Clinical evaluation of algorithms for context-sensitive physiological monitoring in children

M. Dosani1, J. Lim1, P. Yang2, C. Bouse2, J. Daniels1, G. Dumont2 and J. M. Ansermino1*

1Department of Anesthesiology, Pharmacology and Therapeutics and 2Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada

*Corresponding author: Department of Anaesthesia, BC Children’s Hospital, 1L7-4480 Oak Street, Vancouver, BC, Canada V6H 3V4. E-mail: anserminos@yahoo.ca

Background. Subtle changes in monitored physiological signals might be used to guide clinical actions and give early warning of potential adverse events. Automated early warning systems could enhance the clinician’s interpretation of data by instantaneously processing new information and presenting it within the context of previous observations. In this study, we tested algorithms for tracking the behaviour of dynamic physiological systems and automatically detecting key events over time.

Methods. Algorithms were activated in real-time during anaesthesia to run context-sensitive monitoring of six variables (end-tidal $P_{CO_2}$, heart rate, exhaled minute ventilation, non-invasive arterial pressure, respiratory rate, and oxygen saturation), alongside standard physiological monitors. The clinical evaluation included real-time feedback on each change point (change in the physiological trend) detected by the algorithms and the completion of a usability questionnaire.

Results. Fifteen anaesthetists completed the evaluation during paediatric surgical cases. A total of 38 cases were evaluated, with a mean duration of 103 (102) min. The mean number of change points per case was 22.8 (23.4). Sixty-one per cent of all rated change points were considered clinically significant, and <7% were due to artifacts.

Conclusions. The algorithms were able to detect a range of clinically significant physiological changes during paediatric anaesthesia, and were considered useful by participating anaesthetists. These findings indicate that automated detection of context-sensitive changes is possible and could be used by early warning systems during physiological monitoring. Further investigations are required to assess how this information can best be communicated to the anaesthetist.


Keywords: computers; equipment, computers; equipment, expert adviser system; model, mathematical; monitoring, intraoperative

Accepted for publication: February 22, 2009

Current physiological monitoring technologies, combined with clinical information systems, collect large amounts of data at the bedside. Typically, if the value of a monitored variable strays outside a preset range, an alarm will be triggered. Unfortunately, false and unnecessary alarms are so frequent that, in practice, they can represent a nuisance rather than improvement to monitoring.1 Major advances in sensor technology have resulted in increased information available to the anaesthetist. Despite this, the means of extracting important information to advise clinicians of significant events is still underdeveloped.

Early warning systems generate scores from a composite of current observations, which can be used to trigger rapid intervention for patients at risk. Such systems have been implemented within the past decade, but recent investigations have shown poor diagnostic accuracy, partially attributable to the selection of physiological variables and cut-off values, mistakes in the manual calculation of scores, and intra- and inter-rater reliability errors.
Although rigorous investigations have shown improvement in the selection and validation of appropriate physiological variables and cut-off values,\(^3\) the distinction between normal and abnormal values is complex due to the profound variability in physiological measurements between and within patients.

The likelihood that a patient is at risk, based on the information available at that point in time, is rarely absolute. It is more common for this likelihood to follow a probability distribution, with the likelihood of correctly identifying a patient at risk increasing as more information becomes available. The use of previously recorded information, rather than only the current observation, could significantly improve the usefulness of early warning systems both in the operating theatre and in acute care environments. Automated early warning systems can enhance the clinician’s ability to interpret data by instantaneously processing new information and presenting it to the clinician within the context of previous observations.

In this study, we tested algorithms for tracking the behaviour of dynamic physiological systems and automatically detecting key events in the processes over time. After extensive offline testing\(^4\) and refinement, we now report the results of evaluation in real-time during anaesthesia.

**Methods**

**Context-sensitive monitoring**

The mathematical process of detecting trend changes has been extensively studied in many areas such as price-trend prediction and weather forecasting.\(^5\) A similar mathematical process can be used to identify trend changes in physiological monitoring. For context-sensitive monitoring, the threshold for providing an alert to the clinician is not fixed, but is dynamically dependent on the information available before an observation. Predictions based on information from all the previous observations are made for the next observation or observations further in the future. When the prediction accurately describes the observation, the trend is unchanged. Alternatively, when the prediction differs from the observation, the trend is changing. The accumulation of differences between the predictions and the observations is used to identify whether a significant change in trend has occurred. The point at which the accumulation of differences exceeds a threshold is called a change point.

**Physiological models of change**

Physiological observations fluctuate due to the natural rhythms of the human body. The changes in trends, which differ from natural fluctuations that occur over time, are constrained by the physiological processes that produce the observations. For example, heart rate (HR) cannot increase or decrease by 50 beats min\(^{-1}\) in 1 s. The pattern of change for each physiological variable differs (e.g. changes in HR have different patterns than those seen in oxygen saturation). The patterns of change can be captured using statistical models and can be used to distinguish a true change from a normal physiological fluctuation or an artifact.

**Change-point detection algorithms**

The algorithms in this study predict future observations and update the Cumulative Sum (CUSUM) threshold in real-time. The CUSUM test, a method widely adopted in process control and monitoring, is a technique where the differences between successive observed values and a target value (a predicted value in this case) are accumulated. The standard CUSUM test compares the CUSUM values with a threshold. A change point is reached when the accumulated value breaks the threshold. Change points in mean non-invasive arterial pressure (mean NIAP) and oxygen saturation (Sp\(_{O_2}\)) were detected using CUSUM testing on a filtered residual from an Exponentially Weighted Moving Average filter as described in detail elsewhere.\(^6\) The algorithms used for HR, end-tidal carbon dioxide (t\(_{CO_2}\)), exhaled minute ventilation (MV\(_{exp}\)), and respiratory rate (RR) treat the observations as a series of linear segments, the true levels of which are contaminated by noise such as measurement error, artifact, and physiological variations (e.g. respiration).\(^7\)\(^8\) When the current observation follows the previous pattern, the mean and slope from the previous observations are similar to the current observations. However, when the observations start to change into a different pattern, either the mean or the slope will be different. The probability of these relationships is described by a matrix-formatted model called the dynamic linear model.\(^9\) On the basis of this model, an adaptive Kalman filter (a minimum-variance stochastic estimation procedure) is used to estimate the mean and the incremental rate from all previous observations. The standard Kalman filter provides the optimal prediction by minimizing the squared estimation error. Furthermore, it can recursively update the estimates when a new observation becomes available. In the method used in this study, the standard Kalman filter is improved by adapting the filtering process to the signal characteristics. The adaptive Kalman filter extracts the level of the noise and the degree of signal variability, and then adjusts the weights assigned to the observations. For example, if the recent observations are highly corrupted by artifact, less weight will be assigned to the recent observations with more influence given to earlier observations. Since the Kalman filter captures the full monitoring history and assigns more weight to the more recent or more reliable observations, the estimate from the Kalman filter is robust against disturbances.
Predictions of the future observations are compared with actual observations to detect changes. To avoid detecting very brief changes, the Kalman filter is configured to predict a number of observations into the future. The predictions for the same future point from different historical instants are averaged with more weight assigned to those from the more recent observations. When the observations for these points become available, the observations are compared with the predictions, and the resulting differences are cumulated and tested using the CUSUM test. The degree of signal variability, extracted by the adaptive Kalman filter, is used to adapt the thresholds of the CUSUM test. The CUSUM, which is updated with each new observation, is reset once a change point has been detected.

The algorithms were implemented in a Java® software environment. They were engineered to detect events in real-time data with performance optimized, using receiver operating curves, on both simulated and real clinical data (offline). These algorithms compare favourably with other methods of trend detection, such as Trigg’s tracking signal.

Clinical evaluation

Approval for this study was obtained from the Clinical Research Ethics Board of the University of British Columbia and from the British Columbia Children’s and Women’s Hospital Research Review Committee. Informed consent was obtained from all participating anaesthetists.

The software tool running the signal processing algorithms was activated in real-time during anaesthesia to enable context-sensitive monitoring of six variables (tCO2, HR, MVexp, NIAP, RR, and SpO2) displayed on a touch screen monitor alongside current Datex S/5 physiological monitors (Datex-Ohmeda, Helsinki, Finland) using a serial RS232 interface. Each participating anaesthetist attended a 5 min demonstration to familiarize themselves with the software tool and monitor. The anaesthetist activated the software once the patient’s physiological state was considered stable (usually 5–10 min after the induction of anaesthesia) and discontinued the evaluation once the surgical procedure had been completed. Each participant evaluated the software during at least three cases. The clinical evaluation included real-time feedback, using a software interface, on each change point detected by the system and the completion of a usability questionnaire. Characteristic data of participating anaesthetists including gender, age, number of years of anaesthesia experience, and computer proficiency were also recorded.

Change-point detection feedback

An event-feedback system was incorporated into the user interface of the software tool. When a signal processing algorithm detected a change point, the software tool generated an event record and provided a graphical display of the trends with an arrow denoting the specific event (Fig. 1). Using the touch-screen monitor and on-screen keyboard, subjects selected the appropriate change-point classification from a dropdown menu. Classification options included artifact, clinically insignificant, clinically significant, clinically significant with action taken, and clinically significant due to intervention (Table 1). Subjects rated the usefulness of each event by selecting a point on a standard 10 cm visual analogue scale on the user interface. Usefulness ranged from ‘frustrating’ (0) to ‘very useful’ (10). Subjects were given the option of typing freeform text to provide more detailed event feedback. Missed events were also noted.
Usability

A modified Post-Study System Usability Questionnaire (PSSUQ) was administered to participants after their first use of the software tool. The PSSUQ rates the usability of the application across four factors: overall, system usefulness, information quality, and interface quality. The modified questionnaire used in this study included a question supplemental to the standard PSSUQ, which measured respondent agreement with the statement ‘The system was unobtrusive and did not hinder my performance in other duties.’ Additionally, two questions were removed from the standard PSSUQ as the system was not deployed as it would have been in routine clinical practice due to the evaluation component. The removed questions stated, ‘It was simple to use this system’ and ‘I liked using the interface of this system.’ Participants completed a paper version of the survey, responding to each question by ranking their agreement with statements on an 8.5 cm long numerical scale, with possible responses ranging from 1 (strongly agree, good usability) to 7 (strongly disagree, poor usability). Each question also provided a ‘not applicable’ option and a freeform text section for detailed feedback. The score for each factor was determined by averaging the responses to the appropriate questions. Results of the supplemental ‘system obtrusiveness’ question were not included in the questionnaire validation and were thus treated as a separate component of the survey.

Results

Event feedback

Fifteen anaesthetists completed the evaluation of 38 paediatric surgical cases, with a mean duration of 103 (102) min. The mean number of change points per case was 22.8 (317). Sixty-one per cent of all rated change points were considered clinically significant, and 17% of change points were unrated (Fig. 2). The mean usefulness was highest for changes rated as clinically significant with action taken and lowest for changes rated as artifacts (Table 2). Only six significant change points were reported as being missed by the algorithms.

Usability

Thirteen of the participating anaesthetists completed the PSSUQ. Two subjects were treated as outliers based on their characteristic data, as these subjects rated themselves as ‘uncomfortable’ with computers. The overall score was 2.8 including all participants, or 2.5 excluding outliers (Table 3). A low score indicates good usability.
cautery noise detection.12 Furthermore, simultaneous changes in more than one variable, such as RR, MVexp, and \( P_{CO_2} \), could be achieved with the inclusion of automated electrocautery noise detection.12

Table 2 Usefulness [mean (SD)] of each change point on a standard 10 cm visual analogue scale. Possible values range from ‘frustrating’ (0) to ‘very useful’ (10)

<table>
<thead>
<tr>
<th>Number of change points</th>
<th>Artifact</th>
<th>Clinically insignificant</th>
<th>Clinically significant without action taken</th>
<th>Clinically significant with action taken</th>
<th>Clinically significant due to intervention</th>
<th>Overall usefulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{E}_{CO_2} )</td>
<td>212</td>
<td>1.9 (1.2)</td>
<td>1.9 (1.5)</td>
<td>4.9 (1.7)</td>
<td>6.9 (2.0)</td>
<td>4.0 (2.1)</td>
</tr>
<tr>
<td>HR</td>
<td>253</td>
<td>2.7 (1.5)</td>
<td>2.1 (1.6)</td>
<td>5.0 (1.5)</td>
<td>6.7 (1.7)</td>
<td>3.6 (1.5)</td>
</tr>
<tr>
<td>MVexp</td>
<td>145</td>
<td>1.4 (0.8)</td>
<td>2.1 (1.7)</td>
<td>4.8 (1.9)</td>
<td>4.4 (3.8)</td>
<td>4.3 (2.1)</td>
</tr>
<tr>
<td>NiAP</td>
<td>124</td>
<td>1.2 (1.8)</td>
<td>2.2 (1.4)</td>
<td>5.1 (1.7)</td>
<td>3.3 (2.8)</td>
<td>4.2 (1.9)</td>
</tr>
<tr>
<td>RR</td>
<td>86</td>
<td>1.8 (1.6)</td>
<td>1.7 (1.3)</td>
<td>4.9 (1.4)</td>
<td>3.3 (2.8)</td>
<td>2.9 (2.4)</td>
</tr>
<tr>
<td>S( \bar{P}_{O_2} )</td>
<td>48</td>
<td>1.6 (1.5)</td>
<td>1.9 (2.0)</td>
<td>3.1 (4.8)</td>
<td>4.4 (2.2)</td>
<td>3.2 (2.7)</td>
</tr>
<tr>
<td>Total</td>
<td>868</td>
<td>1.9 (1.4)</td>
<td>2.1 (1.6)</td>
<td>4.9 (1.7)</td>
<td>6.0 (2.4)</td>
<td>4.4 (2.0)</td>
</tr>
</tbody>
</table>

Table 3 PSSUQ results. Scores represent the mean response to the questionnaire items in each subscale, ranging from 1 (good usability) to 7 (poor usability). System obtrusiveness is a custom element and not a validated part of the questionnaire.

<table>
<thead>
<tr>
<th>PSSUQ factor</th>
<th>Complete (( n=13 ))</th>
<th>Excluding outliers (( n=11 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>2.8</td>
<td>2.5</td>
</tr>
<tr>
<td>System usefulness</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Information quality</td>
<td>2.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Interface quality</td>
<td>3.1</td>
<td>2.9</td>
</tr>
<tr>
<td>System obtrusiveness</td>
<td>3.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Discussion

The algorithms evaluated in this study detected 22 change points per surgical case, more than 60% of which were considered clinically useful and <7% of which were due to artifacts. This contrasts with traditional alarms, which have been reported to indicate actual risk to patients in only 3% of cases.1 Improved performance could be achieved by integrating change points from several monitored variables and using information from other sources.

Further tuning of the algorithms to minimize artifacts and reduce the detection of clinically insignificant change points is possible. Creating a distinction between controlled and spontaneous ventilation, for example, could account for differences between these modes of ventilation, and could minimize the occurrence of artifacts in \( \bar{E}_{CO_2} \) measurements. Reductions in artifacts in the HR trend could be achieved with the inclusion of automated electrocautery noise detection.12 Furthermore, simultaneous changes in more than one variable, such as RR, MVexp, and \( P_{CO_2} \), due to a change in ventilation, could be fused into a single change point. Twenty-six per cent of the change points were attributed to clinical intervention by the anaesthetist. Although the changes in physiological variables resulting from intervention are significant, communicating this to clinicians is unnecessary. These change points could be suppressed by capturing information about clinical interventions from a clinical information system. The type, duration, and probability of changes seen during induction and emergence from anaesthesia are very different from those seen during the stable period of anaesthesia. The inclusion of additional information about the timing and type of clinical interventions will support the development of models of the typical changes seen in less stable periods of anaesthesia.

As in any diagnostic process, a trade-off exists between false alerts (false positives) and missed changes (false negatives). Any threshold or algorithm can be adjusted to favour one or the other. Recent evidence suggests that false alerts affect compliance and reliance on automated delivery of information significantly more than missed events.13 In this investigation, the algorithms were tuned to be more sensitive than they would be for use in a healthy patient, in order to gain insight into their performance. It is possible to reduce the sensitivity of the algorithms for healthy patients, in order to allow less significant changes to be ignored. The sensitivity can then be increased for patients with more unstable physiology (where small changes may be important), or where the anaesthetist requires additional assistance (prolonged periods of monitoring).

Alternate methods for the detection of clinically important changes in patient status have been described. The temporal context of physiological observations can be assessed using a percentage change or measure of statistical variation. The current clinical practice of using a fixed percentage change as a threshold for identifying an abnormal observation performs poorly because a percentage change is dependent on the level of the comparison baseline. In such a method, the current observation is compared with an average value calculated over an arbitrarily chosen interval (such as the previous 10 min). Furthermore, each observation is equally valued, although in fact the more recent information is usually more relevant. Using a measure of statistical distribution for comparison, such as standard deviation, has similar limitations.

Previous attempts at automated trend detection for early warning systems have been made;14 however, adoption of novel technology should ideally be preceded by evidence of improved clinical outcomes. Randomized controlled trials provide an objective and comprehensive measure of clinical outcomes, but require large numbers of participants and long study periods.15 In this initial investigation, we have instead evaluated the subjective assessment by clinicians of software usability and the usefulness of the algorithms. The algorithms tested in this investigation are not designed to replace clinicians, but instead to assist the anaesthetist in rapidly processing the vast amount of information available from monitoring equipment and to convey this information in a...
meaningful manner, so that rapid intervention can occur. Usability results indicate that anaesthetists consider the software useful in the identification of changes in patient status. Clinical adoption of automated trend detection technologies will require that significant changes be communicated to the clinician in a way that is intuitive and useful. This will require that the user understand the processes used to detect these changes and trust that the information delivered has value in improving clinical care. The rapid increase in the number of sensors collecting information in the typical clinical monitoring environment has exceeded the monitoring ability of even the most experienced users. Automated identification of context-relevant changes could therefore be immensely useful.

Acknowledgements
We thank the anaesthetists at BC Children’s Hospital who volunteered to participate in this study and William Magruder for editorial assistance.

Funding
This work was supported by the Natural Sciences and Engineering Research Council of Canada (grant number CHRPJ 262479—2003); the Canadian Institutes of Health Research (grant number MOP—79418); the Michael Smith Foundation for Health Research [CI-SCH-570(06-1) to J.M.A.]; and the Canadian Anesthesiologists’ Society/Abbott Laboratories Ltd Career Scientist Award in Anesthesia (to J.M.A.).

References
10 Trigg DW. Monitoring a forecasting system. Oper Res Quart 1964; 15: 271–4