$STDM^{n+p}{}_{0:}$ A Multidimensional Patient Oriented Data Mining Framework for Critical Care Research

By

Kathleen Patricia Smith

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Masters of Health Science – Health informatics
in

The Faculty of _Health Sciences (MHSc)__
Program

University of Ontario Institute of Technology
April 2011

Acknowledgements

One of the greatest things I witnessed throughout this journey was the immeasurable and unwavering love and support my family offered. While it was my husband and children that stood front line behind me my whole family took turns in very timely offerings of encouragement. To all of you I love and adore you very much. You will never know how grateful I am and how much I appreciated having the whole team behind me!

To my supervisor and mentor Dr. Carolyn McGregor, who believed in me from the beginning right through the entire Master's journey, even in moments when I had trouble believing, thank you for your unconditional and untiring support and guidance. Unwittingly you have made a profound impact on my life for which I am truly grateful.

Dear Christina Catley and Andrew James your encouraging guidance right through to the end was greatly appreciated. I would also like to thank Mike Eklund for being a part of the supervisory team right from the being.

To Carolyn, the supervisory team and friends Rishi and Agam, this past few years has been an incredible journey and would like to thank you all for being a part of it.

Certification/Statement of Authentication

I, Kathleen P. Smith, certify that this thesis, submitted to fulfill the requirements for the Award of Master of Health Science – Health Informatics, in the Faculty of Health Science, at the University Of Ontario Institute Of Technology, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted, either in full or in part, at this or any other Academic institution to meet requirements for any Award.

……

Kathleen Patricia Smith

List of Publications

The following publications have made contributions in this thesis.

Blount, M., Ebling, M. R., Eklund, J. M., James, A. G., McGregor, C., Percival, N., **Smith, K.P**., & Sow, D. (2010). Real-Time Analysis for Intensive Care - Development and Deployment of the Artemis Analytic System. *IEEE Engineering in Medicine and Biology Magazine* , 110-118.

Catley, C., **Smith, K**., McGregor, C., & Tracy, M. (2009). Extending CRISP-DM to Incorporate Temporal Data Mining of Multi-dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study. (p. 5). IEEE.

Catley, C., **Smith, K.**, McGregor, C., James, A., & Eklund, J. M. (2010). A Framework to Model and Translate Clinical Rules to Support Complex Real-time Analysis of Physiological and Clinical Data. *IHI'10.* Arlington, Virginia, USA.: 2010 ACM.

Eklund, J. M., McGregor, C., & **Smith, K.** (2008). A Method for Physiological Data Transmission and Archiving to Support the Service of Critical Care Using DICOM and HL7. *IEEE EMBS conference.* Vancouver.

McGregor, C., & **Smith, K. P**. (2009). A survey of Physiological Monitoring Data Models to support the Service of Critical Care. *33rd Annual IEEE International Computer Software and Applications Conference* (pp. 104-109). Seattle, Washington: IEEE.

Table of Contents

List of Tables and Figures

“The purpose of computing is insight, not numbers” --Richard W. Hamming (1962)

“The purpose of computing numbers is not yet in sight”-- Richard W. Hamming (1997)

Abstract and Keywords

In the neonatal intensive care unit (NICU) environment, critical care and treatment directly correlate to the multidimensional development of an infant and are influenced by attributes such as gender and gestational age (GA). Recent literature on guidelines developed for neonatal intensive care; do not take the gender or the GA of the infant into account. The exponential activity of a growing neonate in its early stages of life needs to be captured and embedded into algorithms designed to extract patterns of predictive temperament within the NICU domain. The $STDM^{n+p}_{0}$ framework presents an extended multidimensional approach with the ability to create patient characteristic clinical rules. Further defining NICU algorithms, through the extended use of attributes to include gender and GA, and using these new algorithms in clinical decision support systems increases the accuracy and thereby minimizes the risk of adverse events.

Keywords: Neonatal intensive care, critical care, multidimensional algorithms, patient characteristics, clinical decision support systems.

Chapter 1 – Introduction

1.1 Introduction

This thesis presents a multidimensional patient oriented data mining framework used to support critical care research to enable the discovery of physiological stream behaviours that may represent earlier condition onset behaviours than those currently used in evidence based practices. This research extends the $STDM^{n}_{0}$ framework (McGregor C. P., July 2010) and the work of Bjering and McGregor, 2010 (Bjering, 2008) through the incorporation of patient specific attributes, thus enabling tailoring and clustering of physiological stream behaviours based on these patient specific measures. The extended framework, $STDM^{n+p}_{0}$, will include methods for applying temporal abstractions (TAs) representing physiological stream behaviours across multiple patient attribute parameters for multiple patients to enable mining of multidimensional temporal data. This research is demonstrated through a case study context of neonatal intensive care using the conditions of apnoea and nosocomial infection (NI).

The $STDM^{n+p}_{0}$ framework proposes a multidimensional approach that supports temporal abstractions of time series data and deployment of clinical algorithms. The term multidimensional data implies that multiple data elements, each representing a dimension that can vary in value, characterize an item of interest (Harrison, 2008). In the clinical Neonatal Intensive Care Unit (NICU) context, specific dimensions of interest include: gender, gestational age and birth weight. The exponential activity of a growing organism, in this case a preterm infant, in the early stages of life needs be captured and

embedded into algorithms designed to extract patterns of predictive temperament within the NICU domain.

One in ten births around the world is premature (WHO, 2011) and requires admittance into NICUs under the watchful care of neonatologists. For several decades now, NICUs around the globe have utilized medical monitors and life supporting devices to assist in the care of critically ill infants. These monitoring devices sense and output an array of different physiological data readings that, in waveform format, are sampled at over 500 readings a second. Unfortunately, these readings are produced at a rate that is much faster than the human brain can analyze (Catley C. , Smith, McGregor, & Tracy, 2009) (Miller, 1956). NICUs have been lagging in sustainable infrastructure, tools and techniques to utilize this information to its fullest potential to aid real-time clinical management and historical clinical research (McGregor & Smith, 2009). These limitations lead to opportunities for developing a more structured approach to defining complex processing of physiological data streams.

1.2 Research motivation

The motivation for this research was to consider recent computing and information technology (IT) research as applied to physiological monitoring to support real-time data mining of the vast amounts of unused data. Use of data mining in healthcare is increasing as data mining offers a newer approach to analyzing retrospective physiological data streams for clinical research in healthcare. The potential benefit of applying data mining tools on electronically stored physiological data, for improved real-time clinical management and clinical decision support, is significant. Zhang

emphasized the importance of real-time data mining by stating ‘Our expanded system of real-time data collection and algorithm development demonstrated that patient-specific learning in real time is a feasible approach to developing alarm algorithms for monitoring purposes in the ICU (Zhang, 2007)’. Benefit from this research may also be applied for clinical research on stored physiological data streams to deduce new findings for condition onset prediction indicators based on patient characteristics.

Addressing healthcare issues have taken top funding priority for many countries. Both Canada and the United States have planned large funding commitments to healthcare issues over the next couple of years. In Canada, ‘the Canada Health Transfer (CHT) provides long-term predictable funding for healthcare, and supports the principles of the Canada Health Act which are: universality; comprehensiveness; portability; accessibility; and, public administration. The CHT cash transfer will reach $25.4 billion in 2010-11 and will reach over $30 billion in 2013-14 (Canada Department of Finance, 2010)’.

1.2.1 Research Motivation within NICU

Within the NICU context individual patients undergo rapid growth and development leading to changes in individual patient characteristics, such as weight, heart rate (HR), blood pressure, and postnatal age. There is a growing body of research showing examples of the use of data mining and temporal abstractions to demonstrate that a given condition exhibits certain physiological stream behaviours (Catley, Smith, McGregor, & Tracy, 2009). However sensitivity and specificity are not yet near 100 %, which in healthcare can have devastating impact on the individual patient (Sharek, et al., 2006). There is potential to use patient characteristics to gain better understanding of individual patients in retrospective data and improve sensitivity and specificity by

creating subgroups of physiological behaviours. The motivation for this research is to determine whether current examples of clinical research and the associated frameworks can support exploration and clustering based on patient characteristics.

Preterm infants who are admitted to the NICU are monitored through the assessment of a range of physiological parameters as displayed by a variety of devices. The preterm infant may have multiple streams of physiological data being collected throughout the duration of hospitalisation; however, NICU medical records are largely manually prepared paper notes of cumbersome manual data. Often clinicians manually summarize continuously streamed physiological data by single readings at 30 or 60 minute intervals. This hand notation of data is not conducive to recording abnormal behaviours among the multiple streams that frequently occur minute by minute or second by second. Frequent transient falls in blood pressure and blood oxygen content, which may be of critical importance in survival and the indication of the onset of illness, can be left undetected until monitoring device alarms are triggered. Currently in clinical settings, a nurse will mark when an alarm spell or spells have occurred in the clinical information management system (CIMS) if it is perceived to be of significance; i.e. if there is an alarm and the nurse has addressed the neonate (Catley C. , Smith, McGregor, James, & Eklund, 2010). Recent medical research literature has reported that physiological data exhibits early indicators of potentially life threatening conditions such as nosocomial infection (NI) and that these points of interest in the data precede the existing clinical practice detection (Griffin & Moorman, 2001).

When exploring patient specific physiological data streams, identifying trends and patterns to improve real-time clinical management and clinical decision support, a patient oriented approach may assist in minimizing adverse events that occur in the NICU. Findings published on the Institute for Healthcare Improvement website (IHI.org) from a North American NICU study conducted by Paul J. Sharek, et al. in 2006 to investigate and detect incidences of adverse events (AE) resulted in 'One or more AEs contributed to 27 of the 30 patient deaths in this study (Paul J. Sharek et al, 2006).' This report later was adopted as an improvement tool by the Institute of Healthcare Improvement called 'Trigger Tool for Measuring Adverse Events (AE) in the Neonatal Intensive Care Unit (IHI Tool)' (IHI.org, 2010). Within this NICU toolkit, nosocomial infection was listed as the number one 'trigger' due to having the highest positive predictive value for potential AEs (Sharek, et al., 2006).

Similar findings were published in Canadian literature: 'However, 37%–51% of AEs have been judged in retrospect to have been potentially preventable (Baker, et al., 2004).' In 2003-04 over 3 million patients were admitted to critical care units in Canada. Over 40% of these patients were over 63 years of age and 15% were neonatal patients. It is important to note that the average stay for the neonatal patients was almost three times as long as that of those from other units. Intensive care units providing critical care are one of the most costly and accounted for 15.9% of inpatient direct expenses but only 8.1% of inpatient days in Canada between 1999-2000 and 2003-04 (Leeb, Jokovic, Sandhu, & Zinck, 2006) (McGregor & Smith, 2009).

These statistics demonstrate a need to eliminate such adverse events and improve and assist in the ability to care for critically ill patients. The current challenges within healthcare, and in particular critical care, has motivated the proposed extensions to the $STDM^{n}_{0}$ framework (McGregor C. P., July 2010), called $STDM^{n+p}_{0}$ within this thesis.

1.3 Research Aims and Objectives

The aim of this research is to address an open research area, namely to enable multidimensional data mining based on patient characteristics that ultimately can assist in providing clinical support to caregivers as physiological thresholds are being breached. The main research objective is to create a framework to support clinicians as they perform patient oriented clinical research to improve patient outcomes and morbidity via real-time anomaly detection in multidimensional physiological data streams. Four research hypotheses are presented here and addressed in this work:

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multidimensional temporal data.*
3. *The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.*

4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.*

1.4 Contribution to knowledge

The contributions to knowledge within this thesis include;

- Extensions to the previously designed $STDM^{n}_{0}$ multi-agent framework for analysing time series data, to facilitate use of attributes such as gender and gestational age into a multidimensional approach capturing patient characteristic-based temporal abstractions, complex abstractions and relative alignment of these abstractions.
- Design of a framework to enable patient characteristic multidimensionality to temporally abstractive data mining.
- Demonstrating the potential benefit and use of data mining from electronically stored physiological data for improved real-time clinical management and patient centric clinical decision support.
- Demonstrate the potential for clinical research on stored physiological data streams to deduce new findings for condition onset prediction indicators in support of a current ethics approved clinical research study.

1.5 Research method

The design approach was based on a constructive research methodology. Constructive research is most commonly referred to as a computer science research method. The term ‘construct’ infers a new contribution being developed such as a new theory,

algorithm, model or a framework. This method is heuristic in nature in that it involves the evaluation of many trial and error methods to solving a problem or issue that has been identified in the domain. 'To be considered constructive research, the research must combine problem solving and theoretical knowledge', as illustrated in Figure 1-1 Elements of Constructive Research (Kasanen, Lukka, & Siitonen, 1993).

Throughout the data mining processes physiological data and pattern detections will constantly be analysed and re-evaluated in action research cycles towards the goal of developing clinically relevant algorithms. Within the constructive methodology, proposed framework alterations were made to adapt to these re-evaluations. Basing the framework design on constructive action research provides a flexible and adaptable approach for the research within this thesis.

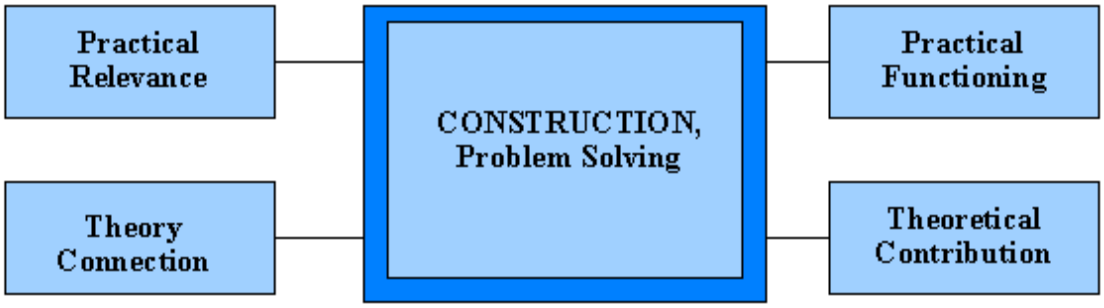


Figure 1-1 Elements of Constructive Research

The authors (Kasanen, Lukka, & Siitonen, 1993) propose six phases of the constructive approach:

1. *Find a practically relevant problem that also has research potential.* This research proposes adding patient characteristic attributes to model the exponential growth and developmental changes of the preterm infant.
2. *Obtain a general and comprehensive understanding of the topic.* Investigate for the purpose of gaining understanding of the NICU domain context from clinicians' knowledge as well as published information.
3. *Innovate, i.e., construct a solution idea*. The proposed research will build on to an existing framework to incorporate patient characteristic attributes into clinical rules.
4. *Demonstrate that the solution works*. The research extensions proposed will be demonstrated in an apnoeic event case study utilized as an early indicator of nosocomial infection (NI).
5. *Show the theoretical connections and the research contribution of the solution concept*. The case study demonstration integrates the explored theoretical connection with the research contribution by offering a solution to further define clinical rules through the adoption of patient characteristics.
6. *Examine the scope of applicability of the solution. (Kasanen, Lukka, & Siitonen, 1993).* Once the research contribution has been demonstrated, discussion will follow regarding the scope of its application.

The Practical Relevance and Practical Functioning portions of the above figure invoke the following steps in constructive research process; these include: Set objectives and tasks, Indentify process model, Select case execution, Prepare simulation, Run simulation, Interpret simulation results and Give feedback. The Theory Connections and

Theoretical Contribution portion of the above figure (Figure 1-1 Elements of Constructive Research) invoke the following steps in constructive research process, those include: Theory connection acquired from different information sources and training materials (processes, literature, articles, work experience etc.), Case research, Theory creating, Theory testing and Theoretical contribution gives new theoretical knowledge that needs scientific acceptance.

1.6 Thesis overview

Chapter 2 presents the literature review, focusing on current frameworks for enabling data mining and temporal abstractions and in particular, entities included within these temporal abstractions that further define temporal condition onset of illness based on patient characteristics such as gender and gestational age. The application domain for this research is that of the NICU and is introduced in chapter 3. This chapter discusses the neonate as a growing and developing physical being both dependant on gender and driven by gestational age that should be reflected in rules that assist in the diagnosis of conditions that may affect these infants. Chapter 4 fully describes the existing multi-agent framework, $STDM^{n}_{0,}$ including integration of relevant aspects of the extended data mining model Cross Industry Standard Process for Data Mining (CRISP-TDM) to support temporal data mining and facilitate null hypothesis testing on real-time series physiological data streams. Chapter 5 details extensions made to the framework design to incorporate the patient characteristic attributes of gender and gestational age, resulting in an extended $STDM^{n+p}_{0}$ framework. The extensions made to each of the agents and their functions are described fully. Chapter 5 also contains the design of the

extended tables to be stored in the $STDM^{n+p}_0$ framework. Chapter 6 demonstrates the extended functionality of the $STDM^{n+p}_0$ framework presented within the NICU context. This apnoea event research case study utilized multiple time series physiological data streams collected from preterm infants enrolled in a collaborative research project.

Chapter 2 – Literature Review

2.1 Introduction

The driving motivation behind this health informatics thesis is to determine whether current examples of physiological data stream based clinical research and associated research frameworks can support data exploration and clustering based on patient characteristics. Within a NICU there is a vast range of available monitoring devices used to display physiological data, most of which have the ability to output this data through serial, USB or other ports. The rate at which these monitors produce data makes it humanly impossible to analyze manually.

The case study context for this research is focused on the NICU environment, where recent research has shown that a change in a patient's condition can be supported by the change across synchronous collected physiological time series data streams such as heart rate (HR) and blood oxygen saturation (SpO_2) (Blount, et al., 2010). Another aspect of this research is combining this synchronistic collected physiological data with asynchronistic clinical data attributes, such as gender and gestational age, to assist in a more accurate analysis of produced readings. Both synchronous and asynchronous data provides important information relating to the growth of the neonate, and need to be captured and embedded into clinical NICU algorithms.

NICU and ICU nurses, doctors, and respiratory therapists are: 'fast moving, hands-on, multitasking, time-pressured professionals. Clinical caregivers use streaming clinical

data generated by bedside devices and off-site laboratories to inform their rapid-fire bedside clinical decision making' (Drummond, 2009). The goal of this research is to support clinical decision making based on patient characteristics; in addition, this would occur in real-time and on location, as opposed to off-site as discussed in Drummond's work. This can be achieved through the use of data mining tools to extract commonalities through high computational pattern recognition. The extracted patient characteristic information can then be used to generate patient characteristic clinical rules which are fed into clinical decision support systems (CDSSs).

This chapter is divided into the following sections: first Knowledge Discovery in Data (KDD), data mining, and analysis of real-time physiological data streams in the NICU and ICU healthcare context and the use of intelligent data analysis (IDA) and CDSS on streaming data. The second portion of the literature review is based on current use of temporal abstractions and their adoption into CDSS (and IDA) practices and to investigate whether frameworks enable the integration of clinical and streamed data to create sub-classifications. The second portion concludes with a current review of literature based on patient characteristics, such as gender and gestational age, and how these parameters have been adopted into temporal abstractions used in clinical rules and their inclusion into CDSS and IDA.

2.2 Knowledge Discovery in Data

Data mining, also referred to as knowledge discovery in data, is the process of analyzing data to extract useful information. The goal of knowledge discovery in data

mining is to search for knowledge contained within the data or databases, with the aim of making better use of the vast volumes of multidimensional physiological data generated and potentially stored within the healthcare domain.

Data mining is fundamentally the union of three techniques: statistical analysis, artificial intelligence and machine learning. These techniques are then combined to study data and find previously hidden trends or patterns within. For this reason, data mining is finding increasing acceptance in the healthcare domain as a way to analyze large amounts of data to discover previously undiscovered trends and patterns (Catley, Smith, McGregor, & Tracy, 2009).

Data mining is the tool used to develop clinical algorithms based on multidimensional physiological data collected from a plethora of devices in the NICU environment. The proposed data mining framework used within this research supports the industry standard data mining approach called *CR*oss *I*ndustry *S*tandard *P*rocess for *D*ata *M*ining (CRISP-DM). CRISP-DM was developed in 1996, with the goal of being industry, tool and application neutral; repeated references to the methodology by analysts have established it as the de facto standard for data mining. Previously published extensions made to CRISP-DM, resulting in CRISP-TDM (Catley, Smith, McGregor, & Tracy, 2009) enabled use of temporal abstractions, in turn, encouraging adaptability for clinical investigations on multidimensional time series data. That research extended the model by integrating Temporal Abstractions (TA) and allows for storage and the inclusion of Intelligent Data Analysis (IDA) based systems.

2.3 Review method

This Data Mining Internet-based systematic review is a focused exploration of peer reviewed literature published between January 2005 and January 2011. A retrospective, phenomenological combined with grounded theory method approach was adopted for the purposes of this review. Publications were sourced from IEEE Xplore digital library, ACM (Portal) digital library, and PubMed databases. The review used keywords such as “physiological data”, “NICU”, “ICU”, “real-time stream data”, “NICU clinical decision support”, “intelligent data analysis”, “temporal abstraction”, “data mining”, “gender”, “gestational age”, “physiological parameters” and “NICU algorithm”. These parameters resulted in a limited number of peer reviewed articles; thirteen articles were identified as being relevant to this research. However, due to the close association that the first five articles have with the ongoing Artemis research (indentified in the first column, Ref #, by ‘A’) and the $STDM^{n}_{0}$ framework utilized within that research, the last eight papers will constitute the primary focus of discussion in this chapter.

The review was not designed to represent a complete coverage of the topic of physiological monitoring as this research has a primary focus of research methods for collection and storage that will enable real-time data mining on physiological data streams. The overall objective was to expose structured formatting standards being used within the healthcare domain for the collection and storage of physiological data. Other review papers have been published addressing differing aspects of physiological monitoring and the collection of data, as well as different tools and methodologies of data mining which unavoidably may result in some small overlap.

The motivation for this research was to consider the ability of recent computing and IT research as applied to physiological monitoring to support real-time data mining. The following diagram (Figure 2-1) illustrates the areas of interest within this systematic review.

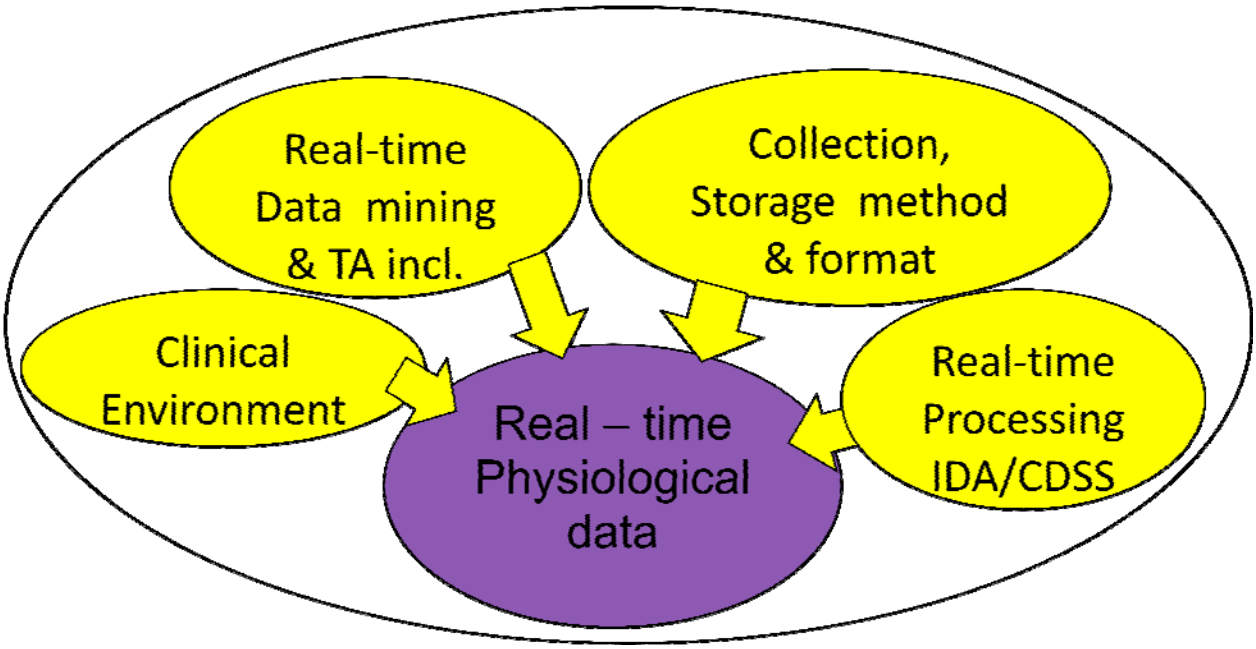


Figure 2-1 Literature Review Design Approach

Four key areas were identified based on the assessment of the relevant research papers regarding real-time physiological data. These four domains were: 1) clinical environment; 2) real-time data mining and temporal abstraction use; 3) data format and method of collection and storage used; and 4) whether real-time processing enabled use of IDA and CDSS. The information extracted from this portion of the literature review has been allocated to four subsections: 1) clinical context; 2) data management (data mining, temporal abstraction, real-time use and storage); 3) current CDSS (IDA)

practices, to investigate whether current frameworks enable the integration of clinical and streamed data to create sub-classifications and 4) patient characteristic multidimensional algorithms.

The key words used by each of the sub sections of the review section 2.4 following are presented in Table 2-1.

Table 2-1 Summary of key words used in review

Section #	Section Name	Key words used in literature review process
2.4.1	Clinical Context	NICU, ICU, physiological data, and real-time stream data
2.4.2	Data Management and Storage	NICU, ICU, real-time physiological data and physiological data storage
2.4.3	Temporal Abstractions used in CDSS (IDA)	NICU, ICU, IDA, CDSS, temporal abstraction and data mining
2.4.4	Patient Characteristic Multidimensional Algorithms	NICU, physiological parameters, gender, gestational age and NICU algorithms

Table 2-1 Key words used in literature review process

2.4 Review

2.4.1 Clinical Context

Within the clinical context the investigation of relevance was the critical care environment and whether the clinician and patient were located outside of the base intensive care unit of the given clinician. This information was captured and presented below in Table 2-2 data format, devices used to collect data, and dimensionality were also included. For the purposes of the review, 'clinical environment' describes the source domain in which the research was conducted as well as where the patient(s) was located at time of monitoring. Dimensionality assessed whether the research

catered to multiple patients, monitoring devices, multiple physiological steams in unison, and/or multiple diagnosis/conditions of interest. Data format considered what type of data was discussed in the specific article, the formatting structure of data received from devices and/or databases.

Table 2-2. Clinical context

Ref. #	Author/year	Clinical Environment	Dimensionality Patients/devices	Data Format
A	McGregor C, Kneale B, Tracy M. 2005	Neonatal intensive care	Multiple patients, multiple streams Philips component monitoring system	An integrated XML-based healthcare framework
A	McGregor C, Stacey M. 2007	Neonatal intensive care	Multiple patients Philips component monitoring system	High frequency distributed data streams
A	Bjering H, McGregor C. 2010	Neonatal intensive care	Multiple patients Philips component monitoring system	Multiple real-time physiological data streams
A	Kamaleswaran R, McGregor C, Eklund M. 2010	Critical care environment	Multiple patients Philips MP70	Synchronous physiological and asynchronous clinical data streams
A	Sun J, Sow D, Hu J, Ebadollahi S. 2010	Clinical data obtained from ICU patients	Multiple patients	MIMIC II database, physiological waveforms
1	Lyman J, Scully K, Harrison J. 2008	Laboratories, hospitals and clinics around the world	Multiple patients Multiple types of data	Many healthcare data types I.e. ECG, text
2	Tong C, Sharma D, Shadabi F. 2005	Diabetes health study Database	Multiple patients	UCI Pima Indian dataset
3	Zhang Y, Szolovits P. 2008	Intensive critical care unit collected physiological data	Multiple patients/readings. Bedside monitoring devices	Digitized waveforms and vital signs
4	Zhang Y 2007	Critically ill patients ICU	Multiple patients/readings. Bedside monitoring devices	Digitized waveforms and vital signs

Table 2-2. Clinical context continued

Ref. #	Author/year	Clinical Environment	Dimensionality Patients/devices	Data Format
5	Harrison J. 2008	Patient clinical and laboratory data	Many different data element i.e. patient, disease etc	Medical data
6	Galarraga M, Serrano L, Martinez I, de Toledo P, Reynolds M. 2007	Personal Health devices for telemonitoring of patients at home & mobile	glucose meters, blood pressure & HR, pulse oximeters, ECG monitors, etc	Digital scales ISO/IEEE 11073 P-of-C med. Device com. Stds.
7	Holmes J. 2007	Hospitals, research facilities and laboratories	Multiple patients multiple types of data	Medical data
8	Verduijn M, Sacchi L, Peek N, Bellazzi R, de Jonge E, de Mol B. 2007	ICU data Cardiac care	Multiple readings BP, HR, TMP & glucose value	Real-time data from ECG monitoring device

Table 2-2 Clinical context

The Clinical environment column indicates whether data collected is based on domain specific clinical and/or medical properties, which for the purposes of this survey mainly consisted of NICU and ICU. For most articles the clinical environment predominantly took place in ICUs with the exception of the article written by Tong (Tong C., 2008) which was a diabetes study and Galarraga et al. (Galarraga, Serrano, Martinez, de Toledo, & Reynolds, 2007) who analyzed patient care outside the hospital setting.

The dimensionality category assessed the ability of the research to cater for multiple patients, multiple physiological steams in unison, and/or multiple diagnosis/conditions of interest. This is an area of significant potential and was mentioned by most authors of papers reviewed.

Data format or representation style considered the encoding given to the data when captured. While many, if not all, of the devices used in the NICU and ICUs have the ability to output the device readings via a serial port or more recently universal serial bus (USB) or network port, the data formats vary greatly from device to device and a general adoption of data format and structure was not mentioned. However, Galarraga (Galarraga, 2007) state that IEEE is developing ten telehealth device standards for controlling information exchange to and from personal telehealth devices and cell phones, personal computers, personal health appliances and other computer engines as a part of the ISO/IEEE 11073 family of standards although there was no formal mention of a healthcare domain wide adoption of such standards.

2.4.2 Data Management and Storage

This section covered a review of storage of data and data mining methodologies, as shown in Table 2-3. 'Transmission/Storage' assessed whether the data was stored persistently and if so, the structural approach used for storage, and whether data transmission was discussed. 'Data Mining/Temporal Abstractions' assessed what mining tool or approach was being applied and whether the use of temporal abstractions was adopted. 'Real- time' assessed if the data was being mined in real-time.

Findings from this section based on the systematic literature review are listed in Table 2-3, summarising physiological data management and storage.

Table 2-3. Data Management & Storage

Ref. #	Author/year	Transmission/ Storage	Data Mining/Temporal Abstractions	Real-time
A	McGregor C, Kneale B, Tracy M. 2005	Transmission of data	N/A	N/A
A	McGregor C, Stacey M. 2007	Framework to enable storage	N/A	Yes
A	Bjering H, McGregor C. 2010	Clinical and physiological stored data	yes - data mining and temporal abstractions	Yes
A	Kamaleswaran R, McGregor C, Eklund M. 2010	Clinical and physiological stored data	N/A	Yes
A	Sun J, Sow D, Hu J, Ebadollahi S. 2010	Clinical data stored in MIMIC II database obtained from ICU patients	Mention of adaptability to data mine	N/A
1	Lyman J, Scully K, Harrison J. 2008	Database & Data warehouses. Star or Snowflake schema	yes MOLAP, ROLAP	yes HL7 messaging
2	Tong C, Sharma D, Shadabi F. 2005	Large diabetic patient database	yes - Multi-Agent System, C4.5 algorithm	Yes
3	Zhang Y, Szolovits P. 2008	Although approval from review board Not clearly defined storage/database	Neural Network and See5	yes data collection and algorithm development
4	Zhang Y 2007	Hosted on a laptop with 2.4 GHz Pent. 4 Processor & 1 GB memory	Neural Network and See5	yes - learning in real-time is a feasible approach to developing alarm algorithms
5	Harrison J. 2008	Clinical databases	Promotes data mining	N/A
6	Galarraga M, Serrano L, Martinez I, de Toledo P, Reynolds M. 2007	Collection discussed but no detail on storage	N/A	yes -, BAN, PAN and HAN

Table 2-3. Data Management & Storage continued				
Ref. #	**Author/year**	**Transmission/ Storage**	**Data Mining/Temp. Abstractions**	**Real-time**
7	Holmes J. 2007	Clinical databases	KDD & IDA in biomedicine	N/A
8	Verduijn M, Sacchi L, Peek N, Bellazzi R, de Jonge E, de Mol B. 2007	N/A	Temporal abstractions	N/A

Table 2-3 Data Management & Storage

It is apparent from the reviewed research papers that structured approaches to storage and related standards are an underdeveloped area where the potential for real-time medical data to support real-time critical care is significant (Zhang, 2007, Harrison, 2008, & Holmes, 2007). Although there was no mention of a storage method in the article written by Verduijn et al. (Verduijin, 2007) one can assume that data was stored in order to enable temporal abstraction analysis.

The potential benefit use of data mining from electronically stored physiological data, for improved real-time clinical management and clinical decision, support is significant. As is the potential for clinical research on stored physiological data streams to deduce new findings for condition onset prediction indicators.

Zhang further emphasized the importance of real-time data mining by stating ‘Our expanded system of real-time data collection and algorithm development demonstrated that patient-specific learning in real time is a feasible approach to developing alarm algorithms for monitoring purposes in the ICU (Zhang, 2007)’. Although all reviewed

articles made mention of data mining with the exception of Galarraga (Galarraga, 2007), it became apparent that the healthcare domain has not adopted a standard methodology or tool to perform data mining processes on physiological data.

2.4.3 Temporal Abstractions used in CDSS (IDA)

In addition to collecting data for use in real-time by care providers, if data is stored it can be used for secondary analysis for other related clinical research. The potential for the secondary use of health data is significant. In an American Medical Informatics Association White Paper published in the Journal of the American Medical Informatics Association in 2007, entitled 'Toward a National Framework for the Secondary Use of Health Data (Safran, et al., 2007)', the urgency for infrastructures to support the secondary use of data in today's data intensive healthcare environment is seen as pivotal to the US Health system.

In support of this need, this section reviews recent health informatics research that applies computing and IT techniques to critical care within the domain of healthcare and medicine. This section presents the findings, based on systematic review of literature published recently in the area of real-time IDA and CDSS used within NICUs and ICUs in table format. Table 2-4 covered the clinical environment, the use of temporal abstraction in IDA/CDSS, real-time IDA/CDSS and if temporal abstractions (data mining) were sub-classification enabled. In this context, sub-classification represents the ability to cluster within a population. For example, not just collecting HR readings for all patients, but also clustering based on a male or female population.

Table 2-4. IDA and CDSS

Ref. #	Author/year	Clinical Environment	Data Mining/Temp. Abstractions	Real-time	Sub-classification enabled
	Apiletti D, Baralis E, Bruno G, Cerquitelli T. 2009	Hospital intensive care units.	yes, 'sliding time window' temporal context.	yes	No
A	Blount M, Ebling M, Eklund M, James A, McGregor C, Percival N, Smith K. Sow D. 2010	NICU	yes, and yes	yes	No
A	Catley C, Smith K, McGregor C, Tracy M. 2009	NICU	yes and yes	yes	No
	Kunac D, Reith D. 2005	NICU medical safety issues	CDSS	yes	No
A	McGregor C. 2010	NICU	yes to all	yes	No

Table 2-4 CDSS & IDA

The 'Clinical Environment' column indicates the CDSS and/or IDA domain specific clinical properties, which for the purposes of this survey, mainly consisted of NICU and intensive care unit (ICU). For most articles reviewed, the clinical environment predominantly took place in ICUs with the exception of the article written by Apiletti, Baralis, Bruno, and Cerquitelli (Apiletti, Baralis, Bruno, & Cerquitelli, May 2009) which presented a framework where after collection within the hospital setting, a patient's physiological data was analyzed to establish a baseline parameter and real-time ubiquitous patient monitoring was enabled via PDA alert transmission outside the hospital setting; however, this framework was not ICU specific.

All articles employed clinical decision frameworks designed for real-time analysis with the exception of the framework proposed by Kunac and Reith (Kunac & Reith, 2005); their framework was designed for the NICU patient population to minimize medication-related adverse events. The framework proposed by Kunac and Reith does enable the intake of patient parameters; however, these parameters have not been utilized in the individual analysis of the patient. By all accounts none of the articles located within this search mention the use or consideration of adoption of patient characteristic attributes to enable sub-classification clustering.

Of the reviewed articles only the $STDM^{n}_{0}$ framework designed by McGregor (McGregor C. P., July 2010) offers an innovative tool that enables mining of multidimensional temporal data as well as incorporating null hypothesis testing to allow clinical research to be conducted on historical physiological data and clinical data.

2.4.4 Patient Characteristic Multi-dimensional Algorithms

Kruger, van Oostrom and Shuster (Krueger, van Oostrom, & Shuster, 2010) published research results confirming that between 28 and 34 weeks postmenstrual age the female gender preterm infants demonstrated HRV indicative of a more mature autonomous nervous system than their male counter parts. Another article of interest entitled ‘Impacts of Age and Gender’ published interesting age and gender results, although the ‘Age’ defined within the article was not ‘gestational age’ Azhim and Kinouchi (Azhim & Kinouchi, 2009). Their research concluded that there are significant gender related differences in the statistical results from their hemodynamic data in flow velocity and pressure when comparing females with their male peers.

Of the literature reviewed under the scope of neonatal care directed at their gestational age, very little was found, although within the last five years neonatal data including gestational age has been collected by the Canadian Neonatal Network (CNN). The CNN has collected extensive data and a particular focus in their 2008 report was data collected regarding gestational age and birth weight; this information is provided in chapter 3.

During this research a neonatal algorithm was found with regards to neonatal resuscitation guidelines published in 2005 by the American Heart Association, shown in Figure 2-2. While this published, Internet-based guideline (UIhealthcare, 2008) does offer a rare model of an openly available algorithm developed for neonates, it treats all neonates with a 'one size fits all' approach and does not allow for adaptation to the growth stages that take place as gestational age increases. Within this article, only once is there an indication of criteria aligning to gestational age and that is with the general statement 'one study of preterm infants (<33 weeks of gestation) exposed to 80% oxygen found lower cerebral blood flow when compared with those stabilized using 21%' (American Heart Association, 2006). Moreover, this statement was quickly dismissed and requested to be viewed with caution due to contradictions from studied animal data collected to date. The only other mention of adaptation to gestational age at birth in these published guidelines was that of 'If a preterm delivery (<37 weeks of gestation) is expected, special preparations will be required' (American Heart Association, 2006).

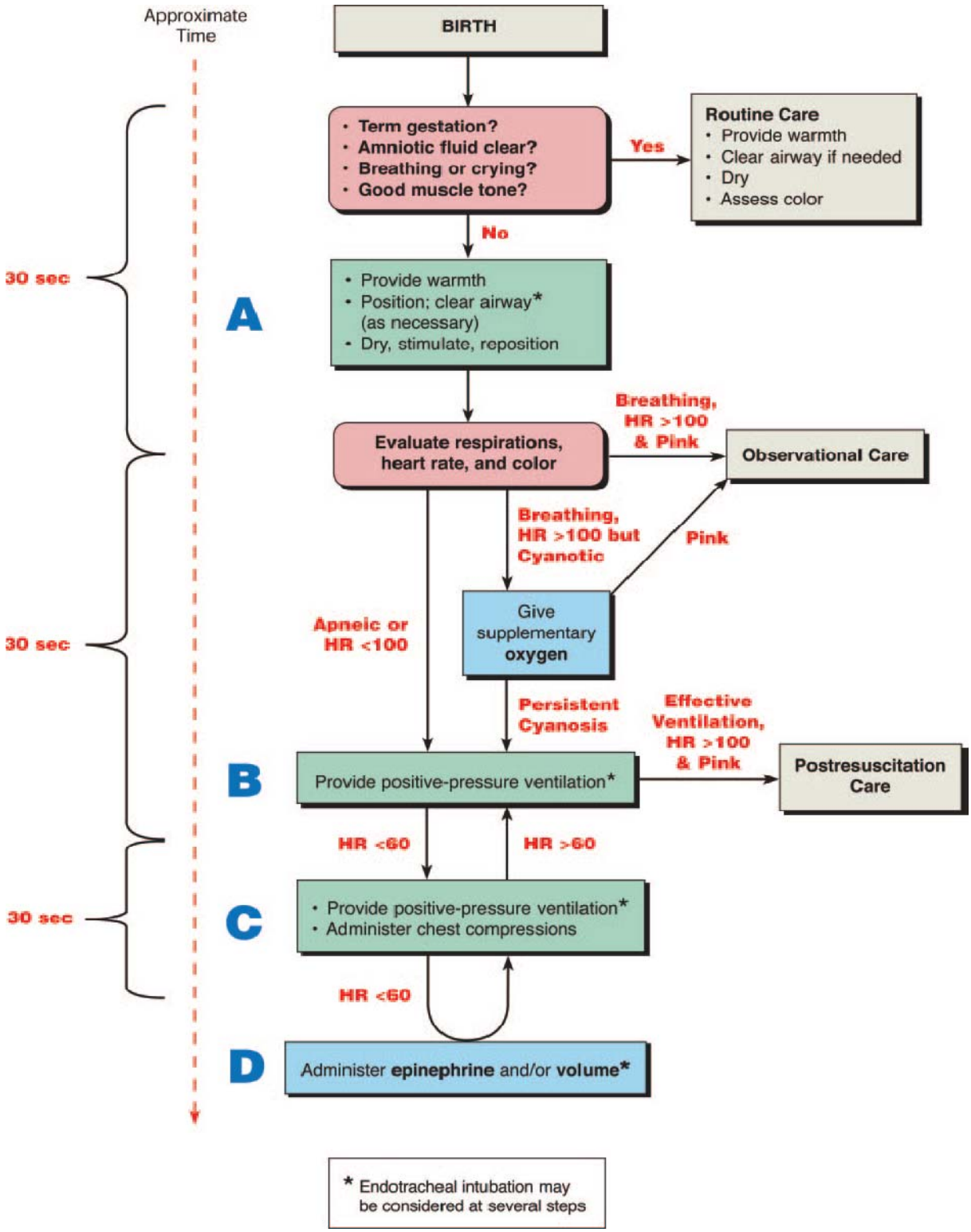


Figure 2-2 Neonatal flow algorithm

Consistently, in several articles, there are very general statements that suggest ‘Preterm babies have immature lungs that may be more difficult to ventilate and are also more vulnerable to injury by positive-pressure ventilation. Preterm babies also have immature blood vessels in the brain that are prone to haemorrhage; thin skin and a large surface area, which contribute to rapid heat loss; increased susceptibility to infection; and increased risk of hypovolemic shock caused by small blood volume’ (American Heart Association, American Academy of Pediatrics, 2006). What is not being defined is that as the fetus is growing and developing and gestational age increasing, the delicate nature and vulnerability is decreasing.

In Figure 2-2 there is mention of an apnoeic occurrence or if HR < 100 but there is little detail as to treatment procedures that should follow. The algorithm does not provide patient characteristic defined parameters, such as whether the HR threshold should be adjusted dependant on neonate gender or gestational age.

Compared with Figure 2-3, Mishra et. al, published a more detailed apnoea algorithm in 2007 (Mishra, Agarwal, Jeevasankar, Aggarawal, Deorari, & Paul, 2007), as seen in Figure 2-3; however, patient characteristic detail was still not incorporated.

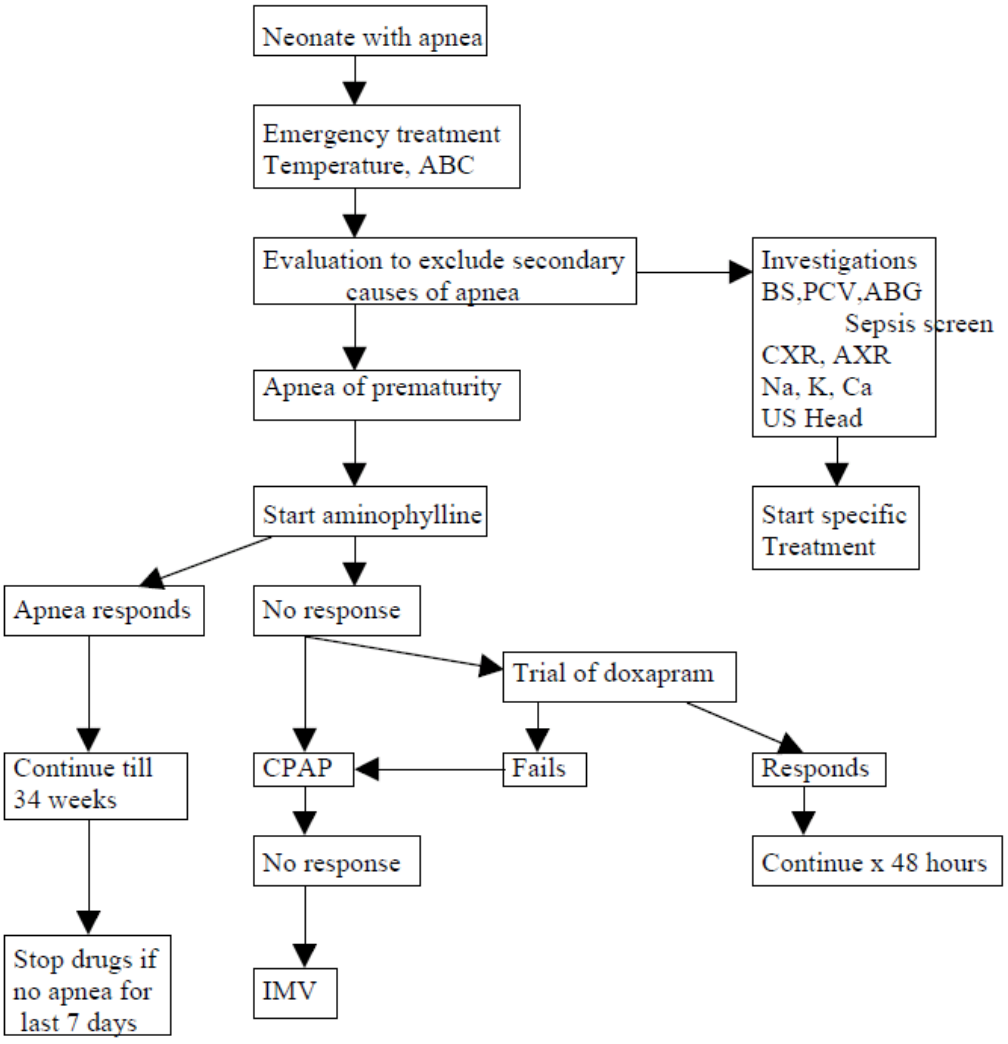


Figure 2-3 Algorithm for management of neonatal apnoea

Although much investigation and research is in the hands of the CNN and their collaborative team members regarding GA and/or birth weight, studies tend to examine these attributes as isolated instances rather than developing algorithms or rules incorporating both.

2.5 Discussion

The domain of healthcare and medicine has been noted as one that has been less receptive than other industry domains in the adoption of new computing and IT approaches (Wu, Wang, & Lin, 2005). However, since 2005 this trend does show promising signs of changing. The real-time physiological data focused on in this literature review can be received at speeds of up to 1000 readings a second. Hence determining a format where this data can be processed in real-time but also stored for persistent storage is not a trivial problem. As stated previously, the only research that provides a detailed mechanism for persistent storage was that of McGregor, but this did not detail the format of the data within the stored tables (McGregor C. P., July 2010).

Opportunities abound to utilize this data for secondary use, yet this will require that the data be formatted and stored in a format accessible by statistical and other data mining approaches. As a result, opportunities exist to consider the input and output of physiological data from a database utilizing a framework to enable a flexible open environment to receive data from the medical devices and have this data utilized for secondary use.

2.6 Conclusion

This chapter has presented a health informatics literature review assessing the ability of recent computing and IT research as applied to physiological monitoring to support the service of critical care. The literature review considered the service of care concept based on an architectural approach, with a focus on the collection, and storage of

physiological data to enable data mining, and temporal abstractions to assist in the creation of clinical rules through IDA to be adopted by CDSS.

While there have been several efforts to utilize computing and IT to create advances in the techniques used for the analysis of physiological monitoring data there has not been a focus on these techniques from the perspective of supporting the service of critical care. Further there has been an absence of research focused on supporting paradigm shifts in the service of critical care to support the use of this approach for storage of data, the data structures within that storage, and also its reuse to support clinical research.

Potential opportunities exist within the NICU domain to adopt a standard when considering storage of data and algorithms not only as they are developed but also how they are developed through use of retrospective clinical data. Once algorithms are developed with the inclusion of patient characteristics there is then the potential for deployment to provide clinical decision support based on real-time streaming of continuous data.

The work presented in this chapter supports our continued research on next generation healthcare solutions to support the service of critical care, real-time clinical management and clinical research, by adopting patient characteristic attributes into diagnosing algorithm. The review demonstrated that such frameworks do not yet exist. These finding are also in support of hypothesis 1 and 2:

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multi-dimensional temporal data.*

Chapter 3 – The NICU environment

3.1 Introduction

The accepted duration of human pregnancy is 40 weeks which is calculated from the last menstrual period and can be calculated as 280 days or approximately 266 days from conception to birth (CIHI, 2009). A birth that takes place after 37 weeks of pregnancy is considered full term.

Unfortunately, not all pregnancies last full term and some infants are brought into this world 'preterm'. According to the 2009 International Committee for Monitoring Assisted Reproductive Technology (ICMART) and World Health Organization (WHO) revised glossary defines a preterm birth as 'a live birth (or stillbirth) that takes place after at least twenty but before thirty seven completed weeks of gestational age' (Zegers-Hochschild, et al., November 2009). The term neonatal defines 'the time interval that commences at birth and ends 28 days after birth' (Zegers-Hochschild, et al., November 2009).

There are two brilliant historical figures well known for their premature births. One is Johannes Kepler, a German mathematician, astronomer and astrologer who 'was born on December 27, 1571, a premature child and according to his own records, the pregnancy lasted 224 days, 9 hours and 53 minutes' (~32 week GA) (Fowler, 1996). The other figure is, 'one of the founders of classical physics, and one of the greatest known scientists of all time, Sir Isaac Newton' whom was also born premature (Terry, 2009). Their contributions to mankind help to prove worthiness of the plight to improve medical attention to all preterm births for the best possible life outcomes.

Premature infants can be born up to seventeen weeks early and may only weigh 450 grams; they can spend three or four months in intensive care and may develop several conditions before discharge. During their time in the NICU, large quantities of physiological data are collected that are for the most part, not being captured, stored or used for potentially obverting the onset of illnesses. Premature infants, by the time they are discharged, can increase in body mass by as much as six times, and have had several medical diagnoses and treatments (McGregor, Kneale, & Tracy, 2007), many of which may have long term implications for the future health of the individual. In addition, 15% of neonatal intensive care admissions are transferred after delivery from smaller regional or remote hospitals without intensive care facilities to larger tertiary referral or Children's Hospitals with NICUs. Similar conditions apply within Australia, New Zealand, Canada and the USA where small non-tertiary units are spread throughout the country (McGregor, Kneale, & Tracy, 2007) (McGregor, Stacey, 2007).

A commonly used device that senses and displays physiological monitoring data within NICUs is the Philips IntelliVue series of monitors. These are monitoring devices that accept multiple sensor modules, and that enable data to be collected and displayed from multiple types of sensors attached to the patient. Data can be extracted from this device via an Ethernet or RS-232 serial port at the rear of the device and each module can produce multiple data streams of the types described below (Blount, et al., 2010):

- numeric: one reading every 1024 milliseconds (ms).
- fast wave: 4 sets of 256 values for every 1024 ms

These devices provide readings ranging between one a second to 1000 readings a second. This data stream format is not catered for within HL7 (McGregor & Smith, 2009).

Stravroudis, Miller and Lehmann refer to the NICU as ‘a clinical environment burdened with challenges that frequently lead to adverse outcomes (Stravroudis, Miller, & Lehmann, 2008). The survival of a preterm baby is dependent on assistance from multiple medical devices and clinician’s interpretation of data readings from such devices. The NICU is an extrauterine environment which is designed to simulate the intrauterine environment, with the main objective of achieving a healthy fully developed human being with minimal lifelong health complications: ‘The severity and critical nature of illness, intricacy of treatment, immaturity of the newborn physiology, difficulty of multidisciplinary care, complexity of communication, and the changing technology that continues to shape and advance neonatal care make neonates a unique and vulnerable patient population’ (Stravroudis, Miller, & Lehmann, 2008).

Preterm births are on the rise (March of Dimes Foundation, 2010) and with modern medical intervention neonates now have an increased chance at survival. Preterm neonates born at $\geq$25 weeks gestation and $\geq$600 g have a survival rate of 60 %. As many as 50% of the survivors have no evidence of severe disability defined as non-ambulatory cerebral palsy, mental retardation, severe visual or hearing deficit, or a combination of these neurodevelopmental impairments upon long term follow-up (Seri & Evans, 2008).

3.2 The Canadian Context

A recently published article in the Globe and Mail revealed that although Canada was once able to boast about its high world ranking for low infant mortality, Canada has recently dropped to twenty fourth place from sixth (Priest, 2010),(WHO, 2011). It was reported a year earlier by the Canadian Press in January of 2009 that 'Preterm births in Canada have jumped a staggering 25 percent over the past 10 to 12 years' and that this trend is not isolated to Canada and results of likeness are found in the U.S. and Europe (Branswell, 2009). Clearly, caring for premature infants is an important goal, both within Canada and internationally.

There are many studied factors that can lead to preterm births, including maternal factors such as age, previous preterm birth, smoker, gestational diabetes, multiple births, as well as social, environmental, and economical factors; however, the focus of this research is of a clinical nature. That being stated, how can we treat, predict and diagnosis illnesses following the occurrence of a preterm birth? With the goal in mind of earlier diagnosis of ailments, possibly avoiding lifelong morbidity, the outcome can only prove to be less costly on the healthcare system. The healthcare cost impact of a preterm birth is considerable. 'For example, a baby born at 750 grams will need, on average $120,000 worth of healthcare before heading home from the hospital when comparing to a full term baby that will cost the system under $1000' (Branswell, 2009).

Across Canada the neonatal-perinatal care system is regionalized. Within each of these regions hospitals are separated into three different levels of care. Patients are referred

or transferred to these different care level facilities depending on their conditions and care level needs. There are three designations that describe the levels of care: level 1 (normal newborn care), level 2 (high-dependency care) and level 3 (intensive care) (CPS, 2006). These levels were further defined as posted by CPS in June of 2006;

Level 1: Basic neonatal care (normal newborn nursery)

Level 1a

- Evaluation and postnatal care of healthy newborn infants; and
- Phototherapy

Level 1b

- Care for infants with corrected gestational age greater than 34 weeks or weight greater than 1800 gram who have mild illness expected to resolve quickly or who are convalescing after intensive care;
- Ability to initiate and maintain intravenous access and medications;
- Gavage feeding; and
- Nasal oxygen with oxygen saturation monitoring (e.g., for infants with chronic lung disease needing long-term oxygen and monitoring).

Level 2: High-dependency neonatal care (special care newborn nursery)

Level 2a

- Care of infants with a corrected gestational age of 32 weeks or greater or a weight of 1500 gram or greater who are moderately ill with problems expected to resolve quickly or who are convalescing after intensive care;
- Peripheral intravenous infusions and possibly parenteral nutrition for a limited duration;
- Resuscitation and stabilization of ill infants before transfer to an appropriate care facility; and
- Nasal oxygen with oxygen saturation monitoring (e.g., for infants with chronic lung disease needing long-term oxygen and monitoring).

Level 2b

- Mechanical ventilation for brief, usually less than 24 hour, durations or continuous positive airway pressure; and
- Intravenous infusion, total parenteral nutrition, and possibly the use of umbilical central lines and percutaneous intravenous central lines.

Level 3: Intensive neonatal care (neonatal intensive care nursery)

Level 3a

- Care of infants of all gestational ages and weights;
- Mechanical ventilation support, and possibly inhaled nitric oxide, for as long as required; and
- Immediate access to the full range of subspecialty consultants.

Level 3b

- Comprehensive on-site access to subspecialty consultants;
- Performance and interpretation of advanced imaging tests, including computed tomography, magnetic resonance imaging and cardiac echocardiography on an urgent basis; and
- Performance of major surgery on site but not extracorporeal membrane oxygenation, hemofiltration and hemodialysis, or surgical repair of serious congenital cardiac malformations that require cardiopulmonary bypass.

Level 3c

- Extracorporeal membrane oxygenation, hemofiltration and hemodialysis, or surgical repair of serious congenital cardiac malformations that require a cardiopulmonary bypass (CPS, 2006).

Within and across Canada there is an organization comprised of a group of researchers, consisting of 29 hospitals and 17 universities, who collaborate on research issues relating to neonatal care, known as the Canadian Neonatal Network (CNN.org). According to the CNN 2008 Annual report there were 13, 401 neonate admissions reported by 26 Canadian NICUs between January 1, 2008 and December 31, 2008. Of those preterm births reported, 4221 were considered very preterm which, is less than 33 weeks gestational age. There were 2830 of the total admissions that were of very low birth weight (VLBW) which is less than 1500g at birth (CNN, 2008).

Data collected during this time period from the twenty six NICU's is shown in tables and figures below:

Table 3-1: Gestational age at birth (CNN, 2008 , pp. 12-50)

Gestational age in completed weeks at birth	Frequency	Percent	Cumulative Percent
<23	12	0.1	0.1
23	51	0.4	0.5
24	177	1.3	1.8
25	257	1.9	3.7
26	279	2.1	5.8
27	355	2.7	8.4
28	428	3.2	11.6
29	498	3.7	15.4
30	616	4.6	20.0
31	685	5.1	25.1
32	863	6.4	31.5
33	888	6.6	38.1
34	1219	9.1	47.2
35	1099	8.2	55.4
36	972	7.3	62.7
37	889	6.6	69.3
38	1190	8.9	78.2
39	1100	8.2	86.4
40	1208	9.0	95.4
41	572	4.3	99.7
≥42	38	0.3	100.0
Total included	13396	100.0	
Total # of missing (GA)	5		
Total # of infants	13401		

Figure 3-1: Gestational age at birth and survival to discharge from participating NICU's

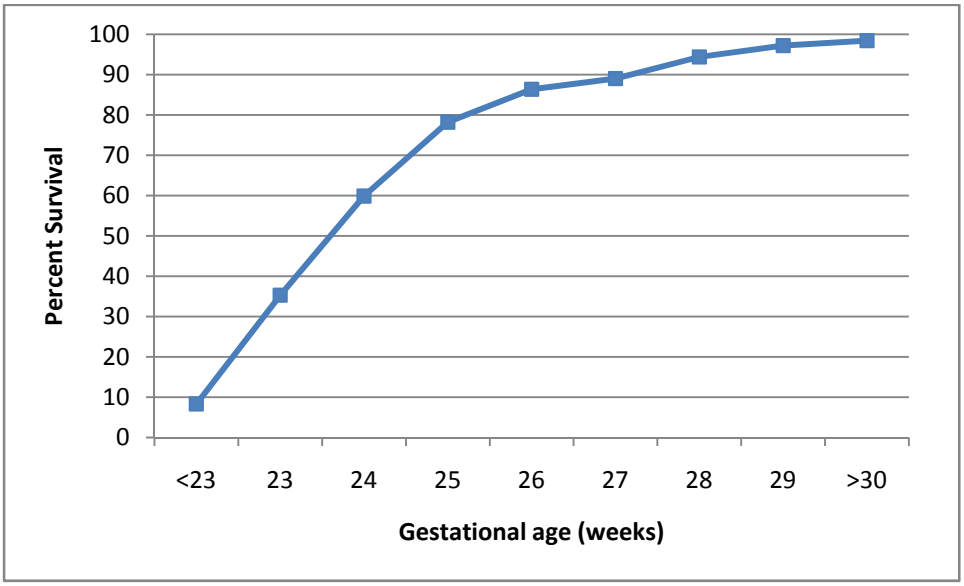
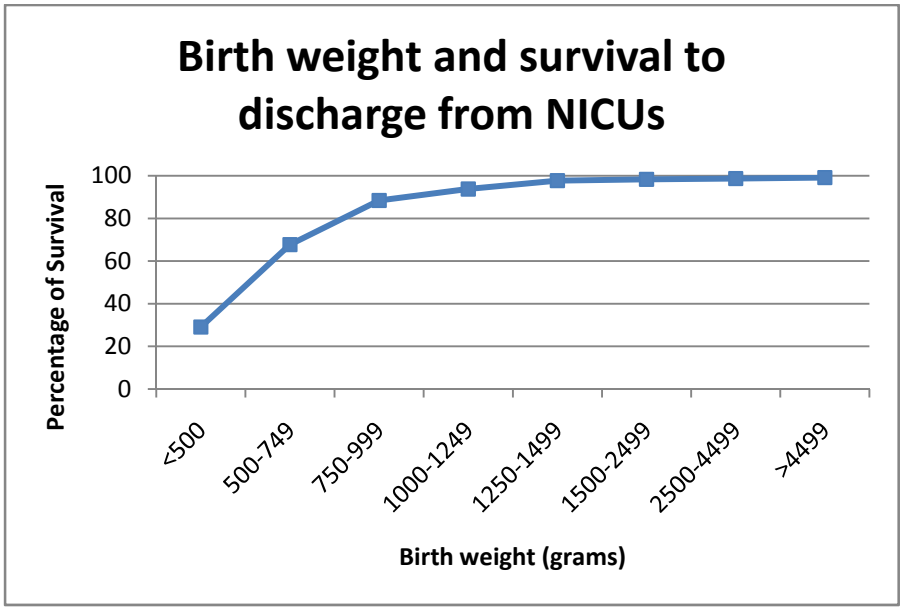


Table 3-2: Birth weight and survival discharge from participating NICU's

Birth weight (grams)	Number of infants	Number of survivors	% survival
<500	31	9.0	29
500-749	402	272.0	68
750-999	679	600.0	88
1000-1249	861	807.0	94
1250-1499	857	837.0	98
1500-2499	4522	4445.0	98
2500-4499	5790	5713.0	99
>4499	219	217.0	99
Total included	13361	12900.0	97
Missing (BW)	40		
Total # of infants	13401		

Figure 3-2: Birth weight and survival to discharge from participating NICU's



The gestational age (GA) and birth weight of concern as unveiled in data provided above would lead one to believe that a birth prior to 30 week GA combined with a birth weight of less than 1000 grams has less chance of survival and increased morbidity. That being said much research is needed in these ranges to further dissect important developmental events occurring during this time and to minimize lifelong detrimental effects caused by illnesses and underdevelopment (CNN, 2008).

Although there are advances being made in the neonatal mortality rates across Canada (Sankaran, et al., 2002), there remains research opportunity to improve critical care by the incorporation of gender and gestational to further refine condition onset detection remains a research area for apnoea of prematurity.

3.3 Medical Devices and NICU monitoring

Physiological monitoring, through a diverse range of devices, is used extensively within intensive care units worldwide. These devices provide an excellent example of high-frequency, high-volume, highly dimensional real-time data which not only have the ability to display physiological data but are able to output this data via serial, ethernet, USB or other ports. Depending on the settings of a multi-module device, for example the Phillips IntelliVue MP70, at the time of use each sensor has the ability to produce multiple data streams, such as numeric, wave, and fast wave (i.e., electrocardiogram, ECG). Figure 3-3 illustrates the myriad of medical devices used within a NICU.

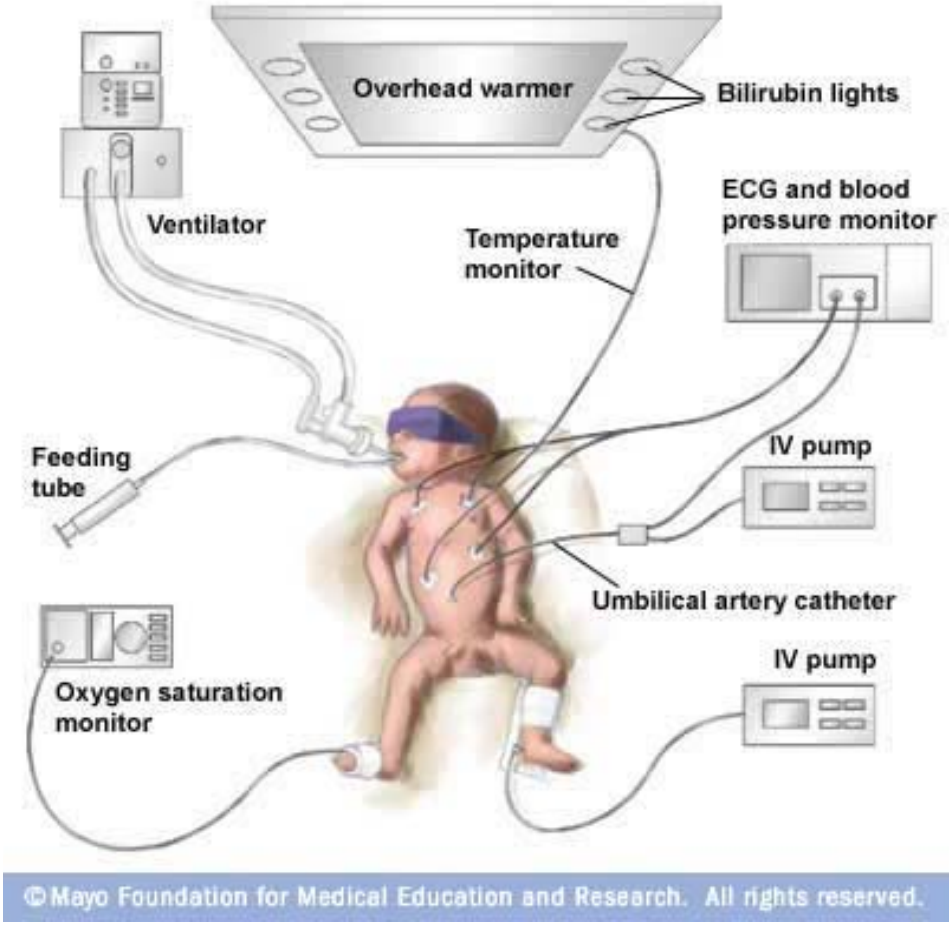


Figure 3-3 NICU medical monitoring devices

3.4 Gender and Gestational Age driven thresholds

Stravroudis, Miller and Lehmann have stated that: ‘Neonates are at further risk for harm from medication errors because of rapidly changing body size parameters (over the course of a hospitalization an infant may double or triple their birth weight); off-label drug usage; inability to communicate with providers; and changing developmental systems affecting drug absorption, distribution, metabolism, and excretion ‘(Stravroudis, Miller, & Lehmann, 2008; Stravroudis, Miller, & Lehmann, 2008). In addition, recent literature states that gender plays a significant role in defining HR differences: newborn male infants have lower baseline HR than newborn females (Nagy & Orvos, 2000) (Krueger, van Oostrom, & Shuster, 2010). These findings suggest that the known gender-related HR differences that are apparent throughout life are also present at the very beginning of life. At such an early stage in life a difference of eight beats per minute may seem like a small physiological value but significant enough that it should be considered when investigating physiological markers for conditions that affect the health and development of the newborn infant (Nagy & Orvos, 2000). The two patient identifiers this research proposes for inclusion are that of gestational age and gender to improve accuracy of diagnosis, treatment and critical care of neonates.

When considering the adoption of gestational age into thresholds, accepted anecdotal evidence matches the mean arterial blood pressure (MBP) to be that of a neonate’s current gestational age (e.g., 24 mmHg for 24 weeks gestational age) (Blount, et al., 2010).

Another physiological value that is related to the neonates' gestational age is that of systolic blood pressure, as seen below in Table 3-3 (UIhealthcare, 2008). Current device thresholds do not incorporate the gestational age trajectory of development when creating alerts.

Table 3-3 Gestational Ages and Systolic Blood Pressure (UIhealthcare, 2008)

	Gestational age, weeks													
Body weight, kg	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
0.80	43	44	44	45	45	46	46	47	47	48	48	49	49	50
0.90	44	45	45	46	46	47	47	48	48	49	49	50	50	51
1.00	45	45	46	46	47	47	48	48	49	49	50	50	51	51
1.10	46	46	47	47	48	48	49	49	50	50	51	51	52	52
1.20	46	47	47	48	48	49	49	50	51	51	52	52	53	53
1.30	47	48	48	49	49	50	50	51	51	52	52	53	53	54
1.40	48	49	49	50	50	51	51	52	52	53	53	54	54	55
1.50	49	49	50	51	52	52	53	53	54	54	55	55	56	56
1.60	50	50	51	51	52	52	53	53	54	54	55	55	56	56
1.70	51	51	52	52	53	53	54	54	55	55	56	56	57	57
1.80	51	52	52	53	53	54	54	55	55	56	56	57	57	58
1.90	52	53	53	54	54	55	55	56	56	57	57	58	58	59
2.00	53	53	54	54	55	56	56	57	57	58	58	59	59	60
2.10	54	54	55	55	56	56	57	57	58	58	59	59	60	60
2.20	55	55	56	56	57	57	58	58	59	59	60	60	61	61
2.30	55	56	56	57	57	58	58	59	59	60	60	61	61	62
2.40	56	57	57	58	58	59	59	60	60	61	61	62	62	63

Add for a post-natal age of:									
Hours	3to7	8to12	13-18	19-24	25-32	33-40	41-54	55-89	90-96
mmHg	1	2	3	4	5	6	7	8	7

* The thick contour includes values pertaining to the area of body/gestational age characteristics of the population in the study*

3.5 GA case study parameter based thresholds alerts case study introduction

Benjamin and Stoll state that early-onset sepsis presents itself in the first 24 hours for over 90% of neonates in the NICU (Benjamin & Stoll, 2006). Some of the signs of sepsis (and nosocomial infection) are often 'non specific and subtle' and can include: temperature instability, lethargy, irritability, apnoea, respiratory distress syndrome, hypotension, bradycardia, tachycardia, cyanosis, abdominal distension, hyperglycaemia, jaundice and feeding intolerance (Benjamin & Stoll, 2006).

Neonates are vulnerable to numerous clinical diagnoses that can have a lifelong state of morbidity; this can depend greatly on their gestational age at birth. One such GA-related illness is respiratory distress syndrome (RDS) that can have detrimental effects on a preterm infant (McMillan, Feigin, DeAngelis, & Jones, 2006). Although alveoli first appear at 28 weeks gestation, lung maturation is usually not adequate enough to sustain extrauterine life without some form of respiratory support until 32 to 34 weeks gestation (McMillan, Feigin, DeAngelis, & Jones, 2006). Infants born prior to 28 weeks gestational age may have experienced more complications due to decreased development of the lung resulting in insufficient surfactant (McMillan, Feigin, DeAngelis, & Jones, 2006).

Periventricular leukomalacia (PVL) is a very common brain injury which occurs most commonly amongst neonate population younger than 32 weeks gestational age at birth. The diagnosis of PVL is concerning for clinicians as a 'significant percentage of surviving premature infants with PVL develop cerebral palsy (CP), intellectual impairment, or visual disturbances' (Zach, Brown, & Kaftan, 2010).

Although there is considerable variation in the incidence and severity of apnoea of prematurity, both are inversely related to gestational age (Mayock, 2009). Any apnoea event in the term infant[1] is considered abnormal. In addition to direct complications, apnoea is a multi-factorial problem and can be a clinical manifestation of conditions such as nosocomial infection.

3.5.1 Apnoeic Definition and background information for case study.

Apnoea is a Greek based word meaning 'without' representing a 'temporary absence or cessation of breathing' (Oxford Dictionary, 2011). Apnoea describes an event of no breathing, which translates into intervals between breaths. Apnoea is very common in preterm infants born up to 35 weeks gestational age and is termed Apnoea of Prematurity (AOP) (Stanford_University, 2010). Apnoeic spells can occur as a result of a number of developmentally immature organ functions resulting from manifestation of prematurity. Approximately 70% of babies born before 34 weeks of gestation have clinically significant apnoea, bradycardia, and/or oxygen desaturation during their hospitalization.

[1] Term is defined as $\geq$ 37 weeks gestational age.

Previous literature has shown a direct correlation between AOP and the infant's stress level (McMillan, Feigin, DeAngelis, & Jones, 2006). Apnoea is associated with and indicative of a multitude of other health concerns the preterm infant may be experiencing such as: infection (NI), anaemia, low oxygen levels, heart or lung problems, temperature problems, overstimulation and feeding problems (NIH, 2010).

AOP can be classified into three groups: central, obstructive and mixed.

- Central apnoea is defined as the cessation of both airflow and respiratory effort.
- Obstructive apnea is the cessation of airflow in the presence of continued respiratory effort.
- Mixed apnea contains elements of both central and obstructive apnea, either within the same apnoeic pause or at different times during a period of respiratory recording (Nimavat, Sherman, Santin, & Protat, 2009).

Central apnoea occurs when the respiratory system does not fire for up to 15 seconds. Obstructive apnoea occurs when respiratory central system is firing but due to a blockage, air is does not flow into the lungs. When this occurs the infant is exerting more effort to breathe and the breathing rate may increase. During this time there will be breathing efforts but no air flow into the lungs and therefore blood oxygen saturation will decrease. Upon recovery, saturation levels will improve. Mixed apnoea begins with obstructive followed by central (Nimavat, Sherman, Santin, & Protat, 2009). It has been shown that: 'Mixed apnea accounts for about 50% of all cases of apnoea in premature neonates; about 40% are central apneas, and 10% are obstructive apneas' (Nimavat, Sherman, Santin, & Protat, 2009).

In this work, three scenarios will be investigated (and demonstrated in chapter 6) for each of these different classifications of apnoea, listed as follows:

1. Occurrence of respiration pause for 15 seconds;
2. Respiration pause continues up to 20 seconds and is accompanied by a decrease in blood oxygen saturation but breathing recommences followed by return to accepted levels of oxygen saturation;
3. Respiration pause continues up to 25 seconds and is accompanied by a decrease in oxygen saturation followed by a decrease in HR.

3.6 Research relevance to NICU environment

Based on the reviewed literature there is an opportunity for gender and gestational age to be incorporated in the creation of patient characteristic multidimensional algorithms to assist in individualizing data mining performed with the intent of early detection at the onset of conditions for each neonate in critical care. Apnoea is a significant clinical event that is seen as a precursor to nosocomial infection and warrants investigation as is demonstrated in the apnoea case study example presented in Chapter 6.

There is clinical motivation that the missing patient characteristic functionality proposed is required within the clinical context. Neonates are susceptible to multiple complications and illnesses and are a population that could benefit from the proposed $STDM^{n+p}_{0}$ multidimensional data mining framework discussed and defined in Chapter 5.

This chapter has introduced the context for the case study to be demonstrated in chapter 6 and supported the motivation for hypothesis 3, which is:

> *3. The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.*

Utilizing a previously developed framework and further extending it by defining physiological behaviours by gender and gestational age is the motivation and focus of this thesis. This chapter has provided the context for the case study demonstration in chapter 6 of this thesis.

Chapter 4 – Design and Methodology of Existing Framework

4.1 Introduction

As mentioned in chapter 1, this research builds on the previous $STDM^n_0$ framework (McGregor C. P., July 2010), described. This chapter will describe in detail the existing Service based Temporal Abstraction Multi-Dimensional Data Mining $STDM^n_0$ framework.

4.2 Existing Temporal Abstractive Multi-Dimensional Data Mining $STDM^n_0$ framework

The framework utilizes components based on research first demonstrated within Foster's (Foster and McGregor 2005) multi-agent system, which in turn was then extended to facilitate the tasks needed in the $STDM^n_0$ framework. Heath's 2006 research (Heath, 2006) then extended the CRISP-DM data mining model to facilitate null hypothesis testing. The *extended* CRISP-DM model was then integrated into the extended multi-agent framework based on Bjering's thesis (Bjering H. , 2008) to complete the tasks of the of the $STDM^n_0$ framework, as shown in Figure 4-1 following.

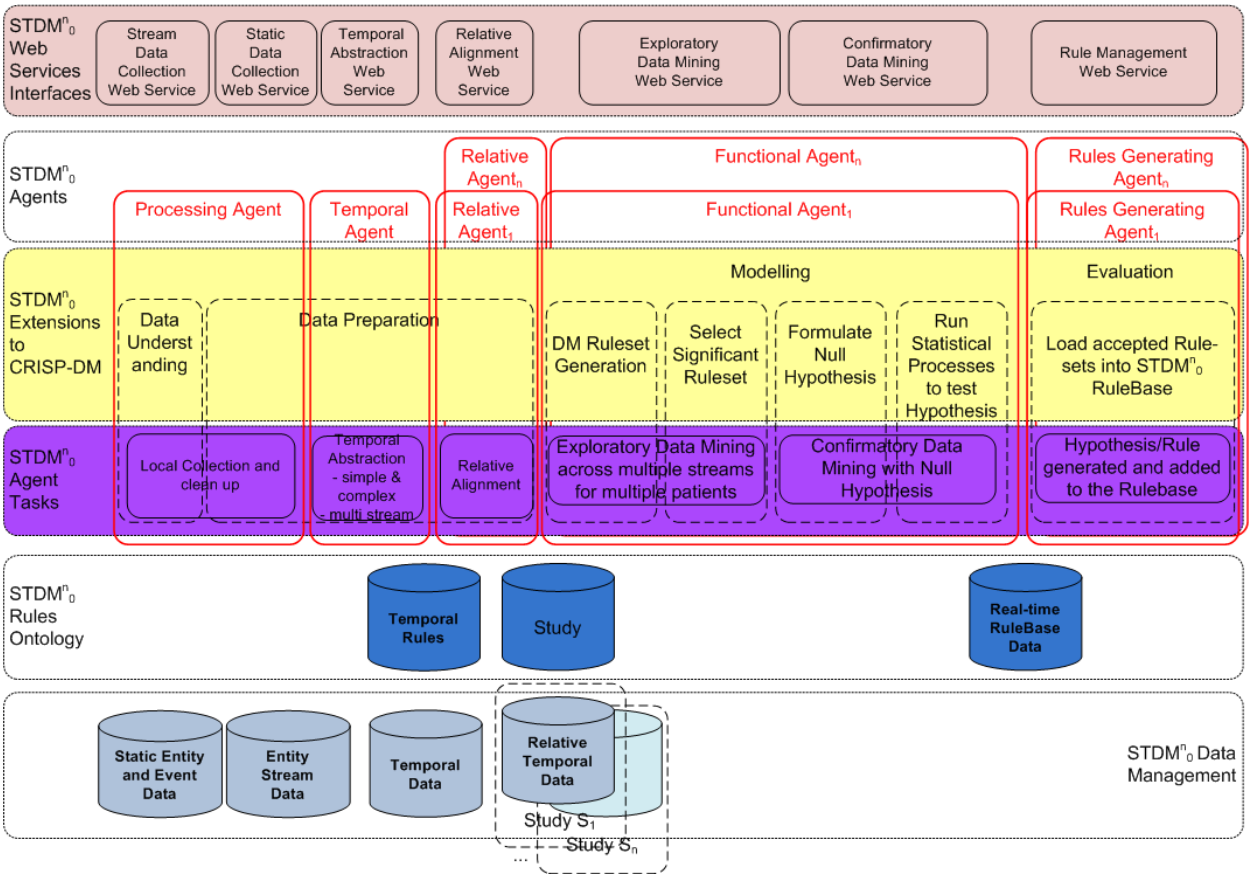


Figure 4-1 STDM$^n_0$ Framework (McGregor, 2010)

4.3 STDM$^n_0$ Background

The research intent behind the existing STDM$^n_0$ framework was to bring together clinical management and clinical research in an environment that would enable the secondary use of data created by the myriad monitoring devices utilized within a NICU. Three layers within this framework are the focus area of this research; the multi-agent layer, the extended CRISP-DM data mining layer which defines the data mining tasks, and STDM$^n_0$ framework task layer. It is within these layers where further extensions are proposed by this research thesis. The following sections discuss the individual framework components in detail.

4.3.1 Processing Agent Background

The Processing Agent performs such tasks as the preparation of data retrieved from and stored in physiological data and clinical data external databases. This data will later be processed by the Temporal Agent and the functional agent. Both the Data Understanding and Data Preparation phases of the extended CRISP-DM version called CRISP-TDM model are supported within this agent (Bjering & McGregor, 2010).

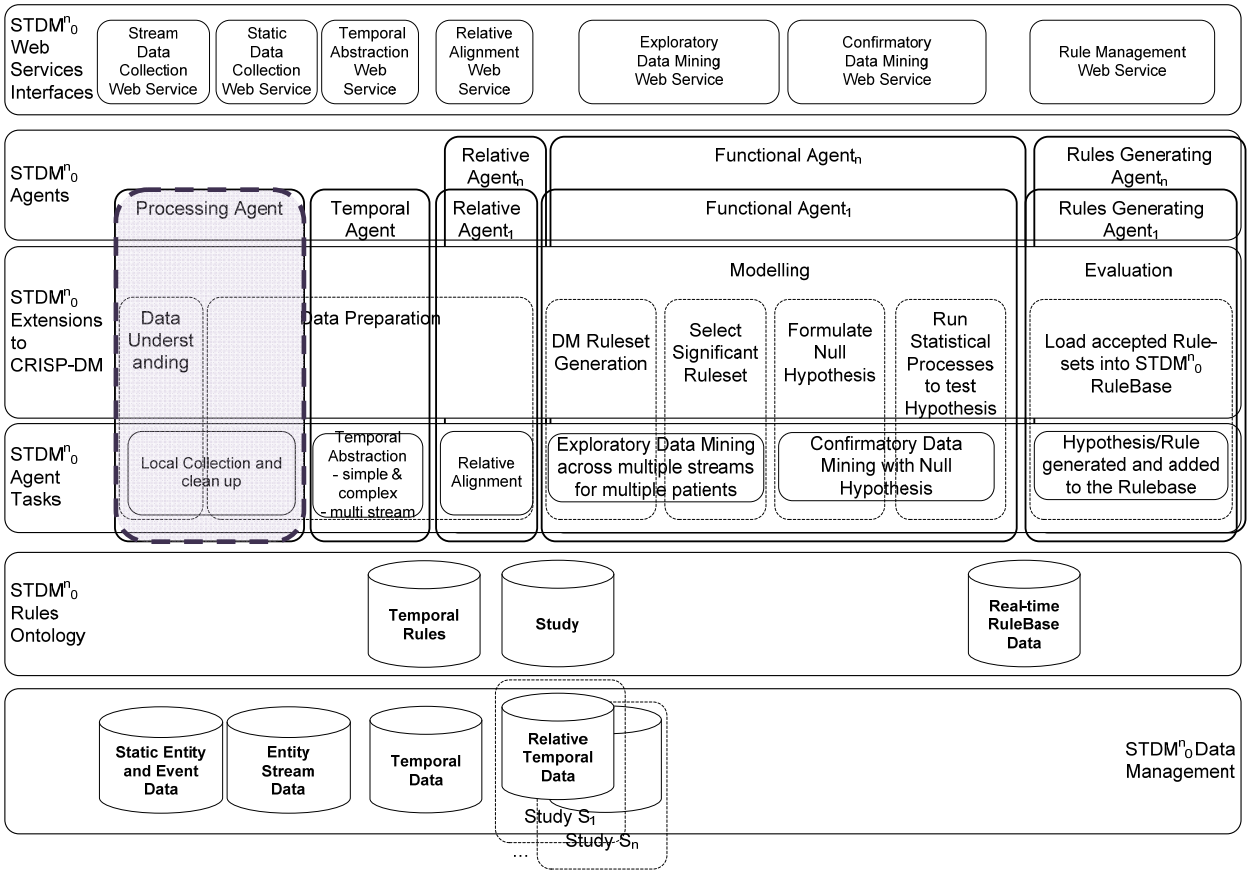


Figure 4-2 Processing Agent in Existing STDM$^n_0$ framework

4.3.2 Temporal Agent Background

The Temporal Agent processes new physiological data entering the framework, creating temporal abstractions as defined by temporal rules in the system. The temporal abstraction process is a preprocessing method before data mining which allows the

temporal aspects and the context of the data to be preserved (McGregor C. P., July 2010).

The Temporal Agent has six main functions:

1. Retrieve the physiological data from the physiological data store for each parameter for each patient
2. Retrieve the relevant abstraction rules from the temporal rules table
3. Apply the rules to the physiological data, creating simple abstractions for individual data streams for individual patients.
4. Store the created abstractions in the $STDM^{n}_{0}$ temporal data store.
5. Create complex abstractions from the simple abstractions created in step 3, according to any rules found in the temporal rules table.
6. Store any complex abstractions created in the $STDM^{n}_{0}$'s temporal data store. (McGregor C. P., July 2010).

Within the extended CRISP-TDM model the temporal agent performs tasks that are part of the preparation phase.

4.3.3 Relative Agent Background

The Relative Agent uses abstractions generated by clinical study investigations of individual patients created by the temporal agent. For instance, pre-diagnosis studies look for new trends and patterns that can be indicative of the onset of a condition within the patient's physiological data streams. Within this agent, realigning the time of abstraction involves the crucial utilization of both start and end times relative to the particular point of interest. The point of interest is the time the patient was diagnosed

with a condition across multiple streams of physiological data in comparison with multiple patients with the same diagnosis.

To enable the detection of particular patterns of these abstractions, at a particular time before diagnosis, realignment of the abstraction relative to the time of diagnosis is necessary. As these abstractions are using absolute time for the start and finish time for each abstraction, it will usually be necessary to give these abstractions start and finish times relative to the particular event of interest. This will enable the comparison and mining of the abstractions, allowing the distance from diagnosis, or other event, to be taken into account as shown in Figure 4-3. Within the extended CRISP-TDM model the Relative Agent performs tasks that are part of the evaluation phase.

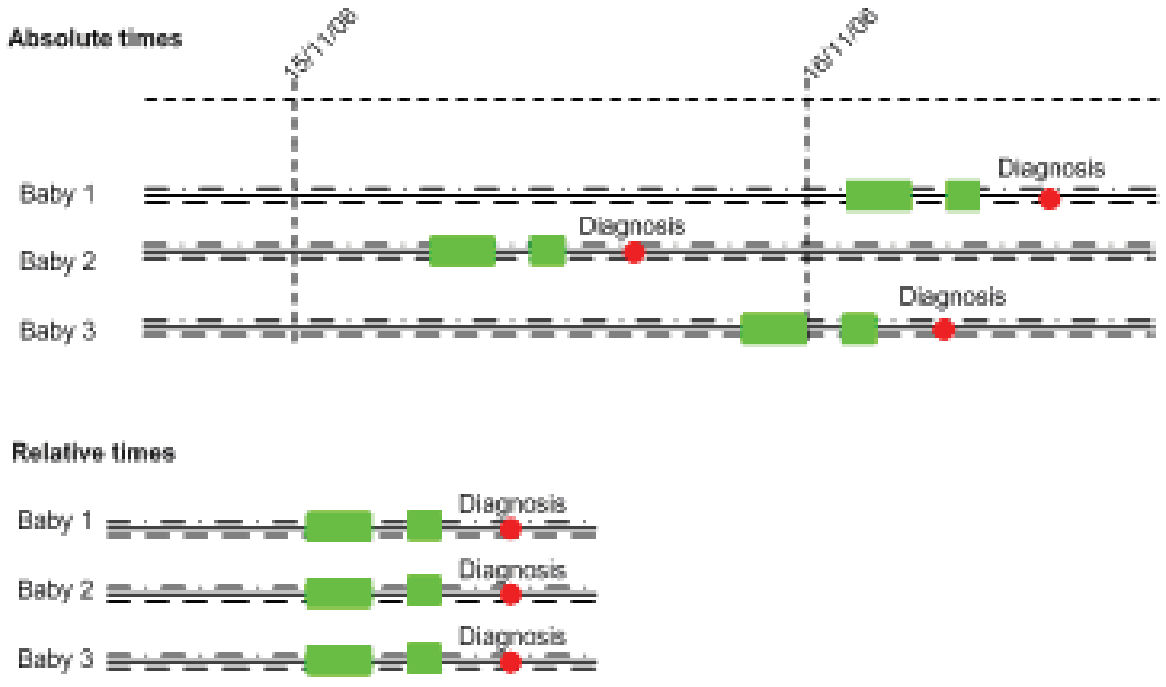


Figure 4-3 Realignment of Abstracted Parameters relative to diagnosis

4.3.4 Functional Agent Background

The Functional Agent is used to facilitate the modeling tasks of the CRISP-TDM model, including rule set generation through exploratory data mining, selecting significant rule sets, null hypothesis formulation and running statistical processes to test the null hypothesis during confirmatory data mining (Bjering & McGregor, 2010).

4.3.5 Rules Generating Agent background

The motive behind searching through historical data is to detect patterns that might lead to new hypotheses that can then be defined as rules for the purpose of intelligent patient monitoring. The function of the Rules Generating Agent is to adopt and translate discoveries through hypotheses created by exploratory data mining made by the functional agent into rules which are then stored within the rules database.

At the completion of exploratory data mining, confirmatory data mining commences. Once the null hypothesis has been rejected, clinicians and researchers will assess whether or not to include the hypotheses into the rules database. Rules created and stored within the rules generating agent will be utilized by an event stream processor in physiological monitoring (Bjering & McGregor, 2010).

4.4 Proposed Extension to Existing Framework

The existing $STDM^{n}_{0}$ framework has been adopted and utilized to provide the knowledge extraction component in an ongoing research collaborative project called Artemis, which has adopted an IBM event stream processing platform called Infoshpere for real-time monitoring for all patients in the NICU.

Although the usefulness of this dynamic framework has proved its worthiness, there is potential to extend its functionality to enable research that can support the exploration and clustering based on patient characteristics. The research focus of these proposed extensions within the existing framework are illustrated in Figure 4-4.

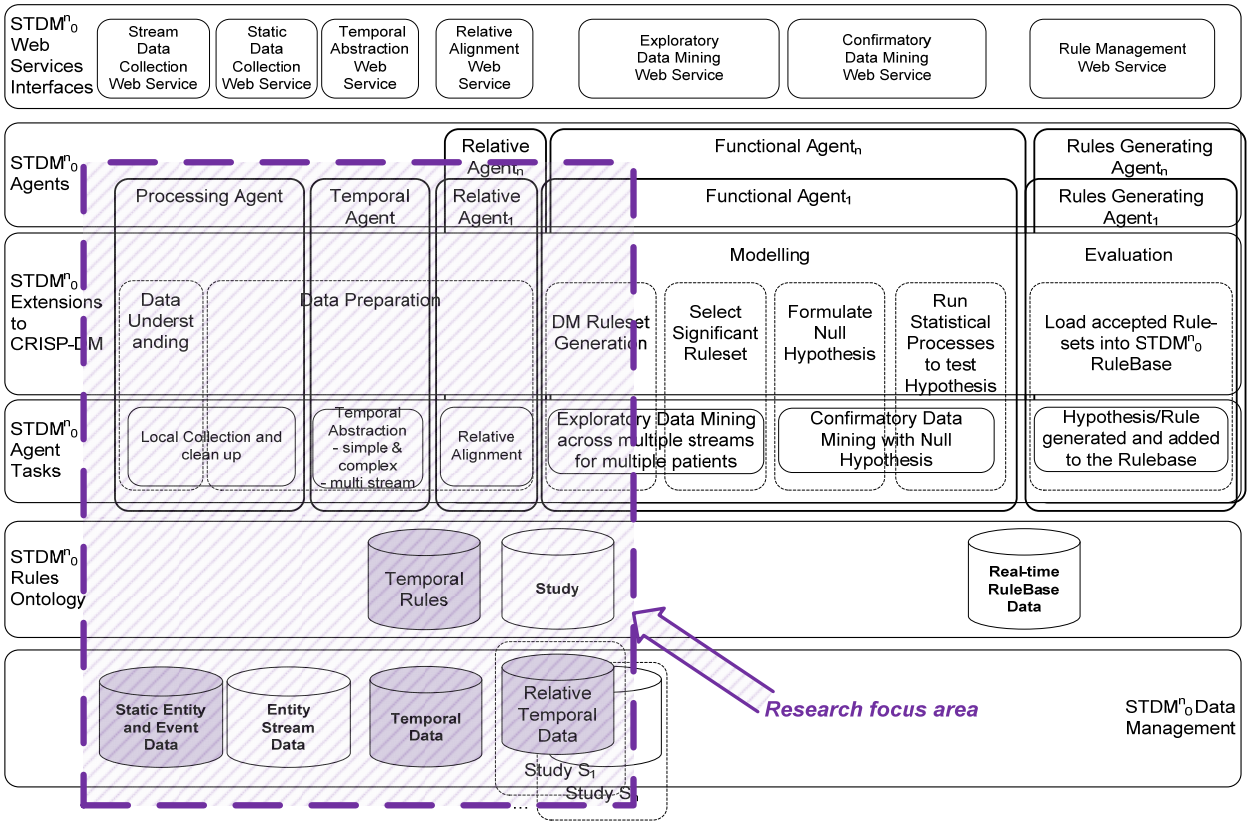


Figure 4-4 Research focus area with existing STDM$^n_0$ Framework

By extending the existing framework there is potential to gain a better understanding of individual patients in the retrospective data that will lead to improvements in sensitivity and specificity through the use of patient characteristics to create sub-groupings of the physiological behaviours as well as temporal abstraction behaviours.

Chapter 5 will discuss adaptions made to create the STDM$^{n+p}_0$ framework by inclusion of attributes in the TA_Rule table within the EntityStream database table and to the

TA_Rule, TA_RelativeTime and Patient tables within the Patient physiological Static database.

Chapter 5 –Defining Extended Multi-Dimensional Framework

5.1 Introduction

As detailed in chapter 1 and chapter 4, this research builds upon the $STDM^{n}_{0}$ framework. This chapter will describe in detail the $STDM^{n+p}_{0}$ extended patient characteristic multidimensional framework that can be defined for clinical research to enable patient specific pre-diagnosing at the onset of illness conditions based on trends and patterns discovered. The approach taken to extending the framework through the use of patient attributes is illustrated in Figure 5-1. One of the main challenges this framework endeavours to meet is diagnosis based on individual patient characteristic attributes. The $STDM^{n+p}_{0}$ framework defines a structured methodology that adds patient attributes, hence the "+p", to the multiple streams of physiological data collected ("n"), enabling individual patient characteristic analysis instead of an undefined patient 'one size fits all' approach. Another challenge the $STDM^{n+p}_{0}$ defined approach addresses is a structured method for creating sub-groupings of the physiological behaviours, as well as temporal abstraction behaviours.

Once these abstractions have been processed by the relative agent they will then progress through to the functional agent. The functional agent will facilitate rule set generation through exploratory data mining, selecting significant rule sets, null hypothesis formulation and running statistical processes to test the null hypothesis during confirmatory data mining (Bjering & McGregor, 2010). Null hypothesis testing is represented by "0" in $STDM^{n+p}_{0.}$ By enabling defined patient characteristic rules this invokes a host of data mining studies to be conducted on not only different genders but at different gestational ages across multiple physiological data stream baseline

acceptable thresholds, providing further insight into accurately diagnosing individual preterm infants. An example of sub-group clustering enabled by the $STDM^{n+p}_0$ is as follows: allow the investigation through static clinical data linked with the physiological HR data of male preterm infants at 35 weeks gestation age or the HR of male neonates at 28 weeks GA. This could then be compared with investigations on female infants of both 35 and 28 weeks GA respectively. The patient characteristic attributes considered and defined within this research are that of gender and gestational age.

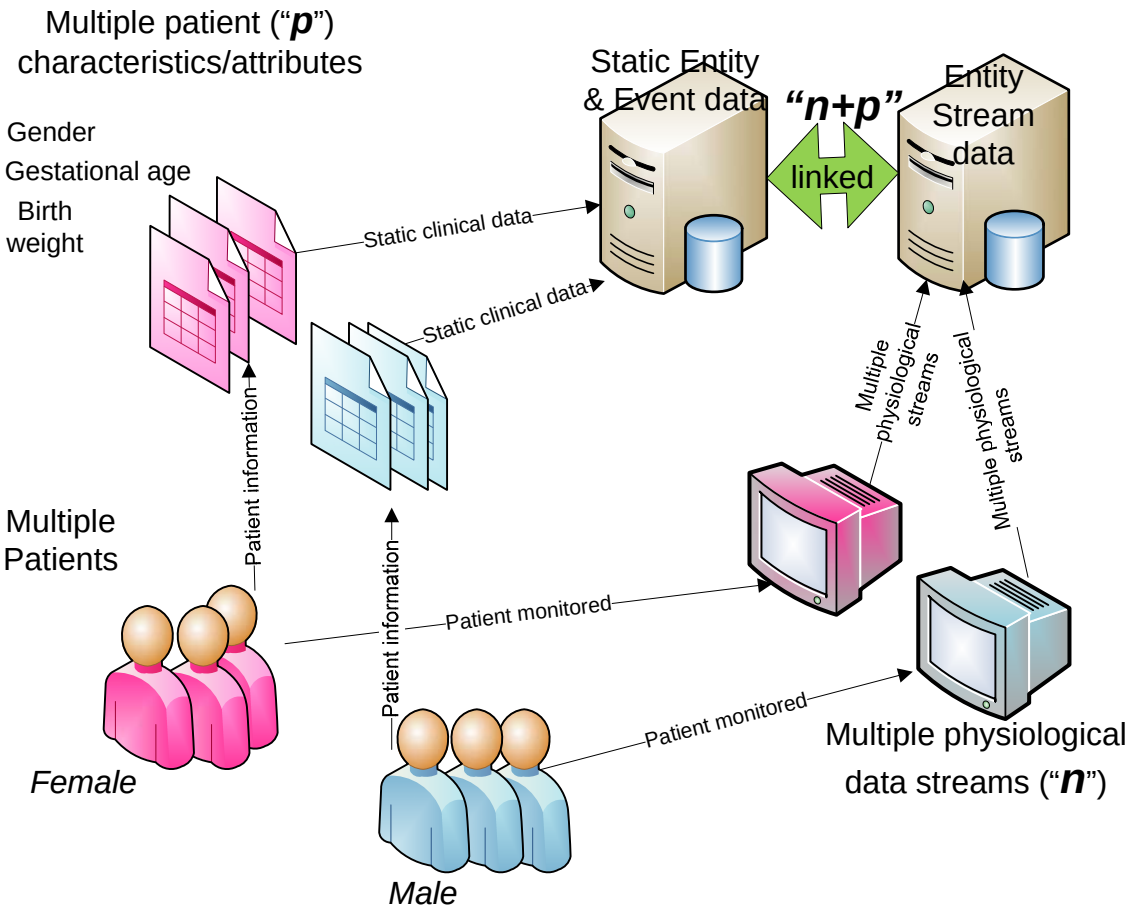


Figure 5-1 $STDM^{n+p}_0$ Approach

This chapter addresses the following research hypothesis:

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multidimensional temporal data.*
3. *The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.*
4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.*

5.2 Proposed Research Framework

This chapter details the extension made to create the $STDM^{n+p}_0$ framework which further adds to the multidimensional nature of the existing $STDM^{n}_0$ framework by extending the analysis of multiple streams of data from multiple patients with asynchronistic, static, patient-centric data. The areas of the $STDM^{n}_0$ framework extended to form the $STDM^{n+p}_0$ framework are presented as follows in Figure 5-2:

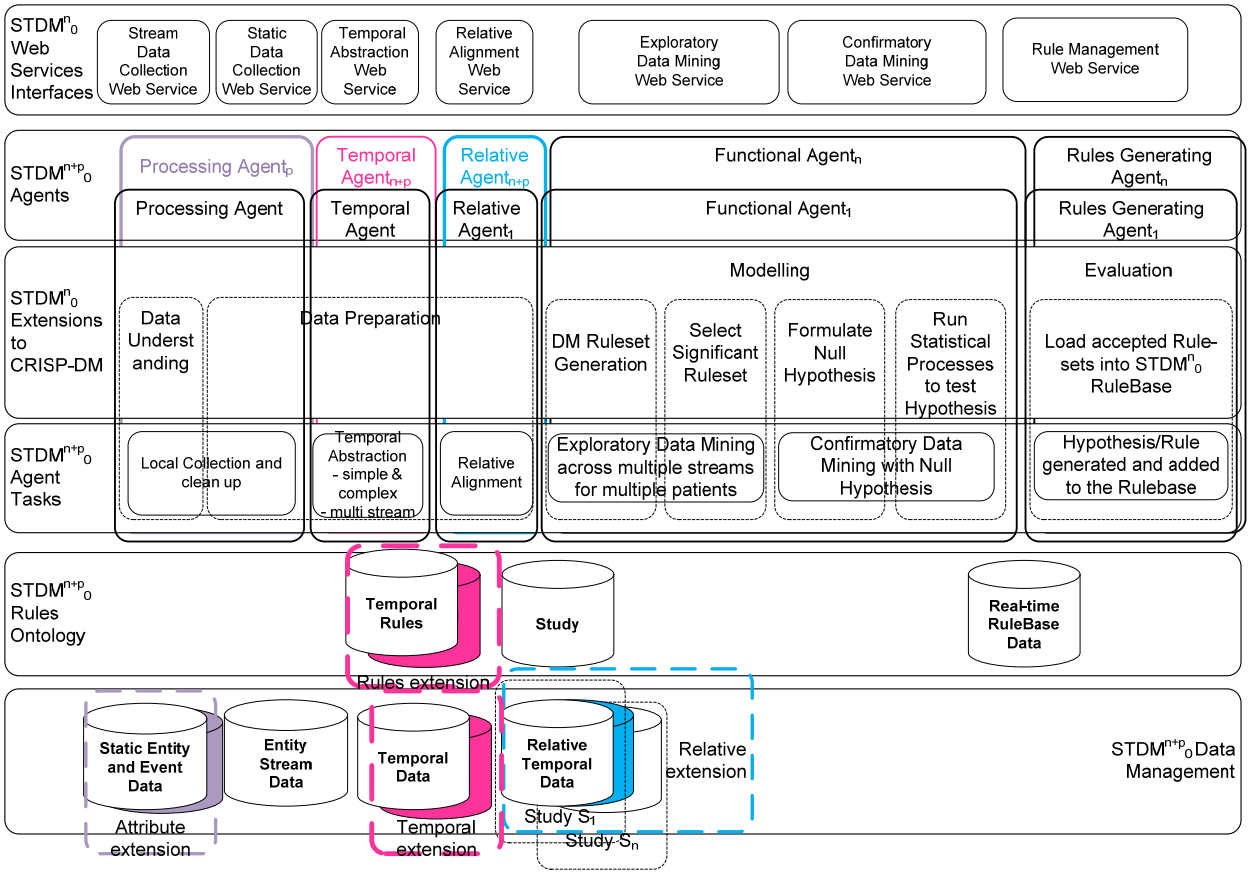


Figure 5-2 STDM$^{n+p}_{0}$ Framework

The Figure 5-1 STDM$^{n+p}_{0}$ Approach illustrates that the Static Entity and Event database interact through extended patient attribute use has now had an impact on other databases within the STDM$^{n+p}_{0}$ framework. There are many different examples of patient-centric data available from the EHR and CIS, including attributes such as: *gender, gestational age, birth weight, birth length and birth head circumference.* Attributes listed can be seen in Figure 5-5 Patient attribute table from STDM$^{n}_{0}$ NICU data model shown following in the Processing Agent section. For the purpose of this thesis we have selected gestational age and gender as attributes that can impact results from clinical algorithms due to their relationship with patient maturity.

The following briefly describes extended tasks completed for the purposes of this research thesis within the layers of the multi-agent systems as they are utilized within the $STDM^{n+p}_0$ framework. Constructing of the $STDM^{n+p}_0$ framework conceptualizes a method of solving research problem areas discovered and defined by research hypothesis 1.

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*

The tasks being extended for purposes of this research are within the Processing, Temporal and Relative Agents.

5.3 Processing Agent within the $STDM^{n+p}_0$ framework

Within the $STDM^{n+p}_0$ framework the Processing Agent performs the task of attaining and preparing physiological streamed data from sensors as well as retrieving static data from their respective tables with their respective databases such as clinical and physiological. Figure 5-3 illustrates all layers and tasks performed within the Processing Agent.

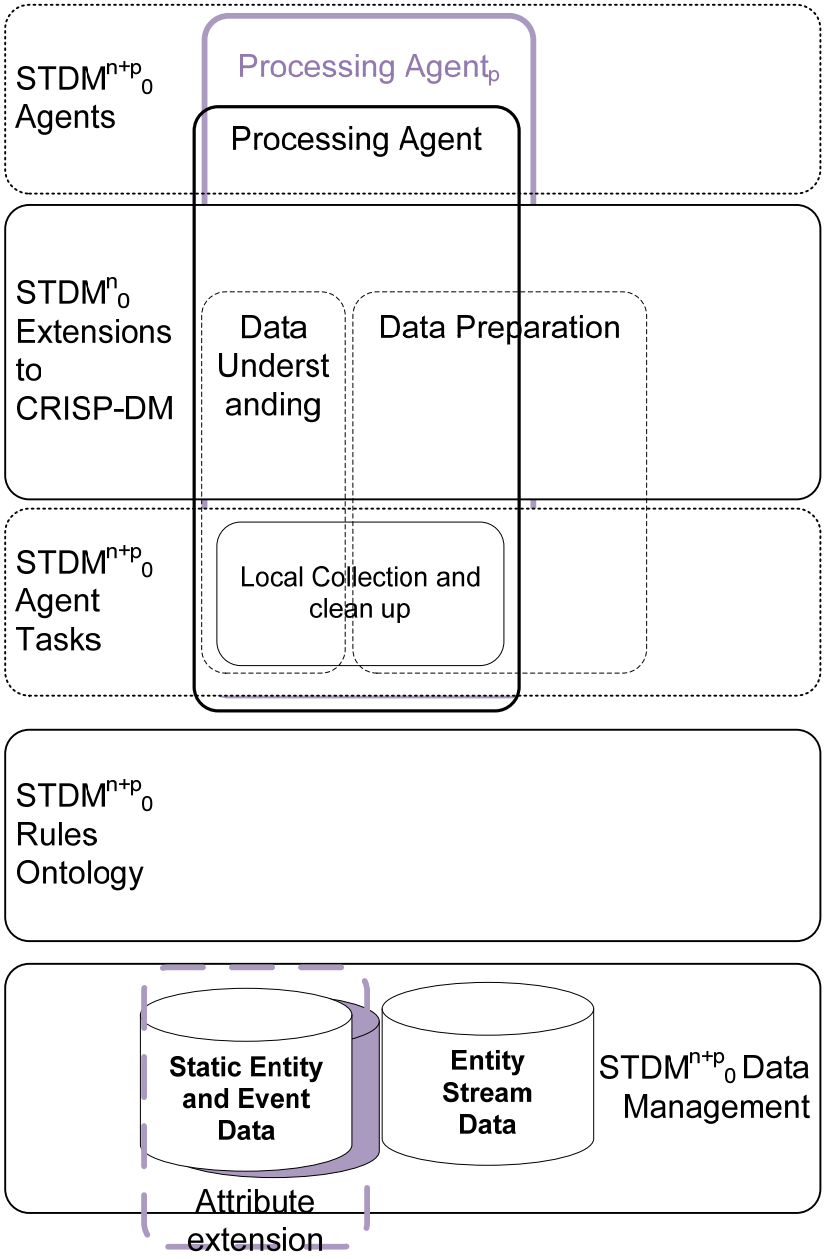


Figure 5-3 STDM$^{n+p}_0$ Processing Agent

The critical step involves the processing and integration of synchronistic collected physiological stream data with asynchronistic, static clinical data within the context of the neonatal intensive care environment. Stream data would represent physiological stream data collected from medical monitoring devices such as RR, SpO_2 and HR. While asynchronous, static or slow moving data within the NICU context would

represent clinical information such as patient ID, date of birth, gender, and gestational age. Although in some rare cases there is a non or 'ambiguous' gender option that will develop into either male or female, this tends to occur in those neonates born within the range of 23-27 weeks (Lehmann, Kim, & Johnson, 2009) .

To understand the static clinical data the $STDM^{n}_{0}$ framework table structures of interest are highlighted in Figure 5-4.

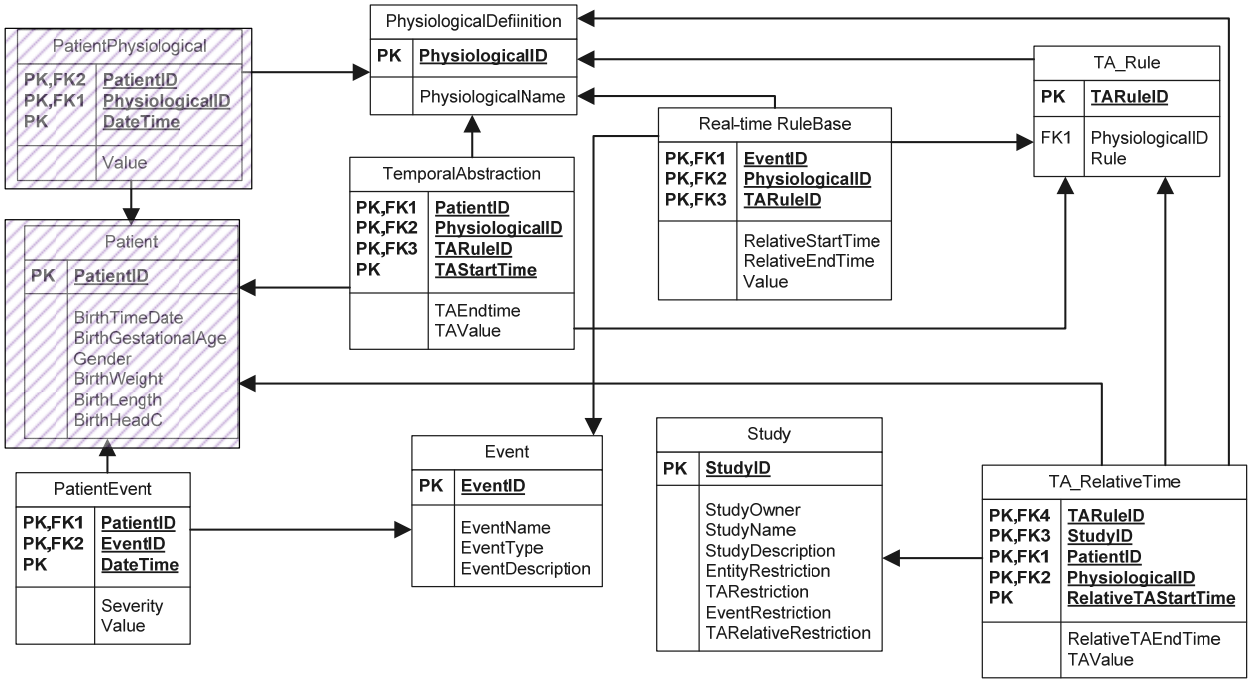


Figure 5-4 Exiting $STDM^{n}_{0}$ table structures

The Patient table contains the attributes of interest for $STDM^{n+p}_{0}$ research that is gender and gestational age at birth (BirthGestationalAge), as seen in Figure 5-5.

Patient	
PK	**PatientID**
	BirthTimeDate BirthGestationalAge Gender BirthWeight BirthLength BirthHeadC

Figure 5-5 Patient attribute table from $STDM^{n}_{0}$ NICU data model

The existing $STDM^{n}_{0}$ table created for the synchronous collected physiological stream data is structured as follows:

PatientPhysiological	
PK,FK2 **PK,FK1** **PK**	**PatientID** **PhysiologicalID** **DateTime**
	Value

Figure 5-6 Patient Physiological table from $STDM^{n}_{0}$ NICU data model

Table structure proposed in $STDM^{n+p}_{0}$ is as follows to incorporate new attributes:

PatientPhysiological-x	
PK,FK2 **PK** **PK** **PK,FK1** **PK**	**PatientID** **Gender** **BirthGestationalAge** **PhysiologicalID** **DateTime**
	Value

Figure 5-7 $STDM^{n+p}_{0}$ Patient Physiological table structure

This table structure will improve results when running temporal abstraction queries in the temporal agent, as demonstrated in chapter 6.

These tables will be stored in the clinical knowledge database for future further refinement of clinical knowledge, first investigated by exploratory data mining and followed by confirmatory data mining. This table structure follows that of a relational database; however, looking to the future there may be clinical advantages in adopting a real-time database structure.

5.4 Temporal Agent within the STDM$^{n+p}_0$ framework

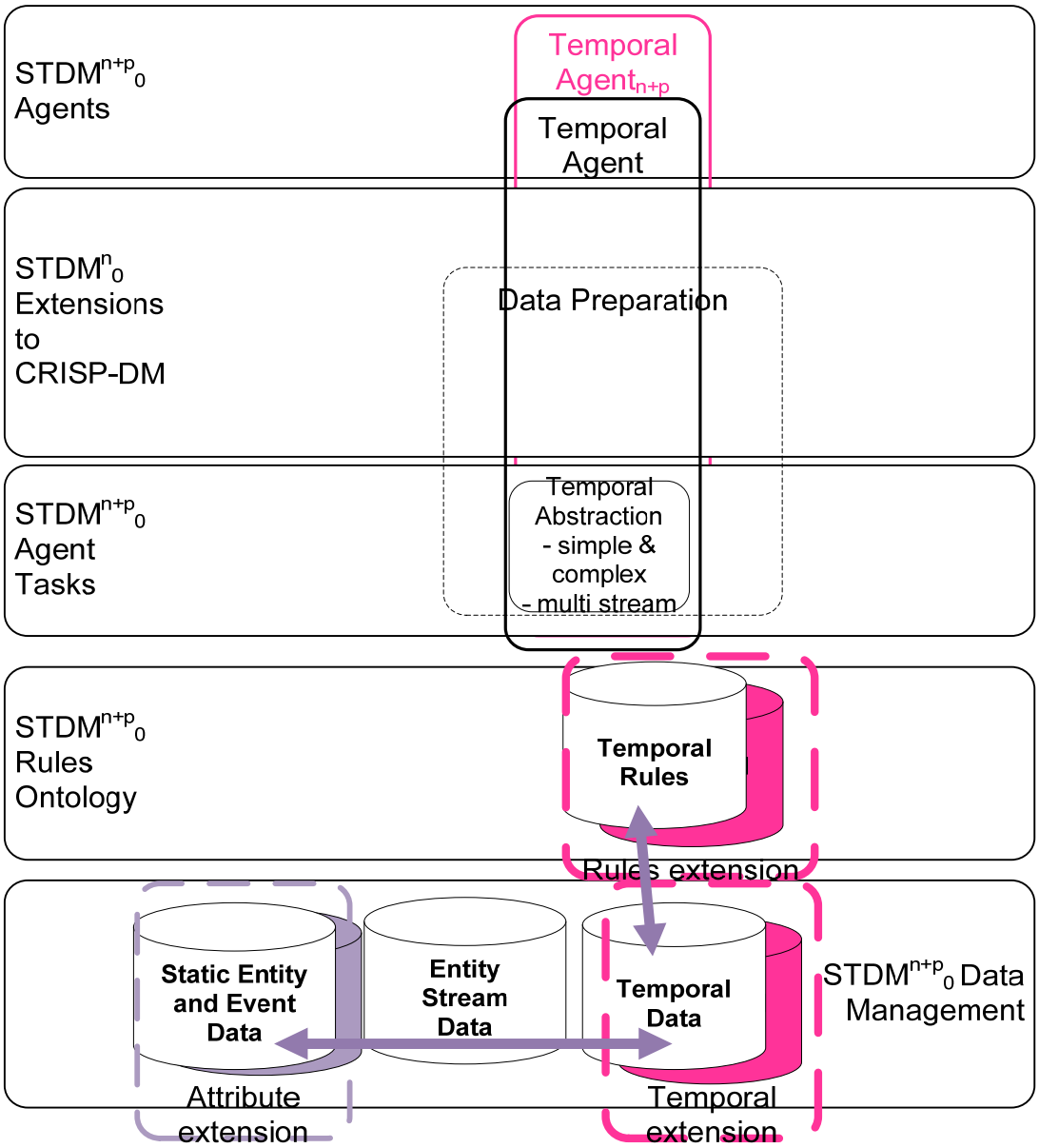


Figure 5-8 STDM$^{n+p}_0$ Temporal Agent

The arrows shown in Figure 5-8 $STDM^{n+p}_0$ Temporal Agent are demonstrating the interaction of the Static Entity and Event Database with the Entity Stream database when creating temporal abstraction in that database. Discoveries made from these temporal abstractions will help drive the temporal rules created. This agent within the $STDM^{n+p}_0$ framework provides a method of solving the research problem areas discovered and defined by research hypothesis 2.

2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multi-dimensional temporal data.*

The Temporal Agent was designed to create new temporal encoded streams by abstracting, at time stamped intervals, behaviours or trends that represent anomalies within that defined stream. Such anomalies, within for example an ECG physiological data stream, can be defined either as a trend such as increasing/decreasing, or as level shifts such as low/normal/high. All thresholds are dependent on the gender and gestational age of the infant the stream is being collected from. Each reading or data point from the ECG stream has the potential to be included in several abstractions. For example, this data point could have been collected while the heart rate was 'increasing' but was still within 'normal' limits. Complex abstraction involves the comparison of abstracted parameters performed across multiple streams. Each abstraction performed is stored in table form within the temporal database.

The existing $STDM^{n}_{0}$ table created for temporal abstractions is structured as follows in Figure 5-9:

TemporalAbstraction	
PK,FK1 **PK,FK2** **PK,FK3** **PK**	**PatientID** **PhysiologicalID** **TARuleID** **TAStartTime**
	TAEndtime TAValue

Figure 5-9 Temporal Abstraction table from $STDM^{n}_{0}$ NICU data model

The table structure proposed in $STDM^{n+p}_{0}$ will look as follows in Figure 5-10 to incorporate new attributes:

TemporalAbstraction-x	
PK,FK1 **PK,FK3** **PK,FK3** **PK,FK2** **PK,FK3** **PK**	**PatientID** **Gender** **BirthGestationalAge** **PhysiologicalID** **TARuleID** **TAStartTime**
	TAEndtime TAValue

Figure 5-10 $STDM^{n+p}_{0}$ Temporal Abstraction table

The existing $STDM^{n}_{0}$ temporal rule table is structured as follows in Figure 5-11:

TA_Rule	
PK	**TARuleID**
FK1	PhysiologicalID Rule

Figure 5-11 existing $STDM^{n}_{0}$ TA_Rule table

The table structure proposed in TA_Rule in $STDM^{n+p}{}_0$ will look as follows in Figure 5-12 to incorporate new attributes:

TA_Rule-x	
PK **PK** **PK**	**TARuleID** **Gender** **BirthGestationalAge**
FK1	PhysiologicalID Rule

Figure 5-12 $STDM^{n+p}{}_0$ TA_Rule table

5.5 Relative Agent within the STDM$^{n+p}_0$ framework

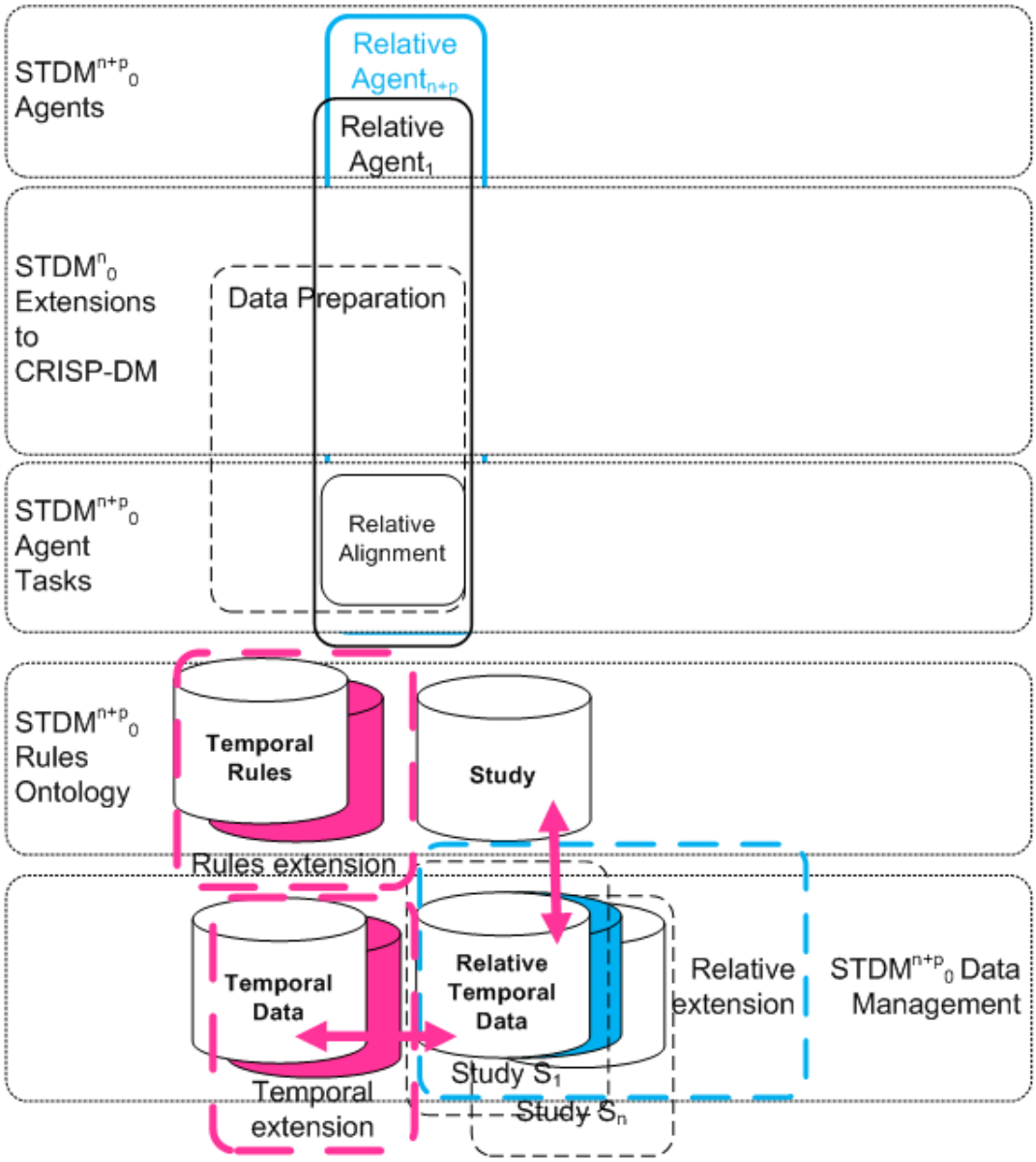


Figure 5-13 STDM$^{n+p}_0$ Relative Agent

The arrows shown in Figure 5-13 STDM$^{n+p}_0$ Relative Agent are demonstrating the interaction of the Temporal Database with the Relative Temporal, which will be driven by the research study of interest. Within the Relative Agent it is proposed that studies performed on temporal abstractions be based on clinical information from individual patients, such as gender and gestational age.

The existing $STDM^{n}_{0}$ TA_RelativeTime table is structured as follows in Figure 5-14:

TA_RelativeTime	
PK,FK4	**TARuleID**
PK,FK3	**StudyID**
PK,FK1	**PatientID**
PK,FK2	**PhysiologicalID**
PK	**RelativeTAStartTime**
	RelativeTAEndTime
	TAValue

Figure 5-14 existing $STDM^{n}_{0}$ TA_RelativeTime table structure

The table structure proposed in TA_RelativeTime in $STDM^{n+p}_{0}$ will look as follows in Figure 5-15 to incorporate new attributes:

TA_RelativeTime-x	
PK,FK4	**TARuleID**
PK,FK4	**Gender**
PK,FK4	**BirthGestationalAge**
PK,FK3	**StudyID**
PK,FK1	**PatientID**
PK,FK2	**PhysiologicalID**
PK	**RelativeTAStartTime**
	RelativeTAEndTime
	TAValue

Figure 5-15 $STDM^{n+p}_{0}$ TA_RelativeTime table structure

5.6 Functional agent

This stage involved the constructing of the framework conceptualized as a method of solving research problem areas discovered and defined by research hypothesis 4 (listed below).

4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.*

Within the Functional Agent the realigned temporal abstractions created in the Relative Agent are further processed. The functional agent is where extended CRISP-TDM (Catley, Smith, McGregor, & Tracy, 2009) modeling tasks occur, which include rule set generation through exploratory data mining, selecting significant rule sets, null hypothesis formulation and running statistical processes to test the null hypothesis during confirmatory data mining. This research has focused on data understanding phase, defining TA abstraction performed through further extending patient centric attributes within algorithms by gender and gestational age.

5.7 Rules Generating Agent

Successful processing within this agent will result in the creation of a new gender and gestational age defined clinical algorithm for the early prediction of disease based on retrospective clinical data, collected within the Processing Agent and stored in the Static Entity and Event database. The Rules Generating Agent involves utilizing the clinical algorithms developed in the Functional Agent to provide patient context-specific intelligent monitoring and alerting on real-time patient data streams. Co-mining, which enables the integration of data mining results with expert knowledge, is also possible, in that additional input may be received in the form of clinician-defined rules.

5.8 STDM$^{n+p}_{0}$ Data Storage

STDM$^{n+p}_{0}$ tables will be structured as shown in figure 5-16 for data storage.

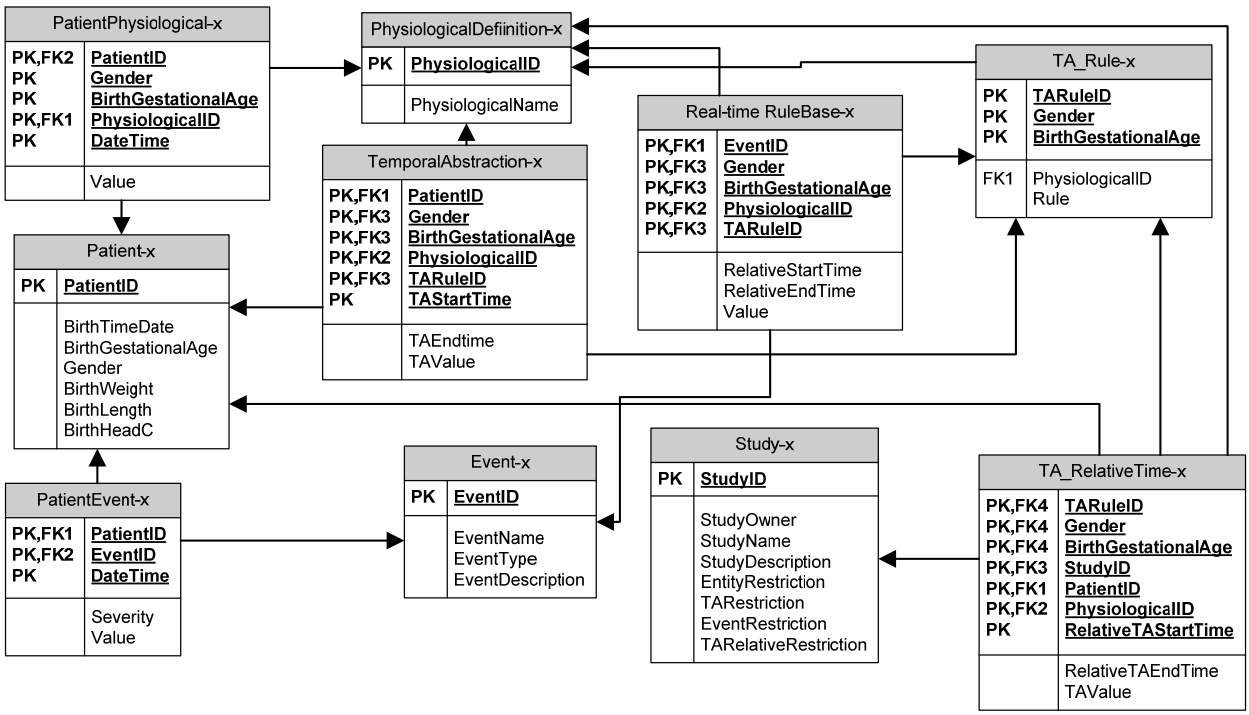


Figure 5-16 STDM$^{n+p}_{0}$ Data Storage

Extensions made to the PatientPhysiological-x, TemporalAbstraction-x, TA_Rule-x and TA_RelativeTime-x tables from the STDM$^{n}_{0}$ framework are discussed in detail above within their respective corresponding agents.

5.9 Summary of Proposed STDM$^{n+p}_{0}$ Methodology

The STDM$^{n+p}_{0}$ framework data collection and flow is demonstrated in Figure 5-17 Summary of STDM$^{n+p}_{0}$ Approach. Within the Processing Agent, multiple streams of

physiological data collected within the Entity Stream database are now linked to the Static Entity and Event database through the utilization of patient attributes such

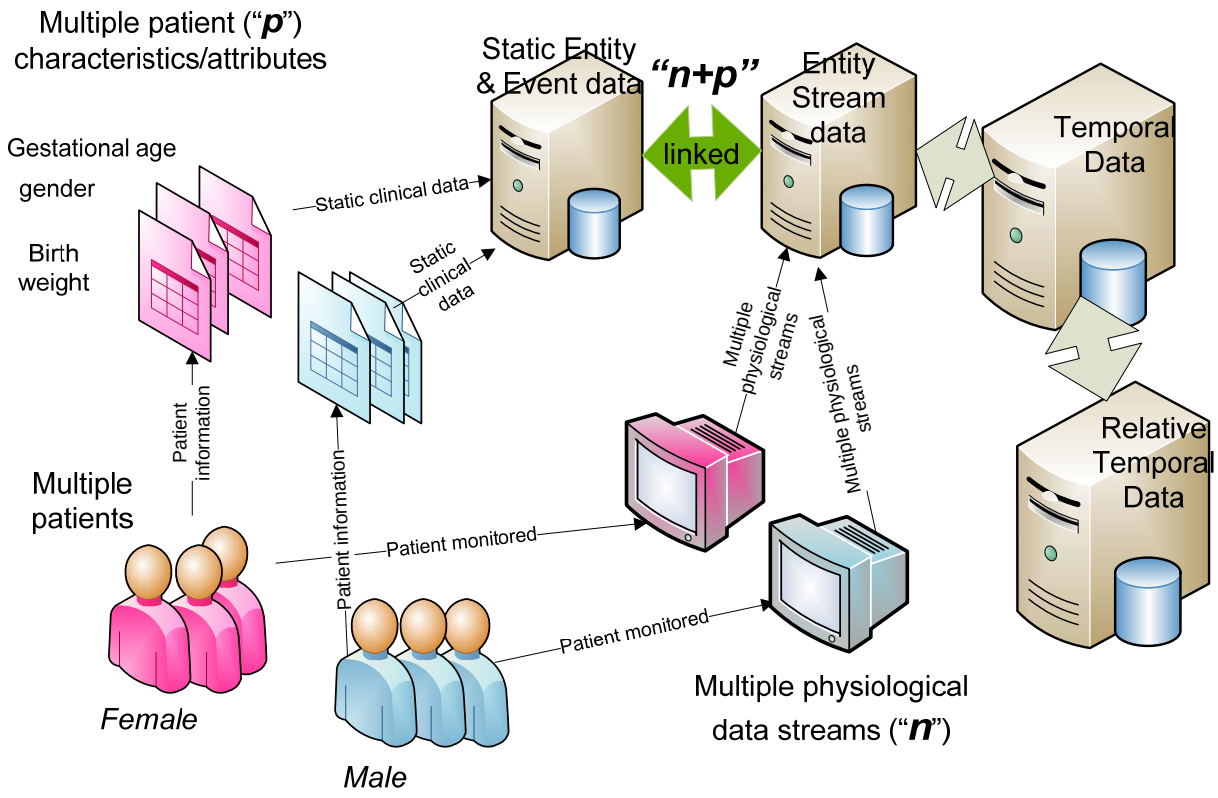


Figure 5-17 Summary of STDM$^{n+p}{}_0$ Approach

as gender and gestational age. This enables a structured format for temporal abstraction queries to be run within the temporal agent once a research study of interest has been defined. In turn, this defined structured format is also carried out within the relative agent through the realignment of these abstractions at a point of interest relevant to the study defined. In terms of workflow ordering, the data is processed using techniques defined in the Functional Agent, and fed through a data mining system using the clinical algorithms developed in the Functional Agent. In the event that this process indicates the potential early onset of a condition of interest, the intelligent patient

monitoring system indicates this knowledge and the results are evaluated. The knowledge gained of pattern detection in physiological data is then encoded to meet HL7 and SNOMED CT standards, and stored as part of the gold standard accepted within clinical databases.

5. 10 Conclusion

This chapter has presented the patient characteristic multi-dimensional adaptation to the $STDM^{n}_{0}$ framework to enabling sub-classifications created by $STDM^{n+p}_{0}$ framework. The chapter explained how the extended CRISP-DM model, CRISP-TDM, was incorporated into the framework while at the same time utilising a more patient centered approach.

This chapter has addressed research hypotheses one, two and four by defining and demonstrating how the patient attributes of gender and gestational age further define the patient characteristic multidimensional data mining structured approach to clinical investigations within the $STDM^{n+p}_{0}$ framework.

Chapter 6 – Case Study within NICU context

6.1 Introduction

This chapter provides a demonstration of the $STDM^{n+p}_0$ framework extensions proposed in chapter 5, through the use of a clinical research study example for neonatal apnoea spells. The extensions result in a multidimensional model incorporating gender and gestational age, to define patient characteristic thresholds for these attributes in relation to thresholds set for the detection of apnoea spells and thereby to assist in the support of clinical research within a NICU context.

The demonstration within the NICU context will provide supporting evidence for the research hypothesis presented in Chapter 1. These research hypotheses were:

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multi-dimensional temporal data.*
3. *The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.*
4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.*

The remainder of this chapter is structured as follows: 1) business understanding; 2) data understanding; 3) a description of all agents from the $STDM^{n+p}_0$ framework.

6.2 Business Understanding

As introduced previously, the context of this case study is neonatal care. The objective of the case study is to support the analysis of apnoeic events in neonates. Apnoeic spells are associated with many conditions including late onset neonatal sepsis. Within this research demonstration the following rule for an apnoea spell utilised: "*A lapse in breathing of a neonate for greater than 15 seconds is of clinical relevance (respiratory rate (RR) <25). At all neonatal gestational ages, a fall in peripheral oxygen saturation less than 85% for greater than 20 seconds combined with a HR of less than 108 bpm (100 bpm for male) is also of clinical relevance"* (Hein, Ely, & Lofgren, April 1998) (Catley C. , Smith, McGregor, James, & Eklund, 2010)**.** To date, as indicated in Chapter 3, HR <100 is the threshold parameter applied (American Heart Association, 2006).

Recent literature, as presented in Chapter 3, states that gender plays a significant role in defining HR differences: newborn male infants have lower baseline HR than newborn females (Nagy & Orvos, 2000) (Krueger, van Oostrom, & Shuster, 2010). These findings suggest that the known gender-related HR differences that are apparent throughout life are also present at the very beginning of life and should be considered when investigating physiological markers for conditions that affect the health and development of the newborn infant (Nagy & Orvos, 2000). The two patient identifiers (Stravroudis, Miller, & Lehmann, 2008) this research proposes for inclusions are that of

gestational age and gender to improve accuracy of diagnosis, treatment and critical care of neonates. The inclusion of gender and gestational age sets the stage for the motivation in choosing an apnoea case study to demonstrate the extended patient characteristic framework.

As such, this retrospective research proposes that the physiological stream behaviour thresholds may be more accurate for an individual baby monitored in real-time if they are adjusted based on patient characteristics, such as gender and gestational age. The object of this demonstration is to extend the apnoea spell research to enable the analysis of apnoea spells in association with gender and gestational age. Specifically, through a demonstration of how the extended $STDM^{n+p}_{0}$ framework enables inclusion of patient characteristics within the analysis of the temporal behaviours of the physiological data streams.

6.3 Data Understanding

Data used within this demonstration was collected and stored through research investment made by the Canada Research Chair program together with an IBM First-of-a-Kind award and resulted in the implementation of the Artemis platform at The Hospital for Sick Children, Toronto, Ontario, Canada. Artemis is a framework to support real-time clinical decision support, together with retrospective clinical research. The goal of the Artemis research project is to provide a flexible platform for the real-time analysis of time series physiological data streams extracted from a range of monitors to detect clinically significant conditions that may adversely affect health outcomes. Artemis supports data collected from multiple physical monitoring devices as well as from the

SickKids NICU's Clinical Information Management System (CIMS) and information from the laboratory information system in the hospital. The Artemis platform supports the ingestion and storage of multiple real-time data streams from multiple patients, while analysing for multiple conditions for the purposes of real-time and retrospective analysis, and data-mining (Blount, Ebling, et al., 2010).

The first implementation of Artemis has been utilizing the Philips Intellivue MP70 Neonatal monitors. These devices produce multiple streams of physiological data collected from each patient at a rate of one reading every 1024ms. This case study demonstration will utilise a reduced data set containing three of these physiological data streams, specifically: electrocardiogram derived RR, SpO_2 and HR (ECG-HR).

This chapter will provide an in depth description on how the collected neonatal raw physiological and clinical data moves through the extended $STDM^{n+p}{}_0$ framework, with particular attention paid to the Processing and Temporal Agents that will support defining the patient characteristic clinical temporal rules where new patient characteristic trends and patterns of apnoea will be unveiled. The $STDM^{n+p}{}_0$ framework is presented in Figure 6-1.

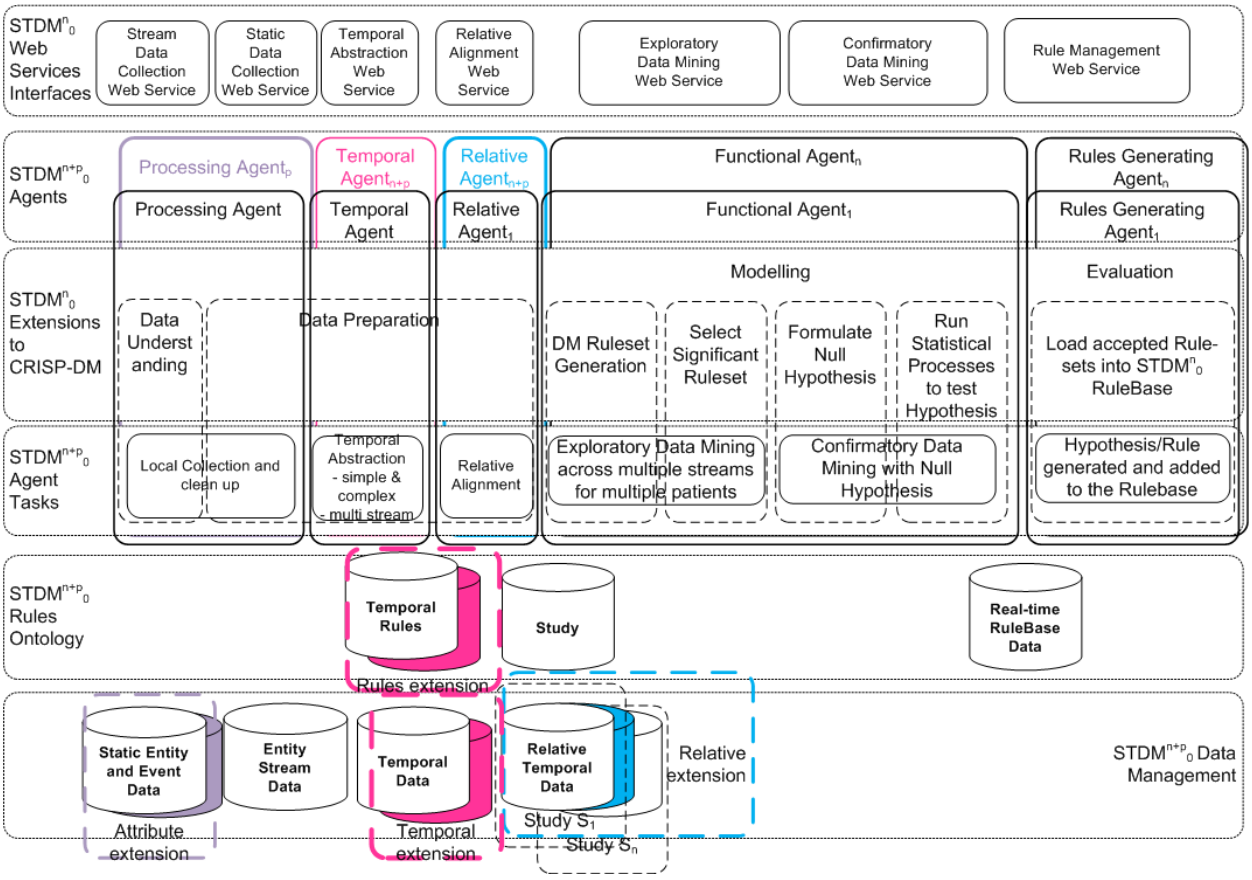


Figure 6-1 Proposed Research STDM$^{n+p}_{0}$ Framework

The study of interest, such as apnoea spells in this demonstration, will determine the parameters placed on the data that in turn produces the temporal abstractions collected. The thresholds applied to the different streams and used by this study are as follows: *RR<25 for greater than 15 seconds, peripheral oxygen saturation (SpO_2) < 85% for greater than 20 seconds combined with a HR of less than 108 bpm (100 bpm for male) for a female neonate of 35 weeks gestational age is all of clinical relevance.* These thresholds are applied to their respective streams to create the temporal abstractions that are then stored within the STDM$^{n+p}_{0}$ data storage framework.

6.4 Processing Agent

The role of the Processing Agent is to initiate access, as well as integration and collection of physiological and clinical data stored from multiple databases. Within this chapter, the mapping of the data from the different de-identified data stores acquired from the NICU will be described in detail and demonstrated. Tasks undertaken by the Processing Agent occur within the Data Acquisition component of the Artemis System Architecture shown in Figure 6-2.

Artemis is a Research Ethic Board (REB) approved collaborative project between SickKids, IBM and UOIT. While the Artemis framework contains components for Data Acquisition, Online Analysis, Data (stream) Persistence, Knowledge Extraction and Redeployment, the demonstration in this thesis has focused on the Data Persistence and Knowledge Extraction components only. The data Persistence is implemented in DB2 V9.7. Some details will also be provided on knowledge extracted which then transferred and deployment occurs for real-time use. The Artemis system diagram is illustrated in Figure 6-2.

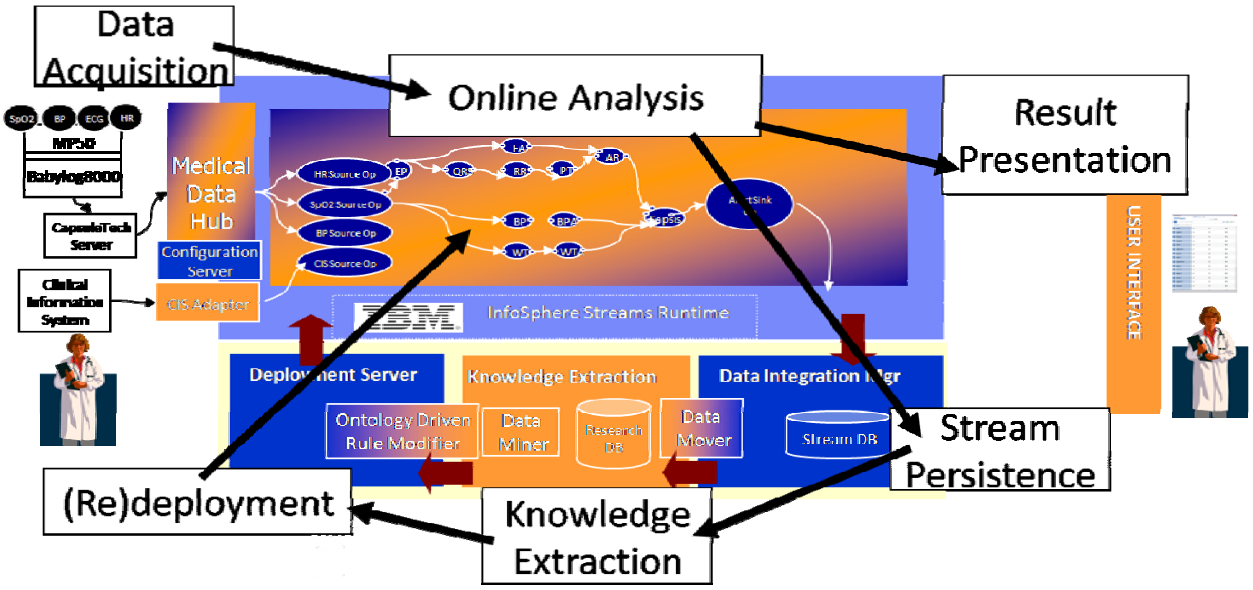


Figure 6-2 Artemis System Architecture

Within Artemis there are two copies of Data Persistence: one used to support the Online Analysis and an incremental replica version of the Data Persistence which receives new data each day from the Online Analysis Data Persistence copy. The Knowledge Extraction copy of the Data Persistence is represented by the Data Management layer of the $STDM^{n+p}{}_0$ framework.

From SickKids there are two main Database storage components within the Data Management layer that will be the focus of this research thesis, those being the Clinical Information Management System (CIMS) and physiological data information management (DIM). Both of these sources of data are stored by the 'Static Entity and Event Data' and the 'Entity Stream Data' Database found within the Data Management layer in the framework provided above.

The CIMS database contains all patient source data and is implemented at SickKids in Oracle. Of primary concern for the purposes of this research is the demographic details and physiological measurements at time of birth (gender and gestational age inclusive).

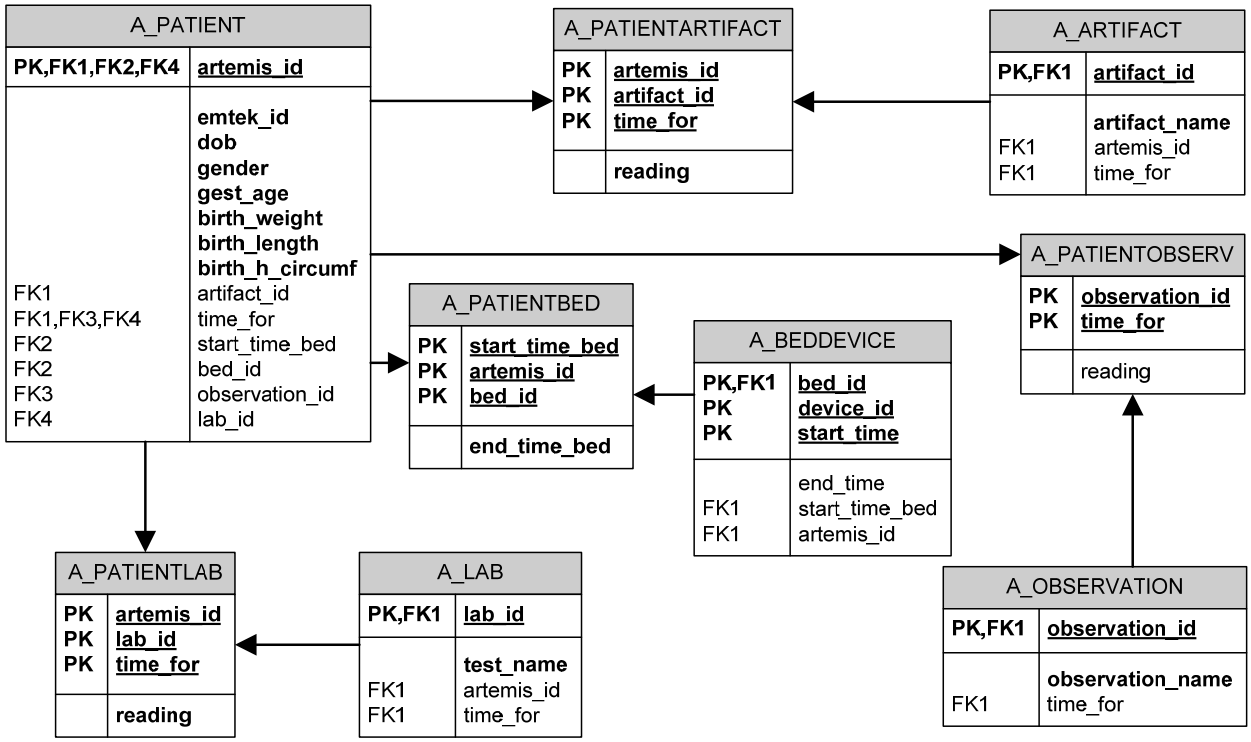


Figure 6-3 Artemis CIMS table structure(s)

As demonstrated in the above CIMS defined table structures, the Artemis project receives de-identifed patient characteristic information in the A_PATIENT table, where the attribute emtek_id has been substituted to attribute artemis_id to maintain anonymity of patients enrolled in the project. The table A_PATIENT is isolated below to show how and where patient information was transferred from SickKids into a mirrored research Artemis database.

A_PATIENT	
PK,FK1,FK2,FK4	**artemis_id**
	emtek_id
	dob
	gender
	gest_age
	birth_weight
	birth_length
	birth_h_circumf
FK1	artifact_id
FK1,FK3,FK4	time_for
FK2	start_time_bed
FK2	bed_id
FK3	observation_id
FK4	lab_id

Table 6-1 A_PATIENT table located within Artemis database

Each preterm infant admitted is registered and clinical data entered into a table and stored within the CIMS database; an example of such a data is shown in Table 6-2. Under the GA column, the gestational age is broken down into weeks plus days.

EMTEKID	Gender	Age (days)	GA (BIRTH)	BW
….1	M	0-3	38+4	3080
….2	M	0-3	40+6	3400
….3	M	0-3	32+6	1850
….4	F	0-3	39+4	3240
….5	M	8+	25+2	770
….6	M	8+	25+5	820
….7	F	8+	36+6	2050
….8	M	8+	36+6	2770
….9	M	8+	25	535
….10	F	0-3	37 +2	1934
….11	M	0-3	31+2	810
….12	F	0-3	34+4	780

Table 6-2 CIMS Patient stored data table

The DIM database contains all enrolled patients' physiological data, collected via medical attached devices such as the MP70. The table structure for this database is presented as follows in Figure 6-4 Artemis DIM table structure(s) below:

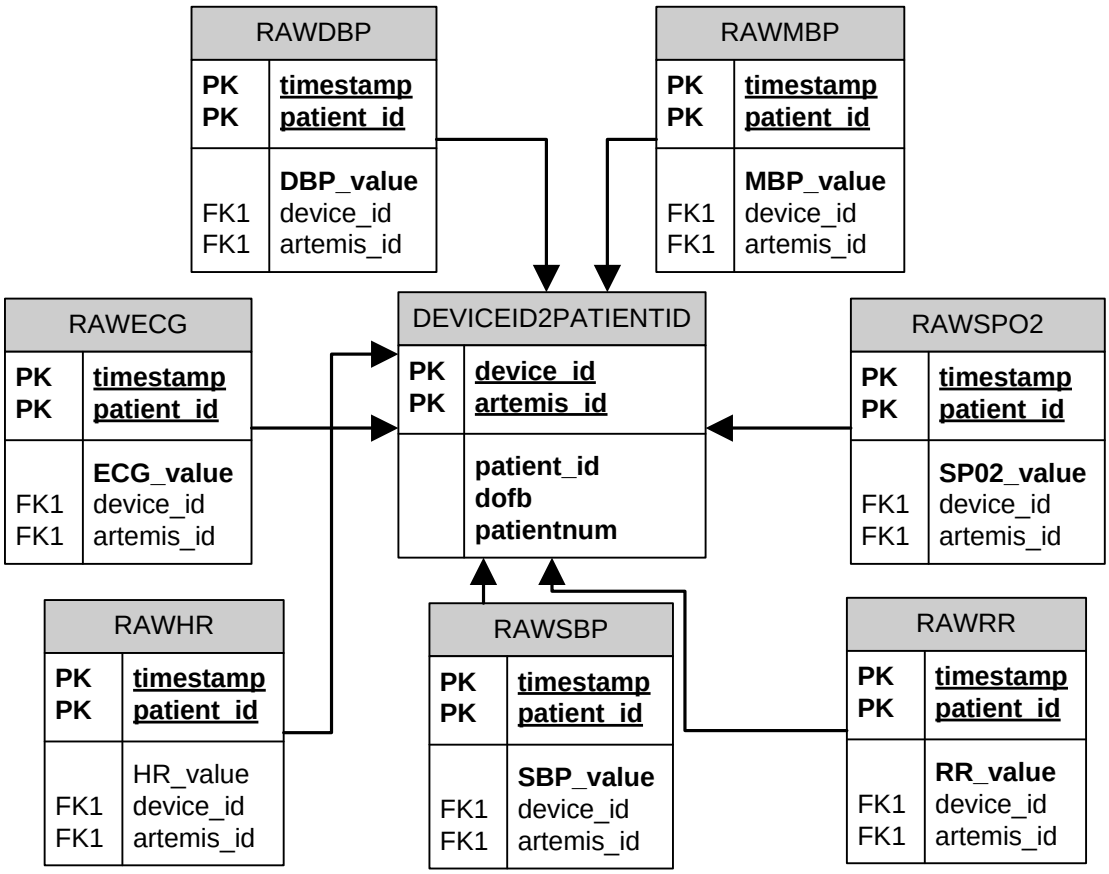


Figure 6-4 Artemis DIM table structure(s)

Each preterm infant enrolled in Artemis has multiple physiological data measurements collected. Each physiological data measurement has a timestamp to the millisecond for every data point collected. The physiological stored files contain the timestamp, patient_id, and the named physiological reading. Figure 6-4 corresponds to the patient's

physiological_id table. Artemis has implemented a horizontal split of the table such that each physiological data reading is on their own.

DateTime (Format: Yyyymmdd hh:mm:ss:000)	Patient_ID	Value (HR)
20090807 11:04:09:021.299	12345	118
20090807 11:04:09:022.323	12345	116
20090807 11:04:09:023.347	12345	115
20090807 11:04:09:024.371	12345	115
20090807 11:04:09:025.395	12345	114
…….	…..	….
…….	…..	….

Table 6-3 Stored HR physiological data

To enable physiological thresholds to be driven by gender and gestational age, additional attributes will be included in TA tables that will be detailed further in the Temporal Agent section.

The first 'A_PATIENT' table in the above CIMS structure demonstrates how information is drawn from and related to the NICU source data regarding the admitted patient. Primarily of interest in this research thesis is the physiological data initially collected such as *dob*, *gender* and *gest_age* shown above in table 6-1.

6.4.1 Mapping Artemis CIMS tables from SickKids to Artemis DIM tables

In order for the data to pass through to the processing agent the following table structures need to be mapped. Mapping creates the links between the two different databases to enable ease of coupling data elements. Mapping from the CIMS database tables to the DIM database table is shown below in Table 6-4 and Table 6-5.

Neonate Artemis (CIMS) A_BEDDEVICE	DIM tables DEVICEID2PATIENTID
artemis_id	*artemis_id*
device_id	device_id
start_time	Value (datetimestamp)
end_time	Value (datetimestamp)
Dob	Dofb

Table 6-4 Mapping Artemis (CIMS) to Artemis DIM table structures

RAW data	Physiological_ID
RAWRR	1
RAWHR	2
RAWSPO2	3
RAWMBP	4
RAWSBP	5

Table 6-5 Mapping RAW data collected at SickKids to Proposed Patient Characteristic database

Once all mapping of valuable input data is finished, the Processing Agent has completed its tasks in preparation for the data to then be passed on to the Temporal Agent.

6.5 Temporal Agent

The Temporal Agent utilizes data detailing gender and gestational age from the patient table that has been placed in data stores by the Processing Agent. The thresholds described in section 6.3 are applied to their respective streams to create the temporal abstractions that are then stored within the $STDM^{n+p}_{0}$ data storage framework. The Temporal Agent uses the rules defined in the temporal rules table to create temporal abstractions from the physiological data that has been collected from the MP70 neonatal monitoring equipment used by The Hospital for Sick Children.

As described in chapter 4, the temporal agent has six main functions:

1) Retrieve the physiological data from the physiological data store for each parameter for each patient based on gender and gestational age
2) Retrieve the relevant abstraction rules from the temporal rules table
3) Apply the rules to the physiological data, creating simple abstractions for individual data streams for individual patients
4) Store the created abstractions in the $STDM^{n+p}_0$ temporal data store
5) Create complex abstractions from the simple abstractions created in step 3, according to any rules found in the temporal rules table.
6) Store any complex abstractions created in the $STDM^{n+p}_0$ temporal data storage.

6.5.1 Abstraction overview demonstration

The TA rules are executed on data for a predefined time. The example used for this demonstration consists of 20 second sample with one reading every 1024 ms. The data used for this demonstration has three main streams of time-stamped physiological readings which have been abstracted separately into simple temporal abstractions. A particular time-stamped physiological reading for a particular patient can be part of several simple abstractions. The following abstractions are designed to demonstrate that minor adjustments made to abstraction threshold parameters produce significantly different result that may be causal to clinically significant outcomes.

Gestational age based on anecdotal evidence suggests that a threshold parameter for mean blood pressure for example: *"Given a hypothetical newborn baby born 5 weeks premature (35 weeks gestational age), a fall in mean blood pressure less than 35 mm*

Hg is clinically relevant." (Catley, Smith, McGregor, & Tracy, 2009). To demonstrate the impact gestational age can have on this one stream, a 20 second sample from one patient is presented in Table 6-6.

Patient_ID	DateTime (Format: Yyyymmdd hh:mm:ss:0)	Physiological_ID	Value (MBP mmHg)
sample 1	20091201 10:59:21.033	4	32.6
sample 1	20091201 10:59:22.057	4	32.7
sample 1	20091201 10:59:23.081	4	32.8
sample 1	20091201 10:59:24.105	4	32.8
sample 1	20091201 10:59:25.129	4	32.7
sample 1	20091201 10:59:26.153	4	33.2
sample 1	20091201 10:59:27.177	4	33.7
sample 1	20091201 10:59:28.201	4	34.5
sample 1	20091201 10:59:29.225	4	35.1
sample 1	20091201 10:59:30.249	4	35.3
sample 1	20091201 10:59:31.273	4	35.7
sample 1	20091201 10:59:32.297	4	36.1
sample 1	20091201 10:59:33.321	4	37.4
sample 1	20091201 10:59:34.345	4	38.2
sample 1	20091201 10:59:35.369	4	39.1
sample 1	20091201 10:59:36.393	4	40.3
sample 1	20091201 10:59:37.417	4	40.5
sample 1	20091201 10:59:38.441	4	42.1
sample 1	20091201 10:59:39.465	4	42.9
sample 1	20091201 10:59:40.489	4	43.1

Table 6-6 MBP gestational age demonstration sample

The abstraction rule applied to the MBP data is based on gestational age the first example will a 35 week old therefore using a threshold a follows:

Low = MBP < 35, and Normal = MBP $\underline{\geq}$ 35 would produce TA results as follows:

Patient_ID	Physiological_ID	AbstractionType	AbstractionValue	ActualStartTime	ActualEndTime
sample 1	4	Level shift	Low	20071201 10:59:21.033	20091201 10:59:28.201
sample 1	4	Level shift	Normal	20071201 10:59:29.225	20091201 10:59:40.489

Table 6-7 TA result on MBP of a GA 35 week old neonate

However, if the same data was retrieved from a neonate with a gestational age of 39 weeks, amending the threshold as follows: Low = MBP < 39, and Normal = MBP $\geq$ 39 would produce TA results as follows in Table 6-8:

Patient_ID	Physiological_ID	Abstraction Type	Abstraction Value	ActualStartTime	ActualEndTime
sample 1	4	Level shift	Low	20071201 10:59:21.033	20091201 10:59:34.345
sample 1	4	Level shift	Normal	20071201 10:59:35.369	20091201 10:59:40.489

Table 6-8 TA results on MBP of a GA 39 week old neonate.

The MBP TA results, as demonstrated, have shown significantly different TA outcomes with only the consideration of gestational age taken into account when defining the threshold rules. This research would like to demonstrate further significant outcomes by incorporating gender into generating the threshold rules defining the algorithms.

6.5.2 Apnoeic Event case study temporal abstraction

The case study is a demonstration of how data collected by the current Artemis pilot could be used through the secondary use of data for new knowledge creation. The following three tables contain a thirty second segment of raw physiological data readings collected every 1024 ms, specifically focusing on RR, SpO_2 and HR, that will used throughout the duration of the case study. To assist in the readers' ability to follow all graphs, tables and charts produced have been colour co-ordinated from this point forward. The colour of the table containing the physiological data will correspond to the

colour of the graphs demonstrating how thresholds are applied to the individual data streams. Values for the three streams are shown in Tables 6-9 through 6-11.

Table 6-9 -RR values

Patient_ID	Gender	GA	DateTime (Format: Yyyymmdd hh:mm:ss:000)	Physiological_ID	RR value
testpatient	F	35	20090807 11:04:10.035	1	28
testpatient	F	35	20090807 11:04:11.059	1	28
testpatient	F	35	20090807 11:04:12.083	1	27
testpatient	F	35	20090807 11:04:13.107	1	25
testpatient	F	35	20090807 11:04:14.131	1	24
testpatient	F	35	20090807 11:04:15.155	1	24
testpatient	F	35	20090807 11:04:16.179	1	23
testpatient	F	35	20090807 11:04:17.203	1	23
testpatient	F	35	20090807 11:04:18.227	1	23
testpatient	F	35	20090807 11:04:19.251	1	22
testpatient	F	35	20090807 11:04:20.275	1	22
testpatient	F	35	20090807 11:04:21.299	1	22
testpatient	F	35	20090807 11:04:22.323	1	22
testpatient	F	35	20090807 11:04:23.347	1	22
testpatient	F	35	20090807 11:04:24.371	1	22
testpatient	F	35	20090807 11:04:25.395	1	21
testpatient	F	35	20090807 11:04:26.419	1	21
testpatient	F	35	20090807 11:04:27.443	1	21
testpatient	F	35	20090807 11:04:28.467	1	22
testpatient	F	35	20090807 11:04:29.491	1	21
testpatient	F	35	20090807 11:04:30.515	1	21
testpatient	F	35	20090807 11:04:31.539	1	21
testpatient	F	35	20090807 11:04:32.563	1	20
testpatient	F	35	20090807 11:04:33.587	1	20
testpatient	F	35	20090807 11:04:34.611	1	20
testpatient	F	35	20090807 11:04:35.635	1	19
testpatient	F	35	20090807 11:04:36.659	1	19
testpatient	F	35	20090807 11:04:37.683	1	19
testpatient	F	35	20090807 11:04:38.707	1	20

Table 6-10: Raw SpO_2 readings

Patient_ID	Gender	GA	DateTime (Format: Yyyymmdd hh:mm:ss:0)	Physiological_ID	Value (SpO_2 %)
testpatient	F	35	20090807 11:04:09.011	2	92
testpatient	F	35	20090807 11:04:10.035	2	89
testpatient	F	35	20090807 11:04:11.059	2	87
testpatient	F	35	20090807 11:04:12.083	2	86
testpatient	F	35	20090807 11:04:13.107	2	85
testpatient	F	35	20090807 11:04:14.131	2	84
testpatient	F	35	20090807 11:04:15.155	2	84
testpatient	F	35	20090807 11:04:16.179	2	84
testpatient	F	35	20090807 11:04:17.203	2	84
testpatient	F	35	20090807 11:04:18.227	2	83
testpatient	F	35	20090807 11:04:19.251	2	83
testpatient	F	35	20090807 11:04:20.275	2	83
testpatient	F	35	20090807 11:04:21.299	2	82
testpatient	F	35	20090807 11:04:22.323	2	82
testpatient	F	35	20090807 11:04:23.347	2	82
testpatient	F	35	20090807 11:04:24.371	2	83
testpatient	F	35	20090807 11:04:25.395	2	83
testpatient	F	35	20090807 11:04:26.419	2	83
testpatient	F	35	20090807 11:04:27.443	2	84
testpatient	F	35	20090807 11:04:28.467	2	82
testpatient	F	35	20090807 11:04:29.491	2	82
testpatient	F	35	20090807 11:04:30.515	2	81
testpatient	F	35	20090807 11:04:31.539	2	81
testpatient	F	35	20090807 11:04:32.563	2	81
testpatient	F	35	20090807 11:04:33.587	2	81
testpatient	F	35	20090807 11:04:34.611	2	82
testpatient	F	35	20090807 11:04:35.635	2	82
testpatient	F	35	20090807 11:04:36.659	2	81
testpatient	F	35	20090807 11:04:37.683	2	81
testpatient	F	35	20090807 11:04:38.707	2	81

Table 6-11: HR readings

Patient_ID	Gender	GA	DateTime (Format: Yyyymmdd hh:mm:ss:0)	Physiological _ID	Value (HR)
testpatient	F	35	20090807 11:04:09.011	3	118
testpatient	F	35	20090807 11:04:10.035	3	116
testpatient	F	35	20090807 11:04:11.059	3	115
testpatient	F	35	20090807 11:04:12.083	3	115
testpatient	F	35	20090807 11:04:13.107	3	114
testpatient	F	35	20090807 11:04:14.131	3	112
testpatient	F	35	20090807 11:04:15.155	3	113
testpatient	F	35	20090807 11:04:16.179	3	112
testpatient	F	35	20090807 11:04:17.203	3	111
testpatient	F	35	20090807 11:04:18.227	3	110
testpatient	F	35	20090807 11:04:19.251	3	109
testpatient	F	35	20090807 11:04:20.275	3	108
testpatient	F	35	20090807 11:04:21.299	3	107
testpatient	F	35	20090807 11:04:22.323	3	106
testpatient	F	35	20090807 11:04:23.347	3	105
testpatient	F	35	20090807 11:04:24.371	3	104
testpatient	F	35	20090807 11:04:25.395	3	103
testpatient	F	35	20090807 11:04:26.419	3	102
testpatient	F	35	20090807 11:04:27.443	3	101
testpatient	F	35	20090807 11:04:28.467	3	100
testpatient	F	35	20090807 11:04:29.491	3	99
testpatient	F	35	20090807 11:04:30.515	3	99
testpatient	F	35	20090807 11:04:31.539	3	99
testpatient	F	35	20090807 11:04:32.563	3	99
testpatient	F	35	20090807 11:04:33.587	3	98
testpatient	F	35	20090807 11:04:34.611	3	98
testpatient	F	35	20090807 11:04:35.635	3	98
testpatient	F	35	20090807 11:04:36.659	3	98
testpatient	F	35	20090807 11:04:37.683	3	98
testpatient	F	35	20090807 11:04:38.707	3	97

A graphical representation of the above three table has been produced to easily view TAs performed on the RR, SpO_2 and HR values plotted above and below the threshold for each of the streams under investigation. The threshold is gender dependant and invoked by the integration of the patient table containing the gestational age, which in turn constantly amends TAs created for the different data streams as determined by the rules for that particular data stream stored in the TA_Rule table.

Abstractions were conducted on the RR readings from Table 6-9, where continuously monitored intervals of RR values at or above a reading of 25 are categorized into 'normal RR' abstraction, and continuous intervals of RR values below 25 are made into a 'low RR' abstraction. Figure 6-5 demonstrates that the values in the table can be reduced into normal and low abstractions:

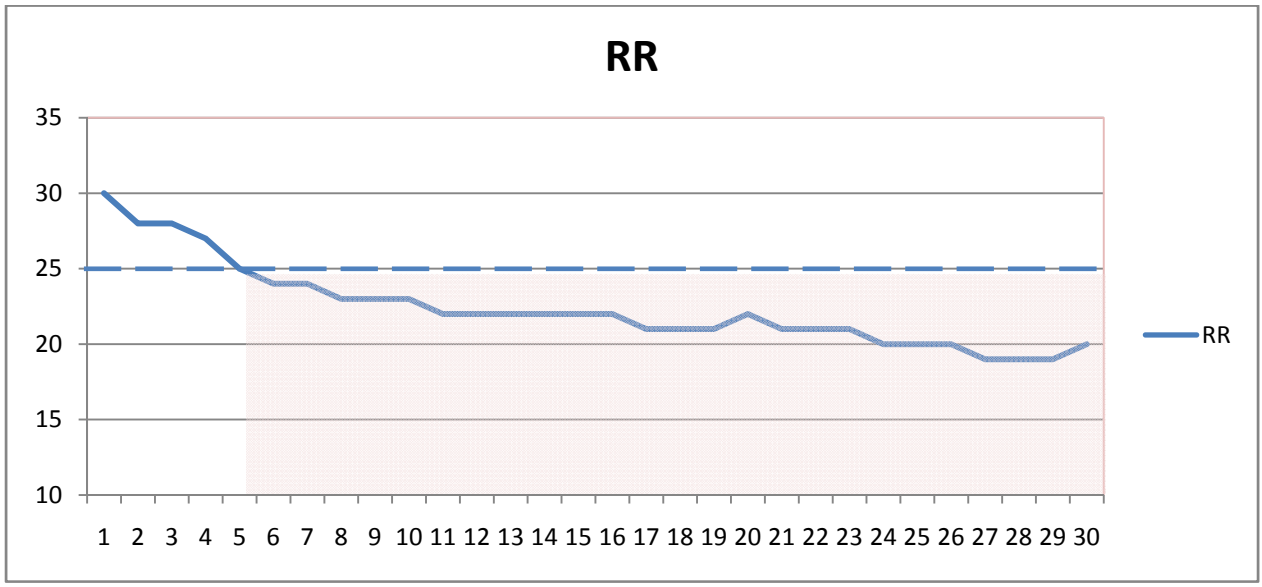


Figure 6-5 Graphing of Respiration Rate values

The rule for this particular abstraction, as presented within this case study, will be:

RR≥25

RR<25

Patient_ID	Gender	GA	Physiological_ID	Abstraction Type	AbstractionValue	ActualStartTime	ActualEndTime
testpatient	F	35	1	Level shift	normal RR	20090807 11:04:09.011	20090807 11:04:13.107
testpatient	F	35	1	Level shift	low RR	20090807 11:04:14.131	20090807 11:04:38.707

Table 6-12: RR Temporal Abstraction

Abstractions were conducted on the SpO_2 readings in Table 6-10, where continuous intervals of SpO_2 values at or above the 85% are categorized into 'normal' abstractions, and continuous intervals of SpO_2 values below 85% are made into a 'low' abstraction. Figure 6-6 was created from the SpO_2 values in Table 2 against the 85% threshold below.

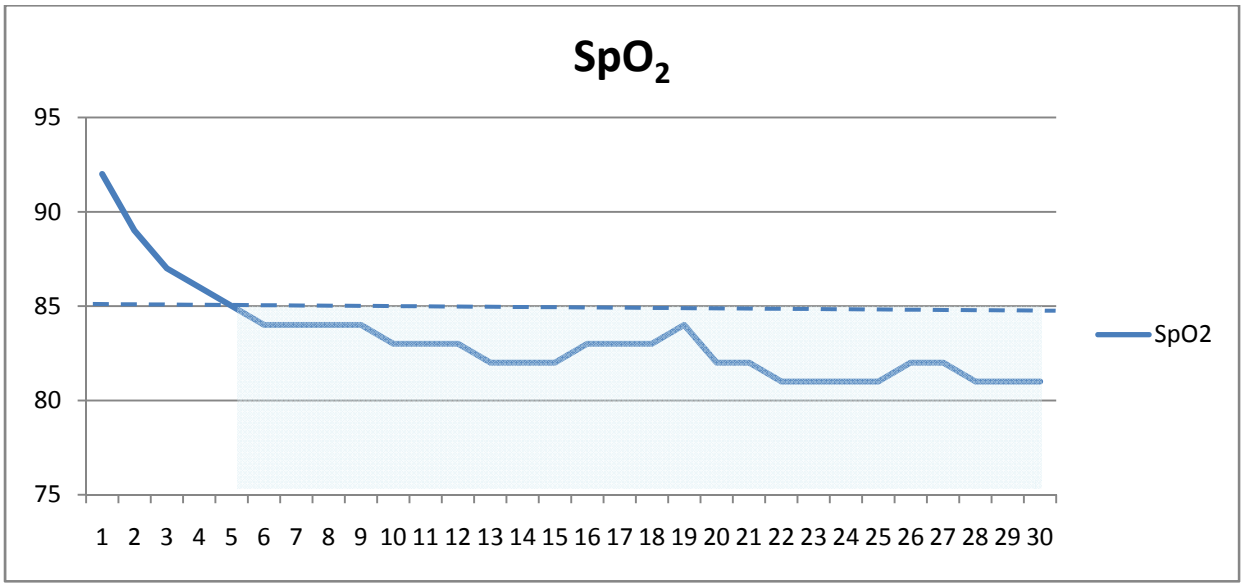


Figure 6-6 Graphing of SpO_2 values against threshold of 85%

SpO_2 threshold of 85% is indicated by the dotted line. SpO_2 readings of 85 and above are seen as normal, and readings below 85 can be problematic to the health and future of the neonate. The rule for this particular abstraction, using 85 as a threshold as presented within this case study, will be:

Low = $SpO_2 < 85$

Normal = $SpO_2 \geq 85$

Here we can see that the first 4 readings in Figure 6-6 are within the normal range, with a start time at 9.011 seconds and end time at 13.107 seconds, creating a 'normal' abstraction. The next readings are below the 85% threshold and therefore would create a 'low' abstraction, starting at 14.131 seconds and finishing at 38.707 seconds.

Table 6-13 represents the condensed data reading(s) outcome following the abstraction rules being applied to the original sample data:

Patient_ID	Gender	GA	Physiological_ID	AbstractionType	AbstractionValue	ActualStartTime	ActualEndTime
testpatient	F	35	2	Level shift	Normal	20090807 11:04:09.011	20090807 11:04:13.107
testpatient	F	35	2	Level shift	Low	20090807 11:04:14.131	20090807 11:04:38.707

Table 6-13: Abstractions created from all SpO_2 readings from Table

6.5.3 Gender comparison HR temporal abstractions

To demonstrate the significance that gender and gestational age can have on HR values present in Table 6-11, Figure 6-7 represents threshold parameters for a **male** neonate of **35** weeks GA against the threshold of 100 (which is gender and gestational age dependant, male and 35 weeks).

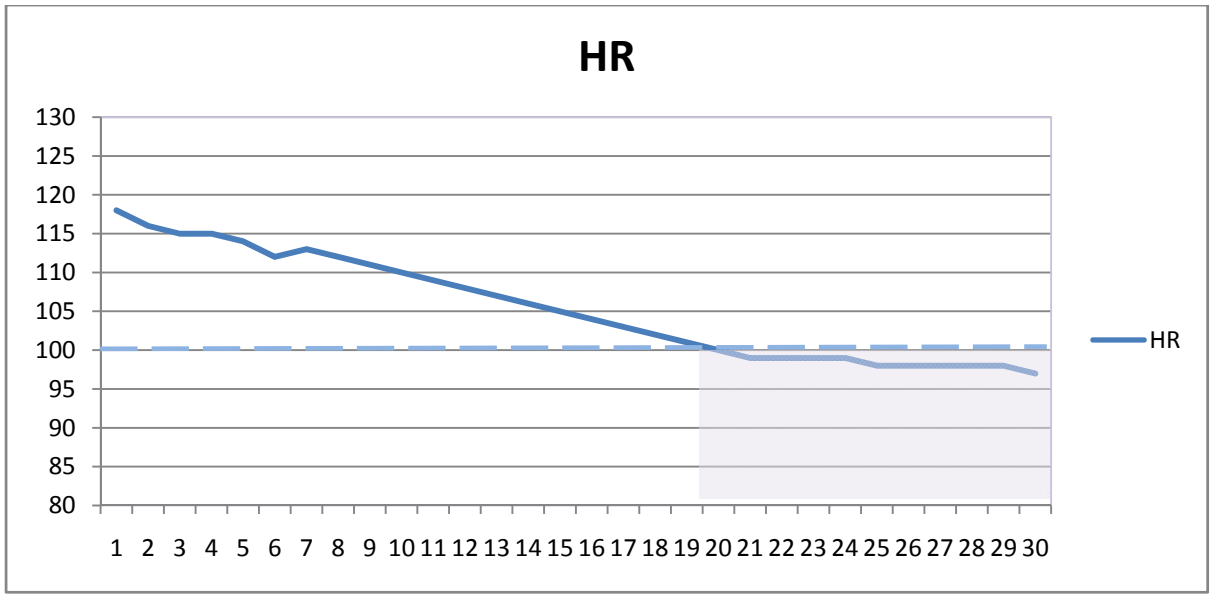


Figure 6-7 Graphing of HR values against a threshold of 100 bpm

The rule used to abstract the HR for a male neonate of 35 weeks GA parameter is:

Low = HR≤100

Normal = HR > 100

Here we can see that the first 19 readings in Figure 6-7 are not within and below the normal range from the first value until values cross the threshold at time of 28.467 seconds, creating a 'low' abstraction. The readings that followed were all below the 100 bpm threshold and therefore would create a 'low' abstraction, starting at time of 28.467 seconds and finishing at 38.707 seconds.

Table 6-14 represents the condensed results from applying abstraction rules to the original sample data from Table 6-11:

Patient_ID	Gender	GA	Physiological_ID	Abstraction Type	Abstraction Value	ActualStartTime	ActualEndTime
testpatient	M	35	3	Level shift	Normal	20090807 11:04:09.011	20090807 11:04:27.443
testpatient	M	35	3	Level shift	Low	20090807 11:04:28.467	20090807 11:04:38.707

Table 6-14: Abstractions created from all HR readings with threshold set at 100

These results differ significantly when compared to those obtained from the threshold that would be applied to a female neonate of 35 weeks gestational age. Figure 6-8 presents the HR values from Table 6-11 against the threshold of 108 (which is gender and gestational age dependant, female and 35 wks).

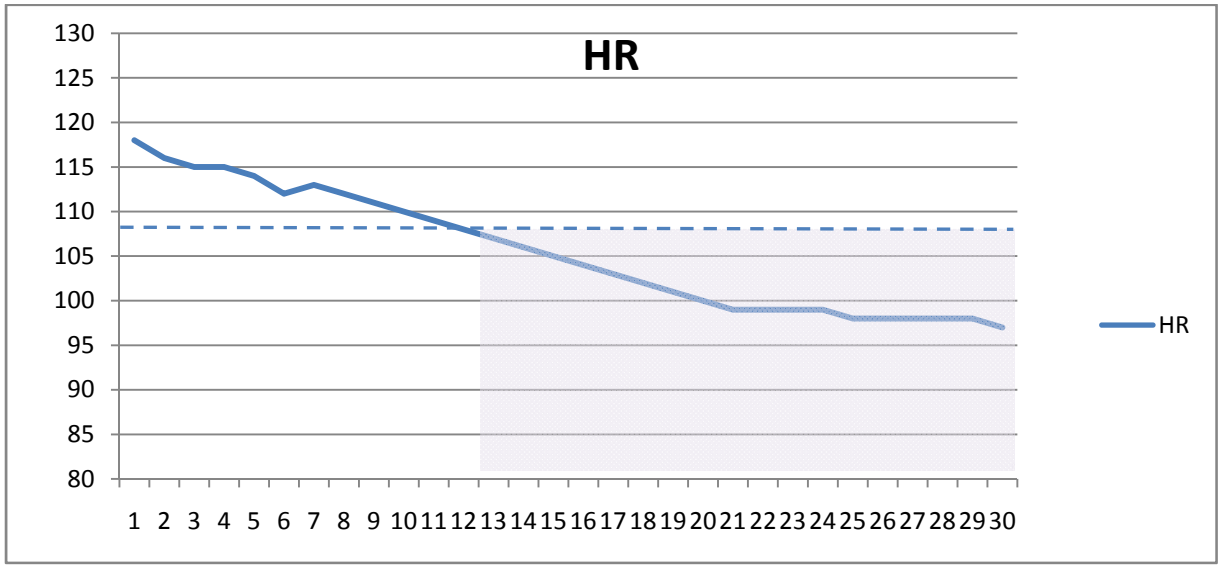


Figure 6-8 Graphing of HR values against a threshold of 108 bpm

The rule used to abstract the HR for a female neonate of 35 weeks GA parameter is:

Low = HR≤108

Normal = HR > 108

Here we can see that the first 11 readings are not within and below the normal range from the first value until values cross the threshold at time of 19.251 seconds creating a 'low' abstraction. The readings that followed were all below the 108 bpm threshold and

therefore would create a 'low' abstraction, starting at time of 20.275 seconds and finishing at 38.707 seconds.

Table 6-15 represents results from applying abstraction rules to the original sample data from table 6-11:

Patient_ID	Gender	GA	Physiological_ID	Abstraction Type	Abstraction Value	ActualStartTime	ActualEndTime
testpatient	F	35	3	Level shift	Normal	20090807 11:04:09.011	20090807 11:04:19.251
testpatient	F	35	3	Level shift	Low	20090807 11:04:20.275	20090807 11:04:38.707

Table 6-15 Abstractions created from all HR readings with threshold set at 108

Temporal abstractions have been shown on three data streams for this case study in ongoing clinical investigation and patient centric research. TAs were created for the different data streams as determined by the temporal abstraction rules.

6.5.4 Complex Abstractions

The next step within the CRISP-TDM methodology is blending abstractions from different data streams to create complex abstractions. Complex abstractions can be created from simple abstractions such as those created above for RR, SpO_2 and HR readings. For example, a complex abstraction can be specified when all streams being monitored are below their respective thresholds. The rule that must hold true for this example is RR <25 for > 15 seconds & SpO_2 <85 AND HR < 108, meaning only intervals where these conditions of interest are met will be included in this particular complex abstraction.

There is one time interval in this sample section of monitoring data that can be used for the complex abstractions where all three parameters have a low abstraction at the same time. These points are t_{20}- t_{30} inclusive and abstractions are summarized and presented as follows in Table 6-16:

Patient_ID	Gender	GA	Physiological_ID	Abstraction Type	AbstractionValue	ActualStartTime	ActualEndTime
testpatient	F	35	1,2,3	complex	CentralApneoaAlert	20090807 11:04:28.467	20090807 11:04:38.707

Table 6-16: Complex Abstractions

These complex abstractions are stored for referencing purposes in the data store.

This complex abstraction is demonstrated in Figure 6-9.

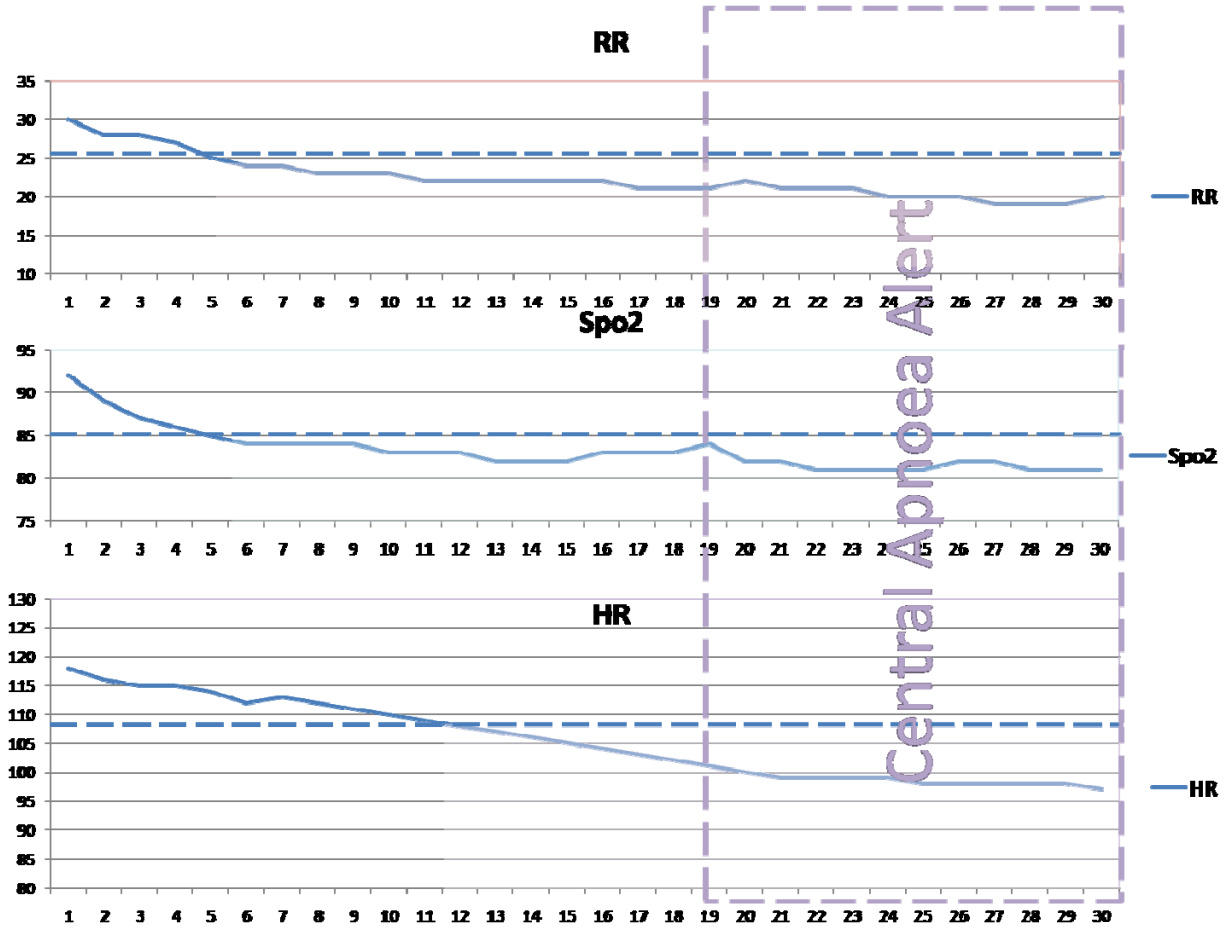


Figure 6-9 Complex abstraction demonstration of 3 physiological data streams

6.6 Relative Agent

Every abstraction created from physiological data in the Temporal Agent can be a part of many different clinical research studies. Both simple and complex abstractions are stored until needed in future studies. The Relative Agent in the $STDM^{n+p}{}_0$ Framework does not commence until a particular study is completed. Once a point in time of interest in the study is discovered, it is advantageous to realign the time of abstractions relative to that particular time. The aim of this case study is to find new trends and patterns that can be indicative to the onset of a condition in the physiological

parameters of the patient pre diagnosis; therefore, the next phase of the CRISP-TDM framework is to realign the time of abstraction relative to the particular point of interest. The point of interest is the time when the patient was diagnosed with a condition across multiple streams of physiological data, in comparison with multiple patients with the same diagnosis.

To enable the detection of particular pattern of these abstractions, at a particular time before diagnosis, realignment of the abstraction relative to the time of diagnosis is necessary. As these abstractions are using absolute time for the start and finish time for each abstraction, it will usually be necessary to give these abstractions start and finish times relative to a particular event that is of interest, such as the time of diagnosis. This will enable the comparison and mining of the abstractions, allowing the distance from time of diagnosis, or another event, to be taken into account (McGregor C. P., July 2010).

Clinical researchers looking for cross correlated changes in the temporal physiological data of patients with a particular condition are interested in changes in this temporal data that may be able to indicate the onset of this condition. They need to be able to identify similar patterns or changes in the data that occur at similar times before diagnosis for multiple patients. Using absolute times for the start and end time of abstractions give absolutely no indication of what time this abstraction takes place in relation to the diagnosis (McGregor C. P., July 2010).

6.6.1 Revisiting Temporal Abstractions

Table 6-10 showed abstractions from a patient's SpO_2 physiological data stream; it can be seen that the start and end times for each abstraction are absolute times recorded by the monitoring equipment in use. The first abstraction has a start time of 20090807 11:04:09.011 and a finish time of 20090807 11:04:13.107. For the purpose of this case study demonstration, if the patient was diagnosed one hour after the start time of the first abstraction recorded, the relative start time for this abstraction would be 00000000_01:00:00.000, exactly one hour before diagnosis. The relative times are created by calculating the difference between the actual times and the time of diagnosis. Trends and patterns in data of interest occur before diagnosis, and therefore data after the diagnosis or event should not be realigned. Table 6-17 contains the relatively aligned temporal abstractions for this particular example.

Patient_ID	Gender	GA	Physiological_ID	Abstraction Type	Abstraction Value	ActualStartTime	ActualEndTime
Testpatient	F	35	2	Level shift	Normal	00000000 01:00:00.000	00000000 00:59:55.904
Testpatient	F	35	2	Level shift	Low	00000000 00:59:54.880	00000000 00:59:30.304

Table 6-17: Relative aligned temporal abstractions

These patient characteristic physiological data parameters will be stored within tables found in the Temporal and Relative Temporal databases. Although static in nature for the purposes of this demonstration, this is an iterative process that is continually deployed on all data as it is being continuously collected.

6.7 Functional Agent

The functional agent is the agent that performs the framework data mining tasks. This is where exploratory data mining is used to detect new trends and patterns in multiple parameters to create hypotheses that can be tested via null hypothesis testing through confirmatory mining. This is demonstrated in the case study, where we are searching for trends and patterns in the temporal abstractions that indicate the onset of apnoea events based on gender and gestational age; such events are also possible cofounders for nosocomial infection.

Considering the case study demonstrated above, further investigations could be placed on the relationship between the various streams of data such as blood oxygen saturation and whether the rule of SpO_2 equaling GA holds true for both genders based on temporal abstractions stored from that study. First exploratory data mining will be exercised to find new hypotheses. An example of such a hypothesis is as follows:

[Breathing pause and {SpO_2<87 (Female) or <85 (Male)} and {HR<108 (Female) or <100(Male)}] $\geq$ 15 seconds $\Longrightarrow$ *Central Apnoea*

Once the hypothesis is formulated the null hypothesis can then be created and tested. A null hypothesis would state that there is no difference in the SpO_2 readings between female and male infants. If confirmatory mining proves the null hypothesis to be correct, the process is discontinued. However, if confirmatory mining proves the SpO_2 readings are in fact different for female and male neonates, the null hypothesis is disproven which warrants further investigations. Clinicians' input and judgment will decide if the

hypothesis is sound enough to be adopted as a rule for an intelligent monitoring system or whether further investigation is required. If it is decided that the hypothesis is of sound nature then it is passed on to the rules generating agent.

6.8 Rules Generating Agent

All rules generated through hypothesis that are created and tested within the Functional Agent and that are clinically approved and adopted by physicians are then stored in the Rule Generating Agent. These rules are available to be used by intelligent monitoring and alerting system such as the one created by Stacey (Stacey, McGregor, & al., 2007) (McGregor C. P., July 2010).

6.9 Future Research Application

The $STDM^{n+p}_0$ framework can be applied to the clinical research subject area of investigation for potential onset indicators for sepsis and other multivariable conditions such as apnoea.

Once thresholds have been derived and hypotheses have been created, tested, and then transformed into rules within this framework, the next stage is enabling the distribution of the framework to interact with the other Artemis locations.

The $STDM^{n+p}_0$ framework will continue to be developed in the Artemis project with a more rich set of actual de-identified data sets from The Hospital for Sick Children, Toronto, Canada, Women and Infants Hospital, Providence, Rhode Island and

Westmead Hospital, Westmead, Australia and multiple other hospitals around the world as the project continues to gain researchers' interest across the globe.

6.10 Conclusion

This research further extended the multi-dimensional $STDM^{n}_{0}$ framework (McGregor C. P., July 2010) to enable the creation of patient characteristic defined rules to be adopted into clinical alerts with the intent of improving patient outcomes.

This chapter was a demonstration of how the $STDM^{n}_{0}$ framework was extended to incorporate gender and gestational age applied within the NICU to further define patient characteristic rules through the extension of the Processing and Temporal Agents' functionality. The extended framework illustrated the adoption of the two new attributes included in the temporal abstractions and relative alignments made to the raw physiological data collected from within a NICU. Once the functional agent's exploratory and confirmatory data mining tasks have been deployed and these newly developed patient characteristic hypothesis it is then up to clinicians' judgment whether to adopt the patient characteristic rule(s) for intelligent monitoring systems (CDSS).

When considering the development of clinical rules that will be adopted into CDSS it important not to adopt a 'one size fits' all approach. This $STDM^{n+p}_{0}$ framework as presented within this case study chapter has demonstrated a way of creating more individualize patient characteristic approach to neonatal treatment of care.

This framework will provide clinical research within the NICU with the flexibility to adjust physiological data thresholds to meet the changing needs of the developing patient being monitored. These thresholds will be patient characteristic derived and based on trends and patterns discovered with the intent of improving patient outcomes.

Chapter 7 – Conclusion

7.1 Summary

The research covered in this thesis has presented extensions to the $STDM^{n}_{0}$ framework (McGregor C. P., July 2010) to enable data collection and clustering based on patient characteristics resulting in the $STDM^{n+p}_{0}$ framework. This research was demonstrated within the context of a NICU in Chapter 6 through the utilization of multiple time series physiological data streams collected from newborn infants enrolled in a collaborative research project including UOIT, IBM and SickKids called Artemis. The extensions demonstrated the inclusion of gender and gestational age into the multidimensional model to define patient characteristic thresholds for these attributes in relation to thresholds set for the detection of apnoea spells and thereby to assist in the support of clinical research within a neonatal intensive care.

Based on recent literature, other than the $STDM^{n}_{0}$ framework included as a basis for this research, there is an absence of architectures with a flexible multidimensional approach to data mining of time series data and integration of null hypothesis that has been adopted in healthcare domain let alone adopted into critical care. However exposed by literature review was the absence of adoption of the gender and gestational age to keep pace with the exponential growth and development that takes place in as an infant matures from 26 to 40 week gestation. Together with this there is an absence of support of the general need to incorporate characteristics of the entity in patient(s) to drive the knowledge discovery.

Gaps exposed by literature review resulted in the following research hypothesis for this thesis:

1. *That a patient characteristic multidimensional data mining framework can be defined for clinical research to enable use of patient attributes when data mining patient physiological data streams.*
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multi-dimensional temporal data.*
3. *The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.*
4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.*

The application domain for this research of the NICU was introduced in chapter 3. This chapter discussed and exposed the neonate as a growing and developing physical being both dependant on gender and driven by gestational age which should be reflected in rules that assist in the diagnosis of conditions that may affect these infants. Chapter 4 described an existing multi-agent framework $STDM^{n}_{0}$ that includes the integration of relevant aspects of the extended data mining model CRISP-TDM to support temporal data mining as well as facilitate null hypothesis testing on real time

series physiological data streams. Chapter 5 described the extensions made and design to incorporate the patient characteristic attributes of gender and gestational age into the $STDM^{n+p}_0$ framework. The extensions made to each of the agents and their functions were described fully. Chapter 5 also contained the design of the extended tables to be stored in the $STDM^{n+p}_0$ framework. Chapter 6 demonstrated the extended functionality of the $STDM^{n+p}_0$ framework presented within the NICU context. This apnoea event research case study utilized multiple time series physiological data streams collected from newborn infants enrolled in a collaborative research project including UOIT, IBM and SickKids called Artemis to demonstrate patient characteristic temporal abstractions and the realignment of the temporally abstracted data.

The research hypothesis generated within this thesis have addressed as follows:

1. *That a patient characteristic multidimensional algorithm framework can be defined for clinical research to enable pre-diagnosing at the onset of illness conditions based on trends and patterns discovered.* Chapter 5 and 6 discuss the design elements to include gender and gestational within the $STDM^{n+p}_0$ framework to enable clustering of patient characteristic data mining on multiple physiological data streams for multiple patients.
2. *The abovementioned patient characteristic framework will include methods for applying temporal abstraction (TA) across multiple parameters for multiple patients to enable mining of patient characteristic multidimensional temporal data.* This research focus fell predominantly on two agents of the existing framework, those being; Processing Agent and the Temporal Agents. Extensions

made to these agents were discussed in Chapter 5 and demonstrated in Chapter 6.

3. *The multidimensional algorithm framework can be applied in a neonatal context clustering patient characteristics by gender and gestational age.* Chapter 6 demonstrated the extended $STDM^{n}_{0}$ framework within the neonatal context on data currently collected by the Artemis project at The Hospital for Sick Children in Toronto with the support and under the supervision of neonatologist Dr. Andrew James. Physiological data streams used within this demonstration were RR, SpO_2 and HR. Apnoea spells are part of an ongoing study as part of a precursor to nosocomial infection (NI).
4. *The hypotheses generated by the patient characteristic framework can be used by a real-time event stream processor analysing the current condition of babies in a Neonatal Intensive Care Unit.* The real-time event stream processor is discussed in Chapter 3 NICU. This adoption within the $STDM^{n}_{0}$ framework is in Chapter 4 and Chapter 5 as well as demonstrated framework concept in Chapter 6. There is ongoing research at present investigating the stream processor within the Artemis project.

7.2 Contributions

The research area contributions to knowledge within this thesis are, specifically;

- Extensions to a $STDM^{n+p}_{0}$ designed multi agent framework for analysing time series data, facilitate use of attributes such as gender and gestational age into

multidimensional inclusion of gender and gestational age driven temporal abstraction and realignment of these abstractions.

- Design of extensions to the $STDM^{n+p}_0$ framework to enable patient characteristic multi-dimensionality to temporally abstractive data mining.
- Demonstrating the potential benefit and use of data mining from electronically stored physiological data, for improved real-time clinical management and patient characteristic clinical decision support is significant.
- Demonstrate the potential for clinical research on stored physiological data streams to deduce new findings for condition onset prediction indicators in support of a current ethics approved clinical research study. As is the potential for clinical research on stored physiological data streams to deduce new findings for condition onset prediction indicators and in support of a current ethics approved clinical research study.

7.3 Future Research

In the healthcare domain and in particular healthcare research there is an inherit ongoing limitation restricted by the lack of availability of data to be analysed as 'secondary use' data for the purpose of developing clinically relevant algorithms for use in this domain (Clarke, 2003).

As the electronic healthcare domain evolves, so too will the data mining software developers use to produce well designed analysis tool to pursue knowledge discovery in real-time physiological data streams (Lyman, 2008). With the interest in healthcare

domain increasing and as issues related to receiving approval of use of data for secondary analysis of health data are resolved, there is optimism that the cost and effort barriers to data mining projects will decrease. The $STDM^n_0$ framework, as proven by the adoption of the Artemis project, is a leading edge architecture capturing all aspects from collection on through to applying newly adopted rules in real-time. The research contained within this document extends its processing capability to patient characteristically define these rules by gender and gestational age.

Within this thesis the research has proposed that the newly defined Patient Characteristic rules are stored within a single physical database, however, as these findings become more defined and adopted as clinical rules there will be a need for multi-centre studies and multi-centre implementation where a distributed functionality option of the framework will be required.

Based on the systematic literature review of this area it is evident that there is exploratory research going on in physiological data analysis area; however, with the onset of the electronic health era, there is still much work to be done in addressing standards for real-time data collection and storage as well as function of databases and data mining methodologies.

The $STDM^n_0$ framework is currently the foundation system behind the Knowledge Extraction component of the collaborative project Artemis that is currently established and collecting data from The Hospital for Sick Children in Toronto, IWK Health Centre in Halifax, Children's Hospital of Eastern Ontario in Ottawa, Women and Infants Hospital

in Providence, Rhode Island, and soon beginning in Shanghai and Shenzhen China. As this project gains momentum and is enabled in more NICU centres worldwide, so too is the ability to collect, analyse and mine more data. Exploratory data mining to further refine and define patient characteristic rules can only prove to assist in the care of these infants being monitored.

7.4 Impact

The intent of this research is to not only to explore data collected from medical devices but also justification for their usefulness in serving a valuable purpose for storing data collected to assist in the improving the provision of better care within the neonatal intensive care context. As indicated by its absence in literature the exponential activity, by nature, of a growing organism in its early stages of life, for example, the preterm infant has not been captured.

Every human by nature is created genetically different resulting in very different characteristics that start from conception and continue throughout life. When caring for critically ill preterm infants it is important to incorporate individual characteristic to assist in shift towards individualized treatments of care when considering developing clinical rules that will be adopted by CDSS.

The $STDM^{n+p}_0$ framework extensions presented within this research thesis enables multidimensional data mining to detect patterns of a patient characteristic predictive temperament within the NICU domain. The $STDM^{n+p}_0$ framework provided a structure

approach to the development of patient oriented trends to be captured, analyzed and finding(s) extracted and embedded into algorithms designed to assist in the predictive trends at the early onset of conditions such NI.

7.5 Conclusion

Unfortunately, even in this day and age, it is still quite common for the vast amounts of real-time physiological stream data to be unexplored and quite often unsaved. This situation leaves little to opportunity for the data to unveil valuable information that could, in future, assist clinicians with earlier recognition and diagnosis leading to better prognosis for the life of the newborn infant.

The extended framework has demonstrated the capturing of patient characteristic attributes in temporal abstractions and the realignment of the abstractions relative to an event in a particular study. The $STDM^{n+p}_{0}$ framework will continue with its investigations into exploring and discovering different patient characteristic trends and patterns across multiple retrospective physiological data streams as part of a collaborative research project.

References:

Abdel-Rahman, Y., Jeremic, A., & Tan, K. (2009). Neonatal Heart Rate Prediction. *31st Annual International Conference of the IEEE EMBS* (pp. 4695-4698). Minneapolis, Minnesota, USA: IEEE.

American Heart Association. (2006). 2005 American Heart Association (AHA) Guidelines for Cardiopulmonary Resuscitation (CPR) and Emergency Cardiovascular Care (ECC) of Pediatrics and Neonatal Patients: Neonatal Resuscitation Guidelines. *Pediatrics Official Journal of the American Academy of Pediatrics* , 1029-1038.

American Heart Association, American Academy of Pediatrics. (2006). 2005 American Heart Association (AHA) Guidelines for Cardiopulmonary Resuscitation (CPR) and Emergency Cardiovascular Care (ECC) of Pediatrics and Neonatal Patients: Neonatal Resuscitation Guidelines. *Pediatrics Official Journal of the American Academy of Pediatrics* , 1029-1038.

Apiletti, D., Baralis, E., Bruno, G., & Cerquitelli, T. (May 2009). Real-Time Analysis of Physiological Data to Support Medical Applications. *Information Technology in Biomedicine* , Vol. 13, No. 3, pg. 313-321.

Azhim, A., & Kinouchi, Y. (2009). Impacts of Age and Gender on Arterial Hemodynamic Functions in Normal Subjects. *IEEE Symposium on Industrial Electronics and Applications* (pp. 446-451). Kuala Lumpur, Malaysia: IEEE.

Baker, G. R., Norton, P. G., Flintoft, V., Blais, R., Brown, A., Cox, J., et al. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *CMAJ* , 1678-86.

Benjamin, D. K., & Stoll, B. J. (2006). Infection in Late Preterm Infants. *Clinics in Perinatology* , 33:871-882.

Bjering, H. (2008). A Framework for Temporal Abstractive Multidimensional Data Mining. School of Computing and Mathematics, University of Western Sydney, Sydney, Australia.

Bjering, H., & McGregor, C. (2010). A Multidimensional Temporal Abstractive Data Mining Framework. *Proc. 4th Australasian Workshop on Health Informatics and Knowledge Management* (pp. Conferences in Research and Practice in Information Technology Vol. 108 pg. 29-38). Brisbane, Australia: Copyright © 2010, Australian Computer Society, Inc.

Blount, M., Ebling, M. R., Eklund, J. M., James, A. G., McGregor, C., Percival, N., et al. (2010). Real-Time Analysis for Intensive Care - Development and Deployment of

the Artemis Analytic System. *IEEE Engineering in Medicine and Biology Magazine* , 110-118.

Branswell, H. (2009). *Preterm births up 25 percent in Canada.* Retrieved April, from The Canadian Press (thestar.com): http://www.healthzone.ca/health/article/579462

Canada Department of Finance. (2010). *Canada Health Transfer*. Retrieved 2011, from Department of Finance Canada: http://www.fin.gc.ca/fedprov/cht-eng.asp

Catley, C., Smith, K., McGregor, C., & Tracy, M. (2009). Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case study. *22nd IEEE International Symposium on Computer-Based Medical Systems, 2009* (pp. 1-5). Albuquerque, NM: IEEE.

Catley, C., Smith, K., McGregor, C., & Tracy, M. (2009). Extending CRISP-DM to Incorporate Temporal Data Mining of Multi-dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study. (p. 5). IEEE.

Catley, C., Smith, K., McGregor, C., James, A., & Eklund, J. M. (2010). A Framework to Model and Translate Clinical Rules to Support Complex Real-time Analysis of Physiological and Clinical Data. *IHI'10.* Arlington, Virginia, USA.: 2010 ACM.

CIHI. (2009). *Too Early, Too Small: A Profile of small Babies Across Canada.* Ottawa: Canadian Institute for Health Information.

Clarke, R. (2003, January 20). *Xamax Consultancy, Canberra, Visiting Professor, CSIS, Uni of Hong Kong, Visiting Fellow, Australian National University.* Retrieved 11 23, 2009, from http://www.anu.edu.au/people/Roger.Clarke/Res/51-NonEmp.ppt: http://www.rogerclark/Res/51-NonEmp.ppt

CNN. (2008). *The Canadian Neonatal Network Annual Report 2008.* Retrieved 2010, from Canadian Neonatal Network: http://www.canadianneonatalnetwork.org/portal/

Comeford, D. (2008, November). Patient Flow Models Essential Elements of Flow. NSW, Australia.

CPS. (2006, June). *Levels of neonatal care.* Retrieved June 2010, from Canadian Paediatric Society (CPS): http://www.cps.ca/english/statements/FN/fn06-02.htm#Level1

CRISP-DM. (2007, January 18). Retrieved October 22, 2009, from CRISP-DM, CRoss Industry Standard Process for Data Mining: www.crisp-dm.org

Drummond, W. H. (2009, October). *Neonatal Informatics Dream of a paperless NICU: Part One: The Emergence of Neonatal Informatics.* Retrieved April 2010, from NeoReviews.org: http://neoreviews.aappublications.org/cgi/reprint/neoreviews;10/10/e480

Eklund, J. M., McGregor, C., & Smith, K. (2008). A Method for Physiological Data Transmission and Archiving to Support the Service of Critical Care Using DICOM and HL7. *IEEE EMBS conference.* Vancouver.

Emergency Department Admissions Algorithm. (2004, May 14). Retrieved April 10, 2010, from the Ottawa Hospital: http://www.anesthesia.org/Civic/Policies/ED_%20Admission_Algorithm_REVISED_May_14_2004.pdf

Fine B., G. B. (2009). Leading Lean: A Canadian Healthcare Leader's Guide. *Healthcare Quarterly Vol. 12 No. 3* , 32-41.

Fowler, M. (1996). *Johannes Kepler.* Retrieved March , 2010, from University of Virginia: http://galileo.phys.virginia.edu/classes/109N/1995/lectures/kepler.html

Galarraga, M., Serrano, L., Martinez, I., de Toledo, P., & Reynolds, M. (2007). Telemonitoring Systems Interoperablility Challenge. *IEEE EMBS conference.*

Garson, G. D. (2008). *Path Analysis.* Retrieved 2010, from http://faculty.chass.ncsu.edu/garson/PA765/path.htm

Griffin, P., & Moorman, R. (2001). Toward the early diagnosis of neanatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics* , vol. 107, no. 1, pp. 97-104.

Harrison, J. (2008). Introduction to the Mining of Clinical Data. *Clin Lab Med* , 28: 1-7.

Harvey, D. (2004, January). Lean, Agile. Retrieved from http://www.davethehat.com/articles/LeanAgile.pdf

Heath, J. (2006). A Framework for an Intellignent Decision Support System (IDSS)m Ubckydubg a Data Mining Methodology, for Fetal-Maternal Clinical Practice and Research. School of Computing and Mathematics. Sydney, University of Western Sydney, Australia.

Hein, H., Ely, J., & Lofgren, M. (April 1998). Neonatal respiratory distress in the community hospital: when to transport, when to keep-Clinical review. *Journal of Family Practice* , http://findarticles.come/p/articles/mi_m0689/is_n4_v46/ai_20582095/.

HL7. (2007). *HL7*. Retrieved October 23, 2009, from HL7, Health Level Seven: http://www.hl7.org/

Ho, T., Kawaski, S., Quang, L., Takabayashi, K., & Yokoi, H. (2004). Combining Temporal Abstraction and Data Mining to Study Hepatitis Data. SIG-KBS.

Holmes, H. J. (2007). Intelligent data analysis in biomedicine. *Journal of Biomedical Informatics* , 40: 605-608.

IHI. (1991). *Institute for Healthcare Improvement*. Retrieved February 10, 2010, from www.IHI.org: http://www.ihi.org/ihi/about

Infoway, C. H. (2009). *Canada Health Infoway.* Retrieved October 15, 2009, from Canada Health Infoway: http://www.infoway-inforoute.ca/lang-en

ISO. (2009). *ISO*. Retrieved October 23, 2009, from ISO, International Standards for Business, Government and Society: http://www.iso.org/iso/home.htm

Jiawei, H., & Kamber, M. (2001). *Data Mining Concepts and Techniques.* San Francisco: Morgan Kaufmann.

K. McGrath, D. B.-T. (March 2008). Implementing and sustaining transformational change in health care: lessons learnt about clinical process redesign. *MJA, Volume 188, Number 6* , 32-35.

Kamaleswaran, R., McGregor, C., & Eklund, J. M. (2010). A Method for Clinical and Physiological Event Stream Processing. *32nd Annual International IEEE EMBS Conference* (p. 4). Buenos Aires, Argentina: IEEE.

Kasanen, E., Lukka, K., & Siitonen, A. (1993). The constructive approach in management accounting researach. *Journal of Management Accounting Reseach* , Vol. 5, pp. 243-264.

Krueger, C., van Oostrom, J. H., & Shuster, J. (2010). A longitudinal Desccription oof Heart Rate Variability in 28-34-Week-Old Preterm Infants. *Biological Reseach for Nursing* , 11(3) 261-268.

Kunac, D. L., & Reith, D. M. (2005). Identification of Priorities for Medication Safety in Neonatal Intensive Care. *Drug Safety* , 251-261.

Leeb, K., Jokovic, A., Sandhu, M., & Zinck, G. (2006). Intensive Care in Canada. *Healthcare Quarterly* , vol. 9, pg. 32-33.

Lehmann, C. U., Kim, G. R., & Johnson, K. B. (2009). *Pediatric Informatics - Computer Applications in Child Health.* Baltimore, MD, USA: Springer Science & Business Media, p. 29.

Lyman, J., Scully, K., & Harrison, J. (2008). The Development of Health Care Data Warehouses to Support Data Mining. *Clin Lab Med* , 28: 55-71.

M. Harry & R. Schroeder. (2005). *Six Sigma: The Breakthough Managment Strategy Revolutionizing The World's top Corporations.* New York, NY: Currency Doubleday (a division of Random House, Inc.).

March of Dimes Foundation. (2010). *march of dimes*. Retrieved 2010, from Funding for prematurity - related research: http://www.marchofdimes.com/advocacy/publicresearch_prematurityrequests.html

Mayock, D. E. (2009). *Apnea.* Retrieved 2010 from NICU-WEB University of Washington Academic Medical Center: http://depts.washington.edu/nicuweb/NICU-WEB/apnea.stm

McGregor, C. P. (July 2010). *Patent No. 089705-0009.* Canada, Gatineau Quebec.

McGregor, C., & Eklund, J. M. (2008). Real-time Service-Oriented Architectures to Support Remote Critical Care Trends and Challenges. *IEEE COMSAC conference.*

McGregor, C., & Smith, K. P. (2009). A Survey of Physiological Monitoring Data Models to Support the Service of Critical Care. *33rd IEEE International Computer Software and Applicaitons Conference* (pp. 104-109). Seattle, Washington, USA.: COMPSAC2009.

McGregor, C., Kneale, B., & Tracy, M. (2007). On-demand Virtual Neonatal Intensive Care units supporting rural remote and urban healthcare with Bush Babies Broadband. *Journal of Network and Computer Application* , 30: 1309-1323.

McGregor, C., Purdy, M., & Kneale, B. (2005). Compression of XML Physiological Data Streams to Support Neonatl Intensive Care Unit Web Services. *IEEE International Conference on e-Technology, e-Commerce, and e-Service* (pp. 486-489). Hong Kong: IEEE.

McMillan, J. A., Feigin, R. D., DeAngelis, C. D., & Jones, M. D. (2006). Oski's Pediatrics: Principes & Practice (Fourth Edition). Philadelphia, PA, USA.: Lippincott Williams & Wilkins.

Medical Dictionary-The Free Dictionary. (2010). Retrieved 2010 November from http://medical-dictionary.thefreedictionary.com/apnea

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Informaiton. *The Psycology Review* , Vol. 63, pg. 81-97.

Mishra, S., Agarwal, R., Jeevasankar, M., Aggarawal, R., Deorari, A. K., & Paul, V. K. (2007). *Apneo in the Newborn.* Ansari Nagar, New Delhi: AIIMS.

Nagy, E., & Orvos, H. (2000 Volume 47 - Issue 6 - pp 778-790-June). *Gender-Related Heart Rate Differences in Human Neonates.* Retrieved 2010 from Pediatric Research Journal: http://journals.lww.com/pedresearch/Fulltext/2000/06000/Gender_Related_Heart_Rate_Differences_in_Human.16.aspx#

NHS. (2007). *NHS Institute for Innovation and Improvement*. Retrieved February 2010, from www.insitute.nhs.uk: http://www.institute.nhs.uk/

NIH. (2010 15-November). *Apnea of Prematurity*. Retrieved 2010 17-Nov from MedlinePlus: http://www.nlm.nih.gov/medlineplus/ency/article/007227.htm

Nimavat, D. J., Sherman, M. P., Santin, R. L., & Protat, R. (2009 26-Oct). *e-medicine.* Retrieved 2010 Nov from WebMD: http://emedicine.medscape.com/article/974971-overview

Okascharoen, C., Hui, C., Cairnie, J., Morris, A. M., & Kirpalani, H. (2007). External validation of bedside prediction score for diagnosis of late-onset neonatal sepsis. *Journal of Perinatology* , 496-501.

Ontario_Health_Coalition. (2009, June 18). Putting Patients At Risk:. Toronto, Ontario, Canada.

Oxford Dictionary. (2011). Retrieved 2011 April from Oxforddictionaries.com: http://oxforddictionaries.com/view/entry/m_en_gb0034080#m_en_gb0034080

pregnancy.org. (n.d.). *pregnancy.org*. Retrieved February 2010, from pregnancy.org: http://www.pregnancy.org/fetaldevelopment

Priest, L. (2010, May 22). Why are our babies dying? *The Globe and Mail; Ontario Edition* . Toronto, Ontario, Canada: The Globe and Mail.

R. Albrecht, S. M. (April 2008). Do Emergency Physicians Admit More or Fewer Patients on Busy Days? *Annals of Emergency Medicine, Volume 51, Issue 4* , 532.

REDESIGN. (2008, November 14). Patient Flow Systems. NSW, Australia.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tan, P. C., et al. (2007). Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatices Association White Paper. *Journal of the Americal Medical Informatics Association* , 1-9.

Sankaran, K., Chien, L.-Y., Walker, R., Seshia, M., Ohlssson, A., Lee, S. K., et al. (2002). Variations in mortality rates among Canadian neonatal intensive care units. *CMAJ* , 166(2) pg 173-178.

Seri, I., & Evans, J. (2008). Limits of viability: definition of the gray zone. *Journal of Perinatology* , 28, S4-S8.

Sharek, P. J., Horbar, J. D., Mason, W., Bisarya, H., Thurm, C. W., Suresh, G., et al. (2006). Adverse Events in the Neonatal Intensive Care Unit: Development, Testing, and Findings of an NICU-Focused Trigger Tool to Identify Harm in North American NICUs. *PEDIATRICS - Official Journal of the American Academy of Pediatrics* , 1332-1340.

Smith, L. (2001). Six Sigma and the Evolution of Quality in Product Development. *Six Sigma Forum Magazine* , 28-35.

Stacey, M., McGregor, C., & al., e. (2007). An Architecture for Multi-Dimensional Temporal Abstraction an dits Application to Support Neonatal Intensive Care. *Engineering in Medicine and Biology Society.* IEEE/EMB.

Stanford_University, S. o. (2010). *Lucile Packard Children's Hospital-Stanford University of Medicine-Stanford Hospital & Clinics*. Retrieved 2010 November from http://www.lpch.org/DiseaseHealthInfo/HealthLibrary/respire/apneapre.html

Stravroudis, T. A., Miller, M. R., & Lehmann, C. U. (2008). Medication Error in Neonates. *Clinics in Perinatology* , 35: 141-161.

Terry, L. (2009). *Premature - Isaac Newton.* Retrieved March 2010, from BellaOnline: http://www.bellaonline.com/articles/art48510.asp

Tong, C., Sharma, D., & Shadabi, F. (2008). A Multi-Agents Approach to Knowledge Discovery. *IEEE/WIC/ACM conference,* 571 - 574*.*

Trochim, W. M. (2006, 10 20). *Reseach Methods Knowledge Base*. Retrieved 10 19, 2009, from Research Methods Knowledge Base: http://www.socialresearchmethods.net/kb/statcorr.php

UIhealthcare. (2008). *Department of Pediatrics*. Retrieved 2010, from University of Iowa Children's Hospital - University of Iowa Health Care:

http://www.uihealthcare.com/depts/med/pediatrics/iowaneonatologyhandbook/general/bloodpressure.html

Verduijin, M., Sacchi, L., Peek, N., Bellazzi, R., de Jonge, E., & de Mol, B. (2007). temporal abstraction for feature extraction: A comparative case study in prediction from intensive care monitoring data. *Artificial Intelligence in Medicine* , 41: 1-12.

WHO 2011. (n.d.). *Health statistics and health information systems*. Retrieved February 2011, from Neonatal mortality: http://www.who.int/healthinfo/statistics/mortality_neonatal/en/index.html

Wu, J.-H., Wang, S.-C., & Lin, L.-M. (2005). What Drives Mobile Health Care? An Empirical Evaluation of Technology Acceptance. *Hawaii International Conference of System Sciences 2005* (pp. 150-154). Waikoloa, Big Island: IEEE.

Zach, T., Brown, J. C., & Kaftan, H. A. (2010 18-Mar). *Periventricular Leukomalacia*. Retrieved 2010 from emedicine from WebMD: http://emedicine.medscape.com/article/975728-overview

Zegers-Hochschild, F., Adamson, G. D., de Mouzon, J., Ishihara, O., Mansour, R., Nygren, K., et al. (November 2009). International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (Who) revised glossary of ART terminology, 2009. *Fertility and Sterility Vol. 5* , 1520-1524.

Zhang Y, & Szolovits, P. (2008). Patient-specific learning in real time for adaptive monitoring in critical care. *Journal of Biomedical Informatics* , 41: 452-460.

Zhang, Y. (2007). Real-time Development of Patient-specific Alarm Algorithms for critical care. *IEEE EMBS conference.*

Appendix

Glossary of terms

Alveoli

The alveoli are the final branchings of the respiratory tree and act as the primary gas exchange units of the lung. The gas-blood barrier between the alveolar space and the pulmonary capillaries is extremely thin, allowing for rapid gas exchange. To reach the blood, oxygen must diffuse through the alveolar epithelium, a thin interstitial space, and the capillary endothelium; CO2 follows the reverse course to reach the alveoli. [http://oac.med.jhmi.edu/res_phys/Encyclopedia/Alveoli/Alveoli.HTML]..

Bradycardia

Slowness of the heart rate, usually measured as fewer than 60 beats per minute in an adult human

Diastolic Blood Pressure

The diastolic pressure is the measurement of force as the heart relaxes to allow the blood to flow into the heart. High diastolic pressure is a strong predictor of heart attack and stroke in young adults.

Explicit knowledge

Is knowledge that has been or can be articulated, codified, and stored in certain media. It can be readily transmitted to others. The information contained in encyclopaedias are good examples of explicit knowledge.

Extracorporeal

Situated or happening outside the body

Hemodialysis

Dialysis of the blood

Hemofiltration

A technique similar to hemodialysis, used for removing waste products from the blood when the kidneys have failed

Hypovolemic

Hypovolemic shock is an emergency condition in which severe blood and fluid loss makes the heart unable to pump enough blood to the body. This type of shock can cause many organs to stop working. (http://www.nlm.nih.gov/medlineplus/ency/article/000167.htm)

Implicit knowledge

Is knowledge that has not stated, but understood in what is expressed.

Intraveneous

Existing or occurring inside a vein, or administered into a vein - used in administering fluids or medicines into the veins

Parenteral

Injected, infused, or implanted - describes drug administration other than by the mouth or the rectum, e.g. by injection, infusion, or implantation

Percutaneous - describes medication that is administered or absorbed through the skin

Periventricular leukomalacia (PVL)

Is the most common ischemic brain injury in premature infants. The ischemia occurs in the border zone at the end of arterial vascular distributions. The ischemia of PVL occurs in the white matter adjacent to the lateral ventricles. The diagnostic hallmarks of PVL are periventricular echodensities or cysts detected by cranial ultrasonography. Diagnosing PVL is important because a significant percentage of surviving premature infants with PVL develop cerebral palsy, intellectual impairment, or visual disturbances.

Postnatal Age

Gestational age is the 'age at birth'. After birth the term postnatal age is used and corrected gestational age (GA). Corrected GA = GA + PNA

Systolic Blood Pressure

The systolic pressure (the first and higher number) is the force that blood exerts on the artery walls as the heart contracts to pump out the blood. High systolic pressure is now known to be a greater risk factor than diastolic pressure for heart, kidney, and circulatory complications and for death, particularly in middle-aged and elderly adults.

Tachycardia - an excessively rapid heartbeat, typically regarded as a heart rate exceeding 100 beats per minute in a resting adult