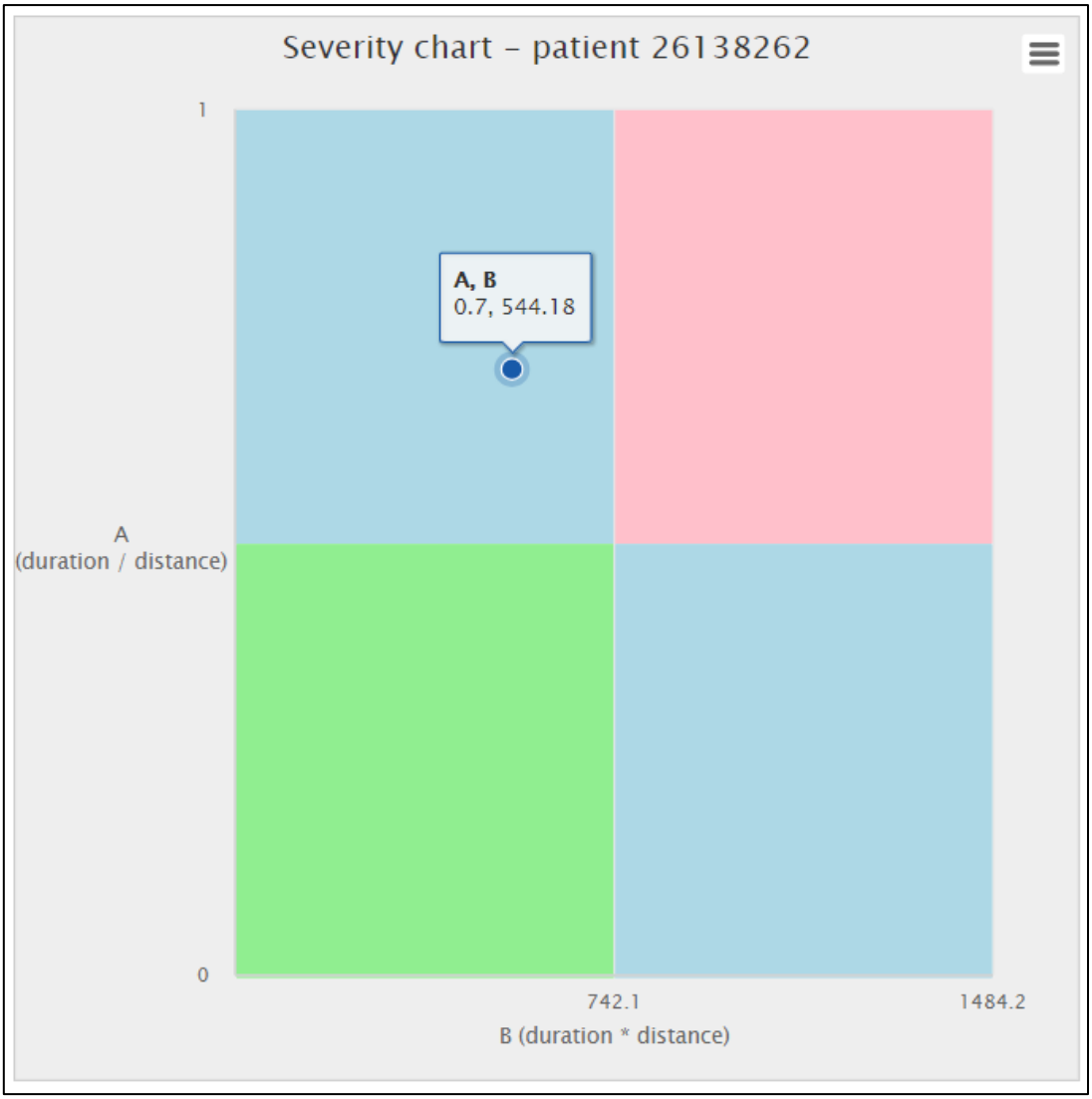

Figure 9.44: Severity chart for patient 26138262

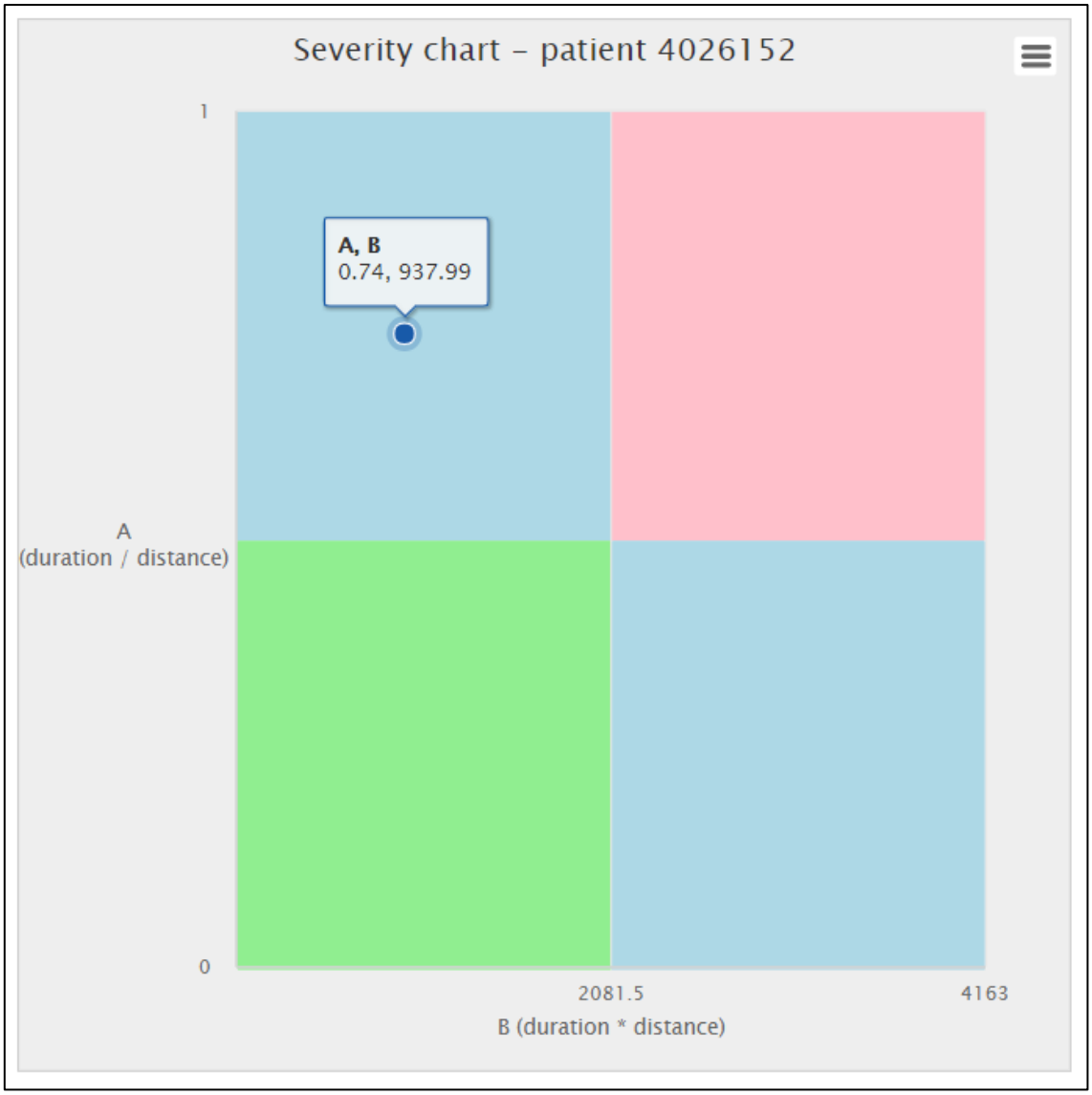

Figure 9.45: Severity chart for patient 4026152

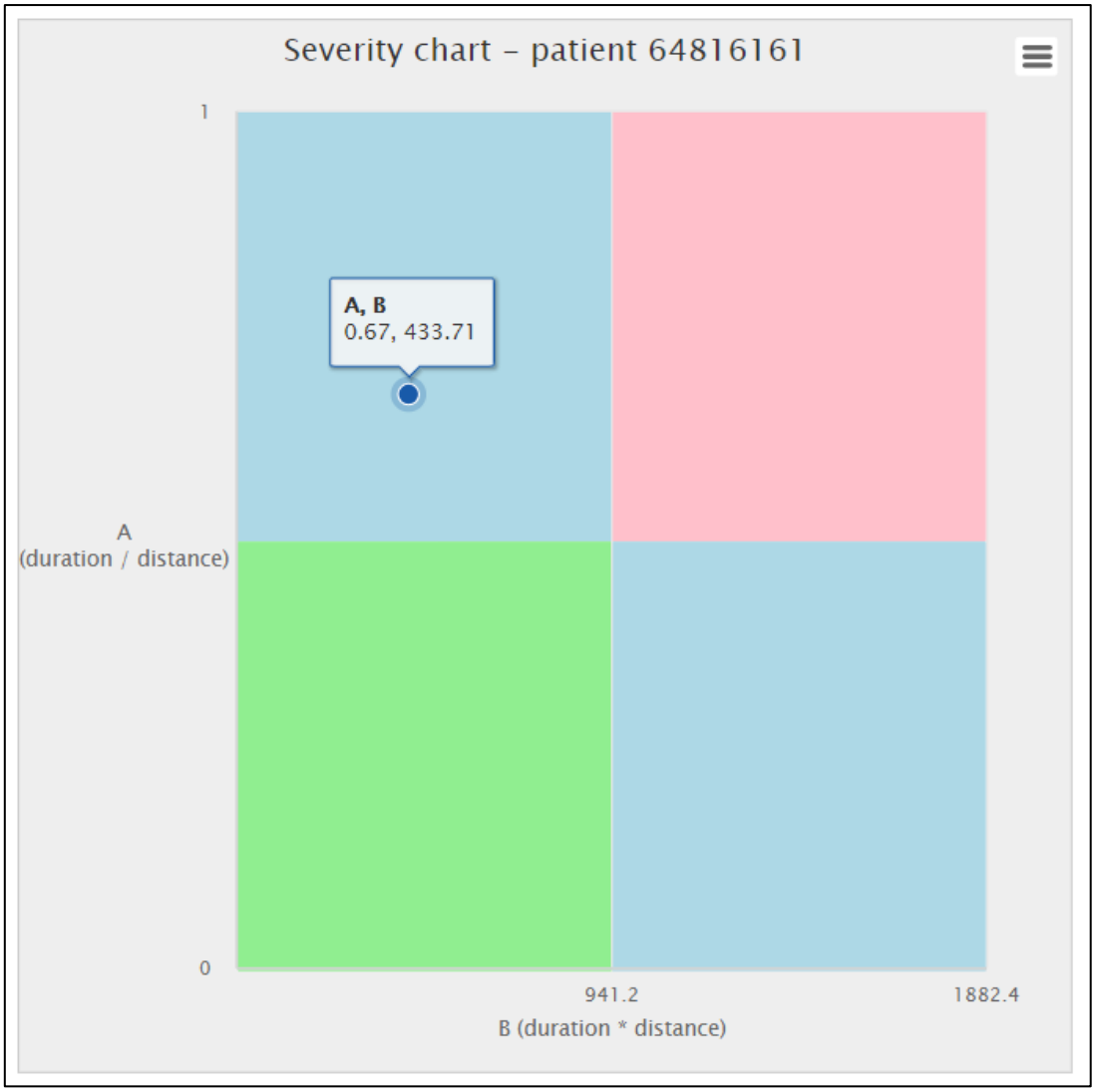

Figure 9.46: Severity chart for patient 64816161

All five of the severity charts generated by these Brain-IT examples, give severity indicators that occupy the top-left (mid-range) quadrant. The interpretation is that despite variations in the adherence information, in aggregate the instances are providing approximately similar output. The likelihood is that this common output is due to a similar volume of treatment annotations within the dataset (as discussed in section 9.2, the association of EUSIG-event and treatments are not numerous), so the output is dominated by the "treatment missing" non-adherence reason.

9.4. Additional dataset information

Two points of additional interest were derived from the Brain-IT dataset. First was the impact of laying the time-window over the physiological/treatment

information, in terms of the events counted. Second was the derivation of the distribution of treatments

9.4.1. Impact of treatment association time-window on event counts

Additional to the event log, the association of a treatment with an event provides a modifying parameter to the overall event number count. An inference that can be made is that this modification indicates a “preferred” event definition that clinicians are more likely to react to. 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, shown in figure 9.47, when compared to the unmodified count in figure 9.1 (section 9.1.3).

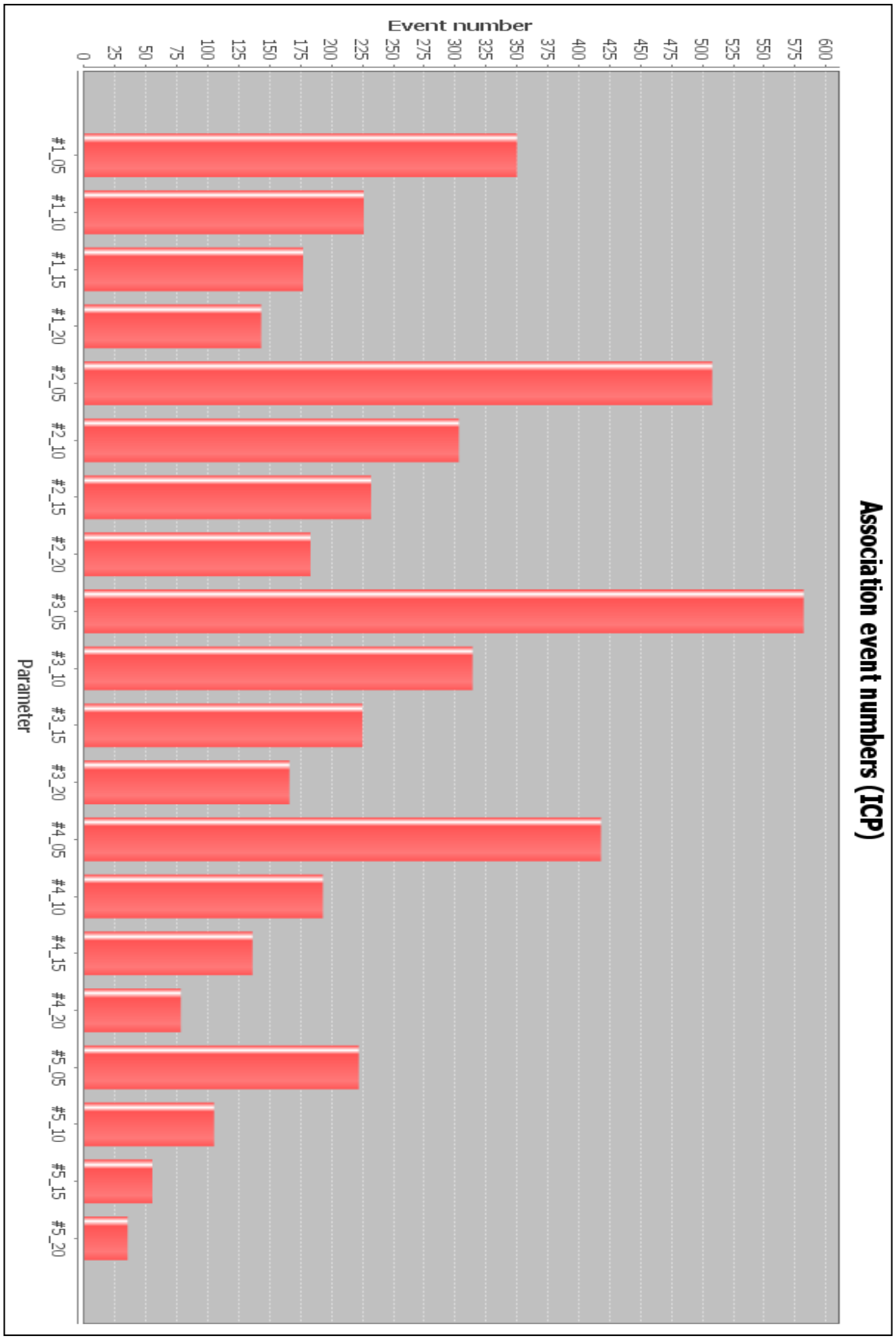

Figure 9.47: Number of ICP events that have treatment associations when a time window of 30 minutes is applied

The graph shape in figure 9.47 is evidently different from the event count numbers in isolation in figure 9.1. It is now the definition of ICP > 20 mmHg with hold-down of 5 minutes that appears to be most numerous, rather than ICP > 15 mmHg, which would suggest that this definition input is triggering a larger number of clinical responses (as administered treatments). A further clinical interpretation of these graphs can be seen in the shape of the distribution as the time-window increases towards the asymptote of infinite time. According to the association algorithm presented in section 5.1.4 the number of events with treatment associations will approach the total event number as the time-window approaches infinity (i.e. with a sufficiently large time-window, every event will have an associated treatment). Figure 9.48 demonstrates this progression of the distribution shape of the ICP events with treatment associations as they move through the other three time-window definitions (10, 15 and 20 minutes). The result is that definition #2 gradually predominates again.

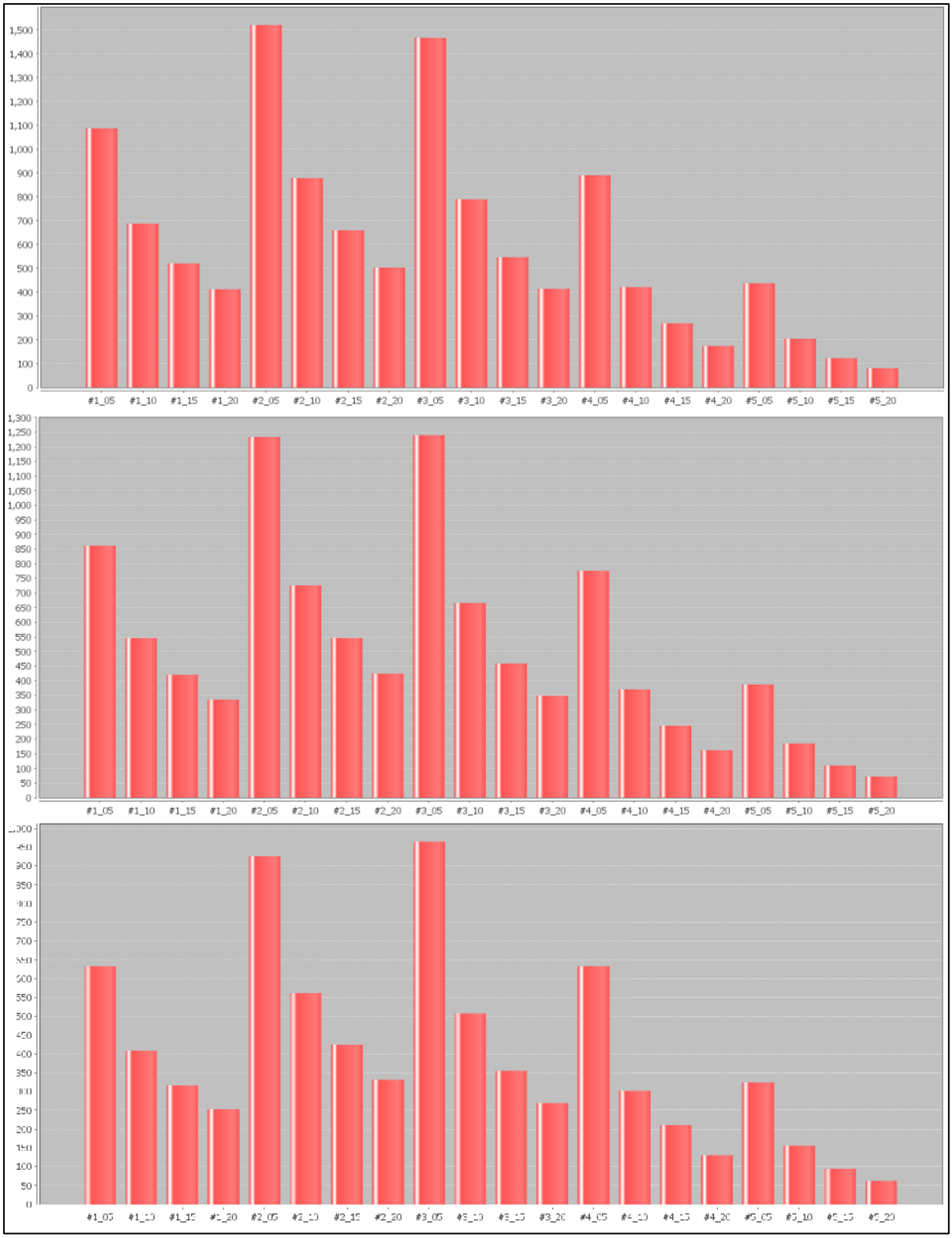

Figure 9.48: The three other time-window definitions (20, 15 and 10 mins) for ICP event and treatment association, with event count on vertical axis and ICP definition/time window on horizontal (labels removed for space). As the time-window increases (top to bottom), the distribution shape reverts back to that of the event count without treatment association.

9.4.2. Treatment distribution

For every combination of EUSIG definition and treatment association time-window, a composition of the treatment categories included in the list can be constructed. This allows a "map" of predominating treatment protocols to be constructed which could be, for instance, later used as a refinement on the specific treatments outlined in a guideline. Figure 9.49 shows the treatment distribution for an ICP threshold of > 20 mmHg with a hold-down value of 5 minutes and a time-window for association of 30 minutes (selected due to its predominance discussed in section 9.2.3). The top three treatments applied in this instance are paralysis (18.2%), sedation (17.2%) and osmotic therapy (16.2%), from an absolute number of 582 events with treatment associations.

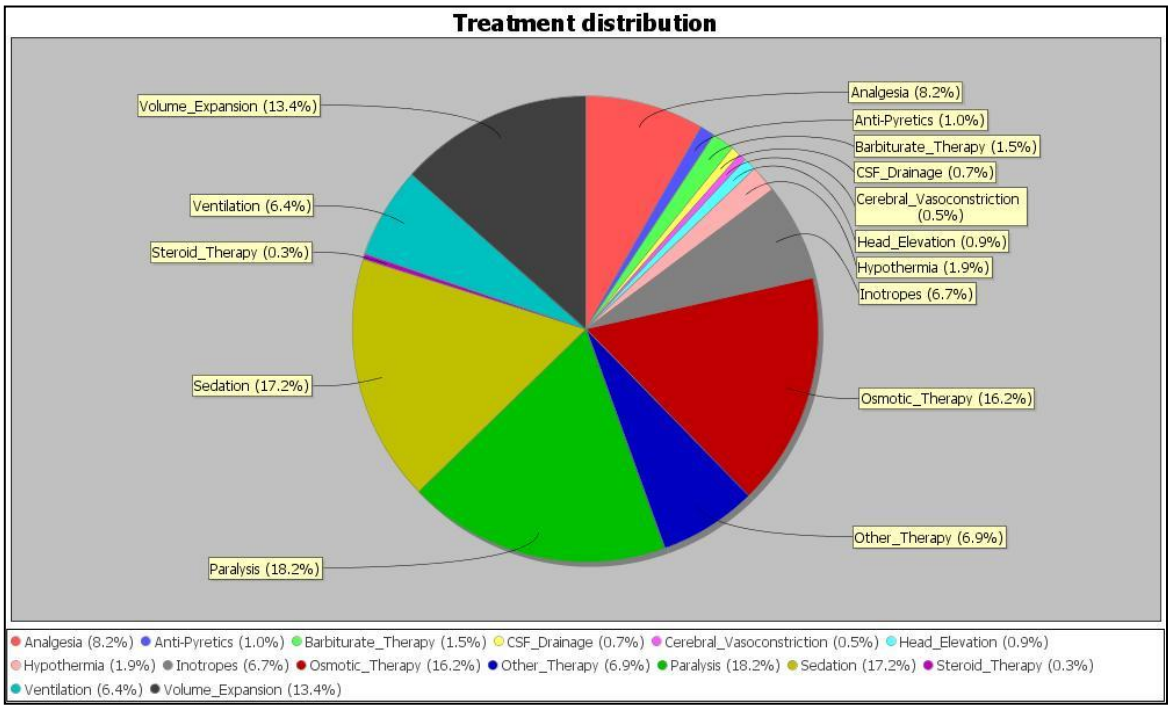

Figure 9.49: Treatment distribution for an ICP threshold > 20 mmHg with a hold-down value of 5 minutes and a time-window of association of 30 minutes

9.4.3. Centre-specific information

Similarly, using the unique centre reference identifier in the dataset, the same contributing information can be used to create a treatment composition for each individual centre. For example, the top three treatments for the centre in Uppsala, Sweden were paralysis (32.9%), analgesia (13.4%) and a joint third place (11.4%) for ventilation, volume expansion and sedation, from an absolute number of 373 events with treatment associations (figure 9.50).

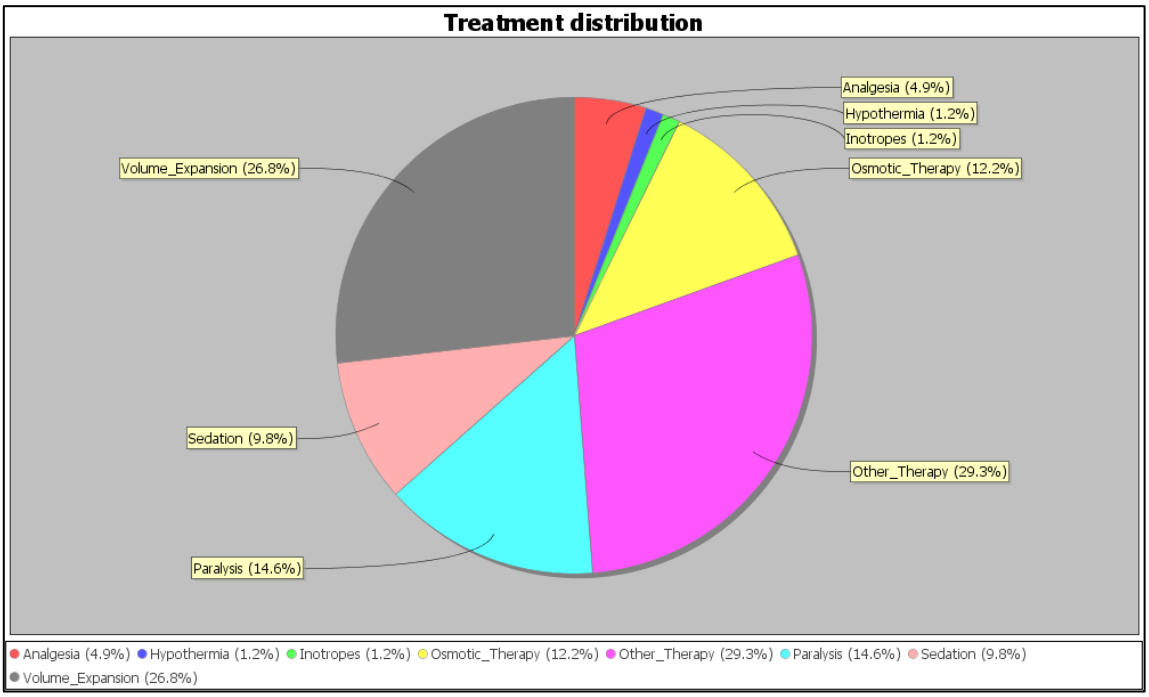

Figure 9.50: Treatment distribution from one centre (with highest coverage: Uppsala, Sweden) for an ICP threshold of > 20 mmHg with a hold-down value of 5 minutes and a time-window of association of 30 minutes

9.5. Discussion

Overall, the physiological data of the Brain-IT dataset is comprehensive and the EUSIG event pattern can be readily extracted. However, the annotation frequency and density in the dataset overall is low, leading to low association numbers with EUSIG events (and all other assertions made must also be qualified with this consideration). This is mainly due to the time of data collection (mid 2000's) where the available technology was manual inputs to PDA Palm Pilots. Modern ICU technologies, such as Philips ICCA show a much higher density of treatment annotations, so a higher association number, and therefore more accurate representation of clinical management can be assessed. This potentially relates to the overall issues confronting the TBI medical community, as the increase in treatment annotations can be used to increase the power of studies and trials.

However, despite this low resolution, clear indications of the adherence output can be found within the dataset, highlighted by the patient examples in this chapter (e.g. treatment presence, dosage, type differences). This suggests that, though requiring refinement, the framework does provide viable output for measuring guideline adherence.

10. Application of framework to non-specialist ICU dataset: MIMIC III

Chapter summary

The MIMIC III dataset was analysed using the adherence framework. The purpose of this was to explore whether a non-specialist ICU dataset could show guideline adherence output to a similar degree as datasets that have a specialist focus, such as Brain-IT.

The dataset had different extraction challenges than Brain-IT. To extract TBI-specific injuries, a survey of the ICD9 codes relating to brain injury had to be performed. The semantic and heterogeneous nature of the repository made the output of this process unpredictable (e.g. using wildcard matching for drug names). However, data was extracted using four physiological codes (two each for ICP and CPP), and three ICD9 codes describing brain injury. From this, 100 subjects were identified, and their physiological output traced. Of this 100, seven had viable treatment annotations (other than blank or null inputs) and event identification and treatment association were performed for these seven patients.

Due to the low-resolution of the dataset (end-hour averaged physiological values), a different approach for event identification and treatment association was also required. In this case, an event was considered to be active if the end-hour data-point was above the appropriate EUSIG threshold, and a treatment that occurred within an hour of this was considered to be associated (c.f. the EUSIG definitions and association windows described in chapter 6).

Some guideline adherence information was captured but the association figures were very low. Combined with the low-resolution of the physiological data, the confidence in results and representation was not high.

The MIMIC III dataset is a collection of non-specialist ICU data collected from 2001 to 2012 across different intensive care units in the United States (Saeed, 2007). The purpose of the collection is a comprehensive repository of de-identified ICU data specifically for secondary studies.

The application to this research was the use of a non-specialist dataset to obtain guideline adherence output in an environment where ICP/CPP management, and specialist management of brain injury, is not (necessarily) the primary focus. The generalisation of this process to other datasets will give feedback to the general viability of the framework.

10.1. Dataset description

The data is provided in a set of zipped ASCII text files, exported from a central database, with most data being timestamped key-value pairs, which are interpreted using a dictionary lookup of ICD9 codes. The data itself is protected against identification, using standard measures such as the use of an anonymous identifier (individual patient records can be distinguished, but the record itself cannot be traced back to a real identity). A further protection implemented by MIMIC is the application of anonymised timestamps – again the individual data points can be associated together for the purposes of analysis but are set at a time in the future (e.g. “01-04-2157” is a standard timestamp in the physiological readings).

A major feature of this dataset is that it is very low resolution – each physiological data-point is an end-hour averaged value. This has an impact on the certainty of association of treatment with event and makes a standard EUSIG event more difficult to extract.

The labels used have a wide range of free-text variations on the full ICD9 dictionary and apply to all the labels of everything occurring in an ICU. Standard physiological terms, such as “ICP” and “CPP” are relatively well defined, but other terms such as “craniospinal fluid” have many different abbreviations and reference terms (e.g. “CSF”, “CS fluid”, etc). This makes categorisation – particularly of relevant treatment annotations – challenging.

10.2. Method

The method to extract guideline adherence information from this dataset was similar to the Brain-IT method, but the low-resolution nature of the MIMIC dataset, required two different features:

- The definition of an active EUSIG pressure event required a single point to be above the threshold (as this implies that the average value over an hour has already been above said threshold). Therefore, in terms of processing, this became hold-down and clear hold-down values of one minute. Clearly, this does not reflect reality but merely marked a non-zero point in the timeline.
- A treatment was considered associated with that event if it occured at any point in the hour following that event time-point. And in this case therefore, the time-window was set to 1 hour.

All the rest of the processing after using these definitions continued in the same way as when processing the Brain-IT dataset. However, of course some information was lost due to the low-resolution temporal data, and this would have an impact on the certainty of the adherence output.

10.3. ICD9 codes and patient numbers

The diagnosis definition file (D_ICD_DIAGNOSES) was queried to extract the relevant ICD9 code. This was done by returning the code associated with any diagnosis description that contains the word “brain”. Three codes were extracted that showed significant results in terms of patient numbers (i.e. greater than two):

- “Traumatic brain hem NEC” = 85300
- “Screen-traumtc brain inj” = V8001
- “Hx traumatc brain injury” = V1552

There were many physiological codes to choose from in the MIMIC III dataset but the relevant ones used here were two each for both ICP and CPP, as extracted from the item definition file (D_ITEMS):

- “CPP” = 92, 227066
- “ICP” = 226, 2205765

Table 10.1 shows the number of subject IDs that were extracted using the three ICD9 codes identified in section 10.3.1. This was done by searching the subject diagnoses files (DIAGNOSES_ICD) for those three codes.

ICD9 code	Patient number
85300	45
V8001	55
V1552	0

Table 10.1: Number of subject IDs (individual patients) with brain injury ICD codes

The biggest file in the MIMIC III dataset (CHARTEVENTS) was then queried using these 100 subject IDs, in order to establish a physiological trace reading for each patient. This file had to be pre-processed in order to be usable, which was done by “chunking” the extracted CSV output into 264 files (an approximate working file size, which could be viewed in Notepad or MS Excel) containing 1 million lines each (a line corresponds to a single physiological data-point for a patient).

Once this processing was complete, each of the 264 files was in turn analysed for the presence of the 100 subject IDs sourced from above. Overall this resulted in physiological output of 1835 entries, filtered down using the four ICP/CPP codes listed above.

10.4. EUSIG-event detection

As before, the first stage of evaluation of this dataset is to detect physiological EUSIG (Edinburgh University Secondary Insult Grade) events. In contrast to the Brain-IT database, coverage cannot be assessed in this dataset as the physiological data is extracted from a set of append-only text files, so no enclosing schema gives information about the available space.

10.4.1. EUSIG-event definitions

Following from the description of EUSIG definitions of ICP and CPP events (table 10.2), 8 threshold definitions are used to cover the range of clinically relevant definitions.

Parameter	Threshold values (mmHg)	Direction	Event Hold-Down (hrs)
ICP	10, 15, 20, 25, 30	Up	1
CPP	50, 60, 70	Down	1

Table 10.2: Definitions of raised ICP and lowered CPP events in terms of threshold values

As mentioned previously, in contrast to the Brain-IT dataset, there is only one (artificial) definition for the hold-down value in this dataset. Therefore, there are 8 ways (one per threshold definition) that a physiological monitoring event can be detected in a dataset.

10.4.2. EUSIG-event counts

The event detection algorithm was applied to the 8 definitions of EUSIG events. Tables 10.3 and 10.4 shows the count of individual events for each EUSIG definition across the 7 patients with viable annotations (with all profiles now stored in the treatment profile database after processing).

Threshold value (mmHg)	Hold-down value (hrs)	Count
10	1	373
15	1	179
20	1	96
25	1	64
30	1	43

Table 10.3: Event count for each ICP EUSIG definition

Threshold value (mmHg)	Hold-down value	Count
50	1	3
60	1	5
70	1	6

Table 10.4: Event count for each CPP EUSIG definition

The number of CPP events recorded was evidently low compared to the count of ICP events. It is uncertain if this is of significance other than the capture process of ICP had been markedly better in this particular patient sample. Also, the

decreasing number of ICP events as the thresholds increase was in contrast to the Brain-IT dataset, which had a peak at the 15 mmHg definition.

10.5. Generation of an event log

As before with Brain-IT, to generate the full event log that can then be turned into a process model, required the association of treatment annotations with physiological events. Referring back to section 5.1 and the relative definitions of the word "event": the convention in this thesis is to use "EUSIG-event" to refer to the pattern in the physiological output described by a threshold, a hold-down and a clear hold-down; whilst "event" (singular) refers to a EUSIG-event *and* an associated treatment. Multiple instances of this latter definition are chained together to form the full patient event log described in this sub-section.

10.5.1. Treatment categories

The list of treatment annotations was extracted using wildcard string-matching (or “regular expression matching”) with the Brain-IT category names as input patterns. This list was found in the item definition file (D_ITEMS) and was then applied to the non-physiological (or “episodic”) event files: INPUTEVENTS, OUTPUTEVENTS and MICROBIOLOGYEVENTS.

Table 10.5 shows the treatment categories available for association in the MIMIC dataset. It also shows the overall count of individual annotations in each category in the dataset.

Treatment category	Count
Tidal Volume (Set)	120
Anti-Embolism [Device]	1058
Tidal Volume (Obser)	122
Minute Volume(Obser)	185
Anti-Embolism [Status]	1054
Minute Volume (Set)	29
Stroke Volume	24
Tidal Volume (Spont)	91
Dilantin	22

Table 10.5: Overall treatment category count (total = 2705)

10.5.2. Treatment association and event log

To show the representation of instances of treatment/event associations, table 10.6 shows an example of the count of events with associated treatments expressed as an absolute number count and a percentage of overall events for the single 1-hour definition possible (see table 10.3 for the absolute ICP event number). The definition selected is the ICP threshold value of 10 mmHg, for a hold-down time of 1 hour (as defined for the MIMIC dataset).

Time-window (hours)	Count (absolute)	Count (%)
1	55	7.3

Table 10.6: Number of events with associated treatments for a 1-hour time-window (only option available in MIMIC dataset)

10.6. Process model comparison - individual output

The following two patients were selected to indicate information relating to different classes of non-adherence that had appeared during the dataset assessment. They were selected from the patients that had treatment associations (and hence varied adherence information) and displayed different characteristics that exemplified the possible outcomes from this framework. Within those parameters and that patient subset, they were structurally representative of the variation shown by the framework over the whole dataset.

The timeline captions indicate the reasons for the pattern of non-adherence for that event (e.g. “treatment missing”, “dosage too high”, etc). As before, each patient has an accompanying table showing the aggregate (total) non-adherence duration and level, as well as an interquartile range table showing the statistical spread of the various instances of non-adherence (one feature repeated visually as an example box-plot diagram). Table 10.7 shows the EUSIG-event definitions used for these examples.

Feature	Value
EUSIG definition	**ICPm** with threshold of **10 mmHg**
Hold-down	**1 min**
Time-window	**60 mins**

Table 10.7: EUSIG definition details that the following two examples are sourced from

10.6.1. Patient 16265

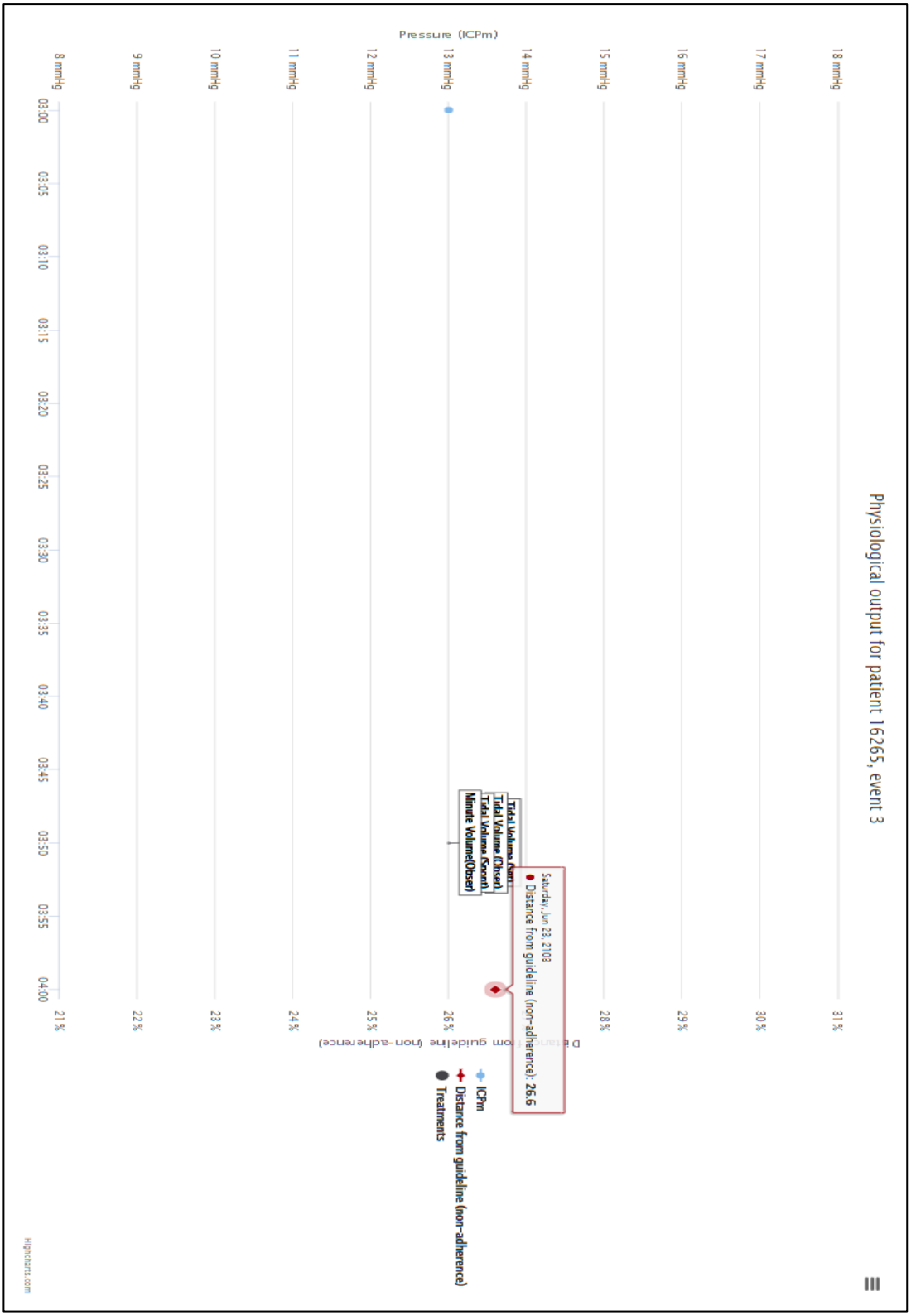

Figure 10.1: Timeline of adherence for patient 16265

Total duration (hrs)	Non-adherence (%)	Contributing reasons (%)
1	15.8	Type (10.0), Time to treatment (5.8)
3	19.8	Type (10.0), Time to treatment (9.8)
1	23.2	Type (10.0), Time to treatment (13.2)
1	23.6	Type (10.0), Time to treatment (13.6)
2	24.8	Type (10.0), Time to treatment (14.8)
1	25.8	Type (10.0), Time to treatment (15.8)
1	26.6	Type (10.0), Time to treatment (16.6)
2	27.2	Type (10.0), Time to treatment (17.2)
1	35.2	Type (10.0), Time to treatment (6.4), Dose (18.8)
142	36.2	Treatment missing (36.2)
1	37	Type (10.0), Time to treatment (8.2), Dose (18.8)
3	38.6	Type (10.0), Time to treatment (9.8), Dose (18.8)

Figure 10.2: Instances of non-adherence for patient 16265

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	15.8	36.2	35.26	36.2	36.2	38.6
Duration (mins)	1.0	1.0	1	1.0	1.0	1.0
Duration / Non-adherence	0.03	0.03	0.03	0.03	0.03	0.06
Duration * Non-adherence	15.8	36.2	35.26	36.2	36.2	38.6

Figure 10.3: Interquartile range of adherence for patient 16265

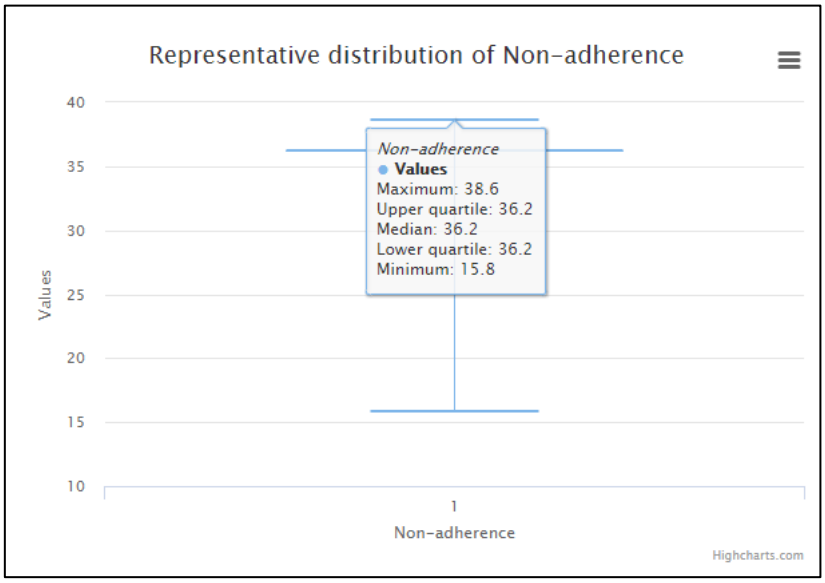

Figure 10.4: Boxplot visualisation of adherence spread for patient 16265

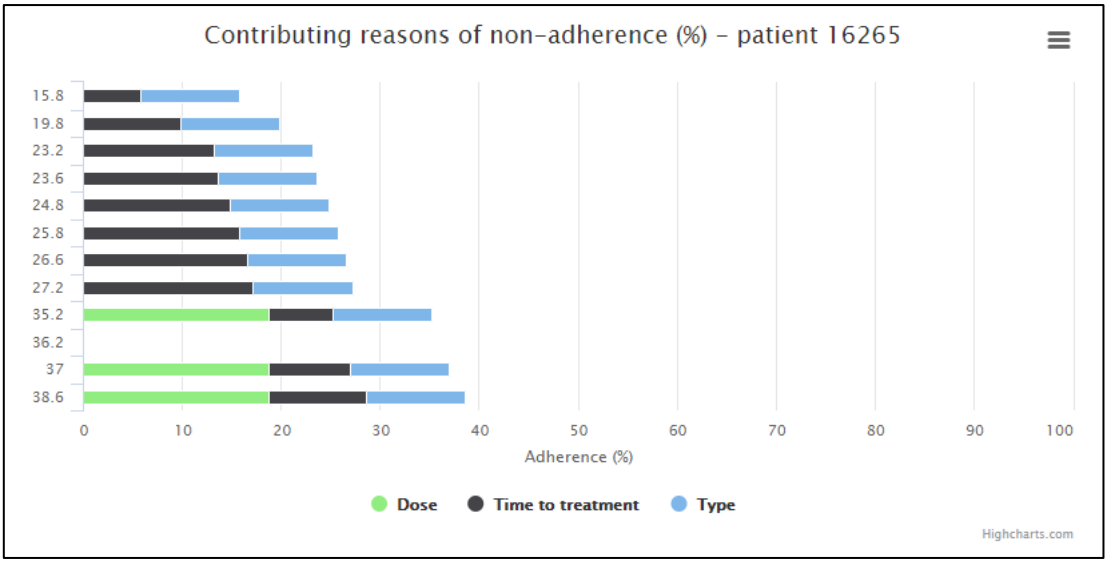

Figure 10.5: Reasons contributing to non-adherence for patient 16265

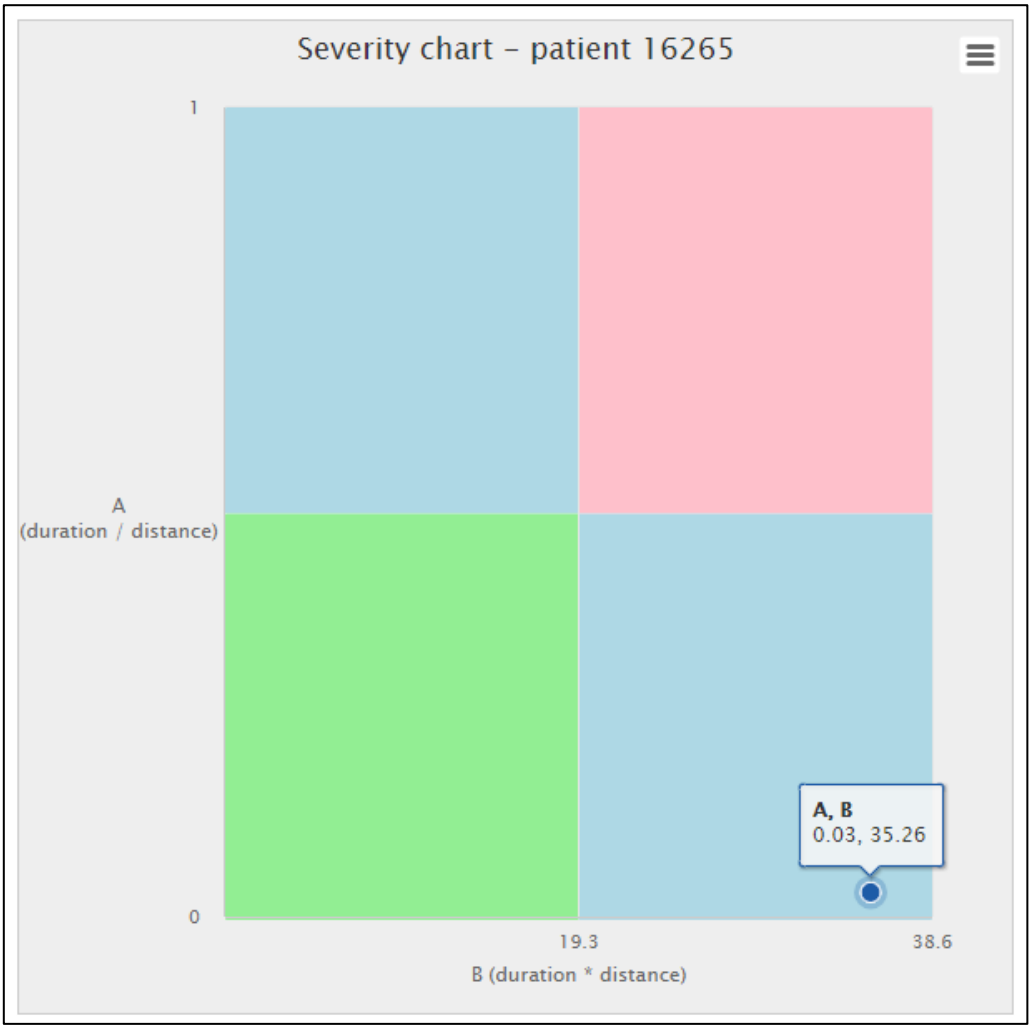

Figure 10.6: Severity chart for patient 16265

10.6.2. Patient 18849

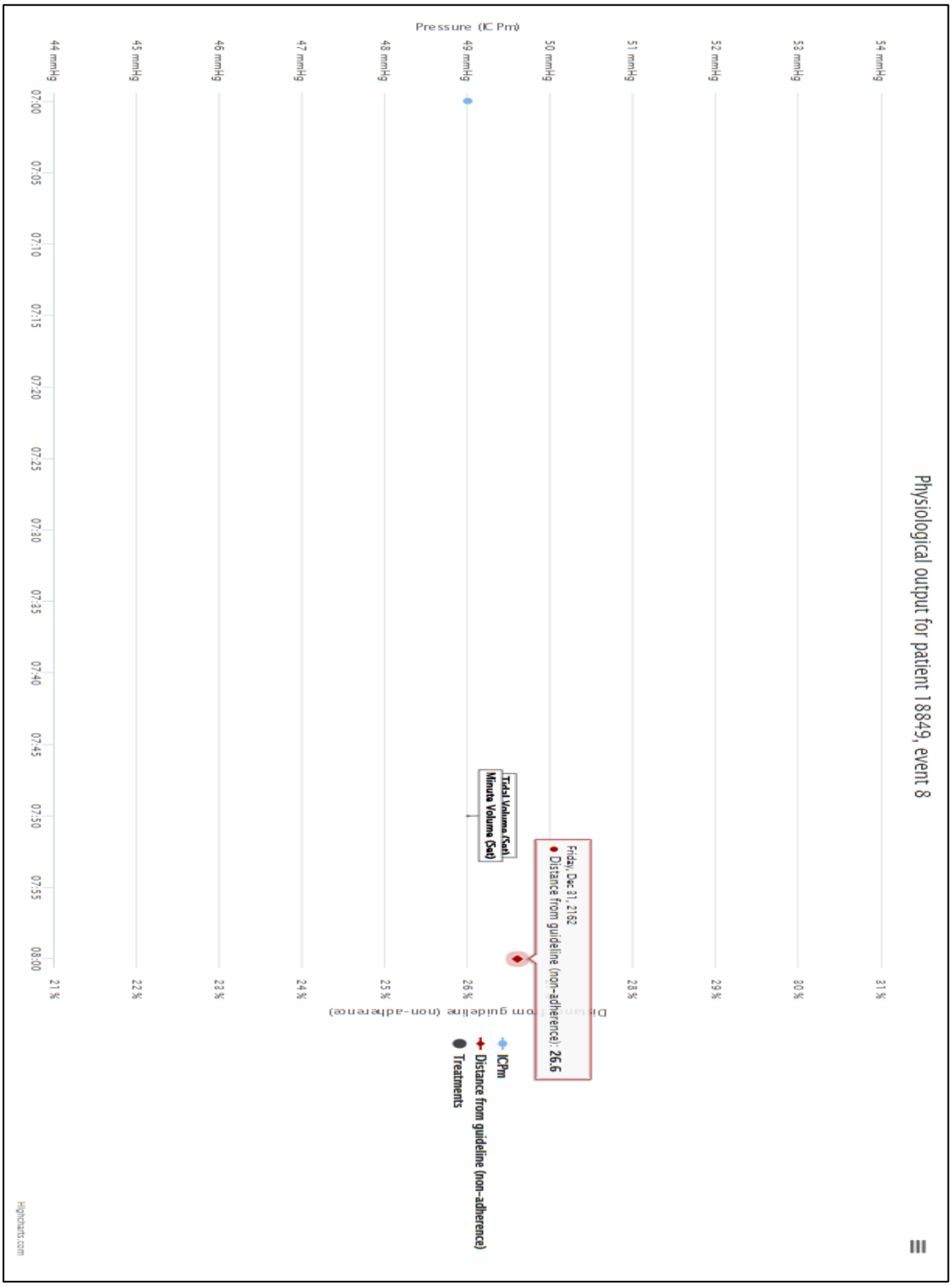

Figure 10.7: Timeline of adherence for patient 18849

Total duration (mins)	Non-adherence (%)	Contributing reasons (%)
1	13.2	Type (10.0), Time to treatment (3.2)
1	15.2	Type (10.0), Time to treatment (5.2)
3	19.8	Type (10.0), Time to treatment (9.8)
1	21.2	Type (10.0), Time to treatment (11.2)
2	23.2	Type (10.0), Time to treatment (13.2)
1	26.6	Type (10.0), Time to treatment (16.6)
1	26.8	Type (10.0), Time to treatment (16.8)
1	28.2	Type (10.0), Time to treatment (18.2)
1	32	Type (10.0), Time to treatment (3.2), Dose (18.8)
80	36.2	Treatment missing (36.2)
1	37	Type (10.0), Time to treatment (8.2), Dose (18.8)
2	38.6	Type (10.0), Time to treatment (9.8), Dose (18.8)
1	43.6	Type (10.0), Time to treatment (14.8), Dose (18.8)

Figure 10.8: Instances of non-adherence for patient 18849

Measure	Min	Q1	Mean	Median	Q3	Max
Non-adherence (%)	13.2	36.2	34.61	36.2	36.2	43.6
Duration (mins)	1.0	1.0	1	1.0	1.0	1.0
Duration / Non-adherence	0.02	0.03	0.03	0.03	0.03	0.08
Duration * Non-adherence	13.2	36.2	34.61	36.2	36.2	43.6

Figure 10.9: Interquartile range of adherence for patient 18849

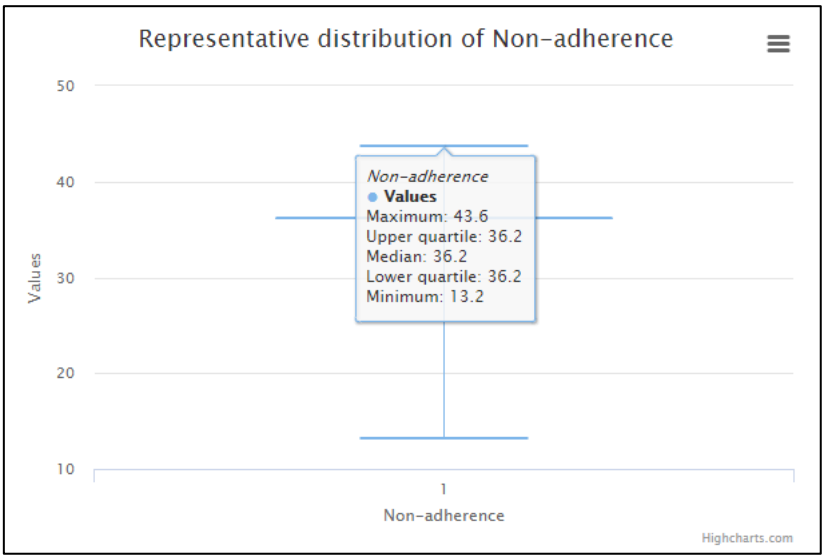

Figure 10.10: Boxplot visualisation of adherence spread for patient 18849

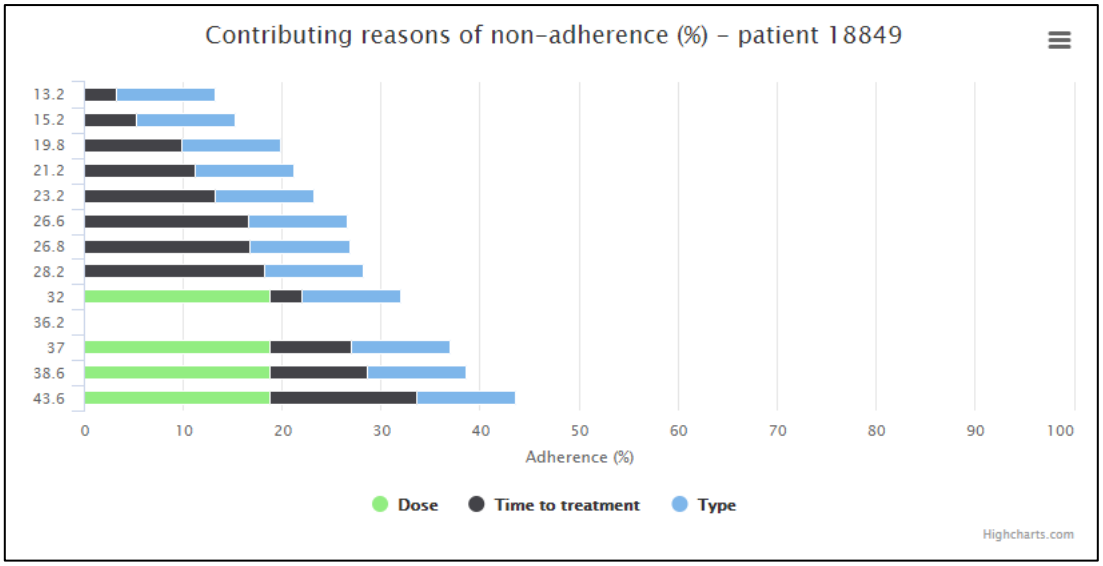

Figure 10.11: Reasons contributing to non-adherence for patient 18849

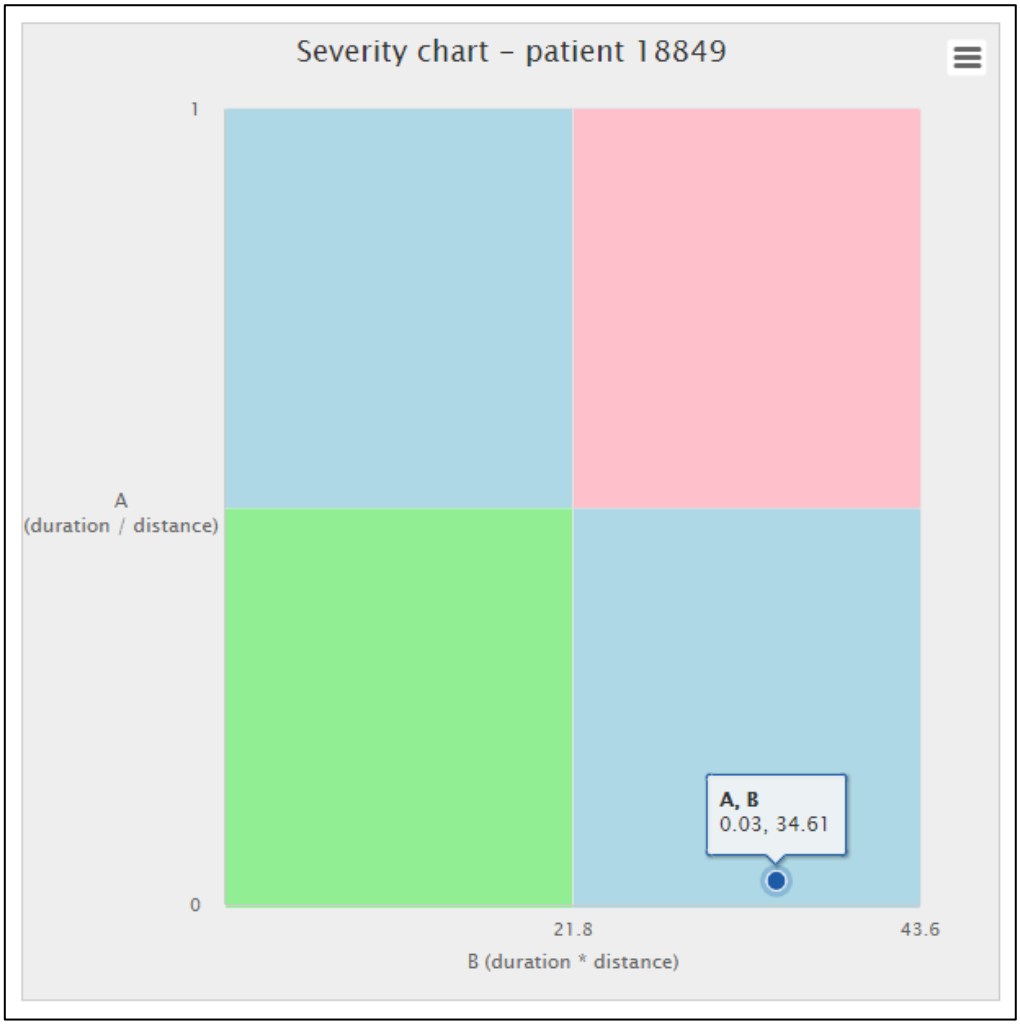

Figure 10.12: Severity chart for patient 18849

10.7. Discussion

The MIMIC dataset poses difficult challenges in terms of data extraction and processing. The low-resolution nature of the physiological data and the difficulty in categorising treatment annotations, make the accuracy of any adherence output very difficult to establish. Of particular note are the low association percentages in table 10.6 (7.3%) and the difficulties in rendering the end-hour summary information on a windowed timeline (as the treatments and physiological data stack one upon the other).

However, as with Brain-IT some adherence information was able to be extracted. With further work it may give more accurate output, but advances in understanding adherence would be much more readily achieved by using richer datasets such as Philips ICCA.

The original reason for using the MIMIC dataset was to find out if the framework could be used on a general ICU dataset. The answer to this is that it can, but with many qualifications and low accuracy. These issues are likely due to the nature of the MIMIC dataset, rather than its representation of a general ICU dataset.

11. Discussion

11.1. Aims of the thesis

In this research work, a novel technological method was developed to quantitatively feedback information about adherence to the Brain Trauma Foundation (BTF) guidelines in a given cohort of ICU patients. The following three hypotheses were formulated:

1. *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*
2. *Having extracted treatment processes, one is able to develop an algorithm to compare those against other treatment processes to establish the degree of similarity between them*
3. *One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence*

In summary, the findings of this research were:

1. A process for the management of ICP and CPP can be derived from the analysis of physiological and treatment data in many ICU datasets. This can be seen from the output of the unit tests (section 8.1), Philips ICCA (section 8.3), Brain-IT (chapter 9) and MIMIC III (chapter 10) datasets. However, a common issue, especially in the large-scale datasets was the low number of event/treatment associations, which directly affected the confidence in the adherence output. This led to the identification of the need for rich and accurate treatment annotations, an estimate of which was attempted through the timing evaluation (section 8.2)

2. The process model derived from physiological and treatment data, can be compared against other processes of a similar nature, which in this case are the BTF guidelines represented in Business Process Model Notation (BPMN). Both sets of models can be compared using the method proposed in (Dijkman, Dumas and Garcia-Banuelos, 2009) – a calculation of distance between process models.

3. The output of this comparison can be constructed into a clinically accessible tool – in this case a web-enabled application. The application allows adherence information to be obtained from clinical management processes. However, a statistical relationship with patient outcome was unable to be detected.

11.2. Interpretation of results

11.2.1. Evaluation

11.2.1.1. Unit testing

The goal of the unit testing was to make sure that the system performed as expected, when given predictable inputs. The use-cases covered general cases contributing to guideline adherence including: whether a treatment was present/not-present, or if a dosage too high/within bounds. And individual cases relating to the BTF guidelines included: different patient ages (BPs), mass lesion/diffuse injury present (ICP), and type conflict between pressors and fluids (CPP). Test patient physiological data was used, with treatment annotations artificially applied to create these use-cases. The results showed that all use-cases covered by the BTF guidelines did indeed produce adherence output and captured clinical management processes.

The main issue in this section was the first appearance of the "default" adherence level which occurs when a EUSIG-event occurs and has no associated treatment. This issue re-occurs in all stages of evaluation, has a strong influence on the overall measurement of adherence, and is largely only solved by the rich and accurate input of annotations to the dataset.

11.2.1.2. Treatment annotation timings

Using the unique feature of the CSO dataset – a physiological and treatment dataset, annotated by a "live observer" that was able to watch and confirm physical actions taken in the ICU – an estimation of the difference between "live" and "non-live" timings of annotations was attempted.

The results showed that when observed in real-time (the "live observer"), the accuracy of timing does correspond well to the physical output of the physiological data at the bedside. Unfortunately, the comparison against the non-live dataset

was inconclusive, with no matching annotations at all in the non-live dataset, it is possible that the discrepancy between manual recording of annotations in a regular ward shift, and the physiological output is so dense, that (correctly) none appeared in this small analysis. However, other reasons such as the type of event observed or the approach to noting those types of event in the ward may be contributing factors.

If repeated, a different set of events would be tracked and attempted to be matched, and a consultation with the ward staff would allow an understanding of which events are priorities in order to more accurately gauge the timing. However, the "no matching events" result from this study could be a key finding in itself in this regard. The events were never annotated, therefore timing overall is inaccurate, which has large repercussions for the performance of the framework in a real ICU setting.

11.2.1.3. Domain expert information

The system was evaluated against the input of three ICU patients that had accompanying patient notes, compiled and distributed by domain experts. The data was processed and adherence output retrieved and compared against the patient notes.

The adherence output generated from the three patients highlighted several issues with the system:

- The drug categorisation issue (mentioned above) affected the quantitative output, as this would be registered as type conflict when compared against the guideline model.
- The adherence output responds poorly to refractory pressure events (constant increase with no positive resolution). This is mainly due to the time-window application issue (mentioned in 11.3.1.1).
- Though the outputs derived from this initial error in processing of refractory events, it also highlighted the issues encountered when variation is low (i.e. only one adherence instance is present). The spread of instances is highly skewed – which is a correct, if unusual, representation - and the severity chart indicates "most severe", though it has reached this conclusion through

an incorrect pathway (believing that there have been no annotations, rather than the refractory nature of the patient context). This issue relates to the use of a single time-window when assessing treatment annotations (see section 11.3.2).

- When a clinical context of intentional non-treatment is encountered (patient #3 had regular transient ICP pressure events which appeared to synchronise with the rise and fall of ETCO2 levels), the system also fails to incorporate this into its adherence evaluation. In this case a notion of baseline context should be incorporated, though it is uncertain how this would be manifest in the technology.

11.2.1.4. Patient outcome

The logistic regression against patient outcome showed no statistically significant effect against 6-month patient outcome (as measured by the extended Glasgow Outcome Score). This was not unexpected as the numerical values associated with the guideline adherence information require calibration and clinical consensus, as well as the Brain-IT dataset only providing around 17% of EUSIG events with associated treatments (though see Appendix F for the reasoning that this number should be roughly double at 34%).

Whilst a relationship of adherence output with patient outcome would have been ideal, the wider context is that this framework captures a clinical management process that is useful for audit purposes, regardless of patient outcome. It is also just one example of a protocol, and though obviously patient outcome is the most important, there may be other end-points that the adherence information does have a significant correlation with.

Finally, in the case of TBI, one of the major reasons for the lack of high-power and clinically significant findings is the lack of correlation with patient outcome. As mentioned in chapter 3, studies such as BEST-TRIP (Chesnut *et al.*, 2012) have indicated that ICP monitoring has no treatment effect on the overall patient outcome. Though this is disputed in the expert community, in this context, the lack of correlation between adherence output and outcome is not necessarily an indicator of failure.

11.2.2. Large-scale datasets

The framework was applied to two large-scale datasets – Brain-IT and MIMIC III – which gave a variety of results that exemplified the individual adherence output that could be derived, and aggregate information about adherence over both cohorts.

11.2.2.1. Brain-IT

Overall, the individual guideline adherence output derived from the Brain-IT dataset is useful but must be considered in the context of low association numbers. The physiological data is comprehensive and the EUSIG event pattern can be readily extracted. However, the annotation frequency and density in the dataset overall is low, leading to low association numbers with EUSIG events (and all other assertions made must also be qualified with this consideration)[18].

This is mainly due to the time of data collection (mid 2000's) where the available technology was manual inputs to PDA Palm Pilots. Modern ICU technologies, such as Philips ICCA show a much higher density of treatment annotations, so a higher association number, and therefore a more accurate representation of clinical management can be assessed. This potentially relates to the overall issues confronting the TBI medical community, as the increase in treatment annotations can be used to increase the power of studies and trials.

However, despite this low resolution, clear indications of the adherence output can be found within the dataset, highlighted by the patient examples in this chapter (e.g. treatment presence, dosage, type differences). This suggests that, though requiring refinement, the framework does provide viable output for measuring guideline adherence over this dataset.

11.2.2.2. MIMIC III

The MIMIC III results were an exercise in applying the same strictures of EUSIG definitions and treatment associations over a very low-resolution dataset. With some liberties taken in the definition of a EUSIG event (hold-down of one minute representing a whole hour of an assumed event) it was possible to chart

[18] There is a known issue with the processing of these associations – detailed in appendix F.1 – which suggests that actually this reported association number (17%) is roughly half of what it actually should be (34%).

physiological output (at a resolution of one reading per hour) and to associate some treatment events, with several TBI patients that had been found in the general ICU population.

The adherence information that this provided again did capture some management processes but was unlikely to provide powerful insight into the nature and duration of non-adherence to the guidelines, as the resolution of the information was just too low. In an age of increasing awareness of digital-privacy it is commendable that the MIMIC team implemented the averaging of physiological data and the transposition of timestamps, but these also conspired to reduce the effectiveness of clinical insight that could be drawn from the data.

However, it was possible to obtain some results. Instances of non-adherence were able to be compiled and a final severity chart was produced, as shown in the example patients in chapter 10. The achievement of the framework to work with almost any dataset is demonstrated here, though the best results are obtained with highly sampled physiological data and high density of treatment annotations. In this work so far, the Philips ICCA system provides that benchmark.

11.3. Evaluation

The following strengths and weaknesses have been identified in this thesis.

11.3.1. Strengths

As identified in the literature review (chapter 3), the system has been developed to provide immediate, detailed and independent feedback on adherence at a level of clinical management. The vision for this application would be to fit into audit procedures, such as weekly meeting to assess compliance (e.g. a technological TBI version of the Surviving Sepsis campaign) or to assist with the review and further development of the guidelines themselves. This level of clinical management is more detailed than the typical inputs to a clinical study and allow a quick assessment of adherence in the live ICU context.

The presentation styles chosen – individual view-charts, interquartile range representation, aggregate information, and clinical severity risk analysis charts – were all carefully chosen as to maximise the utility of the information, in a way that informs a clinician as rapidly as possible about a patient status relative to the

guideline considered. The trace of non-adherence reasons provides a qualitative detail to the overall quantitative non-adherence, so information – whilst hidden for easier viewing – can still be "unpacked" and reviewed if required. These considerations were taken from the combination of concepts brought from the business process management domain and from observing what already worked in a clinical context for the sepsis guideline adherence meetings. The use of the severity chart also allows the quantitative results to be expressed as summary measures for an overall assessment of impact of the guideline adherence.

A key philosophy in the development of this system was that minimal knowledge should be used to identify the patterns of clinical management. Additional contextual requirements could be added "piece-wise" to the knowledge base as necessary, which would be an obvious first step in future work. A major benefit of this approach is that individual insights on adherence can be made in isolation, which can then be combined and built upon, rather than attempting to identify signal from competing sources of noise of other components. This goes some way – though is not a complete solution – to the issue of comorbidities, which guidelines are notorious for not capturing well.

This approach also has the advantage of occupying a clinical work-flow "silently" by using all the data that is already available (rather than either technical developers or domain experts contributing more information to imbue semantic meaning). The knowledge used is the already-agreed standards of repositories that are already in use (i.e. the database schemas supporting Philips ICCA, Brain-IT, MIMIC III, etc).

11.3.2. Limitations

The largest clinical issues raised as motivation for this thesis was the lack of high-powered TBI studies available. It is hoped that a vision of how to build a real-time feedback tool to help with this problem would be useful. However, an underlying factor would be the accuracy of the component parts. This is a major and common issue in the development of many clinical technologies, and often manifests in the issue of clinical interpretation being lost in the opacity of the supporting technology and calculations (e.g. the process model distance calculations outlined in chapter 6). Further validation and refinement of the calculation method and the

quantitative weightings would be required to increase faith in the accuracy of the framework developed in this research work. Until this was achieved and rigorously demonstrated, the technology will experience large barriers to adoption.

A central feature of the framework is the presence or absence of a treatment annotation proximate to the EUSIG events detected. The main Brain-IT dataset provided enough richness of annotations to infer various non-adherence instances. However, the overall percentage of associated events (approximately 17%) could be vastly improved. The ICCA dataset provided a much higher density of associations between treatment annotations and EUSIG-events (approximately 80% in the three-patient sample). A reasonable inference is that the ability to capture treatment annotations has improved since the dates of collection between these two datasets (2006 to 2018).

However, the evaluation of timing accuracy shows that in a typical ICU dataset, even though the density of these treatments is high, they may still be inaccurate, and much work still needs to be done to guarantee the accuracy of these annotation timings. Unfortunately, this variable is also a key component to the accuracy of the guideline adherence output and must be evaluated or qualified whenever the system is used.

Another central feature of this framework is the application of business process techniques and models to the clinical domain, in a manner that is not traditional (e.g. rather than administrative processes, the techniques are applied to the actual clinical management). These tend to be approaches familiar to the informational world of IT, logistics administration, project management, etc, and the primary goal is always the search for efficiencies of process.

The two most obvious presentations of this cross-domain application are the main method of the research (expression of clinical processes as BPMN), but also the use of a risk analysis chart to express clinical severity and impact on outcome when two values are interpreted in combination. Though this feature is potentially useful for expressing clinical outcome of multi-factors in broad strokes, it also runs the risk of over-simplifying complex medical situations. A balance on how to present this should be found through feedback with clinical users.

Two important technical limitations were discovered during the course of evaluating the framework in chapter 8: the failure of the system to effectively deal with refractory EUSIG events – presenting as one catastrophic event in patient #2 in the Philips ICCA sample set – and periods of high physiological volatility, which relates to the issues raised in chapter 6 about how to associate multiple treatments with multiple EUSIG events.

The first could be addressed by the use of multiple time-windows. Currently, a single time-window is overlaid on the output since the start of a single event. However, a more effective solution may be a combination of one time-window at the beginning of the event, then a second moving time-window, which would be evaluated each minute, would likely give enough contextual temporal information to capture single large events, and the more regular patterns (with clear resolution) which make up the bulk of the events assessed in this work. Careful consideration on how to communicate this concept to the end user would be required, as in the practical experience of this work, the concept of a single time-window is possibly still not widely understood amongst the clinical community (as it is also a construct brought from the information theory community).

The second issue has been avoided in this work through the assumption of a one-to-one relationship between event and treatment, and the datasets used so far have been of low enough density for this assumption to hold true. However, as has been noted several times the key to better performance of this framework is the high density of treatment annotations, therefore it is likely this issue will be encountered. Possible solutions include the attempt to capture the explicit target of the treatment (not always available but sometimes – such as in Brain-IT – or can be inferred from patient notes, such as with the Philips ICCA system).

A final limitation of the system is inherent in the nature of guidelines themselves. As detailed in (Greenhalgh *et al.*, 2014) and repeated in many conference presentations on the nature of technology and guidelines, they are inefficient at coping with comorbidities. In this thesis, this has been represented most clearly in the evaluation of the domain expert information with the Philips ICCA dataset (e.g. the transient episodes of ICP events, which the system failed to account for as it was not aware of the ETCO2 outputs and correlation between the two). This

comorbidity problem was considered from the beginning of this research work and a mitigation was provided in the form of only focusing on three guidelines which had features well suited to the technical problem space (quantitative elements with well-defined flow processes). However, future work would include – likely in the form of enhanced domain context – knowledge of general guideline repositories.

11.4. Other considerations

There were several findings and discussion points raised through the course of this research, relating to the adherence framework. These have been grouped into themes that are discussed individually in this sub-section.

11.4.1. Contextual domain knowledge

As mentioned previously, a central part of this thesis is the ability to make these inferences about non-adherence given the minimum knowledge possible. The intention was to add contextual information in a piece-wise fashion, rather than describe everything and attempt to strip away unnecessary "noise". This allows the ability to clearly address each separate use-case of the BTF guidelines in turn and overlay the resulting output. But when context is important (e.g. such as understanding the patient load of pressors when attempting to follow the CPP guideline) this approach requires addition of knowledge.

In this case, an approach using an ICU ontology was attempted (mentioned briefly in chapter 5, and described in full in appendix B, which constituted a short conference paper in its own right[19]). The difficulties involved were largely influenced by the steep learning requirements of ontology implementation, but the main structural issue was one that has been covered in section 3.4.3, namely the desire to "describe everything" that exists in an ICU.

When presented with the BTF guideline as a BPMN diagram (the solution later chosen), a neurosurgical specialist noted that there was a wider context that the diagram was missing. The use of an ontology would go some way to approaching this – encoding the contextual parameters required. But again, careful introduction of only the minimum necessary would be required.

[19] Computer-Based Medical Systems (CBMS), 2012 in Rome, Italy

11.4.2. Technical utility

One re-usable technical outcome of this work is the establishment of what has been called the "treatment profiles" database. The full schema as presented in section 5.2, provides a host of information such as guideline numbering and treatment annotations for a particular dataset. However, one of the main uses, is the extraction of EUSIG pattern information from a particular dataset.

This synthesised information is useful in its own regard (i.e. the same physiological dataset can be viewed using different parameters of the EUSIG definition – threshold, hold-down length, etc), but when combined with the *"EventDetection"* code, with minor edits to the code interface (depending on whether the physio stream is represented as wave-form, minute-by-minute, etc) any time-varying physiological dataset can be analysed and stored to add to this repository. There is great potential for use of this code and repository for future studies, as information about ideal thresholds change or require investigation (e.g. for the establishment of CPPopt levels).

A project that could conceivably help in this regard is the CENTER-TBI initiative, which is believed to collect a lot of high-resolution physiological data and rich treatment annotations.

11.4.3. Weightings and input values

The main source of quantitative information in this framework comes from the application of quantitative values to reasons for non-adherence. These were estimated based on minor feedback from clinicians and an assessment of the relative importance of the component reasons. As these constitute the main components of the quantitative non-adherence output, they will require re-evaluation and refinement for better accuracy.

One approach to understand and refine these values would be to chart them against the different patterns of guideline adherence output and attempt to find optimal minima representing a favoured value.

11.4.4. Detection of clinical behaviour pattern

An unexpected outcome of the initial audit of the Brain-IT dataset was the ability to observe a preferred clinical reaction when assessing counts of events reacted to

as part of a clinical management process. The original chart of counts across several EUSIG definitions was modified when different time-windows were overlaid and the remaining events with associations were counted. The most numerous definition in the original count was 15 mmHg, which shifted to 20 mmHg when the time-window was applied (with a constant hold-down of 5 minutes). This suggested that events on a crossing threshold of 20 mmHg elicited the most frequent clinical management responses. As the time-window was extended from 15 minutes to 1 hour, the pattern gradually returned to the original un-associated version.

It can be concluded from this that even before considering adherence to guidelines the act of associating treatments to EUSIG-events can provide valuable information about clinical management processes.

11.5. Context

11.5.1. Technical

From a technical standpoint, the work is unique in its combined approach to measuring distance between process models, combining this with information from two sources of a different nature (ICU and text guidelines), and implementing this to provide clinical utility. Further review of literature additional to that reviewed in chapter 3, suggest that the state-of-the-art remains as it was in the fields reviewed that are closest to this work.

In the field of knowledge representation, the latest work by Montani is at (Montani *et al.*, 2015) and whilst having an extra module, is virtually the same work as presented a year earlier to the AIME conference. Since the work conducted in (Dijkman, Dumas and García-Bañuelos, 2009), which is referenced throughout this thesis, he has conducted several further iterations on the work of measuring distances between process models. One of these (Dijkman et al., 2012) specifically attempts to improve on the work conducted in the 2009 paper, and the other provides a short overview of the various methods available in addition to the one he outlined previously (Dijkman et al., 2013).

In terms of other work involving clinical guideline formalisms, and their combination with apps, the MobiGuide project – the leading project in this space - concluded in 2016. As also mentioned in chapter 3, this was patient-centred and

focused on the aspect of patient-centred adherence – using “wearable technology” to improve adherence, making use of the features of smart-phones and apps, as covered in section 3.4.2. The latest publication from this group is at (Peleg, Shahar, Quaglini, Fux, *et al.*, 2017b) and whilst novel in the combination of formalisms and physical technology, does not concern the same area of clinical management as this thesis.

11.5.2. Clinical

The project most closely aligned to this work in the clinical sphere is the EU-funded CENTER-TBI project. Many publications have been made by this group over the lifetime to date of the project, with a particular focus on Living Systematic Reviews (LSRs). They cover several areas, such as the need for combining human and machine effort (suggestions for approach) (Thomas *et al.*, 2017), and recommendations for the curation and development of Living Guidelines (Akl *et al.*, 2017).

These publications largely detail the issues and challenges experienced so far rather than outlining a full implementation of procedure. Adherence to guidelines in the TBI space are well covered in particular, as the focus of the PhD thesis of Maryse Cnossen (Cnossen, 2016c). However, these are also manual in approach and do not provide the technological solution as presented in this thesis.

Most instructive in the clinical space has been the recent review of the sixth InTBIr conference in October 2017 (InTBIr, 2017) (a worldwide collaboration of major TBI projects) This reflected upon the progress of many of the leading TBI initiatives world-wide, with particular interest from the point of view of this thesis, on the progress of data development by CENTER-TBI and TRACK-TBI (McMahon *et al.*, 2014).

In particular, challenges expressed by the CENTER-TBI team on the difficulties in curating wide-ranging data from the many patients (approximately 5400 as of the conference date), indicate that the repository has been designed with a desire to “capture everything”, and from the large number of partners involved, across many countries, leads to unmanageable heterogeneity in the datasets. Control was attempted on this from the beginning of the project using the Common Data

Elements (CDEs) inherited from the IMPACT studies (Marmarou *et al.*, 2007), but it is likely these were not specified at a low enough level, and the same data issues now occur.

As such, data curation meetings and recommendations have now been formulated in response to the challenge of controlling this heterogeneity (InTBIr, 2017). An ideal synergy of the work of this thesis – attempted to no avail during the course of the research – would be to apply the framework to the appropriate subset of the heterogeneous data (physiological output and treatment annotations, which can be standardised relatively easily) and attempt to extract adherence output from a well-annotated dataset.

Finally, in the same meeting, a representative of the Brain Trauma Foundation noted that on the development of TBI guidelines there is a lot of progress on literature identification/synthesis and evidence-based recommendations, but not on the development of protocols and algorithms which would assist technologically in the revision and development of guidelines. The framework developed in this thesis would be almost a perfect fit for this vision, and an immediate future step for the continuation of this research will be to contact the BTF and demonstrate the application and how it works with their developed guidelines.

11.6. Future work

There are a variety of avenues of research and refinement that can be pursued as a result of this research work.

Tasks that should be immediately addressed include the technical limitations discovered during the evaluation phases of the work, including:

- Feedback on usability, interface and clinical efficacy of the application from clinical users
- Improved sensitivity to refractory EUSIG events
- Improved processing of multiple treatments and events
- Refinement on the weighting values chosen to calibrate the adherence output
- Refinement of the threshold values of the severity charts

In terms of contextual application, it would be ideal to apply the framework to a subset of the CENTER-TBI dataset and to approach the Brain Trauma Foundation to demonstrate as a possible application of technological algorithms to the feedback and development of TBI guidelines.

12. Conclusions

This thesis presents research into the use of process models to represent clinical guidelines and to calculate a distance between these and real ICU data, as a measure of guideline adherence. The methods are applied to the Traumatic Brain Injury domain, specifically the threshold monitoring guidelines for SBP, ICP and CPP as formulated by the Brain Trauma Foundation (BTF). The work presents the following original contributions to research:

Technical

- The application of process models to neuro-intensive data
- The expression of the Brain Trauma Foundation guidelines as process models
- The application of process model distance calculations to neuro-intensive data (and their use as a guideline adherence measure)
- A novel method of presentation of guideline adherence results
- A novel technological framework: the conversion of text guidelines and clinical data into comparable objects, the implementation of distance calculations to run the comparisons, the implementation of novel presentation techniques

Clinical

- A technological solution to provide direct and detailed information on guideline adherence and clinical management processes of ICP and CPP in neurological ICU data

12.1. The research

Business Process Model Notation (BPMN) was used to represent the process models of both the text guidelines and the physiological and treatment ICU data. A method of calculating distances between process models was then used to compare the two. This framework was then implemented into a web-enabled application to present adherence information on any ICU dataset.

The three hypotheses formulated for this work were:

1. *In high-resolution time-series clinical data, one can extract clinically-valid treatment processes for ICP/CPP management in TBI patients*

2. *Having extracted treatment processes, one is able to develop a method to compare these against other treatment processes to establish the degree of similarity between them*

3. *One can develop a computerised tool that readily quantifies and displays to clinical staff a metric of actual ICP/CPP management protocol adherence*

12.2. Methodology

The main technological concept in this thesis was that of process models – a construct used in corporate and business domains to model time-varying processes and identify efficiencies. The process models were used to measure the adherence of clinicians to specific TBI guidelines (ICP/CPP/BP monitoring thresholds) using physiological and treatment data from bedside machines in neurological ICUs (EUSIG pressure events and corresponding treatment annotations).

Similarly, the relevant guideline texts from the Brain Trauma Foundation (BTF) were represented using Business Process Model Notation (BPMN) so that a comparable process model could be constructed. Then building on previous comparison work between process models, a "distance" between the two models was evaluated and presented as a quantitative metric of adherence, along with the qualitative components making up this metric.

Finally, this model construction and comparison was developed into a web-enabled application that can readily feed-back the non-adherence measurements in a clinical environment for any given cohort of patients that have standard physiological and treatment output.

Four evaluations were completed:

1) Evaluation of the application's functions
2) Evaluation of the typical timing accuracy of treatment annotations
3) Evaluation against domain expert patient notes
4) Evaluation of adherence output against patient outcome

The framework was then applied to two large-scale patient cohorts, one neurological specialist dataset (Brain-IT) and the other a general collection of non-specialist ICU data (MIMIC III).

Adherence "distance" and duration was presented in a variety of ways to communicate as effectively and efficiently as possible how clinical management is affected by guideline adherence. These included minute-by-minute windowing output (single number each minute, with component reasons viewable if desired), list per-patient of all non-adherence instances (also with component reasons) and a summary view using inter-quartile range tables and box-plots (to understand the spread of non-adherence durations).

A key philosophy of the thesis was to minimise the use of subjective judgement information about adherence wherever possible. This was largely achieved but was unavoidable in components such as a clinical "weighting" multiplier added to the reasons for non-adherence.

12.3. Key results

The evaluation of the system showed that adherence output was reliably captured when tested in isolation with data that provided expected outcomes. However, a key point that became apparent was the sensitivity to the accuracy of treatment annotations, in particular their timing, shown by the large "default" adherence instances.

Therefore, a follow-up evaluation was conducted to investigate the typical timing accuracy of an ICU dataset with a "live observer". This study showed that timing could indeed be variable and depended on factors such as the typical ward approach (or culture) towards annotations in general.

To ascertain further insight to the system performance in a real clinical setting, the third evaluation assessed the adherence output against contextual patient notes provided by domain experts. This highlighted adherence output of interest (e.g. repeat dosages administered), and also provided feedback on issues that required addressing (e.g. the system was insensitive to refractory EUSIG events). It also showed more varied adherence output as a result of the higher density of annotations of the Philips ICCA system.

The final evaluation used a logistic regression to investigate if there was a relationship between the adherence output and 6-month patient outcome (captured in the Brain-IT dataset). However, there was no statistically significant relationship observed, once the known indicators for TBI had been accounted for.

All of these evaluations provided insight to the strengths and limitations of the system when deployed into a real clinical setting. Though issues were highlighted, the system was capturing clinical management processes satisfactorily enough to proceed to be applied to the two large-scale datasets (Brain-IT and MIMIC III).

A range of observations on guideline adherence were made in both datasets. A variety of non-adherence patterns could be ascertained and presented in the five methods outlined in the implementation (timeline, instance table, interquartile range of instances, box-plots and severity charts). The coverage of EUSIG events with adherence information in Brain-IT was around 17% (with the caveat that a processing bug suggests the actual figure should have been around 34%), which is low relative to the overall number of events. The MIMIC III dataset had low-resolution physiological data (end-hour averaged) and even lower association values (7%). Therefore, both datasets had issues in terms of coverage and quality. However, even with these limitations examples of many guideline adherence patterns were indicated. With modern data capture methods, such as Philips ICCA, these figures indicating representation would likely be much higher.

In conclusion, the answers to the three original hypotheses were:

1) A treatment process for the management of ICP and CPP can indeed be derived from the analysis of physiological and treatment data
2) This process can be compared against other process of similar nature (in this, the BTF guideline represented in BPMN)
3) The output of this comparison can be constructed into a clinically accessible tool – in this case a web-enabled application

12.4. Future research

Immediate future work on this research would include:

- Feedback on usability, interface and clinical efficacy of the application from clinical users
- Improved sensitivity to refractory EUSIG events
- Improved processing of multiple treatments and events
- Refinement on the weighting values chosen to calibrate the adherence output
- Refinement of the threshold values of the severity charts

In terms of contextual applications, it would be ideal to apply the framework to a subset of the CENTER-TBI dataset and to approach the Brain Trauma Foundation to demonstrate the work as a possible application of technological algorithms/protocols to aid with the development of TBI guidelines.

12.5. Summary

The clinical goals of this research were to investigate possible technological solutions to aid with the challenge of low-power TBI studies, in particular trying to leverage the proximity of data to the ICU source. From a broader perspective, a long-term goal would be the generalisation of the framework to include other clinical domains and management goals beyond the specific management of ICP/CPP in TBI.

To recap, it is believed that this technology is unique and provides an original contribution of providing detailed and information on guideline adherence in the clinical management of ICP and CPP in a neurological ICU. It applies process models to neuro-intensive data, expresses the BTF guidelines as process models, applies a distance calculation between process models in the neuro-intensive domain, presents guideline adherence information in a novel way, and combines all the above technology into one unique framework.

It is hoped that this research and resulting technology represents a new, novel and effective way to capture clinical guideline adherence information in an ICU.

References

van der Aalst, de Medeiros and Weijters (2006) ‘Process equivalence: comparing two process models based on observed behaviour’, In: Dustdar S., Fiadeiro J.L., Sheth A.P. (eds) *Business Process Management. BPM 2006. Lecture Notes in Computer Science*, vol 4102. Springer, Berlin, Heidelberg

van der Aalst, W., Weijters, T. and Maruster, L. (2004) ‘Workflow mining: Discovering process models from event logs’, *IEEE Transactions on Knowledge and Data Engineering*, 16(9), pp. 1128–1142

Aarabi, B., Hesdorffer, D., Ahn, E., Aresco, C., Scalea, T. and Eisenberg, H (2006) ‘Outcome following decompressive craniectomy for malignant swelling due to severe head injury.’, *Journal of neurosurgery*, 104(4), pp. 469–79

Akl, E. A., Meerpohl, J., Elliott, J., Kahale, L. and Schunemann, H. (2017) ‘Living systematic reviews: 4. Living guideline recommendations’, *Journal of Clinical Epidemiology*, 91, pp. 47–53

Allen, B. B, Chiu, Y., Gerber, L., Ghajar, J. and Greenfield, J. (2014) ‘Age-Specific Cerebral Perfusion Pressure Thresholds and Survival in Children and Adolescents With Severe Traumatic Brain Injury’, *Pediatr Crit Care Med*, 15(1), pp. 62–70

Almatar, M., Peterson, G., Thompson, A., McKenzie, D., Anderson, T. and Zaidi, S. (2016) 'Clinical Pathway and Monthly Feedback Improve Adherence to Antibiotic Guideline Recommendations for Community-Acquired Pneumonia', *PLoS ONE*, 11(7): e0159467, pp 1-9

Ansari, M., Shlipak, M., Heidenreich, P., van Ostaeyen, D., Pohl, E., Browner, W. and Massie, B. (2003) ‘Improving guideline adherence: a randomized trial evaluating strategies to increase beta-blocker use in heart failure’, *Circulation*, 107(22), pp. 2799–2804

Azocar, F., Cuffel, B. and Goldman, W. (2003) ‘The impact of evidence-based guideline dissemination for the assessment and treatment of major depression in a managed behavioral health care organization’, *The Journal of Behavioral Health Services & Research*, 30(1), pp. 109–118

Berry, C., Ley, E., Bukur, M., Malinoski, D., Margulies, D., Mirocha, J. and Salim, A. (2012) 'Redefining hypotension in traumatic brain injury', *Injury*. Elsevier, 43(11), pp. 1833–1837

Bratton, S. and Chesnut, R. (2006) 'Guidelines for the management of severe traumatic brain injury. XII. Nutrition.', *Journal of Neurotrauma*, 24(1) pp. 77-82

Brattan, S., Chesnut, R., Ghajar, J., McConnell, F., Odette, H., Hartl, R., Manley, G., Nemecek, A., Newell, D., Rosenthal, G., Schouten, J., Shutter, L., Timmons, S., Ullman, J., Videtta, W., Wilberger, J. and Wright, D. (2007) 'Intracranial Pressure Monitoring Technology', *Journal of Neurotrauma*, 24(1)

Bullock, R., Chesnut, R., Clifton, G., Ghajar, J., Marion, D., Narayan, R., Newell, D., Pitts, L., Rosner, M. and Wilberger, J. (1996) 'Guidelines for the management of severe head injury', *European Journal of Emergency Medicine,* vol 2, pp 109-127

Bury, J., Hurt, C., Roy, A., Cheesman, L., Bradburn, M., Cross, S., Fox, J. and Saha, V. (2005) 'LISA: A web-based decision-support system for trial management of childhood acute lymphoblastic leukaemia', *British Journal of Haematology*, 129(6), pp. 746–754

Cabana MD, Rand CS, Powe NR (1999) 'Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement', *JAMA*, 282(15), pp 1458–1465

Carney, N., Totten, A., O'Reilly, C., Ullman, J., Hawryluk, G., Bell, M., Bratton, S., Chesnut, R., Harris, O., Kissoon, N., Rubiano, A., Shutter, L., Tasker, R., Vavilala, M., Wilberger, J., Wright, D and Ghajar, J. (2016) 'Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition', *Neurosurgery*, 80(1), pp. 6–15

Chesnut, R., Temkin, N., Carney, N., Dikmen, S., Rondina, C., Videtta, W., Petroni, M., Lujan, S., Pridgeon, J., Barber, J., Machamer, J., Chaddock, K., Celix, J., Cherner, M. and Chernix, T. (2012) 'A trial of intracranial-pressure monitoring in traumatic brain injury.', *The New England journal of medicine*, 367(26), pp. 2471–2481

Citerio., G., Piper, I., Chambers, I., Galli, D., Enblad, P., Kiening, K., Ragauskas, A., Sahuquillo, J. and Gregson, B. (2008), ' Multicenter clinical assessment of the

raumedic neurovent-P intracranial pressure sensor: a report by the Brain-IT group', Neurosurgery, 63(6), pp. 1152-1158

de Clercq, P., Hasman, A., Blom, J. and Korsten, H. (2001) ‘Design and implementation of a framework to support the development of clinical guidelines’, *International journal of medical informatics*, 64(2–3), pp. 285–318

de Clercq, P., Blom, J., Korsten, H. and Hasman, A. (2004) ‘Approaches for creating computer-interpretable guidelines that facilitate decision support’, *Artificial intelligence in medicine*, 31(1), pp. 1–27

Clifton, G., Choi, S., Miller, E., Levin, H., Smith Jr, K., Muizelaar, P., Wagner Jr, F., Marion, D. and Luerssen, T. (2001) ‘Intercenter variance in clinical trials of head trauma--experience of the National Acute Brain Injury Study: Hypothermia’, *Journal of neurosurgery*, 95(5), pp. 751–755

Cnossen, M., Scholten, A., Lingsma, H., Synnot, A., Tavender, E., Gantner, D., Lecky, F., Steyerberg, E. and Polinder, S. (2016, a) ‘Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review’, *Journal of Neurotrauma*, 14

Cnossen, M., Polinder, S., Lingsma, H., Maas, A., Menon, D., Steyerberg, EW. (2016, b) 'Variation in Structure and Process of Care in Traumatic Brain Injury: Provider Profiles of European Neurotrauma Centers Participating in the CENTER-TBI Study', *PLoS ONE*, 11(8): e0161367

Czekajlo, M. and Milbrandt, E. (2005) ‘Corticosteroids increased short and long-term mortality in adults with traumatic head injury’, *Critical care (London, England)*, 9(5), p. E21

Dawes, A., Sacks, G., Cryer, G., Gruen, P., Preston, C., Gorospe, D., Cohen, M., McArthur, D., Russell, M., Maggard-Gibbons, M. and Ko, C. (2015) ‘Compliance With Evidence-Based Guidelines and Interhospital Variation in Mortality for Patients With Severe Traumatic Brain Injury’, 90095(10), pp. 965–972

Dellinger, R., Levy, M., Rhodes, A., Annane, D., Gerlach, H., Opal, S., Sevransky, J., Sprung, C., Douglas, I., Jaeschke, R., Osborn, T., Nunnally, M., Townsend, S.,

Reinhart, K., Kleinpell, R., Angus, D., Deutschman, C., Machado, F., Rubenfeld, G., Webb, S., Beale, R., Vincent, J. and Moreno, R. (2013) 'Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock, 2012', *Intensive Care Medicine*, 39(2), pp. 165–228

Dijkman, R., Dumas, M., García-Bañuelos, L. (2009) 'Graph Matching Algorithms for Business Process Model Similarity Search', In: Dayal U., Eder J., Koehler J., Reijers H.A. (eds) Business Process Management. BPM 2009. Lecture Notes in Computer Science, vol 5701. Springer, Berlin, Heidelberg, pp 48-63

Dijkman, R. M., Dumas, M. and Ouyang, C. (2008) 'Semantics and analysis of business process models in BPMN', *Information and Software Technology*, 50(12), pp. 1281–1294

Donald R., Howells, T., Piper, I., Chambers, I., Citerio, G., Enblad, P., Gregson, B., Kiening, K., Mattern, J., Nilsson, P., Ragauskas, A., Sahuquillo, J., Sinnott, R. and Stell, A. (2012) 'Early Warning of EUSIG-Defined Hypotensive Events Using a Bayesian Artificial Neural Network', In: Schuhmann M., Czosnyka M. (eds) Intracranial Pressure and Brain Monitoring XIV. *Acta Neurochirurgica Supplementum*, vol 114. Springer, Vienna

van Dongen B., de Medeiros A., Verbeek H., Weijters A., van der Aalst W. (2005), 'The ProM Framework: A New Era in Process Mining Tool Support' In: Ciardo G., Darondeau P. (eds) Applications and Theory of Petri Nets 2005. ICATPN 2005. *Lecture Notes in Computer Science*, vol 3536. Springer, Berlin, Heidelberg, pp. 444–454.

East, T., Morris, A., Wallace, J., Clemmer, T., Orme Jr, J., Weaver, L., Henderson, S. and Sittig, D. (1991) 'A strategy for development of computerized critical care decision support systems', *International journal of clinical monitoring and computing*, 8(4), pp. 263–269.

Eastridge, B., Salinas, J., McManus, J., Blackburn, L., Bugler, E., Cooke, W., Concertino, V., Wade, C. (2007) 'Hypotension begins at 110 mm Hg: redefining "hypotension" with data', *The Journal of trauma and acute care surgery*, 63(2), pp. 291- 299

Edwards, P. (2005) ‘Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.’, *Lancet*, 365(9475), pp. 1957–1959

Enblad, P., Nilsson, P., Chambers, I., Citerio, G., Fiddes, H., Howells, T., Kiening, K., Ragauskas, A., Sahuquillo, J., Yau, Y., Contant, C. and Piper, I. (2004) ‘R3-survey of traumatic brain injury management in European Brain IT centres year 2001’, *Intensive care medicine*, 30(6), pp. 1058–1065

Ennis, J., Gillen, D., Rubenstein, A., Worcester, E., Brecher, M., Asplin, J. and Coe, F. (2015) ‘Clinical decision support improves physician guideline adherence for laboratory monitoring of chronic kidney disease : a matched cohort study’, *BMC Nephrology*, 16(163), pp. 1–11

Fakhry, S., Trask, A., Waller, M. and Watts, D. (2004) ‘Management of Brain-Injured Patients by an Evidence-Based Medicine Protocol Improves Outcomes and Decreases Hospital Charges’, *The Journal of Trauma: injury, infection and critical care*, 56(3), pp. 492–500

Farquhar, C. M., Kofa, E. W. and Slutsky, J. R. (2002) ‘Clinicians’ attitudes to clinical practice guidelines: a systematic review’, *The Medical journal of Australia*, 177(9), pp. 502–506.

Faul, M., Wald, M., Rutland-Brown, W., Sullivent, E. and Sattin, R. (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.’, *The Journal of trauma: injury, infection and critical care*, 63(6), pp. 1271–1278

Fervers, B., Burgers, J., Haugh, M., Latreille, J., Mlika-Cabanne, N., Paquet, L., Coulombe, M., Poirier, M. and Burnand, B. (2006) ‘Adaptation of clinical guidelines: Literature review and proposition for a framework and procedure’, *International Journal for Quality in Health Care*, 18(3), pp. 167–176

Fox, J. (2017) ‘Cognitive systems at the point of care: The CREDO program’, *Journal of Biomedical Informatics*, 68, pp. 83–95

Franschman, G., Peerdeman, S., Greuters, S., Vieveen, J., Brinkman, A., Christiaans, H., Toor, E., Jukema, G., Loer, S. and Boer, C. (2009) 'Prehospital endotracheal intubation in patients with severe traumatic brain injury: guidelines versus reality.', *Resuscitation*, 80(10), pp. 1147–1151

Ghajar, J. (2000) 'Traumatic brain injury.', *Lancet*, 356(9233), pp. 923–929

Goldberger, J. and Buxton, A. (2013) 'Personalized medicine vs guideline-based medicine.', *JAMA : the journal of the American Medical Association*, 309(24), pp. 2559–2560

Greenhalgh, T., Howick, J. and Maskrey, N. (2014) 'Evidence based medicine: A movement in crisis?', *BMJ (Online)*, 348(6), pp. 1–7

Griesdale, D., McEwen, J., Kurth, T. and Chittock, D. (2010) 'External ventricular drains and mortality in patients with severe traumatic brain injury.', *The Canadian journal of neurological sciences*, 37(1), pp. 43–48

Griesdale, D., Ortenwall, V., Norena, M., Wong, H., Mypinder, S., Kolmodin, L., Henderson, W. and Dodek, P. (2015) 'Adherence to guidelines for management of cerebral perfusion pressure and outcome in patients who have severe traumatic brain injury', *Journal of Critical Care*, 30(1), pp. 111–115

Grimshaw, J. and Russell, I. (1993) 'Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations', *The Lancet*, 342(8883), pp. 1317–1322

Grol, R. (2001) 'Successes and Failures in the Implementation of Evidence-Based Guidelines for Clinical Practice', *Medical Care*, 39(8), pp. 46-54

Grol, R. and Grimshaw, J. (2003) 'From best evidence to best practice: Effective implementation of change in patients' care', *Lancet*, 362(9391), pp. 1225–1230

Güiza, F., Depreitere, B., Piper, I., van den Berghe, G. and Meyfroidt, G. (2013) 'Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury: Development and validation in a multicenter dataset', *Critical Care Medicine*, 41(2), pp. 554–564

Haddad, S. H. and Arabi, Y. M. (2012) 'Critical Care Management of Severe Traumatic Brain Injury in Adults', *Scandinavian journal of trauma, resuscitation and emergency medicine*, 20(1), p. 12

Hamburg, M. and Collins F. (2010) 'The Path to Personalised Medicine', *New England Journal: Perspective*, 363(1), pp. 301–304

Harbour, R. and Miller, J. (2014) 'A New System For Grading Recommendations In Evidence Based Guidelines', *BMJ*, 323(7308), pp. 334–336.

Harris, B., Andrews, P., Murray, G., Forbes, J and Moseley, O. (2012) 'Systematic review of head cooling in adults after traumatic brain injury and stroke', *Health technology assessment*, 16(45), pp. 1–175

Härtl, R. and Stieg, P. (2013) 'Intracranial Pressure Is Still Number 1 Despite BEST:TRIP Study', *World Neurosurgery*, 79(5–6), pp. 599–600

Hemphill, J., Barton, C., Morabito, D. and Manley, G. (2005) 'Influence of data resolution and interpolation method on assessment of secondary brain insults in neurocritical care.', *Physiological measurement*, 26(4), pp. 373–386

Hesdorffer, D. and Ghajar, J. (2007) 'Marked improvement in adherence to traumatic brain injury guidelines in United States trauma centers.', *The Journal of trauma: injury, infection and critical care*, 63(4), pp. 841-848

Heymans, S., McKennirey, M. and Phillips, J. (2011) 'Semantic validation of the use of SNOMED CT in HL7 clinical documents', *Journal of biomedical semantics.*, 2(1), p. 2

Huang, Z., Lu, X. and Duan, H. (2012) 'On mining clinical pathway patterns from medical behaviors', *Artificial intelligence in medicine*, 56(1), pp. 35–50

Hukkelhoven, C., Steyerberg, E., Farace, E., Habbema, J., Marshall, L. and Maas, A. (2002) 'Regional differences in patient characteristics, case management, and outcomes in traumatic brain injury: experience from the tirilazad trials', *Journal of neurosurgery*, 97(3), pp. 549–557

Jones, P., Andrews, P., Midgley, S., Anderson, S., Piper, I., Tocher, J., Housley,

A., Corrie, J., Slattery, J., Dearden, N. and Miller, D. (1994) ‘Measuring the burden of secondary insults in head-injured patients during intensive care’, *Journal of neurosurgical anesthesiology*, 6(1), pp. 4–14

Kaiser, K. and Miksch, S. (2009) ‘Versioning computer-interpretable guidelines: semi-automatic modeling of “Living Guidelines” using an information extraction method’, *Artificial intelligence in medicine*, 46(1), pp. 55–66

Kearns, R. J., Moss, L. and Kinsella, J. (2013) ‘A comparison of clinical practice guidelines for proximal femoral fracture’, *Anaesthesia*, 68(2), pp. 159–166

Lal, P., Williams, C., Georgatzis, K., Hawthorne, C., McMonagle, P., Piper, I. and Shaw, M. (2015) ‘Detecting Artifactual Events in Vital Signs Monitoring Data’, Chief Scientist Office Report, 23rd September 2015

Lazaridis, C., DeSantis, S., Smielewski, P., Menon, D., Hutchinson, P., Pickard, J. and Czosnyka, M. (2014) ‘Patient-specific thresholds of intracranial pressure in severe traumatic brain injury’, *Journal of Neurosurgery*, 120(4), pp. 893–900

Lee, H., Struik, L. and Ahmed, O. (2016) ‘“Sideline Guidelines” by Cleveland Clinic Innovations: a smartphone application for the on field healthcare practitioner’, *British Journal of Sports Medicine*, 50, pp. 1097–1098

Lee, J., Rittenhouse, K., Bupp, K., Gross, B., Rogers, A., Rogers, F., Horst, M., Estrella, L. and Thurmond, J. (2015) ‘An analysis of Brain Trauma Foundation traumatic brain injury guideline compliance and patient outcome’, *Injury*, 46(5), pp. 854–858

Levy, M., Dellinger, P., Townsend, S., Linde-Zwirble, W., Marshall, J., Bion, J., Schorr, C., Artigas, A., Ramsey, G., Beale, R., Parker, M., Gerlach, H., Reinhart, K., Silva, E., Harvey, M., Regan, S. and Angus, D. (2010) ‘The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis’, *Intensive care medicine*, 36(2), pp. 222–231

Lezcano, L., Sicilia, M.-A. and Rodríguez-Solano, C. (2011) ‘Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules’, *Journal of biomedical informatics*, 44(2), pp. 343–353

Likhterov, I., Tuttle, M., Haser, G., Su, H., Bergman, D., Alon, E., Bernet, V., Brett, E., Cobin, R., Dewey, E., Doherty, G., Dos Reis, L., Klopper, J., Lee, S., Lupo, M., Machac, J., Mechanick, J., Milas, M., Orloff, L., Randolph, G., Ross, D., Rowe, M., Smallridge, R., Terris, D., Tufano, R. and Urken, M. (2016) 'Improving the adoption of thyroid cancer clinical practice guidelines', *Laryngoscope*, 126(11), pp. 2640-2645

Lingsma, H., Roozenbeek, B., Perel, P., Roberts, I., Maas, A. and Steyerberg, E. (2011) 'Between-centre differences and treatment effects in randomized controlled trials: a case study in traumatic brain injury.', *Trials*, BMC, 12, p. 201

Lobach, D. and Hammond, W. (1997) 'Computerized decision support based on a clinical practice guideline improves compliance with care standards', *The American journal of medicine*, 102(1), pp. 89–98.

Maas, A., Steyerberg, E., Marmarou, A., McHugh, G., Lingsma, H., Butcher, I., Lu, J., Weir, J., Roozenbeek, B. and Murray, G. (2010) 'IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury', *Neurotherapeutics,* Elsevier, 7(1), pp. 127–134

Mans, R., Schonenberg, M., Song, M., van der Aalst, W. and Bakker, P. (2009) 'Application of process mining in healthcare–a case study in a dutch hospital', *Biomedical Engineering Systems and Technologies*, pp. 425–438

Marmarou, A., Lu, J., Butcher, I., McHugh, G., Mushkudiani, N., Murray, G., Steyerberg, E. and Maas, A. (2007) 'IMPACT Database of Traumatic Brain Injury: Design And Description', *Journal of Neurotrauma*, 24(2), pp. 239–250

McHugh, G., Butcher, I., Steyerberg, E., Lu, J., Mushkudiani, N., Marmarou, A., Maas, A. and Murray, G. (2007) 'Statistical approaches to the univariate prognostic analysis of the IMPACT database on traumatic brain injury', *Journal of Neurotrauma*, 24(2), pp. 251–258

McMahon, P., Hricik, A., Yue, J., Puccio, A., Inoue, T., Lingsma, H., Beers, S., Gordon, W., Valadka, A., Manley, G. and Okonkwo, D. (2014) 'Symptomatology and Functional Outcome in Mild Traumatic Brain Injury: Results from the Prospective TRACK-TBI Study', *Journal of Neurotrauma*, 31(1), pp. 26–33

Miles, A., Loughlin, M. and Polychronis, A. (2008) 'Evidence-based healthcare, clinical knowledge and the rise of personalised medicine', *Journal of Evaluation in Clinical Practice*, 14(5), pp. 621–649

Miller, M., Pasquale, M., Kurek, S., White, J., Martin, P., Bannon, K., Wasser, T. and Li, M. (2004) 'Initial Head Computed Tomographic Scan Characteristics Have a Linear Relationship with Initial Intracranial Pressure after Trauma', *The Journal of Trauma: Injury, Infection, and Critical Care*, 56(5), pp. 967–973

Mokri, B. (2001) 'The Monro-Kellie hypothesis: applications in CSF volume depletion', *Neurology*, 56(12), pp. 1746–8

Montani, S., Leonardi, G., Quaglini, S., Cavallini, A., Micieli, G. (2014) 'Improving structural medical process comparison by exploiting domain knowledge and mined information', *Artificial intelligence in medicine*, 62(1), pp. 33-45

Montani, S., Leonardi, G., Quaglini, S., Cavallini, A., Micieli, G. (2015) 'A knowledge-intensive approach to process similarity calculation', *Expert Systems with Applications*, 42(9), pp. 4207–4215

Morrow, S. E. and Pearson, M. (2010) 'Management strategies for severe closed head injuries in children', *Seminars in pediatric surgery*, 19(4), pp. 279–285

Nelson, A., Allen, K., Golightly, Y., Goode, A. and Jordan, J. (2013) 'A systematic review of recommendations and guidelines for the management of osteoarthritis: The Chronic Osteoarthritis Management Initiative of the U.S. Bone and Joint Initiative', *Seminars in arthritis and rheumatism*, pp. 1–12

Neumann, J.-O., Chambers, I., Citerio, G., Enblad, P., Gregson, B., Howells, T., Mattern, J., Nilsson, P. and Piper, I. (2008) 'The use of hyperventilation therapy after traumatic brain injury in Europe: an analysis of the BrainIT database', *Intensive care medicine*, 34(9), pp. 1676–1682

Nilsson, P., Enblad, P., Chambers, I., Citerio, G., Fiddes, H., Howells, T., Kiening, K., Ragauskas, A., Sahuquillo, J., Yau, Y., Contant, C. and Piper, I. (2005) 'Survey of traumatic brain injury management in European Brain-IT centres year 2001', *Acta neurochirurgica. Supplement*, 95, pp. 51–3

Pandher, P. and Bhullar, K. (2014) ‘Smartphone applications for seizure management’, *Health informatics journal*

Pascual, J., Georgoff, P., Maloney-Wilensky, E., Sims, C., Sarani, B., Stiefel, M., LeRoux, P. and Schwab, W. (2011) ‘Reduced brain tissue oxygen in traumatic brain injury: are most commonly used interventions successful?’, *The Journal of trauma: injury, infection and critical care*, 70(3), pp. 535–546

Patkar, V., Acosta, D., Davidson, T., Jones, A., Fox, J. and Keshtgar, M. (2012) ‘Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance’, *BMJ Open*, 2(3)

Peleg, M., Shahar, Y., Quaglini, S., Broens, T., Budasu, R., Fung, N., Fux, A., Garcia-Saez, G., Goldstein, A., Gonzalez-Ferrer, A., Hermens, H., Hernando, E., Jones, V., Klebanov, G., Klimov, D., Knoppel, D., Larburu, N., Marcos, C., Martinez-Sarriegui, I., Napolitano, C., Pallas, A., Palomares, A., Parimbelli, E., Pons, B., Rigla, M., Sacchi, L., Shalom, E., Soffer, P. and van Schooten, B. (2017a) ‘Assessment of a personalized and distributed patient guidance system’, *International Journal of Medical Informatics*, 101, pp. 108–130

Peleg, M., Shahar, Y., Quaglini, S., Fux, A., Garcia-Saez, G., Goldstein, A., Hernando, E., Klimov, D., Martinez-Sarriegui, I., Napolitano, C., Parimbelli, E., Rigla, M., Sacchi, L., Shalom, E. and Soffer, P. (2017b) ‘MobiGuide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains’, *User Modeling and User-Adapted Interaction*, 27(2), pp. 159–213

Peleg, M., Shahar, Y. and Quaglini, S. (2013) ‘Making Healthcare More Accessible; Better; Faster; and Cheaper: The MobiGuide Project’, *European Journal of ePractice*, 20, pp. 5–20

Perimal-lewis, L., Qin, S., Thompson, C. and Hakendorf, P. (2012) ‘Gaining Insight from Patient Journey Data using Process-Oriented Analysis Approach’, *HIKM '12 Proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge*, 129, pp. 59-66

Perimal-lewis, L., Vries, D. and Thompson, C. (2014) ‘Health intelligence :

Discovering the process model using process mining by constructing Start-to-End patient journeys', *HIKM '14 Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge*, 153, pp. 59–67

Shaw, M., Piper, I., Chambers, I., Citerio, G., Enblad, P., Gregson, B., Howells, T., Kiening, K., Mattern, J., Nilsson, P., Ragauskas, A., Sahuquillo, J. and Yau, Y. (2008) 'The brain monitoring with information technology (BrainIT) collaborative network: data validation results', *Acta Neurochirurgica Supplements*, 102, pp. 217–221

Piper, I., Chambers, I., Citerio, G., Enblad, P., Gregson, B., Howells, T., Kiening, K., Mattern, J., Nilsson, P., Ragauskas, A., Sahuquillo, J., Donald, R., Sinnott and Stell, A. (2010) 'The brain monitoring with Information Technology (BrainIT) collaborative network: EC feasibility study results and future direction', *Acta neurochirurgica*, 152(11), pp. 1859–1871

Prior, M., Guerin, M. and Grimmer-Somers, K. (2008) 'The effectiveness of clinical guideline implementation strategies - A synthesis of systematic review findings', *Journal of Evaluation in Clinical Practice*, 14(5), pp. 888–897

A. Ragauskas, V. Matijosaitis, R. Zakelis, K. Petrikonis, D. Rastenyte, I. Piper, G. Daubaris (2012) ' Clinical assessment of noninvasive intracranial pressure absolute value measurement method', *Neurology*, 78(21)

Roberts, I. and Sydenham, E. (2009) 'Barbiturates for acute traumatic brain injury', Cochrane Database of Systematic Reviews 2021, 12, p1465-1858

Rood, E., Bosman, R., van der Spoel, J., Taylor, P. and Zandstra, D. (2005) 'Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation', *Journal of the American Medical Informatics Association*, 12(2), pp. 172–180

Sackett, D., Rosenberg, W., Gray, M., Haynes, B. and Richardson, S. (1996) 'Evidence based medicine : what it is and what it isn't', *BMJ*, 312(71)

Saeed, M. (2007) 'Temporal Pattern Recognition in Multiparameter ICU Data', *PhD thesis, Massachusetts Institute of Technology*

Shortliffe, E. H. (1987) ‘Computer programs to support clinical decision making’, *JAMA : the journal of the American Medical Association*, 258(1), pp. 61–66

Sinnott, R., Han, J., Hu, W., Ma, X. and Yu, K. (2015) ‘Application of Mobile Games to Support Clinical Data Collection for Patients with Niemann-Pick Disease’, *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 1–6

Smielewski, P., Czosnyka, M., Steiner, L., Belestri, M., Piechnik, S. and Pickard, J. (2005) ‘ICM+: software for on-line analysis of bedside monitoring data after severe head trauma’, *Acta neurochirurgica. Supplementum*, 95, pp. 43–49

Sorrentino, E., Diedler, J., Kasprowicz, M., Budohoski, K., Haubrich, C., Smielewski, P., Outtrim, J., Manktelow, A., Hutchinson, P., Pickard, J., Menon, D. and Czosnyka, M. (2012) ‘Critical thresholds for cerebrovascular reactivity after traumatic brain injury’, *Neurocritical Care*, 16(2), pp. 258–266

Steinman, M., Fischer, M., Shlipak, M., Bosworth, H., Oddone, E., Hoffman, B. and Goldstein, M. (2004) ‘Clinician awareness of adherence to hypertension guidelines’, *The American journal of medicine*, 117(10), pp. 747–754

Stell, A., Moss, L. and Piper, I. (2014) ‘Building an Empirical Treatment Protocol from High-Resolution Traumatic Brain Injury Data’, *HIKM '14 Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge*, 153, pp. 79–88

Stell, A., Sinnott, R. and Jiang, J. (2009) ‘A federated data collection application for the prediction of adverse hypotensive events’, 9th International Conference on Information Technology and Applications in Biomedicine, IEEE, pp. 1–4.

Synnot, A., Gruen, R., Menon, D., Steyerberg, E., Buki, A., Peul, W., Elliott, J and Maas, A. (2016) ‘A New Approach to Evidence Synthesis in Traumatic Brain Injury: A Living Systematic Review’, *Journal of Neurotrauma*

Talving, P., Karamanos, E., Teixeira, P., Skiada, D., Lam, L., Belzberg, H., Inaba, K. and Demetriades, D. (2013) ‘Intracranial pressure monitoring in severe head injury: compliance with Brain Trauma Foundation guidelines and effect on outcomes: a prospective study’, Journal of Neurosurgery, 119(5), pp. 1087–1357

Tarapore, P., Vassar, M., Cooper, S., Lay, T., Galletly, J., Manley, G. and Huang, M. (2016) 'Establishing a TBI Program of Care – Benchmarking Outcomes after Institutional Adoption of Evidence-based Guidelines', *Journal of Neurotrauma*, 33(22)

Taylor, P., Fox, J. and Pokropek, A. T. (1999) 'The development and evaluation of CADMIUM: A prototype system to assist in the interpretation of mammograms', *Medical Image Analysis*, 3(4), pp. 321–337

Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., Glasziou, P., Shemilt, I., Synnot, A., Turner, T. and Elliott, J. (2017) 'Living systematic reviews: 2. Combining human and machine effort', *Journal of Clinical Epidemiology*, 91, pp. 31–37

Watters, W. (2008) 'Defining evidence-based clinical practice guidelines', *American Academy of Orthopaedic Surgeons*

van der Werf, J., Verbeek, H. and van der Aalst, W. (2012) 'Context-Aware Compliance Checking', *International Conference on Business Process Management*, BPM 2012, pp. 98–113

Xing, W. and Ghorbani, A. (2004) 'Weighted PageRank algorithm', *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004*, pp. 305–314

Zheng, Z., Bruns Jr, L., Li, J. and Sinnott, R. (2017) 'A Mobile Application and Cloud Platform Supporting Research into Alcohol Consumption', *Proceedings of 1st International Conference on Medical and Health Informatics*, pp. 48–55.

Books

Crockard A., Hayward, R. and Hoff J. (1992) *"Neurosurgery"*, *Blackwell Scientific*, Boston, 2nd ed, vol 1

Dijkman, R., Dumas, M. and Garcia-Banuelos, L. (2012) "Business Process Graphs: Similarity Search and Matching", chapter 18 in Graph Data Management: Techniques and Applications, ed: Sakr, S. and Pardede, E., *Information Science Reference*

Howick, J. (2011) "The Philosophy of Evidence-based Medicine", *John Wiley & Sons*, p. 15

Panagacos, T. (2012) *"The Ultimate Guide to Business Process Management: Everything You Need to Know and How to Apply It to Your Organization" CreateSpace Independent Publishing Platform*

Payne, S. (2016) *"Cerebral autoregulation – control of blood flow in the brain"*, Springer, Cham

Reilly, P. and Bullock, R. (2005) *"Head Injury: pathophysiology and management"*, *CRC Press*, London, 2nd ed, vol 1

Zobel, Justin (2014) "Writing for Computer Science", London, Springer-Verlag

Websites

BrainTrauma (2018), Available at: https://braintrauma.org [Accessed 15/08/18]

Camunda, (2018), Available at: https://camunda.com/bpmn/ [Accessed 15/08/18]

CENTER-TBI (2018), Available at: https://www.center-tbi.eu/ [Accessed 02/10/18]

ClinicalTrials (2018), Available at: https://clinicaltrials.gov/ [Accessed 02/10/18]

Cnossen, M. C. (2016, c) 'Outcome and Comparative Effectiveness Research in Traumatic Brain Injury: A methodological perspective'. Available at: https://core.ac.uk/download/pdf/154418553.pdf [Accessed on 05/06/19]

Cochrane (2018), Available at: https://www.cochrane.org/ [Accessed 02/10/18]

COSSAC (2014), Available at: http://www.cossac.org [Accessed 30/06/14]

CVPhysiology (2014), Available at: http://www.cvphysiology.com/Blood%20Flow/BF004.htm [Accessed 16/06/14]

Carescape (2018), Available at: https://www.gehealthcare.com/en/products/patient-monitoring/patient-monitors/carescape-v100 [Accessed 02/10/18]

Headway, (2018) Available at: http://www.headway.org.uk [Accessed 20/05/18]

ICCA (2018), Available at: https://www.usa.philips.com/healthcare/product/HCNOCTN332/intellispace-critical-care-and-anesthesia-critical-care-information-system [Accessed 10/09/18]

IJCIIS (2017), Available at: http://www.ijciis.org [Accessed 23/12/17]

InTBIr (2017), "International Initiative for Traumatic Brain Injury Research" – local PDF file, shared by participants, available on request [Accessed 01/10/18]

JENA (2012), Available at: http://incubator.apache.org/jena [Accessed 05/06/12]

Kless (1993) adapted from C. Rolland - Modeling the Requirements Engineering Process, 3rd European-Japanese Seminar on Information Modelling and Knowledge Bases, Budapest, Hungary, June 1993, Public Domain, https://commons.wikimedia.org/w/index.php?curid=25917562

Kuleuvencongres (2018), Available at: https://kuleuvencongres.be/icp2019 [Accessed 02/10/18]

MagicQuadrant (2018), Available at: https://en.wikipedia.org/wiki/Magic_Quadrant [Accessed 10/09/18]

Mobiguide (2014), Available at: http://www.mobiguide-project.eu [Accessed 30/06/14]

NICE (2014), Available at: http://www.nice.org.uk/guidance/CG56 [Accessed 26/06/14]

Ontology (2018), Available at: https://en.wikipedia.org/wiki/Ontology_(information_science) [Accessed 14/08/18]

Carevue (2018), Available at: https://www.usa.philips.com/healthcare/solutions/patient-monitoring [Accessed 02/10/18]

Surviving Sepsis, (2019), Available at: http://www.survivingsepsis.org/Resources/Pages/Surviving-Sepsis-App.aspx [Accessed 06/06/19]

Trauma, (2014), Available at: http://www.trauma.org/archive/neuro/icp.html [Accessed 15/06/14]

OWL (2012), Available at: http://www.w3.org/TR/owl-ref [Accessed 06/06/12]

XES (2015), Available at: http://www.xes-standard.org/ [Accessed 25/01/15]

Citations

Messmer, Bruno T. (1995), *Phd Thesis*, "Efficient graph matching algorithm"

Dijkman, R., van Dongen, B., Dumas, M., Garcia-Banuelos, L., Kunze, M., Leopold, H., Mendling, J., Uba, R., Weidlich, M., Weske, M. and Yan, Z. (2013) "A Short Survey on Process Model Similarity". In: Bubenko J., Krogstie J., Pastor O., Pernici B., Rolland C., Sølvberg A. (eds) *Seminal Contributions to Information Systems Engineering*, Springer, Berlin

Eysenck H J. (1994) "Systematic Reviews: Meta-analysis and its problems" BMJ, 309, p. 789

Ghajar, J., Hariri, R., Narayan, R., Iacono, L., Firlik, K. and Patterson, R. (1995) "Survey of critical care management of comatose, head-injured patients in the United States", *Critical Care Medicine*, 23(3), pp. 560-567

Gruber, T. (2008) "Ontology - Entry in the Encyclopedia of Database Systems", *Springer-Verlag*

Moss, L. – "Explaining Anomalies: An Approach to Anomaly-Driven Revision of a Theory", *PhD thesis*, University of Aberdeen 2010

Sahuquillo J. (2006) "Decompressive craniectomy for the treatment of refractory high intracranial pressure in traumatic brain injury" *Cochrane Database of Systematic Reviews*, Issue 1

Servadei F., Antonelli V., Giuliani G., Fainardi E., Chieregato A., Targa L. (2002) "Evolving Lesions in Traumatic Subarachnoid Hemorrhage: Prospective Study of 110 Patients with Emphasis on the Role of ICP Monitoring", In: Czosnyka M., Pickard J.D., Kirkpatrick P.J., Smielewski P., Hutchinson P. (eds) *Intracranial Pressure and Brain Biochemical Monitoring. Acta Neurochirurgica Supplements*, vol 81, Springer

Appendices

A. Literature Review Methods

The strategy for reviewing literature for this thesis is outlined in this section.

A.1. Research Questions

The main review of literature covered the following research areas:

- Review novel and established tools which encourage and/or monitor adherence to clinical guidelines (technological and non-tech)
- Review common issues encountered when trying to improve adherence to clinical guidelines (general and TBI-specific)

A.2. Eligibility Criteria

- Studies published in English between 1994 and 2018[20]
- Studies which had a clinical implementation and/or evaluation

A.3. Exclusion Criteria

- Studies which were not peer-reviewed (except theses)
- Poster abstracts
- Studies which did not include clinical guidelines

A.4. Search terms with dates

Whilst a general and less systematic search continued throughout the course of the research, the following were periods particularly devoted to organised literature searching, so provide the most relevant reference points.

February 2012 (then re-run in June 2016)

- *"Computerised/Computerized clinical guidelines"*
- *"Computer-Interpretable Clinical/Medical Guidelines"*
- *"Clinical workflows"*
- *"Workflow Patterns"*
- *"Evidence-Based Medical/Clinical Guidelines"*

[20] Though some older papers were referenced, such as (Shortliffe, 1987) published in 1987.

- *“Decision Support Systems/DSS/Clinical Decision Support Systems/CDSS”*
- *“Clinical Guideline Adherence”*

February 2013

- *"Clinical management variation studies" [+neuro, +icu, +tbi]*
- *"TBI RCT reliability studies" [+ expanded acronyms]*
- *"TBI physiological data" [+ expanded acronyms]*
- *"TBI clinical management"*
- *“adherence to BTF guidelines”*
- *“deviation from BTF guidelines”*
- *“why do clinicians not follow the BTF guidelines”*
- *"physiological data analysis icu"*
- *"interpreting ICU data"*
- *"capturing ICU data"*
- *"capturing and interpreting ICU data TBI"*
- *"statistical analysis of icu data tbi"*
- *"time series variation icu data tbi"*
- *“TBI data of low/high quality resolution”*
- *"Therapy intensity level tbi"*
- *"clinical management variability studies icu"*
- *"clinical management variability studies icu tbi"*
- *"quantifying icu clinical management"*
- *"minimising tbi management variability"*
- *"variation in TBI management"*

February 2014

- *"adherence to medical guidelines"*
- *"clinician adherence to medical guidelines"*
- *"physician adherence to medical guidelines"*
- *"medical guideline adherence improvement"*
- *"clinical decision support tools"*
- *"icu clinical decision support tools"*

- *"icu clinical decision support tools traumatic brain injury"*

A.5. Search strategy

Search fields for keywords in the title and abstract of papers were returned. Metadata was recorded using the Mendeley Reference Manager software (www.mendeley.com). Lists and digests of papers based on pre-entered keywords and search behaviour on Google Scholar were also directly received by email weekly. Of the included papers, reference searches for further related papers was then conducted.

Total paper number in Mendeley = **317**
Total paper number referenced in thesis (including books and websites) = **137**

A.6. Sources

A.6.1. Search Engines & Digital Libraries

- Google Scholar
- ACM Digital Library
- IEEE Xplore Digital Library
- Springer Link
- Science Direct
- Pubmed

A.6.2. Journals

- Artificial Intelligence in Medicine
- Journal of Clinical Monitoring & Computing
- BMC Medical Informatics and Decision Making
- Journal of the American Medical Informatics Association
- Journal of Biomedical Informatics
- Health Informatics Journal
- Methods of Information in Medicine
- Journal of Medical Systems
- The New England Journal of Medicine
- The Medical Journal of Australia

- The Lancet
- Journal of Neurotrauma
- Journal of Neurosurgery
- The Journal of Trauma: Injury, Infection and Critical Care
- Journal of the American Medical Association
- Critical Care Medicine
- Acta Neurochirurgica

A.7. Most current literature

A final appraisal of literature was conducted in September 2018 to check on the most up-to-date developments in projects relevant to the research work. This was primarily focused on the CENTER-TBI project, in particular the PhD of Marie Cnossen (which specifically concerned TBI guidelines), and the latest InTBIr meeting (Oct 2017).

B. Knowledge representation through domain ontologies

One of the initial avenues of research work in this thesis was to explore the use of ontologies to achieve the aims of the thesis. This culminated in a short paper that was accepted for publication to the CBMS (Computer-Based Medical Systems) conference proceedings, presented in Rome in 2012, an excerpt of which is presented in this section.

B.1. System architecture

The AMITIE system (Automated Medical Intervention and Treatment Inference Engine) has been developed to identify abnormal physiological events and automatically infer subsequent medical interventions from time series physiological and treatment data. Figure B.1 provides a high-level overview of the system.

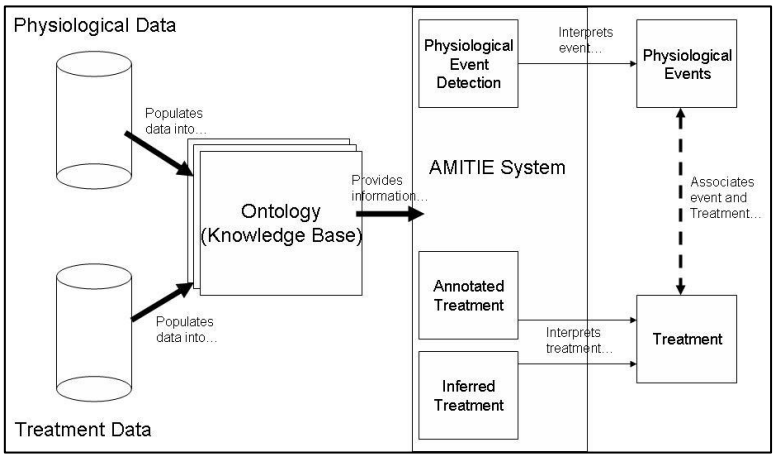

Figure B.1: High-level overview of the AMITIE system

The patient data is explored for instances of abnormal physiological readings. Once identified, the data is further examined to find related interventions given to the patient. In some cases, the interventions are easy to identify as they have been specifically recorded in the patient data-set, with an explicit target noted. However, as described above interventions often have to be inferred. To enable an intervention to be inferred, detailed information can be obtained from domain ontologies. Information such as the known physiological effects of the intervention and other contextual information (e.g. contraindications of a drug) can help to determine whether it is likely that the patient has received the intervention for the abnormal reading. For example, consider a series of abnormally raised

intracranial pressure readings which then return to normal. If the data-set is examined and no intervention is recorded, it may be reasonable to infer from the observation of the patient's temperature decreasing, that the procedure 'therapeutic cooling' has been administered.

In the AMITIE system, a set of ontologies are used to model the domain. An ontology "*defines a set of representational primitives with which to model a domain of knowledge or discourse*." (Gruber, 2008) AMITIE's knowledge base consists of three OWL (W3 – OWL, 2011) domain ontologies which model the medical domain, patient data and human physiology. This knowledge base has been reused from previous work on the EIRA system (Moss, 2010). The following high-level algorithm summarises the functionality of the AMITIE system:

1. Identify an abnormal physiological event (E)
   a. Characterize E into whether it has returned to a baseline ("normal") value or not.
   b. If it has, it is assumed that the patient has been treated and the time period (TP) of E is determined.
   c. Examine TP for instances of annotated interventions
      i. If intervention (I) is noted in data-set: suggest that I has been given in response to E.
      ii. Else, infer non-annotated intervention:
      iii. Identify known physiological effects of possible interventions for E.
      iv. Examine TP for evidence of any of these effects.
      v. If possible intervention (I_p) is found:
      vi. suggest I_p has been given in response to E
      vii. Else, suggest that patient returned to baseline value without intervention.

Issues of negation and ranking have not yet been handled. This algorithm also assumes a one-to-one relationship between event and physiological output. However, it is likely that combinations and emergent factors make this relationship more complex.

B.2. Implementation

The AMITIE system software is written using the JENA API (Apache – JENA, 2011) [incubator.apache.org/jena/]. To obtain the information from the ontology, a set of SPARQL queries have been implemented. These allow for a separation of the inference process from the application code (for later re-use in other medical contexts). They have the following functions and features:

1) The first query obtains all the physiological data above a certain threshold:

SELECT ?timepoint ?physiovalue

WHERE {?x <http://www.owl-ontologies.com/amitie.owl#hasTime>?timepoint .

[...]

?reading <http://www.owl-ontologies.com/amitie.owl#readingParameter>

<http://www.owl-ontologies.com/amitie.owl#"**[physiolabel]**"> .

FILTER (?physiovalue > "**[physioThreshold]**"}.}

Of the variables in the query, “physiolabel” refers to the physiological parameter that is being queried (e.g. heart rate) and “physioThreshold” refers to the value above which an abnormal event is deemed to have occurred (e.g. ICP readings above 20 mmHg are generally considered to require treatment).

2) The second query obtains all the treatments that *are* annotated in the data-set within the time period of a single event.

3) The information required to infer interventions, for when a treatment is not annotated in the data-set, is extracted from the ontology using five queries that interrogate the following features to see if they are present (as they are significant in the event signature):

- *High_Feature*
- *Low_Feature*
- *Increase_Parameter_Change*
- *Decrease_Parameter_Change*
- *Constant_Parameter*

4) The final query obtains other physiological data that may relate to the abnormal event in question but has not been retrieved in the original query. The layout is similar to the first query but is constrained by time series rather than threshold physiological values.

B.3. Results

Using this combination of domain knowledge and patient data, the resulting management of a patient's abnormal event is determined to either be an annotated treatment, a non-annotated treatment, or the patient's vital sign has returned to normal without any clinical intervention. Table B.1 shows the results of the AMITIE system run against three patients selected from the Brain-IT database.

Patient ID	**No. of ICP events**	**Annotated Treatments for each ICP event**	**Inferred Treatments for each ICP event**
15026161	25	0	Ventilation and cerebral vasoconstriction
15127262	209	0	Ventilation and cerebral vasoconstriction
15137626	77	0	Ventilation, induced hypothermia and cerebral vasoconstriction

Table B.1: AMITIE results

B.4. Discussion

Ultimately, this research avenue was abandoned due to similar issues as those quoted in section 3.4.3. (outlining the use of clinical guideline formalisms and ontologies). These included: the over-specification of surrounding context information, only a small sub-section of which was directly relevant to the guideline evaluation (the EUSIG parameter definitions); the difficulty in re-

purposing an ontology for secondary usage; and over-engineering of the solution, which required a large overhead in new technologies that were ultimately unnecessary – for instance, the SPARQL query language over regular Standard Query Language (SQL) to query data repositories, and the use of the Protégé ontology editor over a regular text editor.

Another issue was that the inference engine presented in this paper (AMITIE) claimed to do the detection and association of physiological events and treatments. At a very basic level this was correct, but the implementation of this inference was weak and ultimately required the more robust method implementation that can be found in chapters 4-6 of the thesis.

Though the primary use of ontologies was abandoned, this initial work was instructive in highlighting the features required for the eventual solution. Though the BPMN notation was ultimately chosen to represent the BTF guidelines (see chapter 4), an ontology with more contextual domain knowledge of an ICU would be a possible alternative.

C. Brain Trauma Foundation Guidelines

The guidelines followed for this thesis are outlined in full here. They are comprised of Table 3 ("Thresholds") in (Carney *et al.*, 2016), which is the 4th edition update of the guidelines.

C.1. Brain Trauma Foundation guidelines - 4th Edition

C.1.1. Blood pressure thresholds

Level III

- *"Maintaining SBP at >=100 mm Hg for patients 50 to 69 years old or at >=110 mm Hg or above for patients 15 to 49 or >70 years old may be considered to decrease mortality and improve outcomes."*

C.1.2. Intracranial pressure thresholds

Level IIB

- *"Treating ICP >22 mm Hg is recommended because values above this level are associated with increased mortality."*

Level III

- *"A combination of ICP values and clinical and brain CT findings may be used to make management decisions."*

C.1.3. Cerebral perfusion pressure thresholds

Level IIB

- *"The recommended target CPP value for survival and favorable outcomes is between 60 and 70 mm Hg. Whether 60 or 70 mm Hg is the minimum optimal CPP threshold is unclear and may depend upon the auto-regulatory status of the patient."*

Level III

- *"Avoiding aggressive attempts to maintain CPP >70 mmHg with fluids and pressors may be considered because of the risk of adult respiratory failure"*

C.2. Brain Trauma Foundation guidelines - 3rd Edition

A revision of these guidelines occurred in August 2016 (Carney *et al.*, 2016), during the course of this research work.

Examples of the transition between the 3rd and 4th editions (the 3rd being published in 2007) are shown below.

C.2.1. Blood pressure thresholds

Level II

- *"Blood pressure should be monitored and hypotension (systolic blood pressure < 90 mmHg) avoided."*

Level III

- *"Oxygenation should be monitored and hypoxia (PaO2 < 60 mmHg or O2 saturation < 90%) avoided."*

The differences between these recommendations and those which appear in the 4th edition are quite large: references to hypotension and oxygenation have been removed and replaced with a single consideration of patient age.

C.2.2. Intracranial pressure thresholds

Level II

- *"Treatment should be initiated with intracranial pressure (ICP) thresholds above 20 mmHg."*

Level III

- *"A combination of ICP values, and clinical and brain CT findings, should be used to determine the need for treatment"*

The only difference between these recommendations and those which appear in the 4th edition is with the monitoring threshold: 20 mmHg in the 3rd edition; 22 mmHg in the latest.

C.2.3. Cerebral perfusion pressure thresholds

Level II

- *"Aggressive attempts to maintain cerebral perfusion pressure (CPP) above 70 mm Hg with fluids and pressors should be avoided because of the risk of adult respiratory distress syndrome (ARDS)"*

Level III

- *"CPP of <50 mm Hg should be avoided"*
- *"The CPP value to target lies within the range of 50-70 mm Hg. Patients with intact pressure autoregulation tolerate higher CPP values"*
- *"Ancillary monitoring of cerebral parameters that include blood flow, oxygenation, or metabolism facilitates CPP management"*

For CPP, the differences are structural as well as in the content: the lower limit of the target value of CPP has changed (50 mmHg in the 3rd, 60 mmHg in the 4th). The certainty of the recommendations have changed places (the considerations of ARDS is level 3 in the 4th edition, where it was level 2 in the 3rd, and vice versa with the CPP target information), along with a new sub-category of "IIb" introduced in the 4th edition.

D. Automated process mining

This section details the work invested in attempting to automatically – as opposed to manually – derive process models from the neurological ICU data in this research work. The following excerpt is from a paper that was submitted (unsuccessfully) to the CBMS (Computer-Based Medical Systems) conference in Sao Paolo, 2015. It details the main methods used which were the conversion of the physiological and treatment data to XES format (XES-Standard, 2015), and the use of the PrOM framework (Dongen, 2005) to generate a process model using simple heuristic filtering and the "alpha-algorithm" process miner.

D.1. Conversion of XES standard

The XES standard (eXtensible Event Stream) is an XML representation of event logs that provides a generalized format from which process models can be extracted. Represented by the authors as being simple, flexible, extensible and expressive, the basic structure includes a log, made up of traces, which are in turn made up of events. All entities can have attributes of primitive types familiar to programmers (e.g. string, int, boolean, etc), which can be nested within each other and can have varying scope (e.g. within a trace, within a log, global, etc).

To convert the raw EUSIG information taken from the ICU data stream into XES, the event information is translated to the appropriate extension definitions, with the mapping as shown in table D.1. This is done programmatically, using the event and treatment information from (STELL, MOSS and Piper, 2014) as the input, and an XML file using the XES templates provided by the ProM authors as the output.

EUSIG Feature	Example	XES Attribute
Event start	Timestamp	*lifecycle: transition*
Event end	Timestamp	*lifecycle: transition*
Event type	'ICP'	*concept: name*
Event threshold	'Raised over threshold'	*concept: instance*
Treatment start	Timestamp	*lifecycle: transition*
Treatment end	Timestamp	*lifecycle: transition*
Treatment target	'ICP'	*concept: name*
Treatment type	'Analgesics'	*concept: instance*

Table D.1: Mapping of data-set features from the Brain-IT data-set to the features as specified in the XES standard

Additional to these, two event classifiers are defined in the header of the XES file. Labelled as *'MXML Legacy Classifier'* (using templates from the ProM authors) this classifier counts all objects that match on the *'concept: name'* and *'lifecycle: transition'* attributes. Another classifier – *'Event name'* – matches on *'concept: name'* only. Using these classifiers and the mappings from table D.1, a full log-file is built up of events, which are part of traces drawn from the 262 individual Brain-IT patient information streams. This creates the full "Brain-IT XES log-file".

D.2. Generation of process model

The ProM framework allows an event log file to be imported, filtered and analyzed. The framework has three major components: workspace, actions and views. The workspace acts as the home file directory allowing import and browsing of event log files. The actions section allows selection of a list of "miner" plugins that can be run to filter and analyze the event log. The views section shows the output of all analysis in tabular and visualized format.

In this work, the event log generated by the XES event log file is filtered using the *"Simple Heuristics"* plugin. This filters out events that are irrelevant to the process that is being analyzed or those that have been picked up from the previous conversion incorrectly (for instance any treatment label that has been assigned to

the target due to a transposition in the original database columns). Types of start, end, and all other event and treatment classes are explicitly selected. After this filter has been run the number of processes, cases, events (with classes and types), and originators will be tabulated.

Once this data-set is filtered, the log is mined for a process model using the alpha-algorithm. The alpha-algorithm creates a work-flow net from a work-flow log by examining the causal relationships between tasks within that work-flow (full details of this process can be found in (van der Aalst, Weijters and Maruster, 2004)).

D.3. Results

In general, the process model outputs show a standard relationship between start and end tags for ICP and CPP events throughout all traces (as expected). The interactions of interest are the treatment responses to physiological events. Covering the 13 main treatment categories (listed in the output graphs of (STELL, MOSS and Piper, 2014)), the listing of occurrences allows the strength of the relationships between treatment and physiological events to be evaluated for particular centre traces, individual patient traces, or for the overall trace of the 262 Brain-IT patients.

Table 2 shows an example centre event log which has had the simple heuristic filter applied. When filtered on all of the treatment information, the list showed that the two most frequently occurring treatments were volume expansion and paralysis (accounting for 80% of all instances within the log, the rest being made up of the other 13 treatment categories).

Class	Occurrences (absolute)
CPP	2138
ICP	1638
Volume_Expansion	127
Paralysis	98

Table D.2: Distribution of events and treatments throughout Brain-IT trace log after filtering with simple heuristics

Using these four classes in the filtered event log, a petri-net diagram was then generated using the alpha algorithm plugin (figure 1). The nodes and arrows indicate the direction of dependency in time of each entity. For instance, ICP start and CPP start always lead to ICP complete and CPP complete, respectively (which is intuitive as only well-defined physiological events with start and end tags are included).

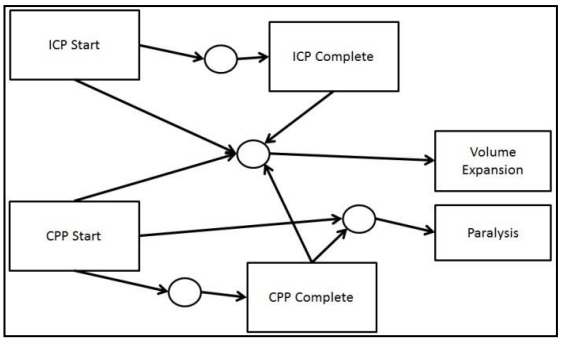

Figure D.1: Petri-net of processes derived from the Brain-IT data-set using the alpha-algorithm on an example center (Vilnius)

In terms of the dependencies of the treatments on the ICP/CPP events, volume expansion occurs in response to ICP start/complete (which indicates that the treatments are given in response to both the start and completion of an event, depending on the clinical circumstances). Volume expansion also occurs in response to CPP event triggers. But according to the event log paralysis only occurs in response to CPP events, not ICP.

A possible medical explanation for this is that paralysis would be applied to ICP events in isolation rather than CPP (paralysis is applied when a patient "fights" a ventilator, their CO2 processing increases, which causes their ICP to rise simultaneously). However, this process model suggests that the treatment is being applied by measuring the CPP instead. Further study to validate this finding would indicate whether this process model does actually show whether they have deviated from the consensus treatment practice for such events.

D.4. Discussion

Ultimately, this work did not produce viable process models, most likely due to the use of heuristics and simple process miner algorithms, both of which are not generally tolerant of "noisy" real-world data. The avenue of research was abandoned as the amount of manual constraints applied to the inputs and outputs

of the processing – in order to produce process models that represented recognisable processes in an ICU – suggested that a manual method (chapter 4) would be just as effective and was also more transparent (rather than depending upon “black box” tools). The trade-offs between both methods in terms of errors and uncertainties (see end of section 5.2) also suggested that either option was a viable path.

E. Mathematical representation of graph-edit algorithms

In this thesis the algorithm used for comparison of process models is an exhaustive one, checking all available possibilities between all states, which is possible due to the constrained nature of the problem space (the categories available for all nodes and edges are low in number by design). This is in contrast to the open problem formulated for all process model solutions in (Dijkman, Dumas and García-Bañuelos, 2009), leading to the designation of that issue as an "NP-complete problem" – one which is of sufficient complexity that finding solutions with proofs that can be verified in polynomial time (i.e. a reasonable time-scale) is not considered likely. The four algorithms they used are reproduced here for future reference, as they may yet be applicable to the research space.

E.1. Greedy

"The algorithm starts by marking all possible pairs of nodes from the two graphs as open pairs. In each iteration, the algorithm selects an open pair that most increases the similarity induced by the mapping and adds this pair to the mapping. The selected pair consists of two nodes. Since each node can only be mapped once, the algorithm removes from the set of open pairs, all pairs in which one of the selected nodes appears. The algorithm iterates until there is no open pair left that can increase the similarity induced by the mapping".

E.2. Exhaustive with pruning

"The algorithm recursively explores all possible mappings, but when the recursion tree reaches a certain size, the algorithm prunes it to keep only the mappings with the highest similarity. In the extreme case, the algorithm is thus exponential, but the pruning parameters will control its complexity. The algorithm starts by initializing the set of unfinished mappings to an empty mapping, with all nodes from the two graphs mapped as 'free' to be mapped. It repeatedly prunes the set of unfinished mappings and performs a step in which finished mappings are added to the set of finished mappings and unfinished mappings are extended with an additional pair of nodes. It repeats this until there are no more unfinished mappings. It then returns the finished mapping with the highest similarity score."

E.3. Process heuristic

"This algorithm is a variation of the exhaustive algorithm. It also builds a recursion tree of possible mappings, but it starts by mapping the source nodes of the business process graphs, then mapping nodes that immediately follow the source nodes, etc. Since it is plausible that nodes closer to the start of a process should be mapped to nodes closer to the start of the other process (and conversely), this should yield a higher-quality pruning. Indeed, the algorithm is more likely to prune mappings with node pairs that are further apart in terms of their distance to the starts of their processes."

E.4. A-star search

"This algorithm is based on the well-known A-star heuristic search, which has been applied to the problem of graph matching in (Messmer, 1996). In each step, the algorithm selects the existing partial mapping map with the maximal graph edit similarity. The algorithm then takes a node n1 from graph G1 that has not yet been mapped and creates a mapping between this node and every node n2 of G2 such that n2 does not already appear in map. Let us say that m such nodes n2 exist. The algorithm then creates m new mappings, by adding (n1, n2) to map. In addition, one mapping is created where (n1, ?) is added to map (? is a "dummy" node). This latter pair represents the case where node n1 has been deleted. This step is repeated until all nodes from G1 are mapped. It can be proven that the result is an optimal mapping."

F. Known issues

The following issues were discovered during the course of writing this thesis. The first was unable to be fixed before the date of submission, the second and third were considered after feedback from examiners at the viva.

F.1. Time formatting issue

Dataset: Brain-IT (chapter 9 – dataset analysis)

Nature: the issue involves the incorrect parsing of date/timestamps. The formatter used was converting 24-hour readings to a 12-hour clock.

Impact: Low/Moderate

50% of EUSIG pressure events in the dataset have their timing offset by 12 hours. The event counts remain accurate, but the number of treatment associations will likely be reported as approximately 50% lower than is actually the case. This affects the consideration of how valid the adherence measurements from the Brain-IT dataset are (i.e. how many useful instances of adherence reporting, which depend on those treatment associations, are present). Currently around 17% in the Brain-IT dataset have associations. This error suggests that, once fixed, the number should be closer to 34%.

Resolution: re-run the event detection program for the Brain-IT dataset with the formatter fixed, re-compile the indexing for the treatment profiles database, then re-run the counts of event/treatment associations.

F.2 Severity chart issue

Dataset: all

Nature: It has been considered that the transformation of the guideline output to metric A (duration / non-adherence) is unnecessary, and the severity of duration combined with non-adherence would be best expressed only with metric B (duration * non-adherence). Requiring modification to the two-dimensional severity charts.

Impact: Moderate

The rationale for this change is as follows:

- If both duration and non-adherence are high, then the output has high severity
- If both duration and non-adherence are low, then the output has low severity
- If either factor has a disproportionate influence on metric B then this will affect the severity accordingly.

This would be a reasonable clinical analogue for assessing severity as expressed by non-adherence to guidelines, and it is considered that entity A does not influence this output at all.

A separate but related consideration is that the severity charts themselves would have most utility in a bedside ICU setting, and not in a retrospective audit situation, as in this latter case, time would be available to go into the more subtle and nuanced depth for the reasons for non-adherence (i.e. clinicians can make their own mind up about the case's severity).

Resolution: express the severity as a linear scale rather than a two-dimensional chart.

F.3 Description of "default" instance value

Dataset: all

Nature: a recalculation of the factors that make up the structural difference between two process models, resulted in distance score reporting as 37.8% instead of 36.2%.

Impact: Low

Wherever the value of 36.2% is encountered in the thesis, this should be 37.8%. The clinical/informational effect is negligible.

Resolution: apply the amended weighting factors to the code and re-process.