Journal of Biomedical Informatics 75 (2017) 14–21

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Evaluating performance of early warning indices to predict physiological instabilities

Christopher G. Scully*, Chathuri Daluwatte

Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration, United States

ARTICLE INFO

Keywords:
Patient monitoring
Early warning
Performance assessment

ABSTRACT

Patient monitoring algorithms that analyze multiple features from physiological signals can produce an index that serves as a predictive or prognostic measure for a specific critical health event or physiological instability. Classical detection metrics such as sensitivity and positive predictive value are often used to evaluate new patient monitoring indices for such purposes, but since these metrics do not take into account the continuous nature of monitoring, the assessment of a warning system to notify a user of a critical health event remains incomplete. In this article, we present challenges of assessing the performance of new warning indices and propose a framework that provides a more complete characterization of warning index performance predicting a critical event that includes the timeliness of the warning. The framework considers 1) an assessment of the sensitivity to provide a notification within a meaningful time window, 2) the cumulative sensitivity leading up to an event, 3) characteristics on if the warning stays on until the event occurs once a warning has been activated, and 4) the distribution of warning times and the burden of additional warnings (e.g., false-alarm rate) throughout monitoring that may or may not be associated with the event of interest. Using an example from an experimental study of hemorrhage, we examine how this characterization can differentiate two warning systems in terms of timeliness of warnings and warning burden.

1. Introduction

Patient monitoring has traditionally involved reporting individual vital signs that describe the current state of a patient with alarm conditions defined by the relationship of individual vital signs to, for example, pre-set high and low limits. This type of alarm system may be included on cardiac, blood pressure, and pulse-oximetry monitors. However, individual vital signs may not provide noticeable changes prior to a critical event [1] and may trigger alarms with low clinical relevance contributing to high alarm rates and alarm fatigue [2]. Warning indices and notification systems that use physiological measurements to produce comprehensive metrics of patient health [3,4] and identify patterns predictive of patient instabilities [5,6] are a promising tool to provide lead time prior to a critical event when effective mitigations can be taken [7]. A sampling of such methods include approaches for identifying need for life-saving interventions in trauma patients [8,9], patients at risk for septic shock in the intensive care unit [10], need for care escalation from step-down units [11–13], and potential heart failure in at-home monitoring [14,15] environments. Patient monitoring environments are known for the number of false alarms [16,17], presenting a need to balance the performance of any new monitoring index and warning system to provide timely warnings prior to an event and low false-alarm rates.

Warning indices are often evaluated as classification systems using metrics such as sensitivity, specificity, and ROC curve analysis for the index to warn prior to an event [18]. Sensitivity (number of warnings generated prior to an event divided by the total number of events [true positives/(true positives + false negatives)]) and a combination of metrics such as false-alarm rate (number of warnings not associated with an event over some unit of time [false positives/monitoring time]) and positive predictive value (number of warnings generated prior to an event divided by the total number of warnings [true positives/(true positives + false positives)]) may be reported to characterize the performance of this type of detection system. As these metrics alone fail to incorporate the timeliness of the detection [19] and patterns leading up to the event into the performance metrics (see below in *II. Potential Warning Index Patterns Leading up to an Event*), they may not always provide a complete characterization of the performance of a continuous warning index to predict a change in patient conditions.

In some scenarios detection of an event far in advance may not be useful or distinguishable from a false positive and detection as the event is occurring or after may be too late for the user to interpret the

* Corresponding author at: U.S. Food and Drug Administration, 10903 New Hampshire Ave., Building 62, Room 1129, Silver Spring, MD 20993, United States.
E-mail address: Christopher.Scully@fda.hhs.gov (C.G. Scully).

http://dx.doi.org/10.1016/j.jbi.2017.09.008
Received 5 May 2017; Received in revised form 18 September 2017; Accepted 19 September 2017
Available online 20 September 2017
1532-0464/

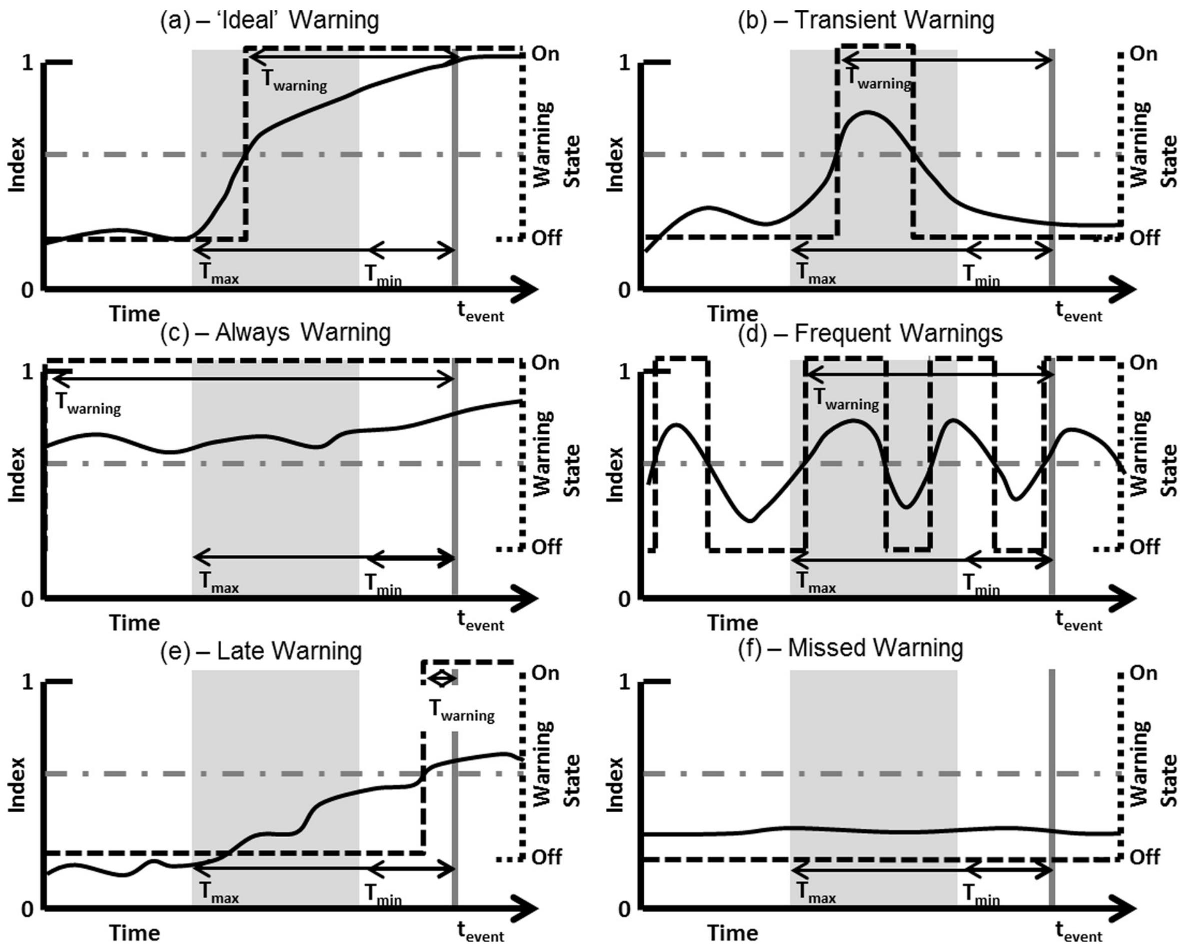

Fig. 1. Potential patterns of a warning index leading up to a time critical health event. (a), (b), (c), and (d) represent scenarios where a warning is provided during a pre-defined meaningful window. (e) and (f) represent scenarios where a warning is not provided during a pre-defined meaningful window. t_{event} is the time that the known critical health event occurs. $T_{warning}$ is the time between the warning activation and t_{event}. T_{max} and T_{min} are the pre-defined maximum and minimum warning times, respectively, relative to t_{event}. Solid black lines indicate the index and dashed lines the warning state that is determined by whether the index is above or below the warning condition threshold (horizontal dash-dot line). The gray box represents the pre-defined window for a warning to be considered a true positive, bounded by T_{max} and T_{min}, and the vertical gray line indicates t_{event}.

warning, reassess the patient, and provide a response. In this article we consider characterization of warning indices performance to detect true positive cases of critical events considering both detection accuracy and timeliness of detection at the same time. We demonstrate the performance information that is provided by comparing two warning system configurations using a normalized shock index for predicting a hemorrhage-induced drop in blood pressure in sheep.

2. Potential warning index patterns leading to an event

The potential patterns of a warning index prior to a critical event presented in Fig. 1 highlight the need for a comprehensive characterization of warning index performance. In these hand curated examples, t_{event} corresponds to the time of the event that has either been defined through an annotation process (e.g., through retrospective analysis of clinical data) or as an experimental endpoint (e.g., a target stopping point in a laboratory study), and $t_{warning}$ corresponds to the time when the warning is first activated. $T_{warning}$ is the warning time (duration) provided from the time the warning is activated until the event ($T_{warning} = t_{event} - t_{warning}$). T_{max} and T_{min} define the interval prior to t_{event} where a warning should be considered meaningful. T_{max} is the longest duration of notice prior to an event that a warning index was designed to provide warning (e.g., 60 min) and T_{min} is the shortest duration of notice prior to an event that a warning index was designed to provide warning (e.g., 1 min). For example, immediately prior to the event it may be expected that a high priority alarm condition is met based on individual vital signs. At this time, the value of an additional warning from an early warning index is likely limited. These values should be defined for the clinical scenario where the warning system is to operate. For the theoretical examples in Fig. 1, a 0–1 index is considered where a warning is triggered when the index crosses a threshold.

Considering a true positive definition as a warning present anytime within the window from T_{max} to T_{min}, examples in Fig. 1a–d would all be considered true positives. Alternatively, a true positive could be defined only when a warning is triggered within the window from T_{max} and T_{min} which would consider Fig. 1a, b, and d as true positives. However, the index and warning pattern of each of these differs which may affect the clinical interpretation. Fig. 1a may represent an ideal scenario for an index representative of patient deterioration where the index is monotonically increasing up to the time of a critical health event, and the warning is initiated within the pre-defined window and stays on until the event occurs. Fig. 1b shows an index that crosses the threshold within the window but then retreats below the threshold a short while later. This leaves a time gap between the end of the warning and the event. This scenario could be appropriate depending on the condition being monitored, but the interpretation of any given warning may suffer for a system that sometimes behaves as in 1a and other times as in 1b. When observed in real-time, the pattern in 1b may initially be considered a false alarm, until the event occurs. In Fig. 1c and d we present two additional patterns that would be considered true positives but may not enable a meaningful response. In Fig. 1c the warning is on