ECRONICON OPEN ACCESS

EC PAEDIATRICS
Research Article

Using Temporal Data Mining on Patient Data for Clinical Decision Making in the Care of the Sick Newborn

Sidhartha Tan[1]* and KP Unnikrishnan[2]

[1]*Department of Pediatrics, Wayne State University School of Medicine, Detroit, MI, USA*
[2]*eNeuroLearn, Ann Arbor, MI USA*

***Corresponding Author:** Sidhartha Tan, Department of Pediatrics, Wayne State University School of Medicine, Detroit, MI, USA.

Received: April 18, 2022; **Published:** April 28, 2022

Abstract

Background: In a neonatal intensive care unit, streaming healthcare data comes from many sources, but humans are unable to understand relationships between data variables. Data mining and analysis are just beginning to get utilized in critical care. We present a case study using electronic medical record data in the neonatal intensive care unit and explore possible avenues of advancement using temporal data analytics.

Case Presentation: Electronic medical record data were collected for physiological monitor data. Heart rate, respiratory rate, oxygen saturation and temperature data were retrospectively analyzed by temporal data mining. Three premature babies were selected and data de-identified. The first case of a urinary tract infection showed nursing ability to synthesize data streams coming from a patient. For the second case of necrotizing enterocolitis, Temporal-Data-Mining analysis of combinations of clinical events based on deviations from the mean showed specific heuristic biomarkers related to events before discovery of necrotizing enterocolitis. Specific sequences 6-event and 5-event in length were identified with nursing unease at clinical deterioration, which were 100- and 87-times unlikely to occur randomly with 99.5% confidence. No such sequences were found in the rest of the 37 days for the second case and entire 133 days of stay in the third case of an uneventful premature baby.

Conclusion: Temporal data mining is a possible clinical tool in providing useful information in the neonatal intensive care unit for diagnosis of adverse clinical occurrences such as necrotizing enterocolitis. There is the possibility of changing the clinical paradigm of episodic watchfulness to constant vigilance using real-time data gathering.

Keywords: *Critical Care; Computer Data Processing; Data Mining; Nursing Informatics*

Abbreviations

EMR: Electronic Medical Record; HMM: Hidden Markov Model; NEC: Necrotizing Enterocolitis; NICU: Neonatal Intensive Care Unit; TDM: Temporal Data Mining

Background

Early identification of clinical problems in the neonatal intensive care unit (NICU) has the most beneficial impact on patient outcome, much more so than in older children, adults, and the elderly. Death or major morbidity affects 49 - 60% of babies in the NICU of developed countries [1], while 13% of NICU premature babies die [1]. There are many diseases that occur without much warning, some of them being fatal, such as necrotizing enterocolitis (NEC), a condition that is akin to intestinal gangrene. NEC occurs in 7% of premature babies < 1500g birth weight and causes death in 20 - 30% [2]. There are presently no methods to identify or even a priori stratify as high-risk, premature newborns who will get NEC. Care of the sick newborn in a NICU is dependent on periodic evaluation of healthcare personnel with the maximum frequency of assessment being every hour; nurses usually conduct examinations every 4-hours and physicians even less frequently.

Newborn babies cannot talk making the clinical care especially demanding. A clinician needs to develop a comprehensive understanding of the sick deteriorating newborn quickly as the patient's condition can change rapidly. Yet, the concept of collecting and analyzing all streaming data in the NICU is still in its infancy. Healthcare personnel seldom utilize real-time data streaming from various sensors in their decision making. Occasional numbers are heeded but a lot of data is wasted, as in adult intensive care units [3].

Three broad categories of data exist in the NICU; the first being patient description, the second being streaming data coming from monitors and medical equipment, and third being textual description of the patient's condition. Data from physiological monitors is difficult to comprehend by the human mind especially the relationships between two or more variables. To gain insight, data mining is helpful to support decision making [4].

Temporal data mining (TDM) discovers implicit, previously unknown and potential useful sequential information from time-stamped and sequential data. The sequential information discovered by TDM can be represented as hidden Markov models (HMM) [5], which has been used previously in neuroscience [6-8], and public health [9]. In an HMM framework, the unobservable or hidden states of a NICU patient can be thought to evolve towards some final endpoint of interest, such as necrotizing enterocolitis (NEC). This concept of state-space evolution has been introduced before for the newborn as in a state of "physiological immaturity" [10]. TDM, however, takes into account the evolutionary nature of the process, non-linearity of biological pathways, and naturally captures the time-evolution of the system. It allows the discovery of unique patterns of non-trivial events that do not occur in controls or during a period of time remote from the diagnosis in the same patient and gets rid of false positives by focusing on specific event-combinations occurring close in time to the actual diagnosis of disease.

Objective of the Study

Our objective was to test the feasibility of using temporal data mining as a tool to identify early biomarkers of necrotizing enterocolitis in the NICU.

Case Presentation

Three cases of premature babies were retrospectively analyzed as a proof-of-principle investigation for nursing interaction with cardiorespiratory and temperature monitors using electronic medical record (EMR) data at Evanston Hospital, NorthShore University Health System, Evanston, IL, which formerly employed both authors.

Ethics, consent and permissions: The hospital's institutional review board gave an exemption approval because of the retrospective anonymous single case review of the collected data.

Heart rate, respiratory rate and oxygen saturation were automatically entered into the EMR with a single click by the nurse. Temperature was manually entered. The definitions used in for analysis of second and third cases are given below:

1. Readings: Heart rate, respiratory rate, oxygen saturation and temperature were downloaded to an anonymous file and analyzed. Hourly readings were categorized into variables.

2. Variables: The mean and standard deviation (S.D.) of the time epoch were calculated.

3. Events: Variables had two possible events, one that exceeded 1 S.D. ("Hi") or one below 1 S.D. ("Lo"). We chose one S.D. based on the premise that early changes would occur with only one S.D. change. Note that such a strategy is different from traditional statistics where possibly 95% confidence intervals could have been used. Events were then analyzed by TDM software, custom written by an author (KPU), to investigate for sequences of events. The data was viewed by TDM as a single long sequence of ordered pairs (E_i, t_i), where E_i denotes the event type and t_i is the time of its occurrence. Sequences are ordered temporal collections of event types. For example, HR-Lo→O_2Sat-Lo→RR-Hi is a 3-event long sequence where an event type HR-Lo is followed some time later by O_2Sat-Lo and then RR-Hi, in that order.

4. Sequences: Events occurring together or immediately following one another. For example, in the data sequence [(A, 1), (D, 2), (E, 4), (B, 5), (D, 6), (C, 10)], the sequence (A→B→C) occurs once. A sequence is considered interesting if it occurs "often enough" in the data. In the discovery of all sequences that occur often, a frequency measure for sequences is defined. The task is to find all sequences whose frequencies exceed a user-defined threshold. For modeling frequent sequences, specialized Hidden Markov Models (HMMs) called Episode Generating HMMs are used to rigorously relate frequent sequence discovery with learning generative models. The likelihood a sequence did not happen by chance is given by the equation [5]:

   $\boldsymbol{\Gamma = T/M + \{(T/}M)(1-1/M)\}^{1/2}\, \Phi^{-1}(1-\varepsilon)$, where Φ is the cumulative distribution function of a normal random variable and ε the type-II (false positive) error we are willing to tolerate. T is the total no: of events in the data, M is the number of channels. A good approximation for Γ is (N x f), where N is the length of the sequence (in figure 1, it would be 6) and f is its frequency. For reasonable values of Φ and ε, if $f > T/N(M+1)$, that sequence is not by chance, with 99.5% confidence. The more the events in a sequence, the more the likelihood of reflecting a hidden state of the baby, and the less likely that was due to chance alone. In our data, sequences of 6 events are about 100 times unlikely to occur randomly, with 99.5% confidence, and sequences of 5 events are 87 times unlikely to occur randomly with 99.5% confidence.

5. Occurrence: An occurrence was a clinical sign or clinical opinion occurring during the hospital course such as NEC.

Results

The first case was a case of positive urine culture of *E. coli* in the urine and urinalysis findings of white cells and positive nitrite. This patient was discovered after work up resulting from the nurse's judgment of something happening. In this case, the parents were the first to bring the baby to the attention of the nurse for tachypnea and use of accessory muscles of respiration. The subsequent nursing note is illustrative of the clinical acumen of an experienced nurse. "Junior had a desat/decel episode (desaturation/deceleration) following nasal suctioning. After recovery I replaced ngt (nasogastric tube) into R. nare and he had another brief a/b (apnea and bradycardia spell, which he had not had before with similar placement). Over the course of the evening he had numerous decels to 100 - 110s with a resting heart rate in the 120s (his usual is the 140s). His sats (oxygen saturation) would also drift down to 88 - 90. None of these episodes were significant enough to set off the monitor, I was just observing him closely and noticed this. Septic w/u (workup) performed" (italics are from authors). The septic workup included blood and urine cultures and the urine culture was positive. What was remarkable for this patient was that no alarms went off and no desaturation or apnea episodes were recorded for that entire day. Figure 1 shows the graphical trend of the heart rate monitor.

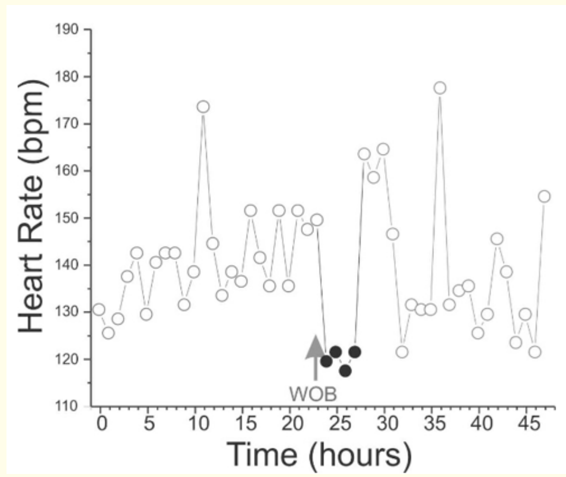

Figure 1: *Arrow shows patient having increased work of breathing (WOB).*

The average and variability of the heart rate pattern in the previous 24 hours was not helpful in identifying deterioration. The notable point was that had the nurse not been observing the patient, no alarms would have gone off and the urinary tract infection would have not been detected till complications of systemic sepsis occurred. The combination of respiratory pattern, oxygen saturation and heart rate changes that constituted the nurse's decision to seek a septic workup piqued our interest. It was felt reasonable to posit that every baby who has a clinical deterioration probably manifests early signs that could only be detected by a discerning eye. We wondered if we could use combinatorial data analytics to identify patients that were about to become sick.

The second case was a stable ventilated patient being fed with full feeds. This baby was discovered to have abdominal discoloration, and abdominal distention on a routine examination by the nurse. The onset was deemed to have occurred suddenly as the previous examination four hours before was normal. The baby developed early shock, and findings of pneumatosis on abdominal X-ray were found, confirming a diagnosis of necrotizing enterocolitis (NEC). Because of the sudden onset of signs of NEC, we retrospectively went back into the chart to find earlier warning signs. In this patient, only two notable clinical occurrences happened before the diagnosis of NEC: 1) 21 hours before, on evaluation of the baby, an experienced nurse felt uncomfortable but did not think her uneasiness was enough to warrant bringing attention of the doctor. She felt "Queasy" and this quote to one of the authors, is used to delineate this occurrence. The difference between this level of discomfort in this nurse and the nurse in case 1 was a judgement made on the vital signs not being as obvious. 2) 42 hours prior, redness was noted by the nurse in the peri-umbilical area of the abdomen, subsequently dismissed by the physician-on-call as inconsequential. We retrospectively investigated 4 data streams of temperature, respiratory rate, oxygen saturation and heart rate data, collected every hour, and divided the events into 8 variables based on 1 standard deviation above or below the mean (See figure 2).

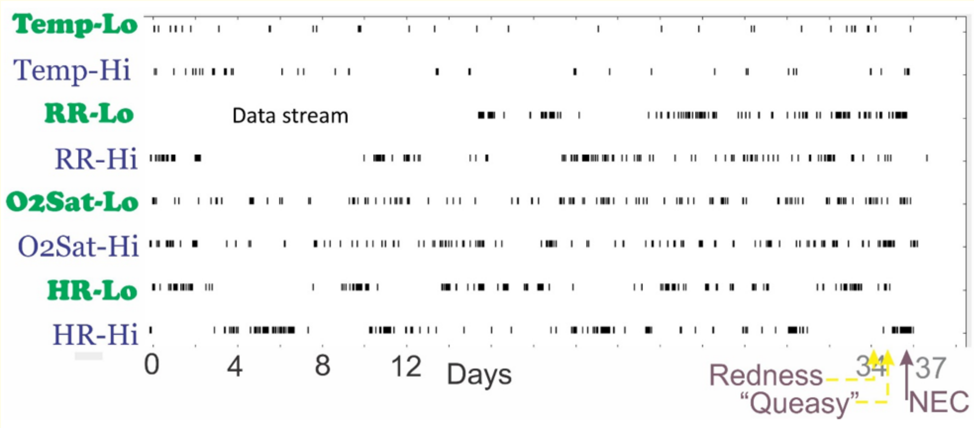

Figure 2: *Eight classes of events: Temperature (Temp), respiratory rate (RR), oxygen saturation (O2Sat) and Heart rate (HR), collected at hourly precision; All events which were either one standard deviation below, "Lo", or above, "Hi", the mean was marked as one tick. Figure showing all ticks occurring during entire hospital course (37.6 days). NEC occurred at the end and the two nursing events chronicled "Redness" and "Queasy" are shown with yellow arrows.*

The entire hospital data set was investigated with TDM software in 5-day blocks. An assumption was made that the direct antecedents of NEC were not likely to occur more than 5 days prior to diagnosis. This assumption is reasonable in clinical practice.

Significant sequences of events were identified by TDM, which would not occur randomly. The longest significant sequences of events discovered through TDM each contained 6 events in the five days prior to NEC diagnosis. What is remarkable that all 15 such sequences, 6 events long, occurred only in the period when the nurse felt "Queasy" about the baby, not at the time or redness or after diagnosis of NEC (Figure 3).

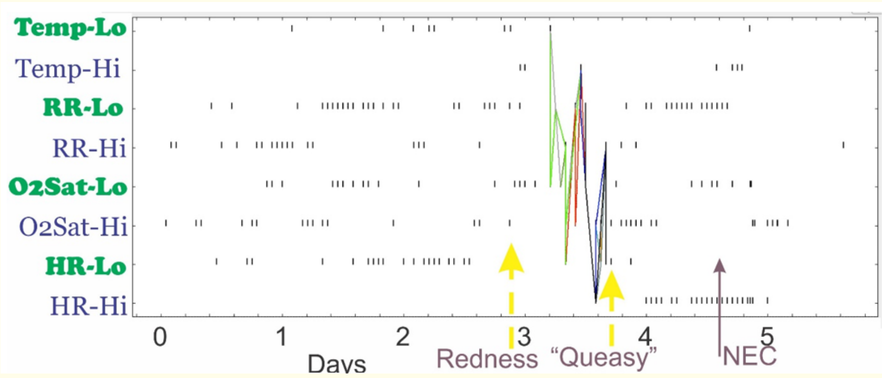

Figure 3: *This figure now narrows the window of analysis to just 5 days prior to onset of NEC. Each tick is either one standard deviation below or above the mean as shown in figure 2. NEC occurred at the end and the two nursing events chronicled "Redness" and "Queasy" are shown with yellow arrows. The sequences of 6 events long were 15 in number and all occurred at about the time of "Queasy". How these sequences are obtained is as follows: The data can be viewed as a single long sequence of ordered pairs (Ei, ti), where Ei denotes the event type and ti is the time of its occurrence. The temporal patterns referred to as sequences in the manuscript are ordered collections of event types. In the data shown above, HR-Lo→O2Sat-Lo→RR-Hi is a 3-event long sequence where an event type HR-Lo is followed some time later by O2Sat-Lo and then RR-Hi, in that order. Longest significant sequences are plotted, with each sequence in a different color.*

The combinatorial occurrences would support the nursing judgment of something happening. The obvious questions that arise from this coincidence are whether these sequences are i) only a deviation from an ongoing stable pattern, ii) reflective of some specific clinical syndrome, or iii) predictive of the disease of interest, NEC. TDM also allows one to examine the various combinations in the sequences (data not shown for brevity).

The next longest sequence found contained 5 events. In the five days prior to NEC, most of the significant sequences consisting of 5-events are clustered around time-points for "Redness", "Queasy", and the final NEC diagnosis (See figure 4).

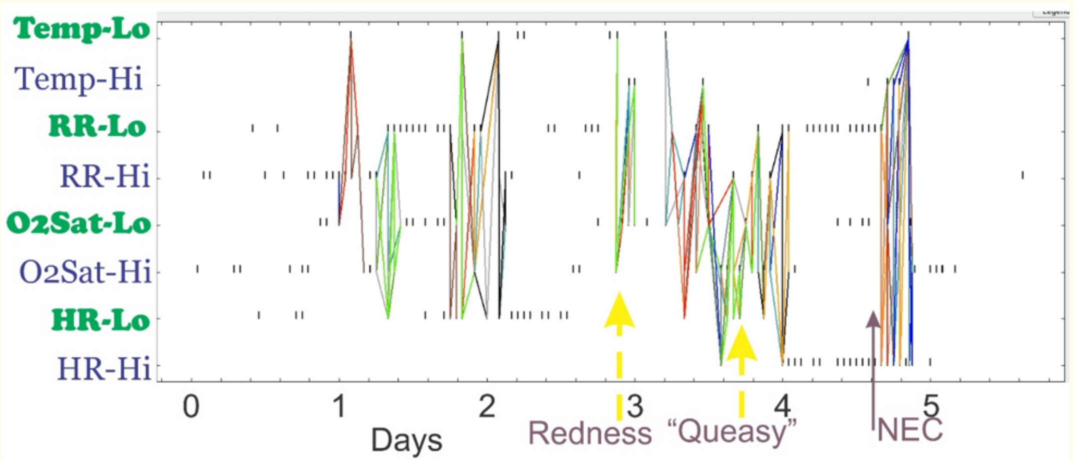

Figure 4: *Figure showing 5 days prior to onset of NEC. Each tick is either one standard deviation below or above the mean as shown in Figure 2. NEC occurred at the end and the two nursing events chronicled "Redness" and "Queasy" are shown with yellow arrows. The sequences of 5 events long are now shown. Clusters are now seen before, just at "Redness", "Queasy" and after NEC was diagnosed.*

Twenty-two significant sequences of 5-events occurred just after the X-ray diagnosis of NEC. Investigation of the sequences reveals an absence of high respiratory rate and low heart rate in the 5-event sequences near NEC, which was not the case with the 6-event sequences near "Queasy".

A way to look at significance with TDM is to generate random sequences based on the range of values found in human newborn patients. The possibility of even three-event sequences is so remote that one can be confident that 5- or 6- event sequences are not false positives. The 6-event sequences at "Queasy" can further be broken down into 5-event sub-sequences, which are different from the 5-event sequences in figure 4. One can see the different pattern of the uni- and bi-directional arrows linking the events in a sequence. When these are plotted across the entire hospital stay, one finds that the 5-event sub-sequences found in figure 3 only occur at 3 other time points starting at 8 days prior to NEC. Developing a cumulative score based on 6-event sequences and 5 event sub-sequences shows the possible identification of a biomarker. Cumulative scores derived from the sub-events of the unique sequences show the predictive value of 6-event sequences near Queasy more so than for the 5-event sequences for NEC (Figure 5).

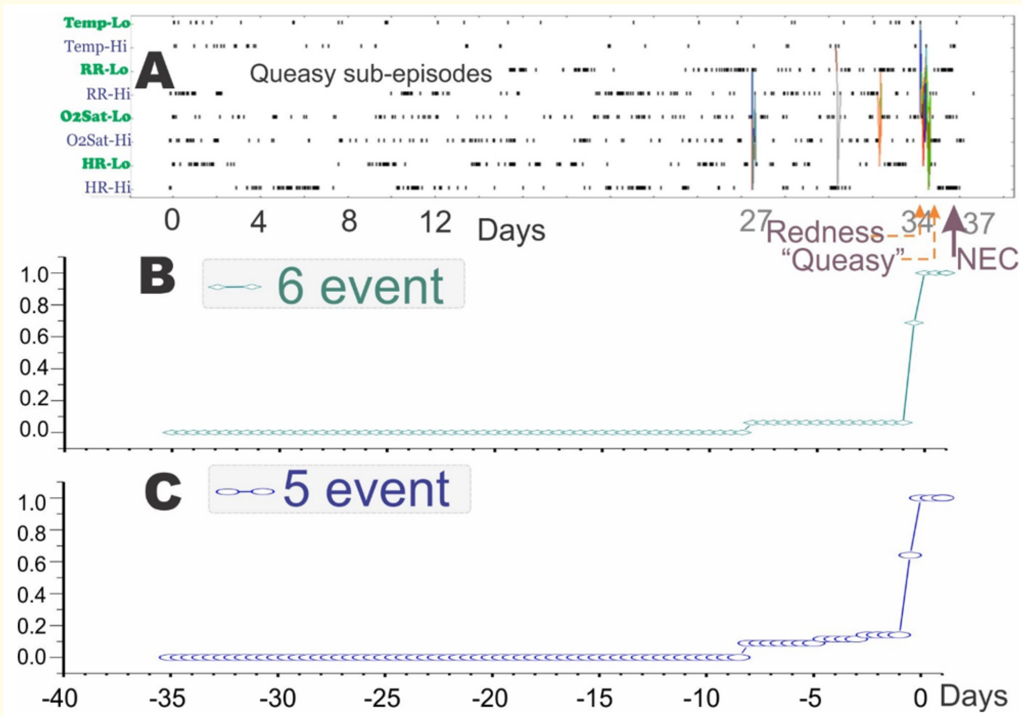

Figure 5: *A. Investigating the entire hospital stay for other events that look like the 6-event sequence shown in figure 3. All the 15 sequences in figure 3 generated 5-event sub-sequences and we investigated if these could be found in any time of the hospital stay. There were only 3 other times that such sub-sequences were found, all of them within approximately one week from the "Queasy" episode, again showing the time-dependent nature of the illness. B. If the cumulative scores of 6-event sequences are plotted one can see that most of the events occur at the "Queasy" period just before NEC. C. For the cumulative scores of 5 event sub-sequences, again numerically one can see that the predominant number of events occurred at the "Queasy" period. This property fits the definition of a biomarker predicting a clinical event.*

There were more of the 4-event sub-sequences early on the hospital course. This would suggest that the 6-event sequences at the time of "Queasy" may actually be a reflection of the appearance of NEC and the 5-event sequences (after the X-ray diagnosis) of consequences of NEC.

Breaking down the entire hospital course prior to NEC into 5-day segments, TDM shows none of the 22 sequences of 5-event near NEC were found elsewhere. There were two times where 7-event sequences were found, one immediately at birth and one 8 days prior to NEC (See figure 5).

The patterns of these sequences were different from each other and to the subsequent sequences of interest around NEC. Except for one sequence that coincided with the later 7-event sequence, the 6-event sequences found near "Queasy" were also unique to that time. At 5 other times, two after birth, and the rest later, 6-event sequences were found but none of these were similar to the sequences at "Queasy". There is thus, a suggestion that certain sequences can be found that are predictive of clinical occurrences.

TDM analysis of 133 days' worth of data from our third case, another premature baby that had only one episode of urinary tract infection (diagnosed by only positive urine culture of *E. coli* without urinalysis findings), showed that the entire hospital course contained no significant 5-event, 6-event, or 7-event long sequences.

Generating a random data set with thousand days' worth of data also did not generate 3-, 4- or higher event sequences, implying that the 6-event sequences may have been unique to that patient of NEC.

Discussion and Conclusions

We present a case study that confirms the validity of nursing clinical acumen using Big Data Analytics. Clinical acumen as defined by us may involve an intuitive understanding of the temporal relationships of various physiological readings in the dynamic assessment of a newborn infant. It is humanly impossible to comprehend the combinatorial relationships of real time data that is being collected. TDM could potentially provide early biomarkers for devastating illness such as NEC. Methods based on TDM are especially useful in n=1 studies in humans as every patient is somewhat unique. This single-case experimental design can be used as empirical evidence because of the time-sensitive nature of the data proximal to the event of interest and reviewed elsewhere [11]. Even though we took the approach of retrospective analysis of EMR data, the implications of our study go beyond just retrospective combinatorial analysis of data that has already occurred. The future would be real-time Big Data Analytics of combinatorial relationships. A recent poll of 20 neonatologists showed that a biomarker needs to satisfy a certain degree of false positives and sensitivity. In future, with Big Data analysis of real-time data of millisecond frequency, the experienced nurse in our case could have insisted on earlier septic work up and earlier stoppage of feeds, which may possibly have prevented NEC or saved a portion of intestines from cell death. However, more cases and controls need to be investigated before there is wide acceptance of the TDM's predictive capabilities.

Big Data is thought to need four characteristics that can be represented by four keywords starting with the letter "V": "Volume" (huge size of data), "Velocity" (speed at which data is generated), "Variety" (number of sources and attributes), and "Veracity" (reliable data) [12]. A review of Big Data usage and analytics in the NICU has not been published before.

Handling of Big Data has already begun in some NICUs, but analytics of high-sampling rate variables are still rudimentary. Newborn heart rate variability and respiratory rate arrhythmia from 3-minute electrocardiography is somewhat prognostic of Bayley Scales of Mental Development at 8 months but not at 12 months of age [13]. Initial Big Data analytical studies in NICU emphasized one parameter of heart rate. Heart rate variability in a 2-hr recording correlated with pain scores after a surgical procedure [14]. Extubation failure in premature infants was associated with decreased heart rate variability in a 60-minute recording before extubation [15]. Using 25 min recordings, reduced heart rate variability was linked to sepsis and heart rate changes included a decrease in sample entropy, all based on altered sensitivity to adrenergic and cholinergic stimuli resulting in fewer accelerations and exaggerated decelerations [16]. Mathematical algorithms based on heart rate variability, moving point average, sample asymmetry, entropy, percentiles and moments led to an index linked to newborn mortality [17] and necrotizing enterocolitis (NEC) [18]. The index was commercialized, called HeRO (MPSC; Charlottesville, VA) using "heart rate characteristics" (HRC). HeRO targets the need for early identification of sepsis in newborn. This need arises because babies deteriorate rapidly, much more rapidly than older children or adults. The HRC index developed from using HeRO has a score ranging from 1 to 5, 1 being normal and 5 indicating a five-fold increased risk of sepsis in the next 24 hours [16]. An increase of > 1 - 2 points over baseline triggers suspicion of underlying late-onset sepsis [16], but is less useful for early-onset sepsis. In some units, the HRC index is used as a deciding piece for clinical workup of sepsis and antibiotics [19]. For ruling out sepsis, HRC index shows a positive predictive value (PPV) of 48% and negative predictive value (NPV) of 57%, and for gram negative sepsis, HRC > 2 has PPV of 4% and NPV of 96% [20]. The increase from baseline of 1 was often reached by the time of diagnosis of surgical NEC but not for medical NEC, with most of the increase occurring after the diagnosis was made for both types of NEC [18]. Anecdotally, clinical utility is less for new admissions with respiratory distress syndrome, persistent pulmonary hypertension or congenital anomalies requiring surgery, because of high scores. Sick babies with severe lung [21] and brain injury often manifest frequent large increases or spikes of HRC index. The spikes are also observed in 8% of non-septic infants [16]. Greater than 7 days prior to sepsis, spikes are observed in 16 - 17% compared to 22-34%, 2 - 7 days prior to sepsis [16]. Monitoring patients with HeRO reduced relative mortality by 22% and absolute mortality by 2% compared to a non-monitored group in a trial of 3,003 premature infants <1500 gram [22]. The exact mechanism of this improvement is not known as there was no decrease in sepsis, nor an increase in blood cultures, nor days on antibiotics [22]. There was no difference in the length of stay amongst all survivors [23].

The disadvantages of only looking at heart rate is that heart-rate variability is affected by hypoxic-ischemic encephalopathy (HIE). Low heart rate variability alone shows a better association with death and EEG/MRI severity in HIE than high HRC index [24]. Variability alone is associated with 2-year neurodevelopmental outcome, with NPV being higher than PPV at 24 hours (90 vs 53%), but reversed at 48 hours of life (63 vs 100%) [25]. Heart rate variability is also decreased with hydrocephalus [26]. Variability in low and high frequency is affected by right and left brain injury respectively [27]. Heart rate variability also changes with sleep [28], massage [29] or kangaroo care [30], and matures with age during sleep in term infants [31]. Cardiorespiratory interaction has been shown [32], is intermittent and increases with age [33]. Thus, it is not surprising that in a study of 2384 NICU infants over 3 years, significant elevations of HRC index by HeRO (≥ 5) were most often not attributable to documented blood stream infection > 3 days of life, suggesting that the HRC index is neither very sensitive nor very specific for late onset sepsis [34].

The use of real time Big Data analytics in the NICU may change the culture of the front-line caretakers into looking at a patient in a more dynamic, more comprehensive and timely manner. Presently the model for patient care in the NICU is episodic care and assessment. Nurses recorded time at the maximum in hourly intervals, with the standard of care being examination every 4 hours. Many of the entries were rounded to the nearest hour. To change the perspective of looking at data from episodic intervals to real-time would require a paradigm shift in outlook of caretakers. This process from evaluating at 4-hourly, hourly, minute or second to millisecond resolution is going to be the future way of taking care of NICU babies. Analysis of real-time data will allow hidden associations to appear that may not only be predictive of adverse events but may shed light on pathophysiological streams. These are likely to become quantitative biomarkers, or at the very least help the frontline caretakers augment their clinical judgment.

Novel approaches to comprehensive hemodynamic monitoring and data acquisition may help in better understanding of physiology in neonates [35]. Looking at more physiological parameters, a study investigated 138 preterm neonates that were 34 weeks gestational age or less and < 2000g in weight, and incorporated in a scoring algorithm, heart rate, respiratory rate and oxygen saturation characteristics in the first 3 hours after delivery [36]. This score predicted high morbidity defined by moderate or severe bronchopulmonary dysplasia, > stage 2 retinopathy of prematurity, > grade 3 intraventricular hemorrhage, NEC or death with 86% sensitivity and 96% specificity [36]. These studies underscore the importance of looking at more than one variable (hence Big Data), not being restricted to a commercially secret mathematical algorithm, and taking into context the condition and time of data-gathering.

Data mining approaches are different from traditional statistical approaches [4]. A simple way to look at data is to take descriptive statistics [37] such as standard deviation and variability [25]. Combining variability with other statistical elements forms the HeRO HRC index, but the mathematical algorithm is fixed once you buy the commercial device. Temporal data mining takes into account the fact that intestinal death in NEC is affected by events that are close to the time more so than the events that are far away in time. The only variables that are far away in time influencing the final event of intestinal death are prematurity and antibiotic usage. Linear models assume binary outcomes that may not be true, where outcomes are either present or absent. NEC, for instance, is affected by ischemia that is in turn affected by hypoxia and the period(s) under hypoxia may have aggregate, evolving, and cumulative effect, which would be impossible to tease out using linear models that do not explicitly account for time. Other methods, such as clustering of data, a form of descriptive data mining, has also been used in adult ICUs [38]. Various mathematical algorithms involving decision tree and other optimization formulas have been used to predict jaundice in newborns [39].

TDM looks at time, in the analysis of databases. TDM discovers useful sequential information from time-stamped data taking into account the time-sensitive nature of pathophysiological processes, allowing for non-linearity and time-evolution of biological pathways. The ability to distinguish between trivial and non-trivial nature of events is a distinguishing feature of TDM. Events remote in time to the clinical occurrences are discounted and the sequential nature of events are considered unique biomarkers, thus getting rid of false positivity of biomarkers.

Similar to the HeRO score, the degree of alarm may be gleaned from the sequences that TDM generates. We can get an idea of the chronological order of the events by TDM. The type of events in each sequence are different between the 6-event and 5-event sequences. Looking at the 7-event sequences, the one at birth is much different from the one at 8 days before NEC. Each sequence possibly has a signature. The uniqueness of the sequence to a hidden state of the baby will be enhanced by real-time data collection of many more variables and the number of events in the signature. It is speculated that the more the variables and the more granular the data (as in millisecond frequency collection), the more likelihood that we will find a multi-event sequence that is unique to a particular hidden state of the baby. Furthermore, real time data collection will overcome shortcomings of hourly collection of physiological data by nurses, as there is always some heuristics in the need to express the overall course of the patient in that hour of observation, making it somewhat qualitative data. Our study did not have that much of a bias as the physiological data was entered into the electronic medical record as a package electronically and the nurses were unlikely to manually change the data. This problem of selection of data from streaming big data exists with all medical record keeping, especially those entered manually on paper.

In summary, real-time Big Data usage is its infancy in the NICU. Our case study suggests Big Data analytics in the NICU could in future provide smart and quantitative tools to augment clinical acumen. More research needs to be done for developing biomarkers for early diagnosis and prediction of adverse events in the NICU. Implementing a culture of Big Data processing and analytics in the NICU or other intensive care units will empower caretakers in changing the clinical paradigm of care from that of episodic observations to real-time continuous observations.

Acknowledgements

NIH R01NS117146 (ST) and R01NS114972 (ST).

Ethics Approval and Consent to Participate

Use of de-identified data in a single case of a disease was deemed to be IRB exempt from informed consent.

Consent to Publish

N/A because of de-identified data.

Availability of Data and Materials

Data N/A (not available) publicly. TDM mining software is publicly available at github.com/patnaikd/tdminer.

Bibliography

1. Horbar JD., *et al*. "Mortality and neonatal morbidity among infants 501 to 1500 grams from 2000 to 2009". *Pediatrics* 129.6 (2012): 1019-1026.

2. Neu J and Walker WA. "Necrotizing enterocolitis". *New England Journal of Medicine* 364.3 (2011): 255-264.

3. Celi LA., *et al*. ""Big data" in the intensive care unit. Closing the data loop". *American Journal of Respiratory and Critical Care Medicine* 187.11 (2013): 1157-1160.

4. Yoo I., *et al*. "Data mining in healthcare and biomedicine: a survey of the literature". *Journal of Medical Systems* 36.4 (2012): 2431-2448.

5. Laxman S., *et al*. "Discovering frequent episodes and learning hidden Markov models: a formal connection". *IEEE Transactions on Knowledge and Data Engineering* 17.11 (2005): 1505-1517.

6. Diekman CO., *et al*. "Statistical significance of sequential firing patterns in multi-neuronal spike trains". *Journal of Neuroscience Methods* 182.2 (2009): 279-284.

7. Sastry PS and Unnikrishnan KP. "Conditional probability-based significance tests for sequential patterns in multineuronal spike trains". *Neural Computation* 22.4 (2010): 1025-1059.

8. Diekman C., *et al*. "Discovering functional neuronal connectivity from serial patterns in spike train data". *Neural Computation* 26.7 (2014): 1263-1297.

9. Unnikrishnan KP., *et al*. "Spatio-temporal Structure of US Critical Care Transfer Network". *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science* 2011 (2011): 74-78.

10. Juve-Udina ME., *et al*. "Newborn physiological immaturity: a concept analysis". *Advances in Neonatal Care: Official Journal of the National Association of Neonatal Nurses* 15.2 (2015): 86-93.

11. Perdices M and Tate RL. "Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued?" *Neuropsychological Rehabilitation* 19.6 (2009): 904-927.

12. Wong HT., *et al*. "The Need for a Definition of Big Data for Nursing Science: A Case Study of Disaster Preparedness". *International Journal of Environmental Research and Public Health* 13.10 (2016): 1015.

13. Fox NA and Porges SW. "The relation between neonatal heart period patterns and developmental outcome". *Child Development* 56.1 (1985): 28-37.

14. Faye PM., *et al*. "Newborn infant pain assessment using heart rate variability analysis". *The Clinical Journal of Pain* 26.9 (2010): 777-782.

15. Kaczmarek J., *et al*. "Heart rate variability and extubation readiness in extremely preterm infants". *Neonatology* 104.1 (2013): 42-48.

16. Fairchild KD., *et al*. "Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial". *Pediatric Research* 74.5 (2013): 570-575.

17. Griffin MP., *et al*. "Abnormal heart rate characteristics are associated with neonatal mortality". *Pediatric Research* 55.5 (2004): 782-788.

18. Stone ML., *et al*. "Abnormal heart rate characteristics before clinical diagnosis of necrotizing enterocolitis". *Journal of Perinatology* 33.11 (2013): 847-850.

19. Hicks JH and Fairchild KD. "Heart rate characteristics in the NICU: what nurses need to know". *Advances in Neonatal Care: Official Journal of the National Association of Neonatal Nurses* 13.6 (2013): 396-401.

20. Raynor LL., *et al*. "Cytokine screening identifies NICU patients with Gram-negative bacteremia". *Pediatric Research* 71.3 (2012): 261-266.

21. Sullivan BA., *et al*. "Infection and other clinical correlates of abnormal heart rate characteristics in preterm infants". *Journal of Pediatrics* 164.4 (2014): 775-780.

22. Moorman JR., *et al*. "Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial". *Journal of Pediatrics* 159.6 (2011): 900-906.e901.

23. Swanson JR., *et al*. "Neonatal Intensive Care Unit Length of Stay Reduction by Heart Rate Characteristics Monitoring". *Journal of Pediatrics* 198 (2018): 162-167.

24. Vergales BD., *et al*. "Depressed heart rate variability is associated with abnormal EEG, MRI, and death in neonates with hypoxic ischemic encephalopathy". *American Journal of Perinatology* 31.10 (2014): 855-862.

25. Goulding RM., *et al*. "Heart rate variability in hypoxic ischemic encephalopathy: correlation with EEG grade and 2-y neurodevelopmental outcome". *Pediatric Research* 77.5 (2015): 681-687.

26. Uhrikova Z., *et al*. "Changes in heart rate variability in a premature infant with hydrocephalus". *AJP Reports* 2.1 (2012): 43-46.

27. Schneebaum Sender N., *et al*. "Effects of regional brain injury on the newborn autonomic nervous system". *Early Human Development* 90.12 (2014): 893-896.

28. Smith SL., *et al*. "Heart rate variability during caregiving and sleep after massage therapy in preterm infants". *Early Human Development* 89.8 (2013): 525-529.

29. Smith SL., *et al*. "The effect of massage on heart rate variability in preterm infants". *Journal of Perinatology* 33.1 (2013): 59-64.

30. Cong X., *et al*. "Kangaroo care and behavioral and physiologic pain responses in very-low-birth-weight twins: a case study". *Pain Management Nursing: Official Journal of the American Society of Pain Management Nurses* 13.3 (2012): 127-138.

31. Yiallourou SR., *et al*. "Maturation of heart rate and blood pressure variability during sleep in term-born infants". *Sleep* 35.2 (2012): 177-186.

32. Peters CH., *et al*. "A continuous wavelet transform-based method for time-frequency analysis of artefact-corrected heart rate variability data". *Physiological Measurement* 32.10 (2011): 1517-1527.

33. Clark MT., *et al*. "Breath-by-breath analysis of cardiorespiratory interaction for quantifying developmental maturity in premature infants". *Journal of Applied Physiology (1985)* 112.5 (2012): 859-867.

34. Coggins SA., *et al*. "Heart rate characteristic index monitoring for bloodstream infection in an NICU: a 3-year experience". *Archives of Disease in Childhood Fetal and Neonatal Edition* 101.4 (2016): F329-332.

35. Soleymani S., *et al*. "Neonatal hemodynamics: monitoring, data acquisition and analysis". *Expert Review of Medical Devices* 9.5 (2012): 501-511.

36. Saria S., *et al*. "Integration of early physiological responses predicts later illness severity in preterm infants". *Science Translational Medicine* 2.48 (2010): 48ra65.

37. Saeed M., *et al*. "Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database". *Critical Care Medicine* 39.5 (2011): 952-960.

38. Joshi R and Szolovits P. "Prognostic physiology: modeling patient severity in Intensive Care Units using radial domain folding". *AMIA Annual Symposium Proceedings/AMIA Symposium AMIA Symposium* 2012 (2012): 1276-1283.

39. Ferreira D., *et al*. "Applying data mining techniques to improve diagnosis in neonatal jaundice". *BMC Medical Informatics and Decision Making* 12 (2012): 143.

Volume 11 Issue 5 May 2022
© All rights reserved by Sidhartha Tan and KP Unnikrishnan.