

CoRAD: Visual Analytics for Cohort Analysis

Rishikesan Kamaleswaran
University of Ontario Institute of Technology
Oshawa, Canada
rishikesan.kamaleswaran@uoit.ca

Christopher Collins
University of Ontario Institute of Technology
Oshawa, Canada
christopher.collins@uoit.ca

Andrew James
The Hospital for Sick Children,
University of Toronto,
Toronto, Canada
andrew.james@sickkids.ca

Carolyn McGregor
University of Ontario Institute of Technology
Oshawa, Canada
carolyn.mcgregor@uoit.ca

Abstract—In this paper, we introduce a novel dynamic visual analytic tool called the Cohort Relative Aligned Dashboard (CoRAD). We present the design components of CoRAD, along with alternatives that lead to the final instantiation. We also present an evaluation involving expert clinical researchers, comparing CoRAD against an existing analytics method. The results of the evaluation show CoRAD to be more usable and useful for the target user. The relative alignment of physiologic data to clinical events were found to be a highlight of the tool. Clinical experts also found the interactive selection and filter functions to be useful in reducing information overload. Moreover, CoRAD was also found to allow clinical researchers to generate alternative hypotheses and test them in vivo.

Keywords— *dynamic visual analytics; case-controlled; relative alignment; temporal data streams; physiologic data streams.*

I. INTRODUCTION

Case-control studies are among the most used research methodologies in clinical research. A case-control study involves isolating retrospective data for patients with a condition of interest, and comparing those features to a sample of individuals without the condition [1], [2]. The goal is to explore correlations across relevant clinical variables. In most cases, cohorts must be relatively aligned to an epoch. The alignment may be a time period when a test result was received, such as a blood result confirming or rejecting a possible infection. The relative alignment process typically involves a large number of manual data cleansing and data preparation activities to align clinical data of each patient to a single and representative scale. Most case-controlled studies use clinical data stored in databases and electronic medical records. Performing case-controlled studies using physiologic data is a challenging task. Physiologic data is often collected at a consistent sample frequency, and appear in their raw form, as arrays of values. This is in contrast to a limited set of discrete clinical variables, such as lab reports, or physical observations.

This paper introduces a novel dynamic visual analytic tool called the Cohort Relative Aligned Dashboard (CoRAD). The CoRAD tool represents an instantiation of the dynamic visual analytic publisher component of a larger framework called the

Temporal tri-event parameter based Dynamic Visual Analytic (TDVA) framework. The CoRAD dynamic visual analytic tool addresses the persistent challenge of enabling case-controlled research using relatively-aligned physiologic datasets. CoRAD further supports the integration of retrospective algorithm-generated output, to enhance the analysis workflow. In addition, CoRAD allows the user to drill through multiple hierarchies of data, from quality of signals, to abstractions and ultimately classifications of relevant events.

To validate the effectiveness of CoRAD in a clinical research case study, a preliminary evaluation was conducted at Neonatal Intensive Care Unit at The Hospital for Sick Children, Toronto. The subsequent sections details related works, problem characterization, task analysis, CoRAD design, evaluation methods and the results of the evaluation.

II. RELATED WORK

A case-control study involves retrospective analysis that separates patients based on the presence of a condition [1]. Case-control studies, among many observational research methods, remain an important aspect of clinical research [2]. Differences are studied and hypotheses are generated based on the analysis, to motivate deeper investigation and more rigorous research. However visualizations that support these efforts in physiologic data remain elusive.

A. Artemis Platform

Artemis is an online analytic platform that was developed to source, analyze, and perform real-time feature detection on multiple physiological data streams, for multiple conditions in multiple patients [3]. Artemis supports the deployment of real-time event stream processing algorithms. In this paper, we use data generated by an algorithm running in the Artemis platform for neonatal sepsis that was executed to detect and classify Heart Rate Variability (HRV) scores between 0 and 60, where zero signifies no variability and 60 demonstrated that the patient's heart rate varied consistently in the hour. The details of the neonatal sepsis algorithm have been previously published [4]. Results from the analysis are then sent to a database and also available for real-time streaming for

visualization. The output are then processed and sent to a platform that was developed using the TDVA framework. That platform produces instantiations called dynamic visual analytic marts, such as the CoRAD.

B. Cohort health visual representations

In the general space of health-based cohort analytics, some recent work has resulted in high fidelity visualizations with a time component. TimeSpan [5] provides an interactive dashboard for identifying door-to-needle time for stroke patients at a large tertiary hospital. LifeLines presents graphical summaries of patient journey [6]. The Cohort Comparison (CoCo) tool, provides a simple interface for exploring statistical correlations across multiple clinical datasets [7]. DecisionFlow presents graphical summaries of patients who developed heart failure relative to a population [8]. VISITORS is a dashboard for analyzing clinical temporal abstractions in oncology patients [9]. EventFlow presents a method to simplify event sequence information to rapidly identify abnormalities [10]. While all of these visualizations introduce cohort analysis of patients using clinical information, there is a need for research in representing temporal abstractions of physiologic data across cohorts, and supporting automated temporal relative alignment, while allowing the user to gain contextual awareness using low and higher-level summarizations of data.

C. Visual analytics of temporal data

Domain specific dynamic visual analytic tools have been shown to perform well in communicating anomalies to the end user. The VisAlert system [11], for example, provides situational awareness for network security analysts. Another system in the same domain is LiveRAC [12], which supports additional exploratory features such as semantic zoom to search through the data set, and allows for side-by-side comparisons between different clusters. However, this system presents a complicated user interface with potential for visual clutter. Director [13] is a visual analytic tool for computer network simulations. It provides a heatmap-based timeline visualization to identify the health of multiple nodes, along with a temporal view of their health deterioration. CloudLines [14] introduces an incremental event visual analytic tool using kernel density estimation (KDE) to amplify signals from highly dense areas and minimize low density areas. The technique is applied to online news stream analytics, and multiple time-series data are used to highlight topic emergence, and when the topic is no longer emerging, a visual decay function is applied to emphasize more popular topics.

While most visual displays are temporally aligned to the most recent epoch, in this paper we present a novel visual analytic tool that uses relative alignment to a real-world independent event. Two heatmap timelines are presented in the main display to allow clinical researchers the ability to visually explore patterns in HRV across multiple patients.

III. PROBLEM CHARACTERIZATION

Sepsis is a form of hospital acquired infection, and remains a serious health problem requiring antibiotic therapy [15].

Currently it is very difficult to detect using non-invasive methods, such as by bed-side monitoring. Clinicians rely on qualitative observational methods for identifying signs on this illness. When sepsis is suspected, blood samples are drawn and required to confirm any diagnosis. However, neither method has been found to be reliable [16]. There is growing body of evidence that shows new pathophysiologic behaviours can be identified earlier using physiologic data. One such case involves the study of reduced HRV as a potential indicator of sepsis [17], [18]. In addition, Flower et al, 2010 [19], present results that indicate periodic cycles of heart rate decelerations, or bradycardias, are common and seen to be clinically correlated with sepsis in addition to reduced HRV and they propose heart rate characteristics as a means to correlate the occurrence of the two together.

McGregor et al., developed an algorithm that produces real-time HRV scoring for neonatal infants [4]. This scoring can be used to identify temporal areas where there is reduced HRV that indicates some sign of illness. A dataset containing HRV information and algorithm-generated classifications of bradycardia as part of McGregor's neonatal spell research are available from a prior study [20]. Data from a total of 47 patients are available, of which 33 patients have sufficient data quality. The goal of this study is to investigate the hypothesis exposed in Flower et al. [19] that periodic cycles of heart rate decelerations together with reduced HRV are common and clinically correlated with neonatal sepsis. This information is presented in CoRAD and we performed an evaluation to test participants' ability to determine sepsis based on Flower's hypothesis. The study was approved by the Research Ethics Boards at The Hospital for Sick Children and at UOIT.

IV. TASK ANALYSIS

Two domain experts were asked to describe specific tasks they perform to conduct hypothesis testing using physiologic data across a cohort of patients. The common tasks were:

- T1 Relatively align temporal abstractions:** Relevant HRV values are filtered and manually aligned to an anchor point. The relative alignment performed manually, can introduce errors, and can be time consuming.
- T2 Import abstractions to a spreadsheet:** Each HRV value is then sorted by the relative aligned time and imported to a spreadsheet manually, this also introduces scope for potential error.
- T3 Graph abstractions:** Once the HRV values were imported into the spreadsheet, line charts and stacked bar graphs were frequently used to visualize the data.
- T4 Identify correlations:** The domain expert would find associations by comparing HRVs before and after the anchor point. Further, the domain expert might highlight multiple patients of interest and investigate patterns between the selections.

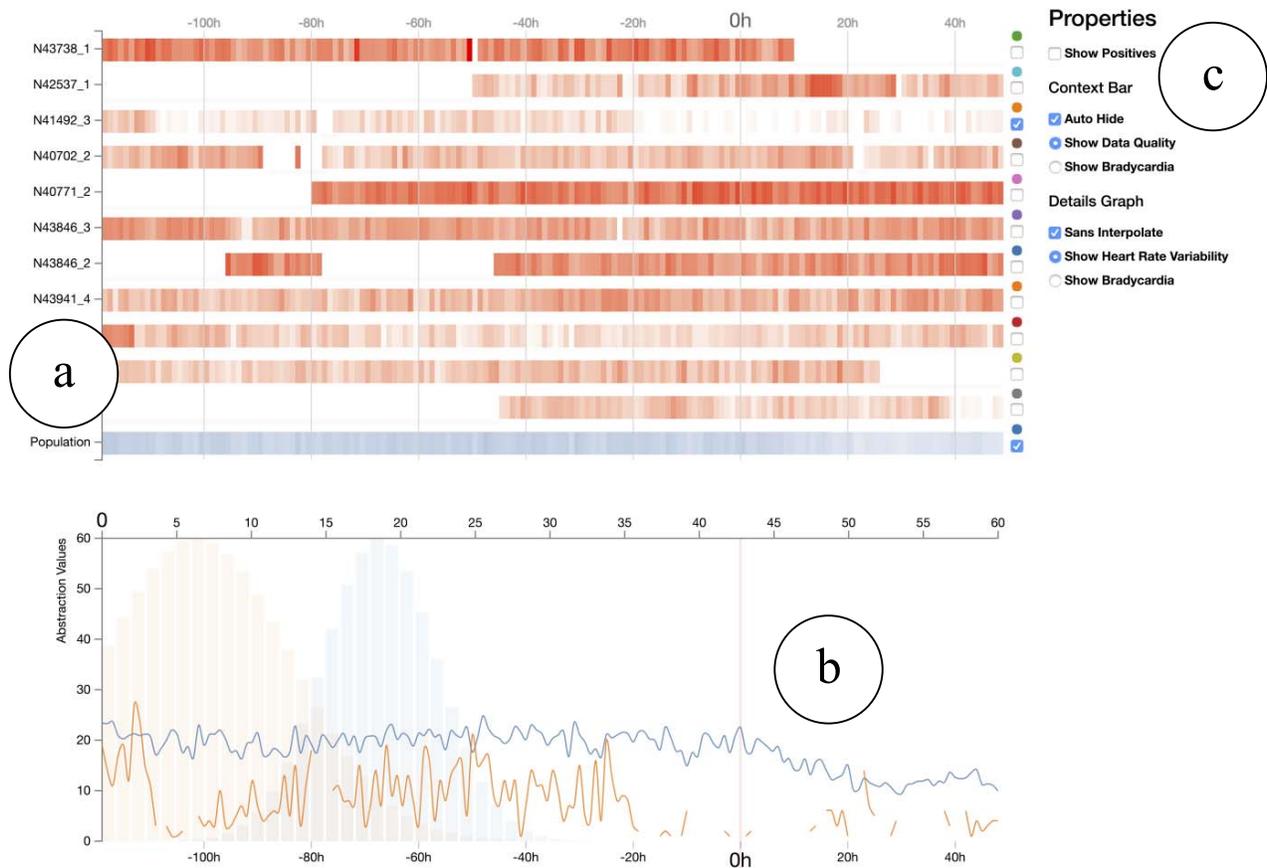


Figure 1: CoRAD provides interactive focus supporting analysis related to events relatively aligned at the zero hour (0h) mark. (a) In this figure all patients are aligned to the y-axis, and the relative-time is marked across the top horizontal position. All patients are coloured using a red scale (lighter means reduced HRV, darker means more variable heart rate), unless the ‘Show Positive’ control is active. The normalization of all results were used to produce the population map coloured in blue. The detailed view on the bottom (b) provides a line-chart view of details including the raw-data, temporal abstraction, or high-level classifications. A multi-coloured histogram is also available and highlights the distribution of HRVs over the entire duration. Each colour is mapped to a patient and the map appears above the selection box in the right. The blue histogram represents an average of the population. (c) Provides a view of the properties control, functions are provided to manipulate the dashboard view interactively.

These tasks were performed manually, and was stated to be time-consuming and error prone. These tasks informed the design of CoRAD and serve as a guide for future research in similar application domains.

V. DESIGN OF CoRAD

We describe CoRAD with its design goals that were informed from the observations and task analysis with domain experts.

DG1 Integrate heterogeneous data: The first task, the relative alignment of physiologic data to clinical data, can involve a mix of numeric, continuous, or ordinal

data types. Our design goal is to unify the representation of these data types for extendibility of CoRAD.

DG2 Single holistic view: Currently most of the current tasks performed are manual, however, the ultimate goal is to collect all important disparate data into a single environment. Patient clinical data is closely associated with the patient’s physiology, which is correlated to the device measuring that data. Therefore the goal is to provide an integrated view of all direct and indirect patient data.

DG3 Details on demand: The user requires access to details, however current tasks limit the degree of data that can

be accessed in a timely manner. Moreover, access to details can be useful in determining the salience of an observation. Our goal is to provide the user convenient access to details on demand.

DG4 Access to statistical tools: Many of the activities performed are by nature, statistical. So our goal is to provide the user with a simple statistical view of the data to assist potential discovery of salient features.

CoRAD is illustrated in Figure 1, and consists of four components, including: the main view (Figure 1a), detail view (Figure 1b), properties view (Figure 1c), and the context bar (Figure 2). The interface was developed using D³ [21]. In this section each component is described in detail.

A. Main View

The main view, illustrated in Figure 1a, consists of several patient bars that utilize an opacity-controlled colour scale to present HRV information to the user. The darker bars reflect higher HRV and the lighter shades denote lower scores. Each patient bar is painted from left to right, where the left most region shows -120 hours relative to the point of interest which in this case was the suspicion of late onset neonatal sepsis. This represents about five days prior to the aligned pivot, the zeroth hour. The right-most side of the heatmap shows information for 48 hours after the aligned pivot. The zeroth hour is marked by a grid line that extends from the top of the main view and repeat every 20 hours. This method of relative alignment, in addition to the context bar support tasks T1 – T3 and DG1 and DG2. Each patient is stacked from bottom up, with the bottom being the population bar. This vertical arrangement provides a convenient means of comparing HRV patterns within their respective relatively aligned epoch. An anonymized patient identification is appended to the left vertical axis.

B. Detail View

The detail view provides an alternative view for selected data from either of the other two views. It consists of a line graph and a histogram. The line graph is a plot of HRV values for an interval selection in the main view. A line graph was previously used to display HRV values [22]. If there are no selections in the main view, the line graph displays HRV values for the entire duration. The user is also able to display the line plot of the average HRV of the population. Having access to this raw data can be helpful in associating discrete values to observations. The line graph supports DG2. For instance, Figure 1b, shows the HRV line graph for patient N41492_3 and the population pinned to the same canvas, while all other lines are set to be transparent. The line graph can be configured to show interpolation, should missing data be present in the dataset. The default option is to avoid interpolation, and make the line transparent when there are missing data.

The detail view also contains a histogram that displays the distribution of HRV values for each selection in the main view. The distribution is a Gaussian plot derived from the mean, and standard deviation of the HRV data for each sample. Should the user select the population, a population mean and standard deviations of HRV's are used based on the values of all 33

patients in the dataset. The availability of the histogram fulfills DG4. The detail view can be altered to higher-level classifications, such as the temporal presence of bradycardia. This view also exposes details about the HRV value and the associated patient when the user selects a single line on the screen. The detail view more specifically supports T4, as it allows the user to directly compare two or more patients within a window of time. The interactive details tooltip allows CoRAD to provide the domain expert details on demand, thus supporting DG3.

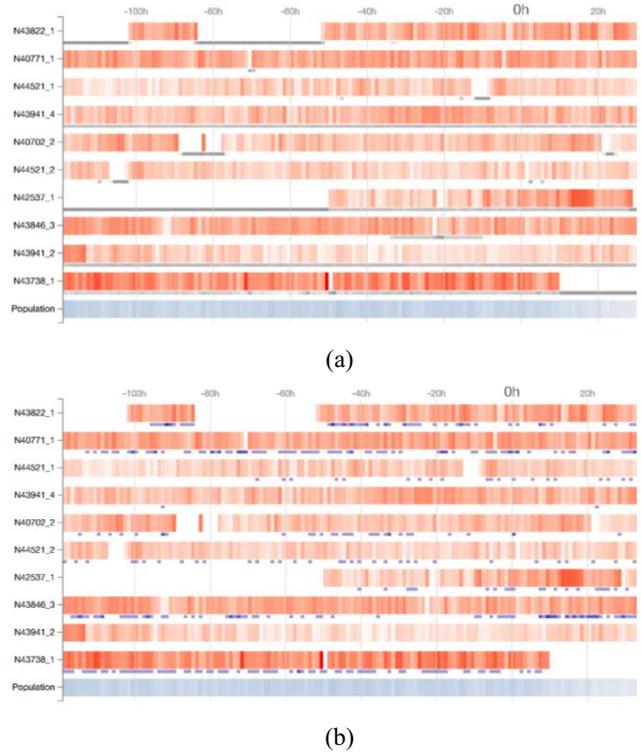
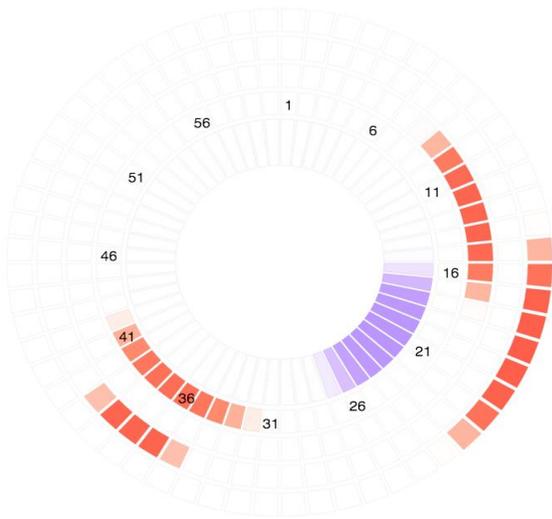


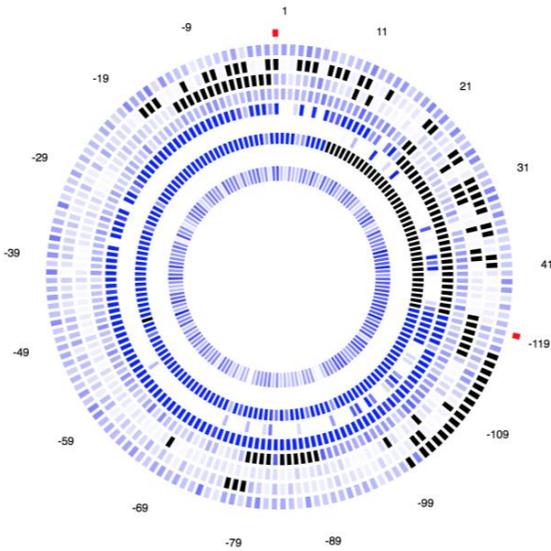
Figure 2: The Context Bar View adds small bars below the main HRV data, with two modes: (a) shows the data quality illustrated using grey fills, the darker fill represents times when the data quality was compromised, and (b) representing bradycardia events illustrated using blue fills, darker regions represent increased number of bradycardia.

C. Properties View

The particular methods by which information is presented in the main and detail views are controlled by the properties view presented in Figure 1c. The first checkbox allows the user to highlight patients that were tested positive, and alternatively to disable the highlighting should the user not want to make positive cases visible. The subsequent selection buttons are grouped according to the views they manipulate. The ‘show data quality’ and ‘show bradycardia’ buttons in the context bar group control the data being represented in the context bar view. The raw data, abstraction and classification selection buttons controls the information visible in the detail view. This view, which can enhance the ability of the domain expert to extract details on demand, supports DG3.



(a)



(b)

Figure 3: Alternative designs for a cohort-based relatively aligned dashboard. (a) A radial graph representing the distribution of HRV scores over 120 hours for each patient. (b) A radial graph representing the temporal trajectory of HRV scores for patients. A red mark is annotated to determine the zeroth hour, as well as the 48th hour.

D. Context Bar View

The context bar resides immediately beneath the patient bar and can represent one of two types of information, including data quality and the presence of bradycardia. The data quality display highlights regions of poor data quality, using a darker shade. That encoding is useful in alerting the user that the red

scale shading of the HRV value is not reflective of the entire hour. This is particularly important as patients are often disconnected from sensors. Identifying data quality issues was an important, but cumbersome task of the analysis process. The context bar is designed to reduce the burden by integrating that information within the main view. Figure 2a, shows the context bar illustrating regions of poor data quality. For instance, patient N43738_1 is shown to have compromised data quality just before the 20th hour and continues until the 48th hour. Meanwhile, N43941_2 is shown to have comparatively better quality throughout the entire duration. The second type of data the context bar can represent is bradycardia data. Figure 2b, illustrates the presence of bradycardia episodes during an hour by affixing a blue box under the appropriate relative time period. To determine the current data represented by the context bar, the user can refer to the properties view to identify the selected option. The user can interactively control the data represented in this layer, hence, providing information on demand.

E. Design Alternatives

Prior to finalizing the visual components of CoRAD, several alternatives were investigated. Among the most prominent alternatives was a radial graph that consisted of two views: a distribution and temporal view. The distribution view illustrated in Figure 3a, consists of a central arc that describes the average distribution of HRV scores for the population, with each ring representing a separate patient. The arc begins as zero at the top of the ring and extends to the 60th mark. Zero represents no variability, while 60 represents variability in each minute of the hour. For the distribution illustrated in Figure 3a four patients are compared to the average of the population. The average of the population has a mean around the 21 mark. However for the patients the first and third ring, a mean for the distribution is observed around 36 mark. Significantly, these patients have had a higher than average HRV scoring recorded during the monitored period.

A temporal radial graph was also constructed to support the identification of abnormal trajectories of HRV values in populations using an average of the population as a baseline. The temporal radial graph illustrated in Figure 3b presents seven patients who are aligned to population average as separated rings at fixed radii from the centre. Opacity is controlled to show regions of higher and lower HRV values. For instance, the first and third patient from the population are seen to have very dark blue rings, signifying higher HRV scores. While the patients in the outer ring have lighter blue rings, signifying reduced HRV. While there has been many forms of radial graphs produced [23], there have been some concerns that have emerged about the interpretation of radial graphs [24], [25]. However, other instances of radial graphs were shown to be successful in identifying trends [26]. The radial visual representations were evaluated in a preliminary study involving two clinical researchers. Both displays required extensive training time to understand, and, the temporal radial graph presented a challenge when interpreting the tail-ends of the monitoring duration. Evaluators had a difficult time observing patterns only in the -120th hour without being influenced by the +48th hour that was within its immediate vicinity.

For these reasons the radial graphs were not selected for the full evaluation. While these challenges show that radial graphs may involve more training, more research needs to be done to further enhance the visual representation to address those shortcomings. In future work, both radial graphs will be evaluated using similar multidimensional datasets.

VI. EXPERT EVALUATION

To determine the usability and usefulness of CoRAD, we conducted an expert evaluation. Two key quantitative values that were measured were accuracy of the verbal statements and task completion.

A. Methodology

The evaluation of CoRAD was conducted with five experts including, clinicians and clinical researchers. A single factor, technique, was varied, with two levels: CoRAD (Figure 1), and stacked bar display (Figure 4). The stacked bar representation is inspired from an alternate design used in the neonatal spells research, however this research involves only the bradycardia episodes [27]. Seven key measures were collected including, demographic information, completion rate, accuracy of response, usability problems verbalized, errors made during the evaluation, posture, and the subjective satisfaction.

The experimental task was to determine and verbalize suspicion of infection for a single patient (a row in CoRAD, a bar in stacked bars). When the participant began the new task they were asked to state “I’m moving to the next patient”, this statement served to mark the end of the former task and the start of a new task. Following exposure to a technique, they were asked to provide feedback on the usability and acceptability of the user interface. The participants were directed to provide their honest opinion of the presented display and to participate in a post-session subjective questionnaire involving a 5 point Likert scale. All verbal discussions, as well as the cursor movements were recorded and transcribed.

Participants received an overview of CoRAD and the stacked bar graph at the start of the experiment, along with the test procedure, and equipment. There was one training scenario consisting of 10 patient datasets. Training consisted of the experimenter reading aloud interpretations of three patient datasets, taking 5 – 10 minutes. Then the participant was provided time to explore the interface and familiarize themselves with the functionality. The 10 patients used in the training set were not included in the evaluation set.

Each *evaluation scenario* consisted of 10 *tasks*. Two evaluation scenarios were carried out for each technique, and repeated for the other technique (data order was randomized). Due to data availability, the same datasets (in random order) were used for the training tasks in both techniques across all

participants. The ordering of technique was counterbalanced to limit learning effects. In summary, from the original 33 datasets, 10 were used for training, and of the remaining 23, 20 were randomly selected and used in evaluation scenarios.

Expert participants were recruited via email. Five experienced staff physicians were selected from a pool of nine qualified personnel. The sample was chosen purposefully to represent the local demographics with respect to age, sex, years of experience, and involvement in physiologic research. Trainees and fellows were excluded from this study. There were a total of 5 (participants) x 2 (evaluation scenarios) x 2 (techniques) x 10 (datasets) = 200 evaluation tasks. Study sessions lasted an average of 45 minutes.

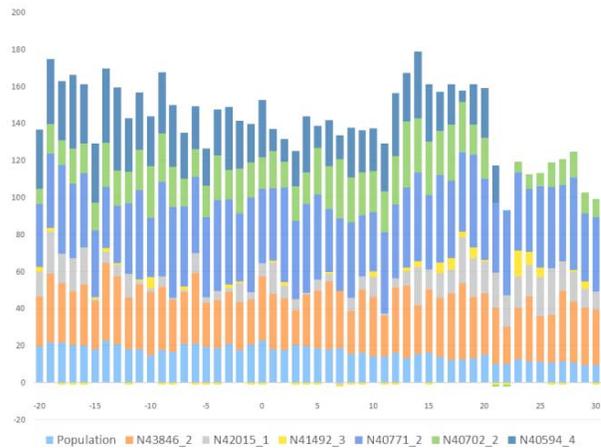


Figure 4: Stacked bar representation used to stack all patients above a population average (bottom). The zeroth mark represents the point of suspicion of infection, and negative numbers illustrate HRV scores in each preceding hours, while positive numbers signify HRV scores in the hours after the event. A bar below the x-axis represents

B. Procedure

A laptop computer with Web site/Web application and supporting software was used in a typical office environment. The participant’s completion of the task was video recorded for aiding transcription and analysis of time to completion. The evaluation was initiated with a brief description of the CoRAD application, and the participant was made aware that the facilitator would be evaluating the application, rather than the diagnostic abilities of the participant. Participants were then prompted to sign an informed consent sheet that acknowledges: the participation is voluntary, that participation can cease at any time, and that the session will be videotaped but their privacy of identification will be safeguarded.

The participant was then asked to complete a demographic and background questionnaire. Once the demographic questionnaire was completed, the participant was introduced to one of the two techniques. In both the training and experiment phases, the participant was frequently asked to think aloud, describing their analysis process. The participant body posture was observed and entries were made to the observation diary. After each the second exposure to each technique, the participant was asked to complete the post-task questionnaire and elaborate on the task session with the facilitator. After all evaluation scenarios were attempted, the participant completed the post-test satisfaction questionnaire.

C. Analysis

Each session was video recorded and transcribed (with field notes). Analysis was ongoing throughout the fieldwork to allow emergent themes to be included into the data collection process. The associated themes and distinctions formed the basis of the coding strategy. Review of the evolving themes contributed to the data synthesis and interpretation. To analyse the accuracy of detection the sensitivity-specificity binary classification method was used. This method is a popular clinical measure for determining the efficacy of an intervention [28]. Average timing was manually determined from the video recording and rounded to the nearest second.

VII. RESULTS

The study yielded data from a total of 200 tasks performed across both conditions (10 datasets \times 4 evaluation scenarios \times 5 participants). This section highlights the main differences in demographics, accuracy of detection of sepsis, task completion, and subjective feedback received from expert participants.

A. Demographic Differences

Five clinical researcher participants were recruited in the study and all participants completed each component to completion. All participants had at least ten years of practice in critical care medicine. Two females and three males were recruited. The average age of the sample was 40 – 50 years of age. The average length of total clinical experience was 18 years. All but one subject reported using the computer multiple times a day for analysis purposes. All participants had at least 15 years of experience working with physiologic data. The average reported score of participants' familiarity with physiologic data was 4 out of 5, where 1 represented minimal familiarity and 5 represented expert proficiency. On the same scale, participants reported their familiarity with HRV as 2.5 out of 5 and knowledge of neonatal sepsis as 3.5 out of 5. Two of the five participants were aware of the hypothesis exploring the link between HRV and neonatal sepsis. The years of experience also did statistically differ in the clinical researcher's familiarity with the relationship between HRV and neonatal sepsis.

B. Accuracy of Detection

Table 1 summarizes the results of the display condition, true positive, true negative, false positive, false negative, and

TABLE I. SENSITIVITY AND SPECIFICITY OF BOTH CONDITIONS

Participant	Condition	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity
1	CoRAD	2	9	4	5	29%	69%
1	Stacked	3	7	6	4	43%	54%
2	CoRAD	2	11	3	4	33%	79%
2	Stacked	0	15	1	4	0%	94%
3	CoRAD	1	13	4	2	33%	76%
3	Stacked	0	13	3	4	0%	81%
4	CoRAD	0	12	5	3	0%	71%
4	Stacked	2	12	2	4	33%	86%
5	CoRAD	2	13	3	2	50%	81%
5	Stacked	1	9	6	4	20%	60%
Average	CoRAD	-	-	-	-	29%	75%
Average	Stacked	-	-	-	-	19%	75%

TABLE II. TASK COMPLETION MEASURES FOR BOTH CONDITIONS

Participant	Condition	Successfully Completed	Errors	Average Time (seconds)	Standard Deviation (seconds)
1	CoRAD	20	3	25	12
1	Stacked	16	0	23	16
2	CoRAD	20	1	9	11
2	Stacked	17	0	5	5
3	CoRAD	20	0	20	7
3	Stacked	19	0	17	6
4	CoRAD	20	0	16	17
4	Stacked	20	0	15	15
5	CoRAD	20	1	15	8
5	Stacked	18	0	27	19
Average	CoRAD	20	1	17	11
Average	Stacked	18	0	18	12

sensitivity and specificity for all tasks performed. True positive refers to the number of true sepsis patients that were correctly identified to be septic. True negative to the correct identification of negative cases as non-septic. False positive refers to the number of patients who were incorrectly identified as positive, and false negative the number of patients who were incorrectly identified as negative. The sensitivity and specificity scores were collected for each condition and an average specificity and sensitivity score was generated.

C. Task Completion

Table 2 summarizes results of the tasks successfully completed, errors, average time in seconds, as well as the standard deviation in seconds. Non-crucial errors occurred in the CoRAD condition that did not obstruct task completion. The error was a result of using an external monitor that did not reproduce colour saturations, hence the normal distribution histograms were less visible. This error was fixed after the first pilot trial by reverting to the laptop monitor.

D. Subjective Feedback

Clinical researchers provided rich subject feedback about the usefulness and utility of both conditions. On the stacked bar representations, clinical researchers noted that as they progressed through each it became progressively difficult to analyse the patient's HRV scoring due to the non-aligned vertical height. The stacked representation was seen to lack the ability to allow the expert to compare a certain temporal range against the rest of the data set. Clinical researchers also noted that using the stacked bar representation required manual scrolling to get a perception of the entire duration of the dataset. The lack of contextual information was noted to be a significant negative of the stacked bar display.

CoRAD was perceptually simpler and easier for the experts to use. The heatmap representation was unanimously noted as being very helpful for analysis. All clinical researchers appreciated having a single view of the dataset. One of the clinical researchers expressed having been confused with the red colour coding, they identified the darker red regions as being more severe. Interactive zooming was heavily used and noted as a positive component. While many experts found the detail view important to their analysis, two experts voiced having options to have the normal distribution appearing as a histogram on a separate display.

The contextual bar was heavily utilized, however three of the five clinical researchers requested to see both bradycardia and data quality at the same time. One clinical researcher found the CoRAD display too cluttered and overwhelming, however that clinician did not use any of the interactive selection and filtering functions. Moreover, that clinical researcher preferred to see a summary graph showing only the most deviant patient. Other clinicians reported high satisfaction with the availability of the interactive selection and filter functions, and stated it helped to reduce excess information. When interactive selections were used, most clinical researchers also used the filter to display key patients of interest in the detail view. A typical workflow is illustrated in Figure 5, where two patients of interest are compared to the population mean in the detail view. In the main view, the user has highlighted an interval of

interest. All clinical researchers stated the highlight function to be useful for determining changes in HRV across multiple patients at the same time, within salient temporal windows. One clinical researcher started the analysis by immediately highlighting a temporal window, and maintained that same window throughout the entire duration of the analysis. That researcher stated that they did not view data in other durations to be relevant.

One clinical researcher stated a desire to see distributions over only a fixed temporal range. That clinical researcher found the display of the average distribution across the entire duration not significantly helpful for completing their task. Researchers used the detail view to confirm their visual suspicions, one subject verbalized: "I am not sure (whether I am correct) visually about these subsets of patients, I want to see them statistically using the detail view. Ah, I see that my visual interpretations were correct".

After both conditions were tested, clinical researchers were asked to state their preference for one display. All experts preferred CoRAD over the stacked bar display. All clinical researchers stated they would utilize CoRAD as one of the applications in their analytic toolkit. Three clinical researchers with significant bed-side research interests expressed an inclination to use CoRAD as a tool as part of their bed-side rounds. One clinical researcher mentioned that after some suggested modifications, such as including a dynamic histogram for the normal distribution, they would see themselves actively using CoRAD.

VIII. DISCUSSION AND FUTURE WORK

An expert evaluation consisting of five domain experts analyzing HRV and bradycardia events was conducted in an attempt to predict the infant's neonatal sepsis status. Results from the expert evaluation revealed several key insights. The demographic differences in this study reveal broad coverage in age, sex, and years of experience. Based on the results that were observed, there seems to be little differences between age, gender, and years of experience to both the accuracy and task completion ($p > 0.05$). The relative low score attributed to familiarity of HRV is significant as this measure has yet to be established as a routine clinical indicator in practice [29]. One clinical researcher mentioned that, while she did not use HRV actively, she had knowledge of its potential relevance.

Accuracy of sepsis detection was reported with sensitivity appearing below 50% for both conditions (Table 1). CoRAD allowed for a 10% increase in sensitivity, however. With respect to the specificity, both the stacked bar and CoRAD displays indicate an identical score at 75%. The low sensitivity score across both displays may support the notion of a weak link between HRV, bradycardia and neonatal sepsis, thereby providing counter evidence against the initial hypothesis for the dataset used by this evaluation [19], [17], [18]. Since the commencement of this research another independent study has also reported low accuracy results for the detection of late onset neonatal sepsis using these two physiological behaviours as part of the heart rate characteristics approach in a three year observational study [30]. Task completion (Table 2) was significantly higher on CoRAD than on the stacked bar display

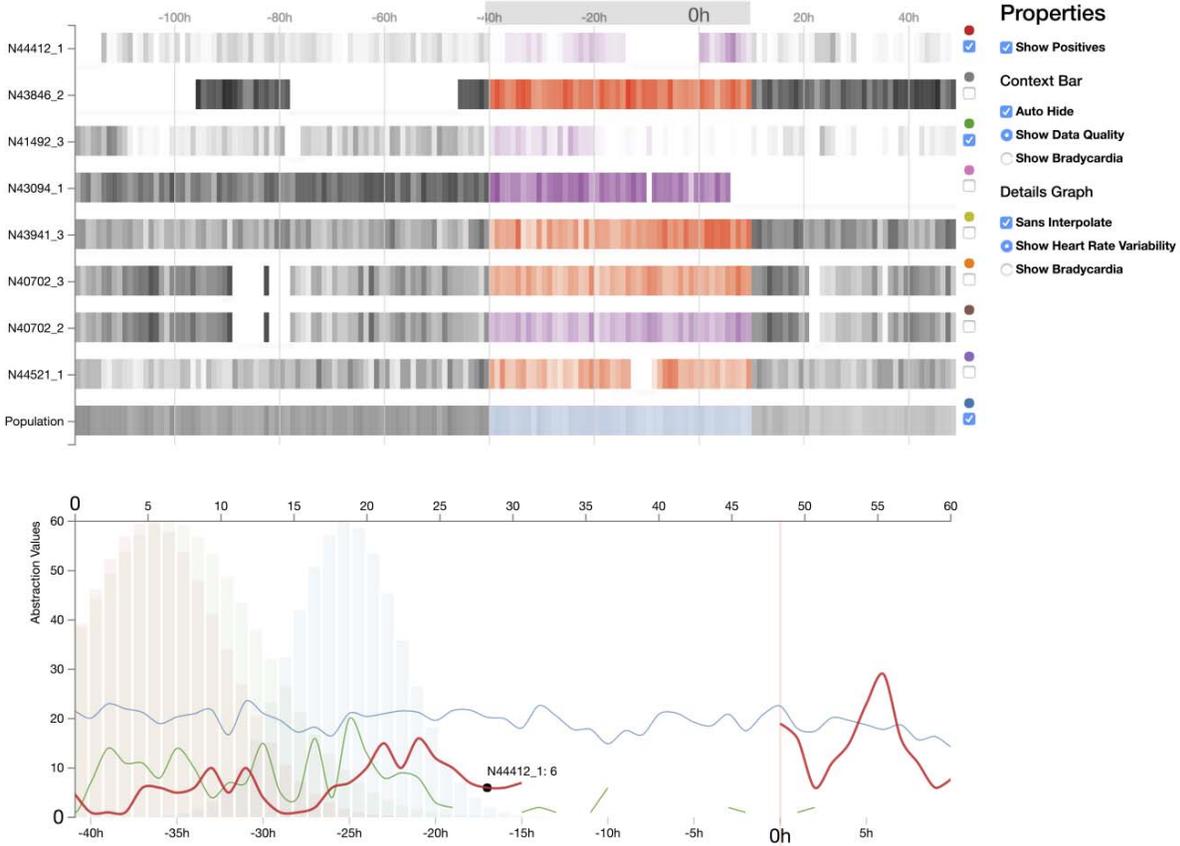


Figure 5: Interactive selection and filtering functions on the CoRAD tool allow clinical researchers to isolate patients of interest. In this figure, the ‘Show Positives’ function is selected, which filters patients based on a positive clinical result for neonatal sepsis. The clinical researcher is shown here highlighting -40 hour to +10 hour two positive cases N44412_1 and N41492_3 in the detail view.

($p < 0.05$). All instances of unsuccessful task completion occurred when these clinical researchers failed to analyse one of the required patients in the display. The omitted tasks were not subsequently identified by the clinical researcher in most cases (8 out of 10), in one instance the researcher spoke aloud to confirm whether they may have missed a patient in their analysis. Most of the omitted tasks appear as patients stacked in the middle or upper region of the representation.

Non-crucial errors were seen early in the evaluation with CoRAD, in particular with colour accuracy with the external display used in a single experiment. The CoRAD display was subsequently shown on another display which produced accurate colour representation. An additional errors were encountered with subject 3 and 7 where the database communication was temporarily timed-out. A refresh of the web page allowed the evaluation to continue. The average time for task completion was not statistically significant between the two conditions (17 vs 18 seconds). Even with the additional

number of interactive manipulations that were performed by clinical researchers, CoRAD still allowed the user to perform their task in the same amount of time. General interest in the tool did not contribute to longer task completion times.

The general subjective feedback shows greater interest in the CoRAD display. A unanimous agreement was present on the integration of CoRAD as an informatics tool that should be deployed as a tool in the hospital analytics suite. In particular, clinical researchers found having the ability to interactively select, filter, and expose details on demand to be helpful to their analysis workflow. Some researchers report using the tool, however with other forms of data, such as electroencephalogram, or an oxygen saturation dataset. The clinical researchers also suggested two major areas for future work. Including having the option to manually change the colour scheme, allow the context bar to represent both data quality and bradycardia at the same time, and separate the histogram view from the details graph. Future work with

CoRAD will address the identified limitations. This study presents early results from a user study of five experts at a single site. Future work will expand then number of participants and include additional sites in the evaluation.

IX. CONCLUSION

CoRAD has shown positive effects in supporting clinical researchers explore patterns across multiple modes of physiologic data using an interactive cohort based visual analytic tool. The CoRAD display was tested in the context of an application by conducting an expert evaluation and experimentation against a control stacked bar display. Exposure to CoRAD within this limited case study, resulted in interest on the part of the clinical researchers to use this tool in other scenarios, such as electrocardiography and oxygen saturation variability. The relatively aligned heatmap allowed each researcher to rapidly identify event details, which was more difficult on the control display. However, open challenges remain in studying alternative visualizations that can be used to display multiple features, such as data quality, and bradycardia without producing visual clutter.

ACKNOWLEDGMENT

We would like to thank our five domain experts for participating in the evaluation and providing valuable feedback.

References

- [1] D. A. Grimes and K. F. Schulz, "Compared to what? Finding controls for case-control studies," *The Lancet*, vol. 365, no. 9468, pp. 1429–1433, 2005.
- [2] K. F. Schulz and D. A. Grimes, "Case-control studies: research in reverse," *The Lancet*, vol. 359, no. 9304, pp. 431–434, 2002.
- [3] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. P. Smith, and D. Sow, "Real-Time Analysis for Intensive Care: Development and Deployment of the Artemis Analytic System," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 29, no. 2, pp. 110–118, 2010.
- [4] C. McGregor, C. Catley, and A. James, "A process mining driven framework for clinical guideline improvement in critical care," *A process mining driven framework for clinical guideline improvement in critical care*, vol. 765, 2012.
- [5] M. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "TimeSpan: Using Visualization to Explore Temporal Multi-dimensional Data of Stroke Patients," 2015.
- [6] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "LifeLines: using visualization to enhance navigation and analysis of patient records.," *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pp. 76–80, 1998.
- [7] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, and B. Shneiderman, "Cohort comparison of event sequences with balanced integration of visual analytics and statistics," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 2015, pp. 38–49.
- [8] D. Gotz and H. Stavropoulos, "Decisionflow: Visual analytics for high-dimensional temporal event sequence data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, pp. 1783–1792, 2014.
- [9] D. Klimov, Y. Shahar, and M. Taieb-Maimon, "Intelligent visualization and exploration of time-oriented data of multiple patients.," *Artificial intelligence in medicine*, vol. 49, no. 1, pp. 11–31, May 2010.
- [10] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, "Temporal event sequence simplification," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2227–2236, 2013.
- [11] Y. Livnat and J. Agutter, "A visualization paradigm for network intrusion detection," in *LAW'05. Proceedings from the Sixth Annual IEEE SMC. IEEE*, 2005, no. June, pp. 17–19.
- [12] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "LiveRAC: interactive visual exploration of system management time-series data," in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. ACM*, 2008, pp. 1483–1492.
- [13] T. H. Yu, B. W. Fuller, J. H. Bannick, L. M. Rossey, and R. K. Cunningham, "Integrated environment management for information operations testbeds," in *VizSEC 2007*, 2008, pp. 67–83.
- [14] M. Krstajic, E. Bertini, and D. A. Keim, "Cloudlines: Compact display of event episodes in multiple time-series," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2432–2439, 2011.
- [15] M. P. Griffin, T. M. O'Shea, E. A. Bissonette, F. E. Harrell, D. E. Lake, and J. R. Moorman, "Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness," *Pediatric research*, vol. 53, no. 6, pp. 920–926, 2003.
- [16] M. R. Hammerschlag, J. O. Klein, M. Herschel, F. C. Chen, and R. Fermin, "Patterns of use of antibiotics in two newborn nurseries.," *The New England journal of medicine*, vol. 296, no. 22, pp. 1268–1269, 1977.
- [17] M. P. Griffin, D. E. Lake, and J. R. Moorman, "Heart rate characteristics and laboratory tests in neonatal sepsis.," *Pediatrics*, vol. 115, no. 4, pp. 937–41, 2005.
- [18] M. P. Griffin, D. E. Lake, E. A. Bissonette, F. E. Harrell, M. O. Shea, J. R. Moorman, T. M. O. Shea, and A. O. Monitoring, "Heart Rate Characteristics : Novel Physiometers to Predict Neonatal Infection and Death," *Pediatrics*, 2005.
- [19] A. A. Flower, J. R. Moorman, D. E. Lake, and J. B. Delos, "Periodic heart rate decelerations in premature infants," *Experimental Biology and Medicine*, vol. 235, no. 4, pp. 531–538, 2010.
- [20] R. Kamaleswaran, C. Collins, A. G. James, and C. McGregor, "PhysioEx: Visual Analysis of Physiological Event Streams," *Eurographics Conference on Visualization (EuroVis) 2016*, vol. 35, no. 3, 2016.
- [21] M. Bostock, V. Ogievetsky, and J. Heer, "D³ Data-Driven Documents," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2301–2309, 2011.
- [22] C. McGregor, C. Catley, and A. James, "Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, 2012, pp. 1–5.
- [23] G. M. Draper, Y. Livnat, and R. F. Riesenfeld, "A survey of radial methods for information visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 5, pp. 759–776, 2009.
- [24] Y. Albo, J. Lanir, P. Bak, and S. Rafaeli, "Off the Radar: Comparative Evaluation of Radial Visualization Solutions for Composite Indicators," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 22, no. 1, pp. 569–578, 2016.
- [25] R. Feldman, "Filled radar charts should not be used to compare social indicators," *Social indicators research*, vol. 111, no. 3, pp. 709–712, 2013.
- [26] D. A. Keim, F. Mansmann, J. Schneidewind, and T. Schreck, "Monitoring network traffic with radial traffic analyzer," in *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, 2006, pp. 123–128.
- [27] C. McGregor, E. Pugh, and A. Thommandram, "A Big Data Based Approach for Visualising Neonatal Apnoea and Spells," 2015.
- [28] B. J. McNeil, E. Keeler, and S. J. Adelstein, "Primer on certain elements of medical decision making," *New England Journal of Medicine*, vol. 293, no. 5, pp. 211–215, 1975.
- [29] P. K. Stein, "Challenges of Heart Rate Variability Research in the ICU*," *Critical care medicine*, vol. 41, no. 2, pp. 666–667, 2013.
- [30] S. A. Coggins, J.-H. Weitkamp, L. Grunwald, A. R. Stark, J. Reese, W. Walsh, and J. L. Wynn, "Heart rate characteristic index monitoring for bloodstream infection in an NICU: a 3-year experience," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, p. fetalneonatal–2015, 2015.