Predicting Adverse Outcomes in Preterm Infants Using Early Bedside Monitor Data

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Physics from the College of William and Mary

Han Truong

Advisor: Dr. John Delos
Senior Research Coordinator: Dr. Irina Novikova

May 2, 2018

Abstract

We use bedside monitor data from the UVA NICU to extract clinical episodes from preterm infants that are detectable through analyzing the signals' numerical values. The three types of clinical episodes we focus on are bradycardia, oxygen desaturation, and central apnea. The primary goal is to improve prediction of bronchopulmonary dysplasia (BPD), a chronic lung disease that preterm infants commonly develop after birth due to aberrant oxygenation and immature lungs. A literature review reveals that birth weight, gestational age, and vent days have high predictive power. We strive to improve prediction further by adding clinical episodes to the mix. In a related study at UVA Medical School, the authors considered desaturation and bradycardia episodes and found that measures of desaturation improve BPD prediction. They defined a baseline clinical model that included BW, GA, and vent days and discovered that adding in desaturation but not bradycardia significantly increases the model's accuracy. In this work, we verify, and in one case, correct their results, and we incorporate statistics about apnea into the baseline model used by UVA and scan for statistically useful parameters. Our secondary goal is to predict other outcomes that are linked to low oxygen levels like IVH, ROP and prolonged-NICU stay using the same models. In agreement with UVA, we found desaturation to be useful when incoporated into the baseline model for predicting primary outcome. In addition, desaturations can be used to predict 2 secondary outcomes: ROP and prolonged stay. Brady episodes did not add to the clinical baseline model for any outcome. Apnea episodes were useful in predicting prolonged stay but not primary outcome.

Contents

1 Introduction **2**
1.1 Background . . . 2

2 Hypothesis **2**
2.1 Data collection and episodes detection . . . 3
2.2 Population of Study . . . 4
2.3 Missing Data . . . 6
2.4 Goal of Study . . . 7

3 Methodology **8**
3.1 Literature Review . . . 8
3.2 Code Library to Manage Infants and Their Clinical Episodes . . . 8
3.3 Data cleaning and checking results . . . 8
3.4 Mechanical Ventilation . . . 8
3.5 Predictive Modeling . . . 9

4 Results **9**
4.1 Average Number of Events per Day . . . 9
4.2 Transitions around Ventilation . . . 10
4.3 Predictive Modeling . . . 13

5 Discussion **18**
5.1 Central Apnea . . . 18
5.2 Future Research Idea: Time from Last Apnea to Desat . . . 18

6 Conclusion **19**

7 Appendix **20**
7.1 Code Library . . . 20

1 Introduction

1.1 Background

Premature infants are highly vulnerable to mortality and morbidities. Predictive models are necessary to inform doctors the risks these infants might experience. Using bedside monitor signals recorded at the UVA NICU, we developed an automated process to quantify the risk for each baby. We focus our study on a subgroup of premature infants known as very low birth weight babies (VLBW). During their NICU stay, these babies experience a lot of clinical episodes and morbidities. However, the association between clinical episodes and morbidities is still unclear. This project aims to further our understanding of this dynamic between these two variables.

The three type of clinical episodes we focus on studying are bradycardia, oxygen desaturation, and central apnea. Clinicians suspect these episodes to be significant in affecting the baby's health and thus require careful analysis.

Bradycardia is defined as abnormally slow heart rate with rate equal to or below 100 beats per minute.

Hypoxemia or oxygen desaturation is defined as low oxygen level in the blood equal to or below 80 percent saturation of hemoglobin. Very frequent episodes of hypoxemia means the baby is unable to breathe independently. Respiratory support is used to elevate low oxygen levels. There are 3 levels of support, listed in the order of decreasing invasiveness: mechanical ventilation, continuous positive airway pressure (CPAP), and nasal cannula. In the worst case, the baby is put on a mechanical ventilator for assisted breathing. The ventilator delivers air to the baby through an endotracheal tube that is inserted into the infants trachea. Alternatively CPAP is a less invasive method which can be used to assists infants in breathing by applying positive pressure to prevent airway closure [12]. The least invasive level of support is attaching the nasal cannula to the baby's nose. The nasal cannula provides oxygen for the baby but does not provide respiratory assistance [7].

Bradycardia and oxygen desaturation can occur separately or together. A brady and a desat are considered to occur together when both the heart rate and the oxygen level start dropping past their respective threshold levels at a time close to one another. When this happens, it is called a BD episode.

Cessation of breathing or central apnea happens when the respiratory system malfunctions and there is no observable breathing activity. Apnea episodes that last for a long time are harmful and require intervention. Nearly all VLBW infants experience apnea due to inherent physiological immaturity or pathologic processes. Apnea events lasting over 10 to 20 seconds may reflect abnormal physiology, impending illness, alarm failure, and/or failure of bedside clinicians to intervene timely [13]. Central apneas may also be accompanied by bradycardia and/or desaturation to become an AB, AD, or ABD episode respectively.

2 Hypothesis

Of the three events, bradycardia, desaturations and apnea (which can occur together or separately), the most important predictor of poor outcomes is desaturation, or hypoxemia. Apneas may cause desaturations, but it is the lack of oxygen, not the cessation of breathing itself, that leads to damage. Likewise it is hypothesized that slow heart rate by itself is not a primary risk. The primary outcomes on which we focus are bronchopulmonary dysplasia (BPD) and death. Secondary outcomes are intraventricular hemorrhage (IVH), retinopathy of prematurity (ROP) and prolonged NICU stay (length-of-stay, LOS).

BPD is defined for babies born before 32 weeks in the following way: [mild] Breathing room

air at 36 weeks' post-menstrual age; [moderate] need for <30% oxygen at 36 weeks' postmenstrual age; [severe] need for >30% oxygen or CPAP at 36 weeks.

ROP is a vasoproliferative disorder of the retina that results in blindness. A premature birth changes the environment under which the retina develops, leading to 'incompletely vascularized retinas with a peripheral avascular zone' according [4].

According to Ballabh, 'pathogenesis of IVH is multifactorial and is primarily ascribed to a) inherent fragility of the germinal matrix vasculature, b) disturbance in the cerebral blood flow (CBF), and c) platelet and coagulation disorders' [3].

2.1 Data collection and episodes detection

The UVA NICU is a quaternary care facility with 45 beds and over 500 admissions per year. There are roughly 2TB of bedside monitor data collected since 2009-2014 stored in William and Marys Sciclone cluster. The data are organized by bed and day, not by patients. A bed assignment table is used to determine which data segment belongs to which baby.

Collected data include waveforms of 240Hz which are electrocardiogram and pulse oximetry signals as well as chest impedance signal collected at 60Hz. The first two waveforms are then used to calculate heart rate and oxygen saturation of hemoglobin in the blood respectively while chest impedance is used to monitor babys respiration activity.

HR and SpO2 values are calculated at 0.5Hz frequency from the EKG and pulse oximetry signals. The signals are all averages, typically of the previous 8 seconds, and not instantaneous at their time mark [12]. Previously, an automated system was developed to detect apnea episodes. This system calculates the probability that the baby is currently undergoing apnea using chest impedance and EKG waveforms and stores it as a separate time-series 4Hz signal called aprob. A random of sample ABD events detected by this signal were clinically validated by clinicians to be accurate for 90% of the time [13].

Our study focuses on using heart rate (HR), oxygen level (SpO2) and aprob signals to extract the clinical episodes of interest.

Bradycardia starts when HR drops below 100 beats per minute (bpm) and ends right before HR goes above this number. There is a minimum duration threshold of 4 seconds between start time to end time for a drop to be counted as a bradycardia episode. Separate episodes are joined together if HR rose above 100 bpm but dropped below again within 4 seconds.

Similarly, desaturation starts when SpO2 drops below 80% and ends right before it goes above this level. The minimum duration threshold for a drop to count as a desaturation episode is 10 seconds. We join episodes if SpO2 rose above then dropped below 80% within 10 seconds.

Apnea starts when aprob rises above 10% and ends right before it drops below 10%. To minimize collecting false positive episodes, we follow current practices on apnea recording when choosing the minimum duration threshold of 20 seconds [8][5][11]. In addition, the weighted duration over duration ratio has to be 0.8 or above for the event to be counted as an apnea episode. Weighted duration is the area under the apnea probability curve from start of episode to end of episode calculated using Riemann sum method. Figure 2 shows how duration and area under threshold is calculated for one desat episode.

BD episodes occur when bradycardia and desaturation start within 30 seconds of each other. AB/AD episodes occur when apnea starts 60 seconds or less before the start of a bradycardia or saturation respectively. ABD episodes occur when apnea starts 30 seconds or less before the start of BD episodes. Figure 1 shows how an ABD episode look like as collected waveforms.

These minimum duration cutoffs were selected by UVA doctors based on the belief that very brief episodes are less likely to be clinically important.

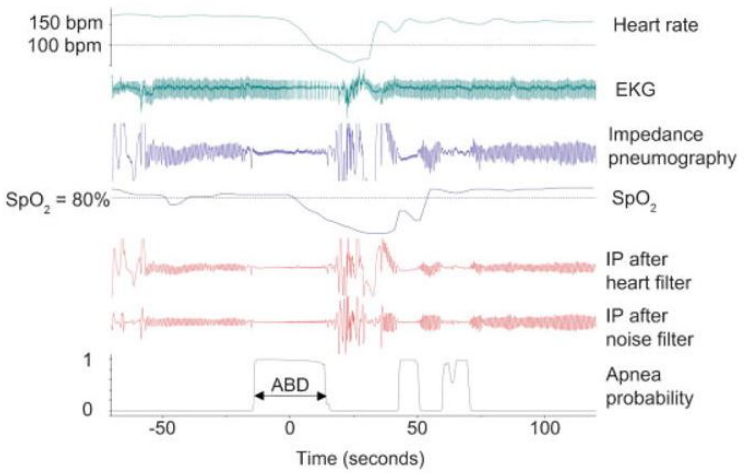

Figure 1: Algorithm-detected episode of central neonatal apnea. This constellation of signals is consistent with an episode of central apnea with bradycardia and desaturation. Time 0 is the center point of the apnea event. (picture and caption extracted from [15]); IP = Chest Impedance

After collecting information about duration and number of clinical episodes, we perform additional analysis to quantify the percentage of total time that a baby spent undergoing all episodes and the area under the critical threshold. For a baby with n desat episodes in the first 28 days after birth:

$$\text{percentage time spent in desat} = \frac{\sum_{i=1}^{n} T_{end}(i) - T_{start}(i) + 2}{\sum_{j=1}^{28} \text{amount of SpO2 data in day } j} = \frac{\sum_{i=1}^{n} duration(i)}{\text{total SpO2 data (secs)}} \quad (1)$$

$$\text{area under threshold per day} = \frac{\sum_{i=1}^{n} 80 * duration(i) - \text{area under SpO2 of desat episode } i}{\text{total amount of SpO2 data (days)}} \quad (2)$$

$$\text{mean area} = \frac{\sum_{i=1}^{n} \text{area under threshold}}{n} \quad (3)$$

$$\text{mean length} = \frac{\sum_{i=1}^{n} duration(i)}{n} \quad (4)$$

$$(5)$$

2.2 Population of Study

Patients being studied are VBLW infants with gestational age ranges from 23 to 33 weeks. Gestational age is measured in weeks from the first day of the mother's last menstrual cycle to the baby's birth date. A normal pregnancy term ranges from 38 to 42 weeks. An infant with a lower gestational age is considered more pre-term. Post-menstrual age (PMA) is also used commonly to describe an infant. It is equal to gestational age plus chronological age. Infants with birth weight of less than 1500 grams fall into the very low birth weight (VLBW) group. About a quarter of all preterm infants that have recorded bedside data are VLBW [10]. A demographic summary of the originally proposed population is listed in Table 1.

We exclude babies with anomalies that is expected to significantly affect breathing behavior. These include cardiac, pulmonary and chromosomal anomalies. In addition, we exclude babies with

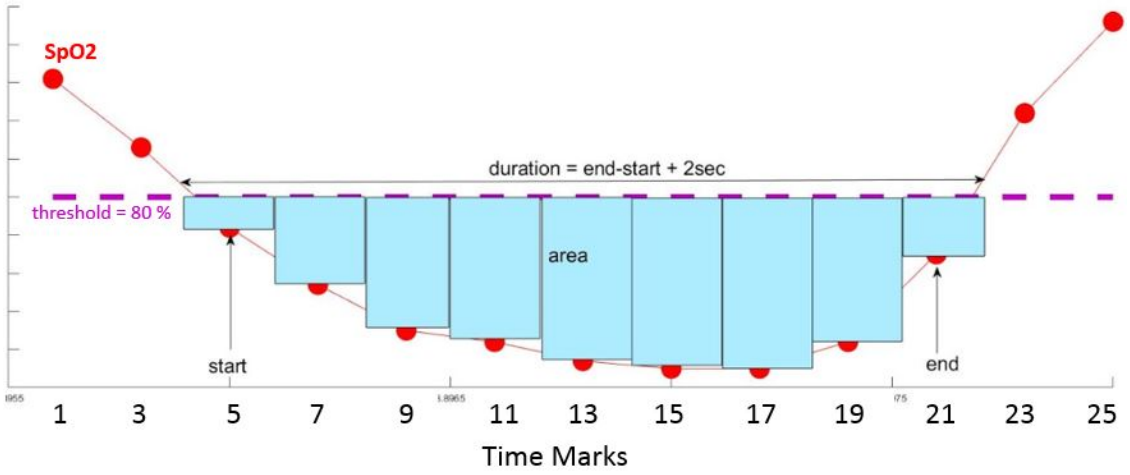

Figure 2: For this desat, each data point has a time width of 2 seconds. Duration is the difference between end and start time mark plus 2 seconds. Area under threshold is the sum of the areas occupied by the blue rectangles pictured

Variables	Median (IQR) or Number of Babies(%)
Sample Size	414 (100%)
GA in weeks	28 (24-32)
BW in grams	1020 (570-1470)
Length of Stay in days	62 (42-93)
Death or BPD	150 (36.23%)
Prolonged Stay (>40w PMA)	55 (13.29%)
severe IVH	32 (7.73%)
severe ROP	42 (10.14%)

Table 1: Demographic and Health Variables for Patient Population (GA: gestional age; BW: birth weight; BPD: bronchopulmonary dysplasia, IVH: interventricular hemorrhage; ROP: retinopathy of prematurity; PMA: post-menstrual age)

less than 7 days of pulse oximetry data available for analysis due to issues regarding missing data described in section 2.3.

In order to maximize lead time for appropriate intervention, we only analyze data collected in the first 28 days of a babys stay in the NICU. The median gestational age for 502 infants originally proposed for our study was 28 weeks. They usually stay in the NICU until they reach 40 weeks post-menstrual age which is around 3 months after birth.

Demographic and health records were provided by doctors from UVA. Available clinical information includes birth weight, gestational age, and time logs for mechanical ventilation support. In addition, we have information about health outcomes of each baby.

We categorize the outcomes into two levels: primary and secondary. The primary outcomes we want to predict are death and bronchopulmonary dysplasia (BPD). BPD is the leading cause of lung disease in US infants. In a longitudinal study from 1997, VLBW infants diagnosed with BPD has poorer developmental outcomes the non-affected infants during their first 3 years of life after controlling for other risks [14]. Babies develop aren't born with BPD but develop it later on. BPD can be mild, moderdate, or severe. The diagnosis depends on how much supplemental oxygen a baby needs when reaching the original due date [1]. Secondary outcomes include morbidities that

might be associated with aberrant oxygenation such as severe intraventricular hemorrhage, treated retinopathy of prematurity, and prolonged length of stay in the NICU. In the medical records, the outcomes were reported in a binary fashion as yes or no.

The seriousness of these outcomes, especially BPD, emphasizes the importance need for an accurate predictive model. Since the intervention treatments are not appropriate for all VBLWs, it is better to identify the highly susceptible babies to allow clinicians more "lead" time for intervention.

2.3 Missing Data

The medical database is not complete and there are random gaps (order of several seconds to hours) of missing time-series data. We do not know the cause behind these gaps. The bed assignment table was recorded by caretakers and are subjected to human error. Whenever there is ambiguity in the bed assignments (e.g. two patients assigned to the same bed at the same time or data for a bed that has no patient assigned to it at that time), this data segment cannot be used for analysis [12]

In the UVA study [7], doctors selected 502 VLBW babies that were patients at the UVA NICU at some point. Scanning through Sciclone reveals that our database is less complete than theirs. 86 babies did not have any bed assignment information. 2 babies have less than 7 days of medical data in their first 28 days. In total, 88 babies were excluded. We narrow down our population to the 414 babies with sufficient data.

The apnea detection system requires that the CI and EKG signals must have sufficient quality. Motional and other artifacts sometimes can "contaminate" these signals and make them uninterpretable by the algorithm [9]. In addition, we are only interested in aprob during no mechanical-ventilation periods (see section 3.4 for explanation). As a consequence, the amount of available aprob data is very limited. The total amount of data we have for all SpO2,HR, and aprob signals is listed in Table 2. Complications arise when we try to collect clinical episodes if there is missing

Signal	Number of Days (% to expectation)
Expected	414·28 = 11592
SpO2	8764.2 (75.6%)
HR	8082.6 (69.7%)
Expected Off-Vent Time	7899.4
Aprob	5876.3 (74.39%)

Table 2: Amount of Available Bedside Data on William and Mary's Sciclone Cluster

data. All babies in our study have missing data within the designated study period. If theres a gap in data before the signal drops below threshold level, we consider the episode to start at this time mark. If the signal is below threshold prior to a data gap, we label this time mark the end of an episode. To handle situations where there is a gap in the middle of an episode, we separate them into 2 cases. If the gap is more than an established gap threshold, the last time mark before the gap is considered the end of the current episode and the first point after the gap is considered the start of a new episode. If the gap is less than or equal to the gap threshold, we ignore the gap and interpolate the signal by averaging the last point before and first point after the gap. This interpolated value is used during calculation for area under threshold.

Different gap thresholds is used for different type of clinical episodes. Table 3 list the values that we chose for our algorithm. These values were established based on previous studies that work with desaturations and apneas [10][9].

Figure 3 shows an example of how the algorithm processes missing SpO2 signal during desats collection. In the first event, missing data ($\leq$ 6 seconds) is ignored and interpolation is used to

Episode Type	Gap Threshold
Apnea	3 secs
Brady	6 secs
Desat	6 secs

Table 3: Missing Data Gap Threshold for Clinical Episodes

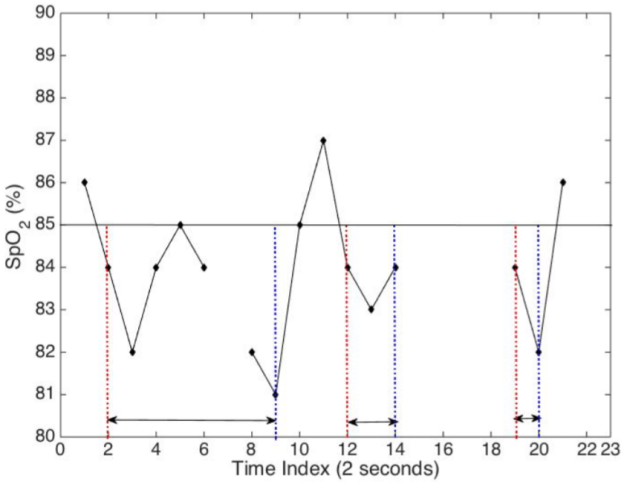

Figure 3: An example of a (fabricated) SpO2 signal for 42 seconds (2 seconds intervals) displaying how data gaps are handled. There are 3 desaturation episodes, with T_{start} and T_{end} shown by red and blue dotted lines respectively.

calculate the area under the threshold of this episode. The second and third episode is separated by a gap ≥ 6 seconds. Figure is extracted from [10], a previous study that attempt to analyze desat episodes for correlation with ROP. The episode collection algorithm from the author's work was reused for this project. However, we choose a threshold value of 80% instead of 85% for desats.

Moreover, 41 babies have no aprob signal during periods of off mechanical ventilation in our database. We choose to only analyze apnea episodes for specific reason mentioned later in section 2.3. During parts that require apnea analysis, the 41 babies with no aprob data are excluded and only 373 babies that meet the stated sufficient data requirement will be studied.

2.4 Goal of Study

With time-series data about heart rate, oxygen level, and apnea probability, we search for any correlation between occurrence of these episodes and the specified adverse outcomes of patients.

We focus on developing a predictive model that can assess each patient in the study for the primary and secondary outcomes mentioned previously.

In addition, we will assess the predictive power of each type of clinical episodes being studied (A, B, and D) as well as their dynamics (BD, ABD, AB, AD) predictive power with regards to patients health outcomes.

3 Methodology

3.1 Literature Review

Previously, the UVA team has analyzed vital signals data for desat and brady episodes to build a predictive model for infants clinical outcomes [7]. We extend their study by introducing another type of clinical episode, central apnea, into the predictive analysis process in order to reconfirm their results and evaluate our hypothesis regarding apnea's predictive usefulness. However, UVAs database is more complete than ours. With the available data on our end, we are only able to conduct our analysis on 414 out of 502 babies that the UVA team used in their study. Nevertheless, we check our results against theirs to externally validate their findings and to determine if the additional analysis we perform will be applicable to their dataset.

Results from UVA provide us with great insights on the variables we should focus on in our study. In agreement with an online BPD risk calculator made by the NICHD (cite), gestational age (GA) and birthweight (BW), the number of days on mechanical ventilation, and Respiratory Acuity Score have high predictive power. When it comes to physiological variables, desaturation significantly increases predictive power when combined with other demographic and respiratory support variables above in a multivariable logistic regression model. In addition, desaturation episodes correlate negatively with birthweight and gestational age. This intuitively makes sense as babies that born closer to normal term have less respiratory issues. In contrast, bradycardia and BD events did not add predictive value.

3.2 Code Library to Manage Infants and Their Clinical Episodes

In total, there were 449427 desats, 86961 brady episodes, and 59509 apnea episodes. The episodes are recorded in a 7-column matrix and organized by bed and time (Figure 11). We use a bed assignment table to link each episode with the baby it belongs to. Due to the large number of events and missing or incorrect bed assignment information, querying episodes is a much more complex of a task then expected. An electronic library was developed for this purpose. Details about the library and example can be found in the Appendix.

3.3 Data cleaning and checking results

Data cleaning was the biggest part of this study. It is very easy to miscount and miscalculate each baby's statistics due to the large number of clinical episodes and complications that arise from missing data. In addition, the ventilation time logs themselves have errors. Some ventilation windows are overlapped each other (Figure 4). We use several testing procedures to check our results for consistency and accuracy.

3.4 Mechanical Ventilation

Since number of days on mechanical ventilation is an important parameter in determining BPD outcome, we strive to quantify the effect mechanical ventilation has on clinical episode occurrences by calculating the number of episodes per day of an average infant. For this part, we focus our attention only on desat and brady episodes. Chest impedance (CI) during mechanical ventilation looks very different from CI patterns of normal breathing. As a consequence, the aprob signal that is generated using CI signal becomes unreliable. Apnea episodes detected during ventilation have a great chance of not being real apnea episodes and are excluded from this part of our study.

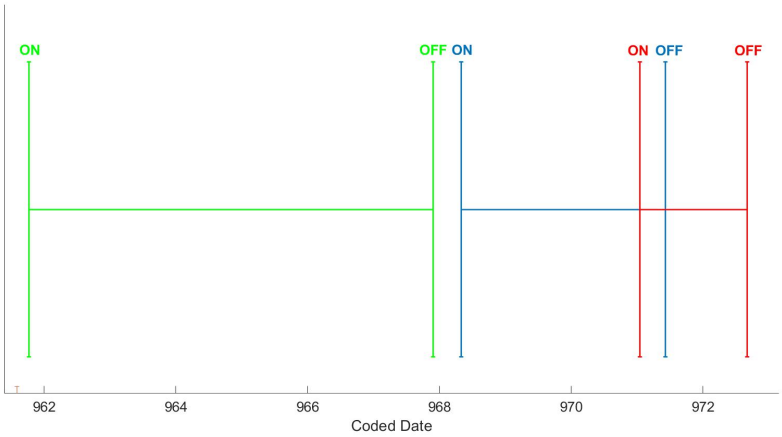

Figure 4: There are 3 ventilation windows in this picture. Each window starts at the ON time mark and ends at the OFF time mark. The last 2 windows, red and blue, overlap each other. When querying for on-vent only clinical episodes, episodes enclosed by the overlapping region is susceptible to being counted twice.

3.5 Predictive Modeling

We develop our statistical models by screening for parameters of interest and use multivariate logistic regression to derive the corresponding model coefficients. Afterwards, we use a receiver operating characteristic curve (ROC) to assess the predictive power of each model.

4 Results

4.1 Average Number of Events per Day

Referring to Table 4, VLBW infants spent a lot more time in desat compared to time spent in brady or apnea. Desat occurrence is the most common at over 50 episodes per day, followed by brady occurrence at around 11 episodes per day, and rare BD occurrence of around 2 episodes per day. Only apnea episodes that occur during non-ventilation periods are counted for as apnea detection becomes inaccurate during mechanical ventilation like discussed in sec 3.4. The occurrence rate was calculated using the formula:

$$\frac{\sum_{b=1}^{414} \text{number of episodes baby b had}}{\sum_{b=1}^{414} \text{amount of data baby had (days)}} \tag{6}$$

Figure 6 suggests desat and BD episodes occurrence increase overtime in the first 4 weeks after birth with rapid increase in the first 3 weeks. Brady occurrence fluctuates but remains relatively stable throughout the first 4 weeks. In non-ventilated time windows, apnea occurrence starts out at an average of 9 episodes per day and decreases gradually over the first week but then rises in the second week and eventually decrease again to 6 episodes per day (Figure 7) by the end of the fourth week. This agrees with previous conceptions that apneas become less frequent as infants are maturing[12]. There are also small fluctuations in AD,AB and ABD occurrence within the order of less than 1 episode per day.

We continue to dissect these occurrence rates further by separating them into categories based on when the babies were on the ventilator or off the ventilator. In their original study, the UVA

Episode Type	Occurrence Rate (episodes per day of data)	Percent Time Spent in All Episodes
Desat	50.91	3.68%
Brady	10.75	0.13%
BD	2.32	NA
Off Vent Only	Off Vent Rate	Percent Off Vent
Apnea	7.29	0.34%
AD	1.56	NA
AB	1.08	NA
ABD	0.54	NA

Table 4: Characteristics of Clinical Episodes in VLBW infants (NA: not applicable)

group made an attempt to do this same task. Their primary results, later removed from the official paper [7], were reported in a Venn diagram shown on the left side of Figure 5. They concluded that episode occurrence was lower during ventilation. We tried to replicate their results on our end but could not do so. Our version of the same Venn diagram is on the right side of Figure 5.

There is a significant disagreement between the two Venn diagrams in Figure 5. Our diagram shows a puzzling result that the episode rate is higher during ventilated periods. In the official paper from UVA [7], there is another plot on episodes progression from first day to the 28th day after birth (left side of Figure 6) which shows episode occurrence is higher during ventilation, totally contradicting the trend in their preliminary Venn diagram but in agreement with our Venn remake. We decided to replicate their progression plot (right side of Figure 6). This time our results agree with theirs as pictured in Figure 6, with episodes occurrence rate being higher during ventilation.

After communicating with the UVA group, both teams agree that UVA's primary Venn diagram is incorrect, which leads to it being removed from the published paper [7]. However, the correct result is alarming: babies have more desaturation episodes during mechanical ventilation which might be interpreted to mean that mechanical ventilation actually causes harm to infants. This goes against general medical understanding. The effect observed here is not due to mechanical ventilation but can be most likely attributed to the fact that the on-ventilator calculation was done on a different population from the off-vent calculation. Only 278 babies out of the 414 in the study were put on mechanical ventilation at some point within their first 28 days after birth. The on-ventilation results represent statistics for only a subgroup of babies being studied while the off-ventilation results represent statistics for all 414 babies. First, babies included in the on-vent analysis have lower birth weight and gestational age as listed in Table 5. In addition, they have naturally higher desat occurrence than the overall population. Calculation of the general episodes rate listed in Table 5 factors in both ventilated and non-ventilated time windows. However, brady and BD occurence are roughly equal in both groups. This makes sense why our Venn diagram on the right in Figure 5 only implies that desat occurs more frequently during on-ventilation out of the 3 episode types. Between the disagreements and the unexpected results on both our end and UVAs end, this is not a good way to study this issue. We need to another way to quantify the effect of mechanical ventilation on clinical episodes occurrence.

4.2 Transitions around Ventilation

We decided to look at the number of brady, desat, and combined BD episodes around the transition between on and off ventilation instead. Patients are put on a mechanical ventilator only when they suffer from severe respiratory issues. By analyzing medical data around the transition period, we can quantify the effect ventilation have on patients physiological conditions more accurately.

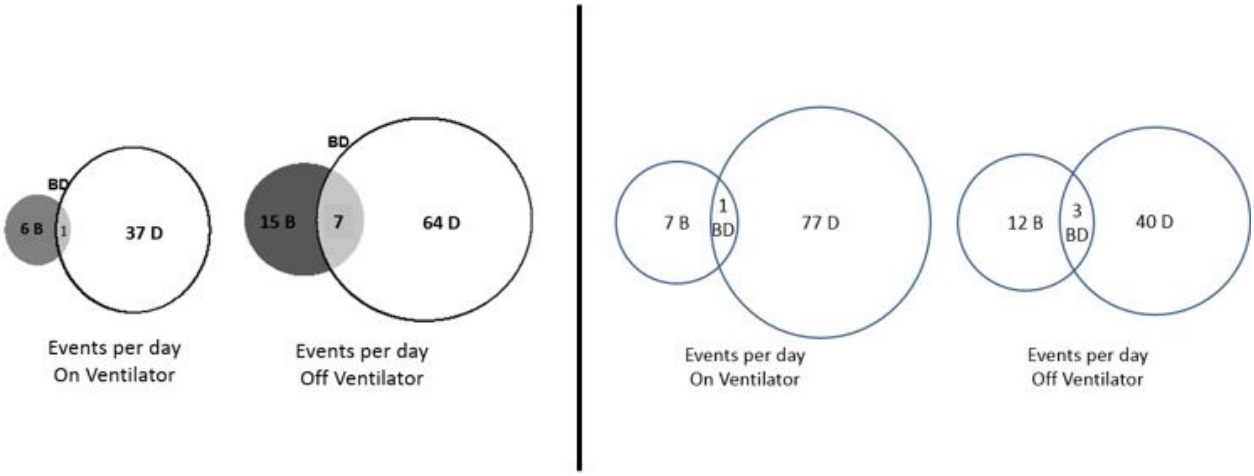

Figure 5: Episodes Venn Diagram - Average Number of Events per Day off Data for On-Ventilation and Off-Ventilation; UVA version is on the left [7] and ours is on the right

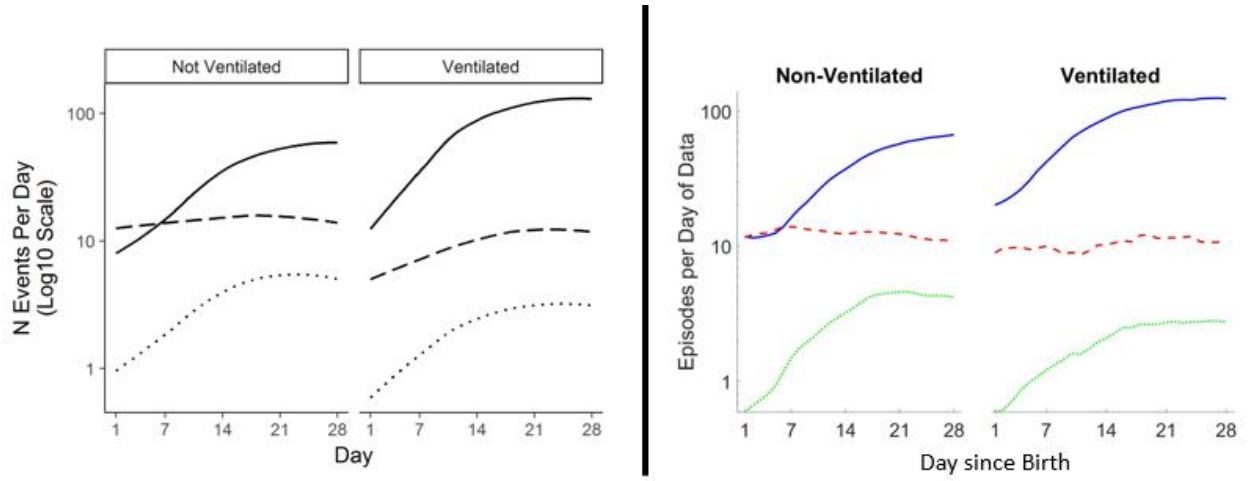

Figure 6: Episodes Occurrence Trend - For infants on mechanical ventilation (left) and off (right), mean number of episodes of bradycardia (dashed line), desaturation (solid line) and BD (dotted line) are shown from day 1 to day 28 after birth; each data point represents a 7 days average occurrence rate that include the current day of age plus 3 neighboring days on each side; UVA version is on the left [7] and ours is on the right

Infant Group	Ventilated at least Once	Have Time Windows of No-Ventilation
Size	278	414
BW in grams: median (IQR)	880 (490-1270)	1020 (570-1470)
GA in weeks: median (IQR)	26 (23-29)	28 (24-32)
General episodes rate	count per day of data	count per day of data
Desat	67.3665	50.91
Brady	10.0656	10.75
BD	2.63	2.32

Table 5: Characteristics of Clinical Episodes in the two different infants group being analyzed for on-ventilation episodes occurrence (ventilated at least once) and off-ventilation occurrence (have time windows of no-ventilation). The ventilated at least once group have distinctly higher desat rate (on and off-vent time included) than the other group.

We conduct our analysis for 278 babies that were put on mechanical ventilation at some point within the study period. An infant in this group on average spent around 10 days of data on ventilation and 10 days off ventilation. 121 out of 278 babies were put on ventilation more than once within the study period of 28 days after birth. On average, these babies have around 12 days of data on vent and 9 days off vent.

We look at two regions near the ventilated time windows: the 3 days surrounding the time mark when ventilator is turned on (on-transition region) and when it was turned off (off-transition region). Occurrence rates were calculated by counting only episodes occurred in these regions. Ventilated and non-ventilated windows ≤ 3 days long are excluded from this analysis.

In Figure 8, we can see episode rates drop for on-transition region. Looking at off transition region for 3 days after ventilation usage stop, we see that number of episodes per day are still lower than they originally were 3 days before the start of ventilation. Thus, mechanical ventilation lowers episodes occurrence and has long lasting effects even after usage is stopped.

In addition, it is noteworthy to point out that episode rate increases during the ventilated time window. This is consistent with the general trend seen in Figure 6 where episode rate increases throughout the first 28 days. Majority of the ventilated time windows are at least 6 days long. This can shown by comparing the middle and bottom histogram in Figure 9. The middle histogram is the distribution of ventilated windows of at least 3 days long with respect to its baby's day of age. The right histogram is the same distribution for ventilated windows of at least 6 days long. Both histograms look very similar. In the bottom histogram, there is a slight concentration of windows from day 7 to day 21 where desats and BD occurrence rate rise very rapidly (shown in Figure 6). According to this trend, a 6-days long ventilated window would see a steep increase in episodes occurrence rate between the first 3 days and last 3 days which is observed in Figure 8.

Because of the general increasing trend for desat and BD episodes, their associated numbers in Figure 8 are highly dependent on when the baby was put on ventilator and the duration of the ventilated window. To clarify what is going on, let's imagine a hypothetical baby that was put on the ventilator from day 7 to day 14 after its birth. Assuming that mechanical ventilation does not affect episodes occurrence in anyway, we will observe that desat occurence increases through both the on and off transition regions. According to the non-ventilated episodes trend in Figure 6, the baby would go from having 15 desats per day before being put on the ventilator to 30 desats per day after ventilator is turned off. If the baby was ventilated from day 7 to day 21, the baby would actually go from 15 to 50 desats per day. If mechanical ventilation actually reduces episodes occurrence, the ventilator is essentially "fighting" against this natural increasing trend to reduce episodes occurrence seen in Figure 8. This issue does not affect brady episodes as much since their occurrence are relatively stable over the course of study.

The distributions of ventilated and non-ventilated windows with respect of day of age (top and middle and histograms in Figure 9) are nearly but not exactly uniform. Ventilated windows (middle histogram) tend to slightly concentrate on the first week of age while non-ventilated windows (top histogram) are slightly concentrated around the fourth week of age. Calculations from the non-ventilated windows will be a bit higher as episode occurrence rates peak at the fourth week. For this reason, interpreting the results from the exact numbers is not recommended. We do not know for sure if desat occurrence is actually cut by half during the on transition region Figure 8 suggests. However, it is obvious that mechanical ventilation reduces episodes occurrence.

4.3 Predictive Modeling

We built several multivariate regression models and assessed their predictive power to identify parameters that are clinically important for prediction. The parameters are calculated based on clinical episodes collected from each baby during their first month of age. In prior work done by UVA, a baseline clinical model using demographic and clinical variables was established. We generating a new model by adding a new parameter X to this base model and fit it with recorded health outcomes using logistic regression:

$$ln\left(\frac{\pi_{\text{sick}}}{\pi_{\text{not-sick}}}\right) = \beta_0 + \beta_1 \cdot \text{Birthweight} + \beta_2 \cdot \text{Gestational Age} + \beta_3 \cdot \text{Ventilated Days} + \beta_4 \cdot X \quad (7)$$

Afterwards, we evaluate the predictive power of each model using the area under the receiver operator characteristic curve (ROC AUC). Table 4.3 show our results for parameters associated with desats, brady, apnea, or combined episodes. Parameters that significantly add to the baseline model ($p<.05$) have their AUC, β_4 and p-value shaded in gray. We determined whether adding a parameter will add predictive power to the baseline model by testing if its coefficient value is equal to 0 using Student's t-test:

$$H_o : \beta_4 = 0 \quad (8)$$

$$H_a : \beta_4 \neq 0 \quad (9)$$

Our level of confidence α is 0.05 and parameters with significant p-values have a 95% chance of being important. The AUC value signifies the proportion of correct predictions made by its associated model. Adding the D% time (percent time spent in desat) to the baseline model helps it predict BPD outcome correctly for 369 infants (89.16%) in the training data set. In addition, for each percentage increase in D% time, we estimate that the odds of having BPD will increase by $33.51(e^{\beta_4} - 1 = 0.3351)$ while holding other parameters fixed.

Moreover, these models can be used to estimate outcome risks. The log-odd that a preterm infant born at 24 weeks gestational age, weighted 900 grams, spent 9 days in mechanical ventilation and 3% of time in desaturation is

$$ln\left(\frac{\pi_{\text{bpd}}}{\pi_{\text{no-bpd}}}\right) = \beta_0 + \beta_1 \cdot 800 + \beta_2 \cdot 24 + \beta_3 \cdot 9 + \beta_4 \cdot 3 \quad (10)$$

$$= -.8667 - 1.3799 \cdot 800 - .0015 \cdot 24 + .0615 \cdot 9 + .2890 \cdot 3 = -.5859 \quad (11)$$

The odd of developing BPD is

$$\frac{\pi_{\text{bpd}}}{\pi_{\text{no-bpd}}} = e^{-.5859} = .5566 \quad (12)$$

and the risk of BPD is

$$\pi_{\text{bpd}} = \frac{.5566}{1 + .5566} = .3576 \text{ or } 35.76\% \quad (13)$$

Shaded areas in Table 4.3 show there is strong evidence that desat parameters are useful for predicting in the primary outcome. In addition, desat parameters can predict 2 other secondary outcomes, ROP and prolonged stay in the NICU. Apnea, brady, and other combined episodes do not significantly contribute to the baseline model for predicting the primary outcome. However, apnea episodes contribute information toward prolonged stay prediction.

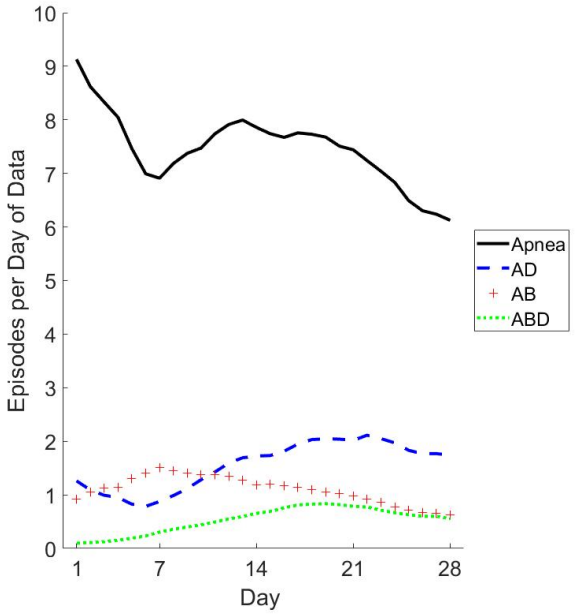

Figure 7: Apnea Occurrence Trend from day 1 to day 28 after birth

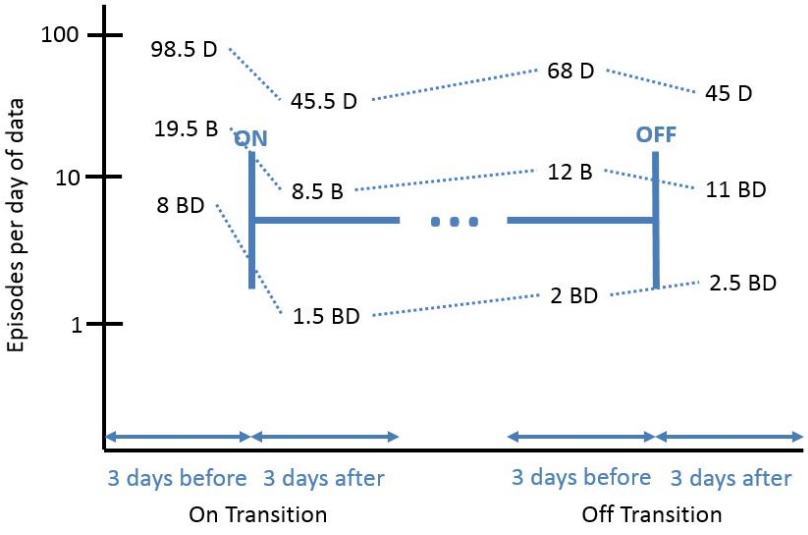

Figure 8: Episodes Occurrence Rate around Ventilation Windows

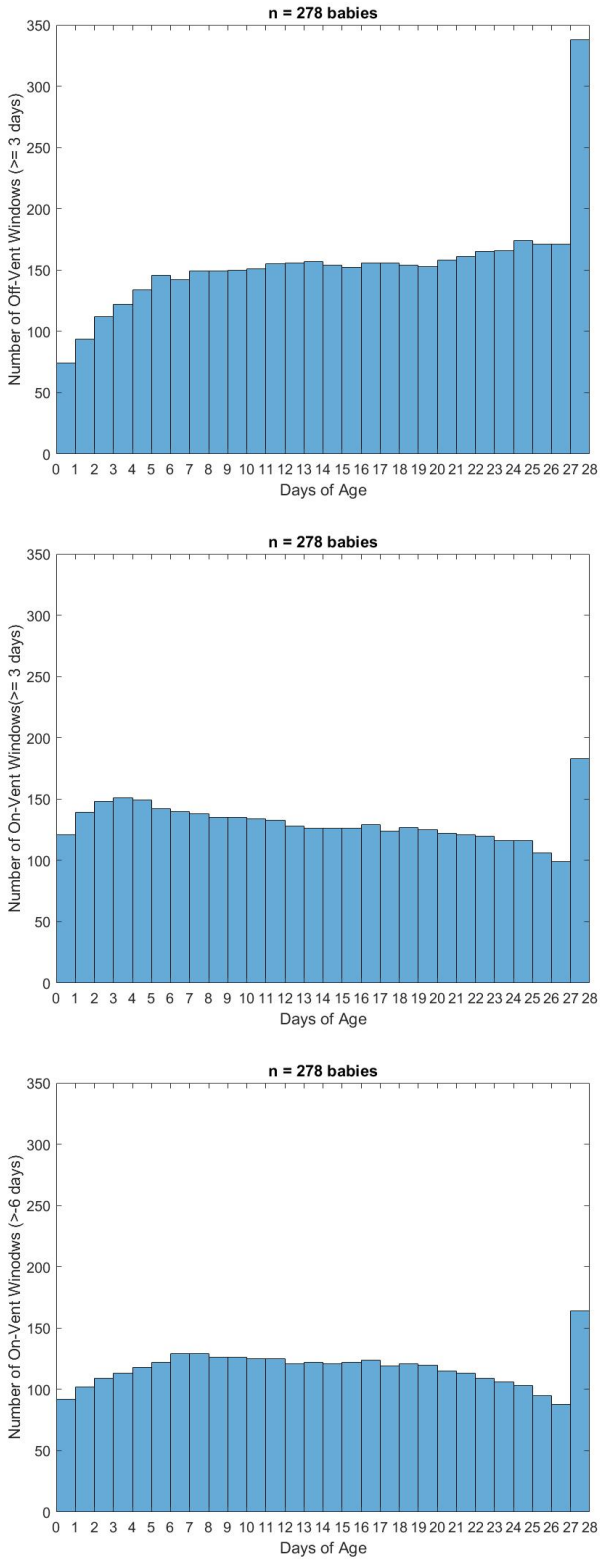

Figure 9: Distribution of Ventilation Windows with Respect to Infants' Day of Age - TOP: Non-Ventilated Windows of At least 3 days long, MIDDLE: Ventilated Windows of At least 3 days long, BOTTOM: Ventilated Windows of At least 6 days long

	BPD or death			severe IVH			treated ROP			Prolonged Stay		
Baseline model AUC	.8612			.7945			.8911			.8038		
Parameters added to base	AUC	β_4	p-value	AUC	β_4	p-value	AUC	β_4	p-value	AUC	β_4	p-value
D count	.8784	.0156	5.702e-4	.7929	.0013	.8337	.9093	-.0219	.0078	.8018	.0036	.4429
D area	.8824	2.997e-5	1.611e-4	.7936	-7.396e-6	.4546	.9031	-2.204e-5	.0397	.8055	8.6733e-6	.1908
D % time	.8916	.2890	1.185e-5	.7934	-5.5479	.5290	.9008	-18.6968	.0383	.8090	.0996	.0655
D mean area	.8769	.0036	3.611e-5	.7924	-3.3662e-4	.8028	.8917	.0012	.3193	.8196	.0026	.0055
D mean length	.8857	.0402	8.950e-6	.7933	-.0014	.9233	.9095	.0386	.0022	.8335	.0357	1.620e-4
B count	.8637	.0089	.5918	.7912	.0080	.6835	.8886	.0135	.4816	.8073	-.0193	.4264
B area	.8635	2.898e-5	.7193	.7952	-2.332e-4	.1670	.8871	1.389e-4	.2544	.8164	-2.1697e-4	.1037
B % time	.8648	1.2049	.3609	.7913	.7019	.5828	.8896	.8112	.5216	.8116	-2.2977	.2641
B mean area	.8629	.4698	.4698	.8073	-.0050	.1139	.8905	-1.7702e-4	.9368	.8082	-.0030	.1693
B mean length	.8644	.0363	.3773	.8023	-.1917	.1276	.8906	.0280	.7623	.8076	-.1225	.1616
A count	.8632	.0036	.7557	.7891	-.0127	.4881	.8916	-.0022	.8873	.8365	-.0578	.0076
A area	.8635	1.6162e-4	.6261	.7933	-6.129e-4	.2845	.8909	2.9875e-5	.9441	.8386	-.0017	.0082
A % time	.8640	.1433	.5802	.7943	-.5492	.2317	.8909	.0315	.9241	.8370	-1.2723	.0094
A mean area	.8649	.0233	.2047	.7934	-.0163	.4686	.8897	.0064	.7524	.8115	-.0282	.0974
A mean length	.8650	.0184	.2310	.7960	-.0206	.2742	.8901	.0072	.6817	.8074	-.0204	.1585
BD count	.8650	.0475	.3999	.7953	-.0958	.3583	.8889	.0601	.4487	.8024	-.0258	.7468
ABD count	.8632	-.0887	.4473	.8040	-.3937	.1514	.8920	.0911	.4630	.8278	-.4717	.0420
AB count	.8626	-.0094	.8795	.8174	-.3462	.0687	.8890	.0781	.3172	.8471	-.5346	.0032
AD count	.8630	.0117	.7111	.7918	.0037	.9263	.8908	-.0011	.9802	.8152	-.1054	.0983

Table 6: Predictive performance of different logistic regression models for primary outcome and secondary outcomes of 373 VLBW infants

5 Discussion

5.1 Central Apnea

Babies aren't born with BPD but usually develop it after suffering from respiratory distress syndrome (RDS). RDS is a breathing disorder that primary affects preterm infants with immature lungs either not fully developed or cannot make surfactant. Surfactant is a liquid that coats the inside of the lungs and plays an important role in keeping them open during inhalation. The lungs collapse without surfactant requires more effort to breath, possibly leading to not insufficient oxygen intake [1]. Therefore, early warning signs of BPD might not central apnea because chest activity can still be present when suffering from RDS. However, central apnea episodes is still harmful as it can creates desaturation episodes and can be a signal of a prolonged NICU stay passed the original due date.

5.2 Future Research Idea: Time from Last Apnea to Desat

Using our current definitions, ABD occurrence is quite rare compare to independent desat or brady episodes. Furthermore, UVA group has shown that ABD do not occur more frequently in BPD infants [6]. Analyzing in the dynamic between apnea and desats episodes can potentially enhance BPD prediction.

Screening through all desats and apneas, we observe that the time between start of the last apnea before a desat and the start of a desat is continuously distributed and has a peak around 30 seconds (Figure 10). Moreover, the median time between apnea and desat was 60 seconds. Under the assumption that apnea and desats are independent events, this distribution should look like exponential decay distribution. When adjusting the y-axis into the natural log scale, the observed distribution is not linear. Therefore there is some possible correlation between apneas and desats. This topic can be studied more to reveal any potential in BPD prediction.

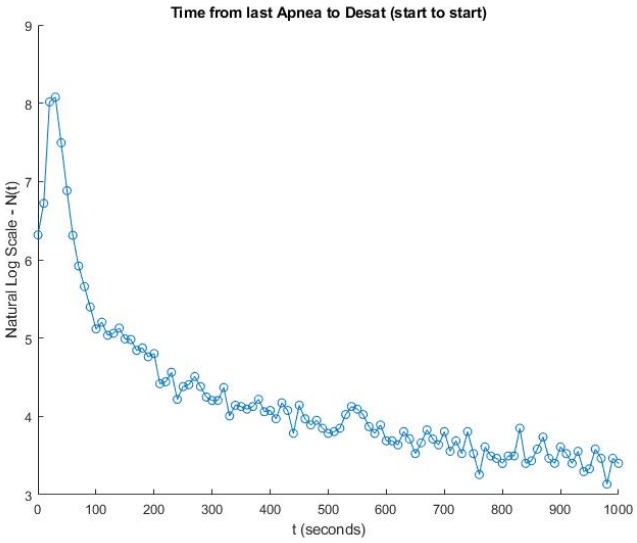

Figure 10: Distribution of time between start of the last apnea before a desat and the start of a desat

6 Conclusion

A major aim of this study was to continue to develop new prediction models for health outcomes of VLBW infants in UVA NICU. The primary goal is to use early medical data of NICU stay to predict an infants risk of BPD, a chronic lung disease that is burdensome and damaging to long term health. In addition, we attempted to predict other morbidities that might be linked with aberrant oxygenation. Previous work used demographic and respiratory support variables to calculate these health risks. Our study look at additional variables extracted from readily available medical data that has potential power to contribute in predictive models. Three types of clinical episodes were studied here for this purpose. Two of them, desats and bradys, have been previously studied by the UVA group. Here we externally validate their findings. Additionally, we look at apnea episodes since they have not been studied before in this context.

Desaturations play an important role in outcome prediction. This result agrees with UVA. As hypothesized and previously proved by UVA, bradycardia and BD do not contribute predictive information. Mechanical ventilation lowers desats, bradys and BDs occurrence even after usage stopped.

Apnea episodes, which were previously believed to be highly correlated with desats and health outcomes, have a much more complex dynamic with these variables than expected. Many desats occur independently without apneas, and all VBLW infants experience lots of apneas regardless of their health outcomes. Thus, apneas do not add power to predictive models. Missing data presents the biggest challenge for us throughout the study. We cannot capture the exact statistics for episodes occurrence, as no infants has a complete set of medical data. Using a larger cohort along with a more complete database will certainly reduce errors and allow for better statistics. Future research can look at interactions between apneas and desats and try to predict BPD outcome solely instead of our combined primary outcome (BPD or death).

7 Appendix

7.1 Code Library

Originally, there were 5 types of tables: bed assignment, patient's outcome, ventilation table, event tables, length tables. We received the first 3 tables from UVA. Event tables consist of desat, brady, and apnea tables. They contain information about all events experienced by all the 414 infants being studied. Example of a desat table is in Figure 11. Length tables consist of SpO2, HR, and aprob tables. They contain information about how much available data there is in a day (Figure 12. To query events, we have to cross reference all of these tables. For example, to get all desats that occured during ventilation from babies that develop BPD, I would need to:

1. Go to outcome table and get IDs of babies with BPD
2. For each baby, look up their ventilated time windows from the ventilation table. Get the start and end time of each window. Since the ventilation record their time in seconds with respect to baby's birthdate, convert start and end time into recorded date.
3. For each ventilated time windows, go to desat table and find the according desats using patient ID and time marks

The first column in desat tables already contain the patient ID that a desat belongs to so no need to look up the bed assignment table when querying. However, that is only because a column was added in later by cross referencing with the bed assignment table through a separate script. Overall, lots of code needed to be written to perform a very simple-sounding task. If we want to calculate desat occurrence rate in the scenario above, we need to know the total amount of SpO2 data available in ventilated windows from babies with BPD. Thus, we would need to repeat the procedure again using the SpO2 table for step 3. The length tables do not have the patient ID column so we would need to do an extra step of cross-referencing before doing step 3 to obtain the total amount of SpO2 data available.

To make this library, I took advantage of a popular programming paradigm called **objected-oriented-programming** (OOP). For a detailed explanation of how OOP works in MATLAB, refer to [2]. In OOP, we have an **object**. This object contains values that are unique to it. In our context, a patient can be represented as an object identifiable by the patients ID. Lets say this object is stored as a variable named *baby*. To make *baby* useful, it needs more information inside, such as clinical episodes, bed information, and time spent on vent. In addition, the object should also contains a specific set of functions that it can perform. These functions can only be called if you have a reference to the object. For example, I have a function called *getEventsOnVent()* that will return a matrix of events. Typing this method in the command line does not make any intuitive sense and will return an error because MATLAB does not know which patient to retrieve events from. Calling *baby.getEventsOffVent()* will return events that belong to *baby*. To put it another way, an object contains a set of values and functions that is only accessible through that object.

Each kind of object has a template that defines the specific type of methods and values it can contain. In OOP terminology, this template is called a **class**. In my library, a Baby class outlines all the values and functions a *baby* object can use. In the file all_classes_clean.mat, there are 414 *baby* objects for each patient stored in a vector called *Babies*. To get *baby* object number 300, use the command *Babies(300)*. Using a similar logic, a ventilation window can also be represented by a *vent* object constructed from the Vent class. Each *vent* object has a unique status (ventilated or non-ventilated), start and end time as well as the baby it belongs to. This is the same case for a bed assignment window with a *bed* object and associate Bed class. Similar to Babies, there are

Vents and *Beds* vectors in all_classes_clean.mat that include all ventilation and bed windows within the first 28 days of the infants' chronological age.

Information from bed assignment, ventilation time, and health outcomes can all be stored into these objects by linking a baby with its beds and vents, and bed and vent can be linked back to the baby they belong. Note that a *baby* can have multiple *vent* and *bed* objects but each vent or bed object can only belong to one baby. Now we don't have to worry about those 3 tables. An illustration of the dynamic between the different classes is in Figure 13.

After setting up and linking the objects together using read_save_new_system.m, we don't need to use the bed, ventilation, or outcome tables anymore. The remaining tables are event and length tables. Event tables are wrapped around a class called EventsSingleton and saved as a variable called *EventsData* in all_classes_clean.mat. Length tables are wrapped around LengthsSingleton and saved as *DataLengths*. Baby, Bed, and Vent classes all contain functions that will query appropriate events from inputs requested (*queryBaby* and *queryTime*) as well as data lengths (*queryLength*). These classes have access to event and length tables inside of *EventsData* and *DataLengths*. To simplify querying further, Baby, Bed, and Vent have dummy variables (Desats, Bradys, Apneas) that will call query functions and return all the events from those *baby*, *bed*, and *vent* respectively. MATLAB documentation called these **dependent variables**. I find illustration is the best way to depict what I am trying to explain so take a look at Figure 14.

There is one more class worth mentioning, the Utilities class. This class contains functions that I usually use for analysis. I find it convenient to have all commonly used functions in one single script for easy reference hence the name Utilities. One of the purpose of Utilities functions is to detect combined events such as BD and ABD. Vent, Baby, and Bed contains dummy variables for BD and other combined events that call upon query functions to get desats and bradys separately then call *Utilites.matchBD(bradys,desats)* afterward and return the results from matchBD() as their BD (Figure 15).

To convert from 5 original tables into this all_classes_clean.mat, load the 3 scripts read_save_new_system.m, load_apneas_to_newsystem.m, and make_length_apnea.m. To use this new system, run the setup.m script that will load all_classes_clean.mat and initialize variables for everything to work . It is very important to run this script before doing any analysis. Note: concepts that can be Googled are in bold, functions name are italicized, scriptname and filename is underlined.

	1	2	3	4	5	6	7
1	1089	28	1982926	1982938	14	14	79
2	1089	28	1986878	1986896	20	40	78
3	1089	28	1999470	1999484	16	58	76
4	1089	28	2002724	2002742	20	34	77
5	1089	28	2025028	2025050	24	90	75
6	1089	28	2045652	2045684	34	162	74

Figure 11: A group of events in the desat table. Each row represents one desat. From left to right, the columns are patient's ID, bed, start time, end time, duration, area under threshold, and minimum SpO2 value

	1	2	3	4
888	3	24881	67651200	67701158
889	3	0	67651200	67651200
890	3	0	67651200	67651200
891	3	-1	67737599	67737599
892	3	18510	67872510	67910398
893	3	13468	67910400	67939198

Figure 12: A group of lengths in the SpO2 table. Each row represents one length. From left to right, the columns are bed, length of signal, start time, and end time. The length of signal column is encoded so that : > 0 - amount of available data in seconds; 0 - no data; -1 - file for this date not found on Sciclone; -2 - SpO2 variable not found in file; -3 - no bed assignment information in this time period defined by start and end time

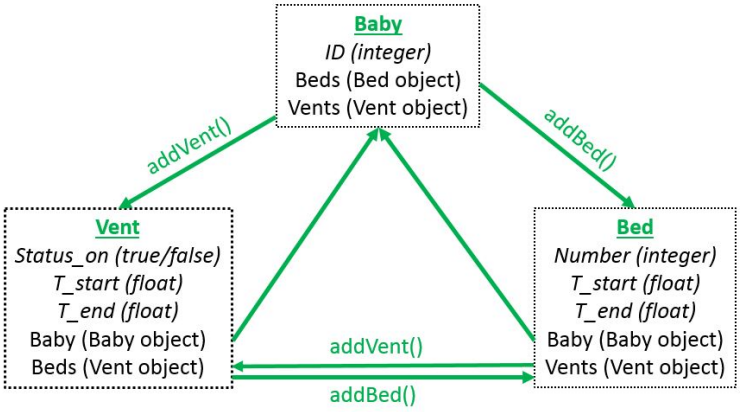

Figure 13: Diagrams for Baby, Bed and Vent. Italicized variables are unique for different objects of the same class. A *baby* can have multiple *vent* and *bed* objects but each *vent* or *bed* can only belong to one baby. Example command to connect *vent* and *baby* to *bed*: *vent.addBed(bed)* or *baby.addBed(bed)*

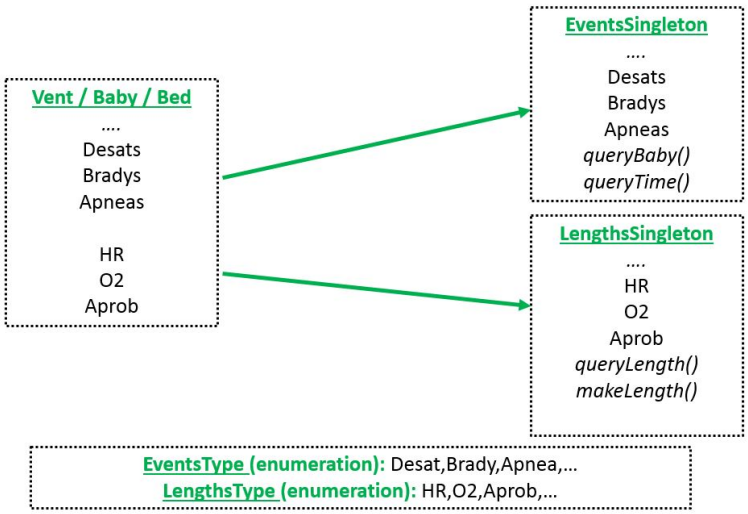

Figure 14: Function calls to query events are italicized. Command to get desats from a bed object: *bed.Desats*. To get all desats occurred from day 1 to day 10: *EventsData.queryTime(EventType.Desat,day1,day10)* where EventsData is an EventsSingleton object and EventType is an **enumeration**, a type of data structure (refer to MATLAB official documentation for more details)

```
function q = get.BDs(obj)
    % version check
    [ver,~] = Utilities.matchBD;
    if ver == obj.pBD_ver
        q = obj.pBDs;
    else
        disp('requeue BD')
        [ver,q] = Utilities.matchBD(obj.Bradys,obj.Desats);
        obj.pBDs = q;
        obj.pBD_ver = ver;
    end
end
```

Figure 15: Body of get.BDs() in Baby class. This function gets called whenever we refer to *baby.BDs*

References

[1] Bronchopulmonary dysplasia. https://www.nhlbi.nih.gov/health-topics/bronchopulmonary-dysplasia.

[2] Object-oriented programming in matlab. https://www.mathworks.com/company/newsletters/articles/introduction-to-object-oriented-programming-in-matlab.html.

[3] P. Ballabh. Intraventricular hemorrhage in premature infants: Mechanism of disease. *Pediatric Research*, 67(1):1–8, 2010.

[4] J. Chen, A. Stahl, A. Hellstrom, and L. E. Smith. Current update on retinopathy of prematurity: screening and treatment. *Current Opinion in Pediatrics*, 23(2):173–178, 2011.

[5] R. Darnall, J. Kattwinkel, C. Nattie, and R. M. Margin of safety for discharge after apnea in preterm infants. *Pediatrics*, 100:795, 1997.

[6] K. Fairchild, M. Mohr, A. Paget-Brown, C. Tabacaru, D. Lake, and J. Delos. Clinical associations of immature breathing in preterm infants: part 1-central apnea. *Pediatr Res.*, 80(1):21–27, March 2016.

[7] K. Fairchild, V. Nagraj, B. Sullivan, J. Moorman, and D. Lake. Oxygen desaturations in the neonatal period predict development of bronchopulmonary dysplasia. *American Journal of Respiratory and Critical Care Medicine*, 2018.

[8] N. Finer, R. Higgins, J. Kattwinkel, and R. Martin. Summary proceedings from the apnea-of-prematurity group. *Pediatrics*, 117(1):47–51, 2006.

[9] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, and J. B. ... Delos. A new algorithm for detecting central apnea in neonates. *Physiological Measurement*, 33(1):1–17, 2012.

[10] S. Maddox. Quantitative analysis of vital signs of premature infants, April 2016.

[11] R. Martin, A.-S. JM, and B. TM. Apnoea of prematurity. *Paediatr Respir*, 5:377–382, 2004.

[12] M. Mohr. *Quantitative Analysis of Periodic Breathing and Very Long Apnea in Preterm Infants*. Physics, The College of William and Mary, Grand Island, New York, January 2016.

[13] M. Mohr, B. Vergales, H. Lee, M. Clark, D. Lake, A. Mennen, and J. Delos. Very long apnea events in preterm infants. *Journal of Applied Physiology*, 118(5):558–568, 2015.

[14] L. Singer, T. Yamashita, L. Lilien, M. Collin, and J. Baley. A longitudinal study of developmental outcome of infants with bronchopulmonary dysplasia and very low birth weight. *Pediatrics*, 100(6):987–993, 1997.

[15] B. D. Vergales, A. O. Paget-Brown, H. Lee, L. E. Guin, T. J. Smoot, C. G. Rusin, and J. ... Kattwinkel. Accurate automated apnea analysis in preterm infants. *American Journal of Perinatology*, 31(2):157–162, 2014.