

Supporting Clinical Trials to Predict Adverse Events in the Brain Trauma Domain

Anthony Stell¹, Richard Sinnott¹, Rob Donald², Iain Chambers⁴, Giuseppe Citerio⁶, Per Enblad⁷, Barbara Gregson⁸, Tim Howells⁷, Karl Kiening⁹, Pelle Nilsson⁷, Arminas Ragauskas¹⁰, Juan Sahuquillo⁵ and Ian Piper³ (on behalf of the Brain-IT Group)

¹Melbourne eResearch Group
University of Melbourne
Melbourne, Australia

²Dept of Statistics
University of Glasgow
Glasgow, UK

³Dept of Clinical Physics
Southern General Hospital
Glasgow, UK

⁴Dept of Medical Physics
The James Cook University Hospital
Middlesbrough, UK

⁵Neurosurgery
Vall d'Hebron University Hospital
Barcelona, Spain

⁶Neuroranimazione
Hospital San Gerardo
Monza, Italy

⁷Dept of Neurosurgery
Uppsala University Hospital
Uppsala, Sweden

⁸Dept of Neurosurgery
Newcastle General Hospital
Newcastle, UK

⁹Neurosurgery
Ruprecht-Karls-Universität Hospital
Heidelberg, Germany

¹⁰Telematics Sc. Lab
Kaunas University of Technology
Kaunas, Lithuania

Abstract

There are many serious and acute physiological conditions about which we have incomplete medical knowledge. To address this and develop effective treatments it is often the case that a wealth of clinical data is required for collection, analysis and feedback. Whilst such data often exists it is typically held in a variety of different formats and locations. This paper describes the EU FP7-funded Avert-IT project (www.avert-it.org), which has developed an integrated, real-time physiological data infrastructure (ICUnet) to address the specific issue of prediction of hypotensive events in the brain trauma domain. This system has been used to support a major multi-centre clinical trial. In this paper, the implementation and application of the ICUnet system is described, followed by the design and results of the clinical trial.

1. Introduction

Hypotension is abnormally low blood pressure (BP) and occurs unpredictably in patients managed in intensive care. If left untreated it can be particularly harmful in unconscious head injured patients who have impaired ability to regulate cerebral blood flow and can impact on long-term outcome. The treatment of hypotension in many intensive care units is common. Most physiological monitoring systems are designed to provide notification “when” a hypotensive status has been detected, through measurement of the blood pressure (i.e. once

the event is already underway). This in turn means that the clinician must intervene clinically often through administration of vasoactive or cardiogenic drugs. If a system were able to predict such an event in advance over a clinically useful timescale (e.g. at least 15 - 30 minutes ahead), then a clinician could intervene and commence a more moderate intervention. It has been estimated that the cost saving in both human and financial terms would be around 1600 euros per patient per day across the EU25 [1].

Hypotension is an example of an abnormal physiological condition, but there are several others that can be identified using similar bedside monitored signals (e.g. BP, ECG), such as hypertension, raised intracranial pressure or Tachycardia [2]. A clinical data platform that can standardize the collection, detection and prediction of such adverse physiological events would have great use both as a clinical and research tool. In order to collect, analyse and feed back the results of this data in a meaningful timescale, the AvertIT ICUnet platform was developed. This system interfaced with a multitude of clinical organisations (intensive care units – ICU) across Europe to gather and standardise data into a uniform format, where it could subsequently be streamed into a central repository for analysis. As well as the data-gathering component, a key challenge has been to use these data sets for prediction of upcoming adverse hypotensive events. In addition to ICUnet, the AvertIT project has two other distinct components:

- A Hypo-Predict engine trained on a Bayesian neural

network to alert clinicians to imminent hypotensive events;

- Supplementary web-based software solutions that allow input of patient management data not routinely input or collected electronically, e.g. procedures (e.g. pressure area care) periodically administered by nurses.

Figure 1 shows the relationship between these components.

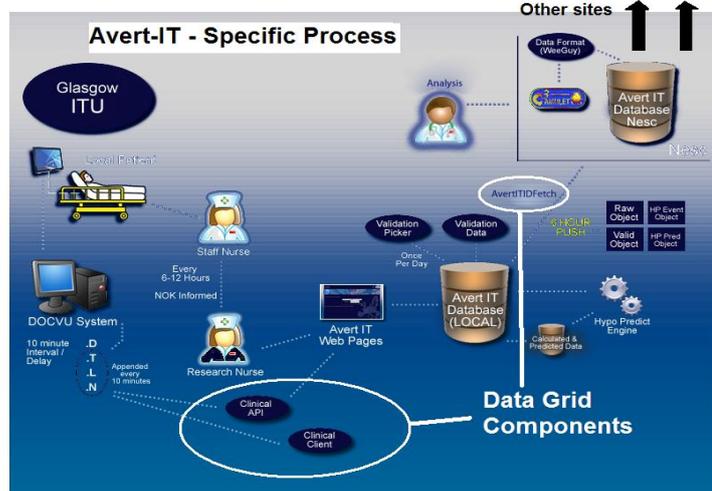


Figure 1: AvertIT Workflow Schematic

The focus of this paper is to describe the ICU net data infrastructure and to present the results of the clinical trial that it was designed to support. For an outline of the operation and details of the full Avert-IT systems, please see [12].

2. Clinical Background

2.1 Hypotensive event definition

There are various definitions of what constitutes a hypotensive event – largely based around the definition of the event threshold (when the event starts), the event hold-down (how long the event lasts), and the clear hold-down (when the event is over), in measurements of the patient’s blood pressure. The various definitions used by the partner sites involved in Avert-IT are shown in Table 1 (BPm = mean blood pressure, BPs = systolic blood pressure, CPP = cerebral perfusion pressure).

The first phase of the project was to identify a common definition that all sites agree upon to use for the training of the Hypo-Predict engine. The definition agreed upon was that published in the Edinburgh University Secondary Insult Grade (EUSIG) paper (threshold of 90 mmHg systolic or 70 mmHg mean arterial pressure, for a hold-down of 5 minutes) [2]. The justification for this definition is described in an earlier project paper [3].

Centre	Measure & Threshold (mmHg)	Event Hold-Down (min)	Clear Hold-Down
Uppsala	BPs < 100	2	BPs > 100;5 min
Glasgow	BPm < 70	5	BPm > 70;5 min
Kaunas	BPs/BPd < 90/50	5	BPm > 70;5 min
Heidelberg	CPP < 50	5	CPP > 60;5 min
Monza	BPs < 90	5	BPs > 90;10 min
Barcelona	BPs < 90	5	BPs > 90;15 min

Table 1: Avert-IT partner site threshold definitions

2.2 Identifying hypotensive event causes

There are a number of parameters that could potentially be used to predict the onset of a hypotensive event that are commonly monitored in an intensive care unit. These include:

- Heart rate (HRT)
- End tidal carbon dioxide (ETCO2)
- Mean blood pressure (BPm)
- Systolic blood pressure (BPs)
- Diastolic blood pressure (BPd)
- Blood oxygen saturation (SaO2)
- Temperature (TC)

The primary aim of the project was to discover how these parameters could be incorporated into a prediction system that would have a high probability of accuracy (and a clinically relevant degree of sensitivity and specificity) whilst minimising the numbers of false positive results, i.e. predictions of adverse events when no hypotensive event actually occurs.

In the context of current clinical practice, it is accepted that the development of such a predictive system, no matter how accurate, cannot completely replace clinical judgment. Rather it can and should only be regarded as a decision support system for clinicians. Thus there is also scope to incorporate wider contextual and demographic information to ascertain a patient’s likelihood of having a hypotensive event, e.g. factors such as their physical condition (e.g. age, severity of injury), or any associated medical treatments they might have been receiving for existing conditions.

A multicentre database analysis has shown that, due to the differing measurement procedures, it was common for hypotensive events to be missed [8, 9]. Our further analyses [4] have shown that this is due in part to how hypotensive events are detected either when based solely upon systolic BP or when considering both systolic BP and mean BP. Therefore the results of this project have already contributed significantly to the evidence that mean blood pressure should

have at least equal weighting in importance as systolic pressure when attempting to detect hypotensive events.

2.3 Clinical Decision Support

To predict adverse events, a decision-support module that takes account of the current data to predict future (hypotensive) events was required. Balancing the project requirements and the various advantages and disadvantages of the different predictive approaches, e.g. genetic algorithms, case based reasoning, multivariate analysis; a Bayesian approach to training an Artificial Neural Network (a BANN) was adopted. The motivation for selection of the BANN approach was its efficiency for the classification and modelling of highly nonlinear, and probabilistic, relationships between parameters, which themselves are expected to be a major aspect in the clinical inputs involved. Though the comparison of other methods would be of great benefit, given the limited time-scale of the project it was decided to focus upon one method (BANN) most likely to yield useful results. It is upon this platform that the Hypo-Predict engine was developed (full details of this process will be provided in future publications).

2.4 Physiological, treatment and demographic data

The ability to harness the wealth of clinical data available in the neurointensive care domain is a pre-requisite for the development of a BANN solution. However, there are wider ramifications for the construction of such a research platform. The AvertIT project focused on the traumatic brain injury (TBI) domain, but the data collected and methods applied could be applicable in other pathophysiological populations (such as patients with Stroke or Sub-Arachnoid Haemorrhage). The different types of data collected broadly fall into three categories:

- Physiological – this provides the most detailed minute-by-minute information on a patient's status;
- Treatment (or Episodic) data – this is data that is not available at regular intervals but still has an impact on the status of the patient at any given time;
- Demographic data – this is data only entered once per patients (eg: age, sex, centre, type of injury) and is important in providing wider contextual information about the status of a patient.

Whilst these definitions are not yet fully standardized, much progress has been made by the Brain-IT consortium (www.brainit.org) to settle on best practice standards in the neurointensive care domain. The Avert-IT project has adopted the BrainIT data model directly. It comprises a standard ontological description that can be shown in properly marked up XML files with textual descriptions of the terms and their usage in traumatic brain injury management.

3. ICUnet

The ICUnet system has three main components as indicated in Figure 1: *ClinicalAPI*, *ClinicalClient* and *AvertITIDFetch*. The

typical scenario of interaction between these components is as follows:

- The patient enters the hospital system as per normal procedure for that ICU. A local identifier is assigned to the patient, and their next-of-kin are informed.
- The patient is identified as being a potential candidate for the Avert-IT study. A consent form is generated and signature obtained from family.
- Upon successful receipt of consent, the local Avert-IT research nurse enters the patient ID into the study using either the web application for manual entry, or the *ClinicalAPI* module of the ICUnet application for point and click selection entry of patients currently in the ICU.
- This ID is assigned a study number and an entry is created in an AvertIT specific file.
- The data from the existing hospital intensive care monitoring and information systems is now parsed by the ICUnet system, rendered into a standard SQL format and input to a local Avert-IT database. It should be noted that every ICU had completely heterogeneous IT systems in place, e.g. Philips DocVu, Draeger, Wardwatcher.
- The HypoPredict system runs as a service in the background and does real-time analysis on the data in the local Avert-IT database. The results of this analysis are fed back into separate tables within this local database.
- Every hour, the *AvertITIDFetch* module contacts the central web service and uploads the latest anonymised data to the main repository database. This includes all anonymised patient data and associated predicted events, to be used for independent evaluation by clinicians to adjudicate on the accuracy of the predictions. One hour intervals were considered an acceptable latency time, given the volume of information being passed through and the bandwidth constraints of the web service XML messages.

3.1 Implementation

The ICUnet system was implemented with the following specifications. In each clinical centre hardware provisioned included:

- Sun Fire X4140 server to house the local database, ICUnet and HypoPredict applications (64GB memory, 2x 2.7GHz quad-core AMD Opteron processors, 8x 2.5" SAS drives)
- Dell E4500 Laptop for bedside data input
- Wireless router for connectivity within the ICU
- Operating systems of Windows Vista and Windows Server 2008. (Most advanced Windows OS's at time of development allowing use of advanced security features and in line with prevailing hospital IT

policies). It was found that having virtual private network support greatly increased the efficiency of work and decreased the time spent debugging issues.

3.2 Support, maintenance and security

The system was supported with a variety of features: log-file uploads, regular automated email updates, a Trac server, master-to-master database synchronization, WS redundant failover, and data volume throttling. Given the nature of the data involved, security was essential. The ICUnet system included capabilities such as: encryption of data streams using WSIT technology; authentication including assertion of user identity; advanced firewall configuration to allow selective streaming of data; WS access through port 80, and administrative and programmatic security. Several tools were also incorporated into the AvertIT infrastructure to provide automatic report generation. These have allowed the rapid creation and dissemination of results – described in the next section – to be analysed and returned to the clinical community involved with the project.

4. Avert-IT Clinical Trial

The ICUnet was built to support the gathering of physiological data for the purposes of predicting hypotensive events. That data, along with the prediction software output of the Hypopredict engine, were stored in the central repository and analysed in the subsequent clinical trial. What follows is the analysis of predicted versus actual events over the trial period, with all the relevant supporting data (given space constraints), focusing on the sensitivity and specificity of the prediction engine.

4.1 Trial Design

The clinical trial was based upon a two-stage trial design. Stage I was used to provide an estimate through a *small pilot sample of patients* ($n = 30$) that the AvertIT technology meets the basic clinician's minimum requirements for sensitivity and false positive rate (FPR) for prediction of arterial hypotension. It also provided a systematic mechanism for gathering clinical domain knowledge to potentially improve the sensitivity and FPR should it fall short of requirements by allowing for a BANN retraining step, before the second stage. Stage II was a Phase II sequential clinical trial with a cohort of 46 patients used to assess definitive early evidence of efficacy that the AvertIT systems for prediction of arterial hypotension. In discussions with the clinicians and ICU intensivists, it was agreed that a primary end-point of an event prediction sensitivity of $> 30\%$ and $FPR < 10\%$ in the clinical population would be suitable and clinically relevant. The low sensitivity value of 30% must be viewed in the context of current practice where the clinical teams have no warning of these events. It is noted that ICUs have many alarms associated with their real-time bedside monitoring equipment, and thus accuracy of prediction and reliability were key criteria on these restrictive selective criteria.

4.2 Sample Size Calculations

After the small pilot sample of $n=30$ in stage I, for stage II, a two-step Phase II clinical trial was adopted using the methods detailed in Hanfelt [10] which itself was a modification of Simon [11]. Using the sample size detailed in the [10], a step 1 sample size of 13 was identified. Assuming a positive outcome to step 1, a second step comprising 33 patients was identified giving an overall sample size of 46 for the trial. Using results from [10], the following parameters were used:

- $\alpha = 0.1$, $\beta = 0.2$ (80% power) (alpha/beta and power referring to the ability to reject the null hypothesis in the trial)
- $P_0 = 0.20$ (must have at least 20% sensitivity to be considered)
- $P_1 = 0.35$ (sensitivity must have at least 35% to be clinically useful)

Therefore from [10] $p_1 - p_0 = 0.15$ which gives us a sample size design of: 2/13, 12/46. The sample design figures were interpreted as: Step 1 requires collection of 13 patients, if greater than 2 patients showing more than 35% sensitivity then the trial proceeds to collect the full 46. Providing more than 12 of the 46 showing more than 35% sensitivity a successful result can be declared, which provides evidence of early efficacy for the BANN prediction technology.

4.3 ROC Confidence Intervals

Three methods of obtaining confidence intervals for the ROC curves calculated (Receiver Operator Characteristic curve – a graph of true-positive rate versus false positive rate, which shows overall sensitivity) were considered: Normal theory (where the sample distribution is assumed to have a probability distribution function that follows a bell curve); Student-t based (used where sample sizes are small and the standard deviation is unknown) and bootstrap calculations (where accuracy is assigned by re-sampling from an approximating distribution). The density plots for sensitivity and specificity did not meet all the assumptions for a Gaussian distribution, which is a requirement for the first two methods therefore the bootstrap method was used.

The bootstrap method takes observed measurements and creates a new sample by randomly selecting values from these measurements with replacement. This process is repeated many times – typically several thousand. For each sample the mean is calculated and stored and then, at the end of the process, the samples are sorted to produce a distribution. The confidence interval is then calculated by taking the required upper and lower quantile from this distribution. By using a simple piece of R code [13] a 95% confidence interval is obtained, by using the 0.0275 and 0.975 quantiles.

5. Results

30 patients were recruited in Stage I and 39 patients in stage II. On comparison of the demographics (age, sex, Glasgow Coma Motor Score on admission) between the two cohorts it

was found both datasets were similar so both sets of data were combined into one of 69 patients.

5.1 Stage I Results

Table 2 shows the results from the Stage I study together with normal, student-t and bootstrap point estimates and confidence intervals for the analyses *without* false positive suppression. False positive suppression when used was based upon simple clinically defined guidelines such as a minimum required pulse amplitude (BPs – Bpm).

Threshold	Sensitivity	Bootstrap		Normal 95 +/- C.I.	Student-t 95 +/- C.I.	Specificity	Bootstrap		Normal 95 +/- C.I.	Student-t 95 +/- C.I.
		Low	High				Low	High		
0.1	63.64	55.44	71.33	8.08	8.43	60.28	52.22	67.73	7.8	8.14
0.2	55.71	46.95	64.26	8.78	9.16	73.94	66.83	80.35	6.84	7.14
0.3	49.28	40.16	58.67	9.29	9.69	82.5	76.23	87.98	5.94	6.2
0.4	39.29	29.93	49.05	9.62	10.04	88.51	83.36	92.85	4.8	5.01
0.5	27.87	19.01	37.58	9.37	9.78	92.67	88.58	96.04	3.78	3.95
0.6	19.53	10.96	29.02	9.13	9.53	95.37	91.78	98.02	3.21	3.35
0.7	8.64	2.12	17.26	7.72	8.06	98.33	93.64	99.06	2.86	2.99
0.8	5.67	0.63	13.72	6.84	7.14	97.74	94.66	99.54	2.72	2.84
0.9	3.63	0	10.55	6.53	6.81	98.33	95.63	99.8	2.51	2.62

Table 2: The results from the Stage I study together with normal, student-t and bootstrap point estimates and confidence intervals for the analyses *without* false positive suppression.

Table 3 below shows the results from the Stage I study together with normal, student-t and bootstrap point estimates and confidence intervals for the analyses *with* false positive suppression.

Threshold	Sensitivity	Bootstrap		Normal 95 +/- C.I.	Student-t 95 +/- C.I.	Specificity	Bootstrap		Normal 95 +/- C.I.	Student-t 95 +/- C.I.
		Low	High				Low	High		
0.1	57.69	47.79	66.57	9.53	9.94	68.74	62.01	74.98	6.66	6.95
0.2	54.45	45.14	63.6	9.29	9.69	79.82	74	85.04	5.61	5.85
0.3	46.03	36.41	55.87	9.72	10.14	87.13	82.28	91.36	4.63	4.83
0.4	40.09	30.91	49.65	9.45	9.86	92.57	88.85	95.61	3.45	3.6
0.5	28.51	19.9	37.92	9.13	9.53	95.8	93.35	97.76	2.27	2.37
0.6	19.47	11.01	28.88	9.06	9.45	98.19	96.9	99.15	1.18	1.23
0.7	8.77	2.29	17.37	7.7	8.04	99.29	98.74	99.72	0.51	0.53
0.8	5.82	0.76	13.9	6.86	7.16	99.66	99.39	99.87	0.24	0.25
0.9	3.63	0	10.55	6.53	6.81	99.94	99.89	99.99	0.06	0.06

Table 3: The results from the Stage I study together with normal, student-t and bootstrap point estimates and confidence intervals for the analyses with false positive suppression.

Figure 2 below shows the ROC curve results for all 30 patients without false positive suppression. The blue line is the line of identity. The ROC curve from the research phase is shown by the black line - where the existing BrainIT database was used, in order to conduct the BANN training. The point estimates for mean sensitivity and false positive rate at the threshold values 0.1 to 0.9 are shown on the solid red curve.

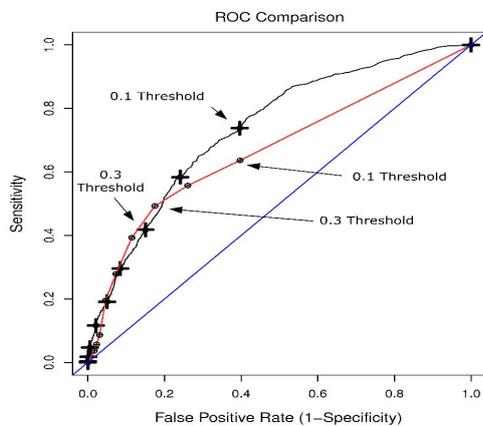


Figure 2: STAGE I - ROC point estimates, no false positive suppression

The study results are very close to the research results particularly in the range of interest with a warning threshold setting of between 0.3 and 0.4 giving a specificity range of 46.03% and 40.09% and a specificity range of 87.13% and 92.57% respectively.

Figure 3 shows the same ROC analysis but for the data with false positive suppression running.

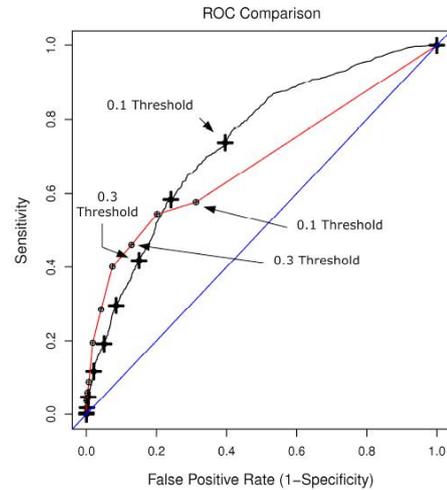


Figure 3: STAGE I - ROC point estimates with false positive suppression

As seen on comparison of the ROC results (Tables 2 & 3) with and without false positive suppression, the false positive suppression analyses, although simple, do appear to be improving the results as seen by the left shift in the ROC curve. Using an alarm-triggering threshold of 0.4 gives results for sensitivity of 40.09 % and specificity of 92.57 %, both of which fall within the minimum sensitivity/specificity required as targets set by the clinical staff. A similar improvement in the results was seen both with Stage II results and with the combined Stage I & Stage II analyses. Consequently, for the sake of brevity, the remainder of this paper only shows the results with false positive suppression analyses.

5.2 Stage II Results

Figure 4 below shows the recruitment rate for the 39 patients required in the Phase II study categorised by centre.

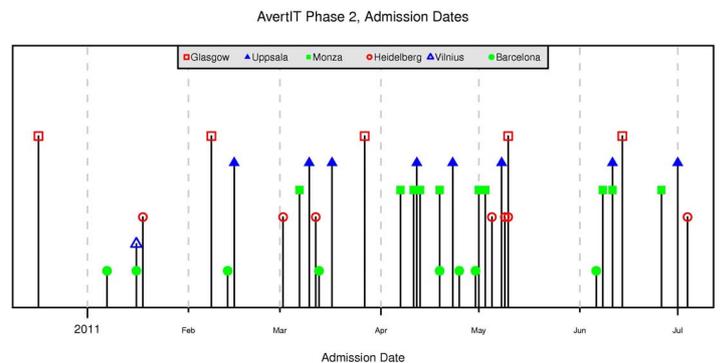


Figure 4: STAGE II recruitment rate by centre.

Table 4 below shows the sensitivity results from the Stage II study together with normal, student-t and bootstrap point estimates.

Threshold	Sensitivity	Bootstrap		Normal	Student-t
		Low	High	95 +/- C.I.	95 +/- C.I.
0.1	56.28	49.67	62.70	6.55	6.76
0.2	50.93	43.15	58.63	7.86	8.12
0.3	44.27	35.54	53.21	8.94	9.23
0.4	35.52	27.04	44.60	8.88	9.17
0.5	27.81	18.92	37.48	9.35	9.66
0.6	19.96	11.67	28.96	8.84	9.13
0.7	14.94	7.98	23.28	7.82	8.08
0.8	8.08	2.06	16.01	7.19	7.43
0.9	5.44	0.13	13.13	7.00	7.23

Table 4: The results from the Stage II study together with normal, student-t and bootstrap point estimates.

Figure 5 below shows the ROC curve results for all 39 Stage II patients. The dotted black line shows the ROC curve from the research phase, and as before, the point estimates for mean sensitivity and false positive rate for the prospective clinical data at the threshold values 0.1 to 0.9 are shown on the solid red curve. Again, the study results are very close to the research results particularly in the range of interest with a warning threshold setting of between 0.3 and 0.5 giving a sensitivity range of 44.27% and 27.81% and a specificity range of 83.76% and 94.12% respectively.

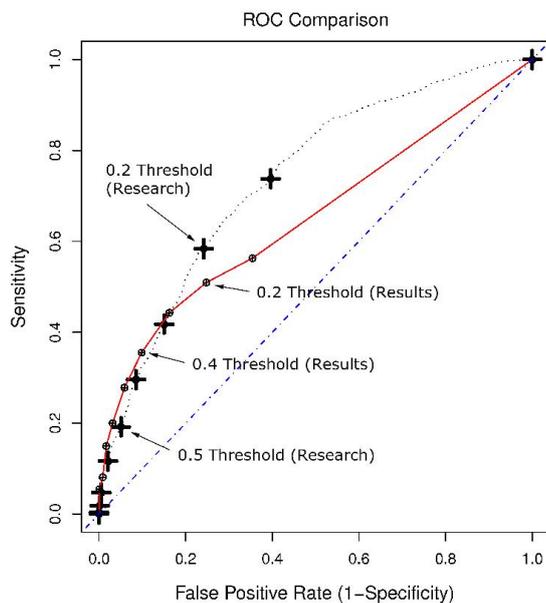


Figure 5: STAGE II - ROC point estimates with false positive suppression

With these results, we can conclude that the Avert-IT technology can indeed predict at least 1 in 3 hypotensive adverse events in a live clinical intensive care environment.

6. Conclusions

In this paper we have outlined the AvertIT ICUnet data infrastructure that allows the real-time acquisition and streaming of physiological, episodic and demographic data from neurosurgical centres across Europe. Using this platform a major multi-centre clinical trial across six major brain trauma ICUs across Europe has been conducted, to attempt to predict hypotensive events in brain trauma patients. The clinical trial design has been outlined and the results have been presented showing that the Hypo-Predict system can predict at least 1 in 3 hypotensive events in a live clinical ICU environment. The AvertIT project was completed in October 2011.

7. Acknowledgements

The authors would like to acknowledge the EU FP7 program (grant number 217049), kindly funding the Avert-IT project. We would also like to acknowledge the work of the BrainIT group investigators and participating centres to the BrainIT dataset without whom this work would not have been possible.

8. References

- [1] Avert-IT project proposal Advanced Arterial Hypotension Adverse Event prediction through a Novel Bayesian Neural Network (available at <http://wiki.avert-it.org/wordpress>)
- [2] Patricia Jones, Peter Andrews, Douglas Miller, et al, *Measuring the burden of secondary insults in head-injured patients during intensive care*. Journal of Neurosurgical Anesthesiology, 6:4-14, 1994.
- [3] Stell, Sinnott, Jiang et al, *Federating distributed clinical data for the prediction of adverse hypotensive events*, e-Science All-Hands conference 2008
- [4] Tarassenko et al., *BioSign: multi-parameter monitoring for early warning of patient deterioration*. 3rd IEE International Seminar on Medical Applications of Signal Processing (2005/11199), p71-76
- [5] Geoffrey Manley, et al *Hypotension, hypoxia, and head injury* ARCH SURG, 136:1118-1123, 2001.
- [6] A. Marmarou, R.L. Anderson, J.D. Ward, et al. *Impact of ICP instability and hypotension on outcome in patients with severe head trauma*, Journal of Neurosurgery, 75:S59-S66, 1991.
- [7] T.P. Howells, I.R. Piper, P.A. Jones, M. Souter, and J.D. Miller - *Design of a research database for the study of secondary insults following head injury*. Journal of Neurotrauma, 12:471-472, 1995.
- [8] Zanier E, Ortolano F, et al - *Intracranial pressure monitoring in intensive care: clinical advantages of computerized monitoring over manual recording*. Crit. Care Med. 2007
- [9] Corrie J, Piper I, et al - *Microcomputer based data recording improves identification of secondary insults in head injured patients*. British Journal of Intensive Care, June 1993. 226-233.
- [10] Hanfelt J, Slack R, Edmund M, Gehan A - *A Modification of Simon's Optimal Design for Phase II Trials When the Criterion is Median Sample Size*, Controlled Clinical Trials 1999;20:555-566
- [11] Richard Simon - *Optimal two-stage designs for phase II clinical trials*, Controlled Clinical Trials 1989;10:1-10
- [12] Stell, A.J.; Sinnott, R.O.; et al. *A Distributed Clinical Data Platform for Physiological Studies in the Brain Trauma Domain*. IEEE e-Science Conference Proceedings 2010, Brisbane, Australia, 2010
- [13] R, statistical computing - <http://www.r-project.org>