NeoCAM

Development and Evaluation of a Neonatal Cerebral Autoregulation Monitor

H.B.A.Borghols

February 2018

Master's Thesis

Human-Machine Communication
Dept. of Artificial Intelligence,
University of Groningen, The Netherlands

Internal Supervisor:

Dr. F. Cnossen (Department of Artificial Intelligence, University of Groningen, Groningen)

External Supervisors:

Dr. E.M.W. Kooi (Department of Neonatology, University Medical Center Groningen, Groningen)

K. van Amsterdam, MSc. (Department of Anaesthesiology, University Medical Center Groningen, Groningen)

university of groningen

umcg

Contents

Abstract

Cerebral autoregulation is the mechanism that ensures that cerebral blood flow is constant during fluctuations of mean arterial blood pressure. In sick and preterm babies this mechanism is sometimes absent or not functioning properly. Absence of cerebral autoregulation is linked to adverse cerebral outcome and cognitive difficulties later in life. There are currently no commercially available solutions to monitor the state of cerebral autoregulation in patients.

This thesis describes the design and evaluation of a monitor for cerebral autoregulation. The main aim was to establish whether the addition of a cerebral autoregulation monitor would enhance the situation awareness of physicians. The design of the monitor was based on requirements that followed from a literature study on measures of cerebral autoregulation and on situation awareness and clinical reasoning. A working prototype of the cerebral autoregulation monitor was implemented and used in an evaluation study. In this study, eight neonatologists participated in a task where they assessed and made a treatment plan for simulated cases. The participants first performed this task on a monitor that was based on the standard monitors the neonatology department uses. Afterwards, they performed the same task with the standard monitor and the new cerebral autoregulation monitor. After completing the task, the participants were interviewed about their experiences with the cerebral autoregulation monitor and their opinion on using it in clinical practice.

It was concluded that the presence of the cerebral autoregulation monitor makes physicians more aware of cerebral autoregulation and promotes reasoning about it. However, it also enhances the uncertainty of the physicians while making a treatment plan. This can be explained by the novelty of the monitor. All physicians were positive about the potential use of a cerebral autoergulation monitor in clinical practice in the future. With further research into the use of cerebral autoregulation values in treatment, the cerebral autoregulation monitor could become a valuable clinical tool.

Chapter 1

Introduction

The neonatology department (NICU) is the intensive care department for sick and preterm babies (born before 37 weeks of gestation (Blencowe et al., 2012)). One of the consequences of being born preterm is the risk of neurodevelopmental problems later in life due to cerebral damage (Mitra et al., 2014). The mechanisms behind this damage are multifaceted and the damage is sustained over a longer period of time. However, one factor that seems to play a major role is cerebral autoregulation dysfunction.

Cerebral autoregulation is the mechanism that keeps cerebral blood flow (CBF) between acceptable margins when there are sudden changes in the cerebral perfusion pressure (CPP). CPP is the net pressure gradient causing cerebral blood flow. The CBF determines the amount of blood in the brain and therefore the amount of oxygen in the brain. When cerebral autoregulation is absent, CBF will rise or fall to harmful upper or lower levels during fluctuating CPP. Both a low and a high CBF can cause severe damage such as cerebral hypoxic-ischemic damage and intra- or periventricular haemorrhage (Mitra et al., 2014; Tsuji et al., 2000). The difference in the relationship between CBF and CPP given the presence or absence of cerebral autoregulation can be seen in figure 1.1.

The relationship between CPP and CBF can be used to measure the state of cerebral autoregulation. To measure cerebral blood flow, the neonatology department of the University Medical Centre of Groningen (UMCG) uses near infrared spectroscopy (NIRS, Villringer, Planck, Hock, Schleinkofer, and Dirnagl 1993). This does not measure blood flow directly but measures the oxygen saturation of superficial brain tissue (rc_{SO_2}). Rc_{SO_2} can be used as a surrogate measure for CBF, assuming a stable oxygen consumption according to the Fick principle (van Bel, Lemmers, & Naulaers, 2008; Wolf & Greisen, 2009). Mean arterial blood pressure (MABP) is most commonly used as a surrogate measure for cerebral perfusion pressure (Tsuji et al., 2000).

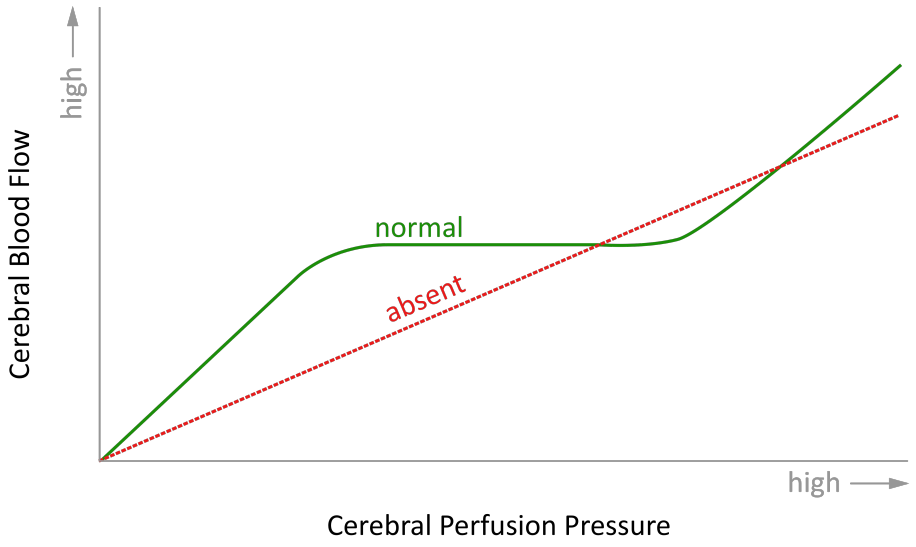

FIGURE 1.1: The difference in the relationship between CBF and CPP depending on the presence of cerebral autoregulation. The green line shows the relationship between CBF and CPP when cerebral autoregulation is functioning. The red line shows the relationship between CBF and CPP when cerebral autoregulation is not functioning properly.

The most accurate way to get a measure of cerebral autoregulation is still under debate in the literature. The combination of a surrogate for CBF and a surrogate for CPP are most commonly used as input parameters. The debate in the literature centers around which type of signal-processing technique to use to determine the strength of the relationship between the signals representing CBF and CPP. There are authors who use correlation-based methods and authors who use coherence-based methods (Caicedo et al., 2012; Kooi et al., 2017; Mitra et al., 2014; Tsuji et al., 2000; Wong et al., 2008), both of which have their own advantages and drawbacks. It will be necessary to choose one of the methods in designing a cerebral autoregulation monitor. Therefore the different approaches will be reviewed as part of this thesis.

Presently there are no commercially available monitors for cerebral autoregulation. As a consequence, physicians have to mentally construct an assessment of the state of cerebral autoregulation to be able to reason about it. This is a cognitively demanding task that involves integrating multiple, continuously changing parameters into a coherent representation of the patient state. This makes the process error-prone, and may divert mental resources from other tasks (Drews, Musters, & Samore, 2008). Due to the lack of possibilities to measure or monitor cerebral autoregulation, physicians may not actively use cerebral autoregulation in their reasoning at all.

Situation awareness (Endsley, 1995) is described as a general understanding of the task environment, how the current state of that environment arises from the past and the ability to make predictions about the future of the environment. It is crucial in decision making in dynamic systems, such as making decisions about the treatment of patients. Having a correct representation of the state of the patient's cerebral autoregulation is important in achieving situation awareness. A well-designed cerebral autoregulation monitor could enhance the situation awareness of physicians at the NICU. This thesis describes the design of a cerebral autoregulation monitor and a study evaluating whether the monitor enhances situation awareness, how the monitor influences clinical reasoning, the attitude of the physicians towards the monitor, and the usability of the monitor.

1.1 Structure of the thesis

The thesis can be divided in two parts. The first part of the thesis will provide necessary background for the second part of the thesis. The first part consists of a situation assessment of the NICU. This is followed by a comparison of methods to measure cerebral autoregulation. Finally there will be an overview of situation awareness and clinical reasoning literature focusing on how these concepts relate to patient monitoring.

The second part of the thesis will start with a chapter on the design of the cerebral autoregulation monitor. The remaining chapters will discuss the evaluation study. This will include a description of the methodology, an overview of the results of the evaluation study and a discussion of these results and the possibilities for further research.

Chapter 2

Situation Assessment of the NICU

Before designing a new interface it is important to consider the environment the interface will be used in as well as the interface's future users. This chapter will describe the environment of the neonatal intensive care unit (NICU) at the University Medical Centre Groningen (UMCG) and its daily operation. Two days were spent following physicians at the NICU in their daily work and one day was spent following the nursing staff to get an overview of the environment of the NICU. In this chapter, the environment and purpose of the NICU will be described, followed by the way it operates and ending on the technological possibilities and limitations of the NICU in relation to this project.

2.1 The environment of the NICU

The NICU is a unit for intensive medical care specialized in neonatal medicine. The organization of the NICU is similar to other intensive care units. Compared to general care units, ICUs often have a significantly higher patient to care-provider ratio (Drews, 2013). The patients on intensive care units are also monitored continuously and often by different means than in general care units, such as invasive blood pressure monitoring using arterial lines. At the time of this study there were two wards in the NICU, each containing a maximum of 12 patients.

There are three types of staff working in the NICU: physicians, nurses and additional staff. The tasks of the physicians consist of diagnosing the patients, assessing their progress, prescribing medication, documenting, and performing certain procedures such

as placing central lines. The nurses' task is to continuously monitor the patients, administering medication to the patients, feeding the patients, performing procedures that don't require a physician and closely documenting patient status. The additional staff consists of all other personnel required to support the core groups of physicians and nurses. Most of the staff in the NICU is specifically trained to work there. As the UMCG is a university hospital, there is also constantly new staff being trained. A closer look at the daily routine and tasks of physicians and nurses follows below.

2.2 Physicians in the NICU

During the day-shift, a section of the NICU is typically staffed by one attending neonatologist and two neonatology residents. Their day starts with the handover. This is a meeting in which the physicians of the outgoing shift brief the physicians of the incoming shift on what happened during the shift and on the status of the patients. This is followed by rounds, in which the physicians visit all patients and brief each other on their status. The residents both take the lead on half of the patients while the attending physician supervises them.

After rounds, the physicians retreat to the physicians' room where they go over each patient in more detail, together with the nurse(s) responsible for that patient. A plan is then formulated on how to treat a patient during the next 24 hours. Physician's activities afterwards vary from day to day, depending on the ever changing needs of the patients and the other activities that a physician is involved in such as performing procedures, visits of parents or administrative tasks. At the end of the day, the physicians do rounds again, visiting all patients to assess their status and they end the shift with the handover to the physicians of the next shift.

In addition to these routine activities, the physicians often have to deal with incidents that do not fit neatly into a schedule. For example: attending and reacting to acute detoriations in patients, attending births of high-risk patients, and being present while transporting patients to or from other regional hospitals.

2.3 Nurses in the NICU

Unlike the physicians, the nurses stay with the patients the entire day. They take part in discussing the plan for the patients they are responsible for and receive instructions of what to do during their shift. Apart from that, they are generally monitoring and taking care of the patients. This includes feeding the patients and changing their diapers.

They keep logs of relevant variables (depending on the medical state of the patient) on paper lists. Each variable that needs to be recorded is required to be sampled at its own frequency. This means that some variables are supposed to be recorded once every fifteen minutes while other variable are recorded once every three hours, with the sampling frequency of most variables falling somewhere in between these extremes. The nurses are also in charge of keeping track of all the medication that is administered and of administering it. The nurses sometimes have some type of schooling before the handover at the end of the day.

2.4 Monitoring in the NICU

The NICU makes use of many machines to monitor and support patients. This includes vital parameter monitors, infusion pumps, respiratory support systems and more. Only the monitoring equipment is relevant for this study. Each patients' vital parameters are monitored using a Philips IntelliVue MP70 monitor. When at risk for disturbances in cerebral oxygenation, this might be extended with a Somanetics INVOS oximeter 5100 monitor for NIRS and a ventilation machine. The Philips MP70 monitor displays waveforms for ECG, plethysmogram (oxygen saturation in the blood), capnogram (amount of CO_2 in respirated air), and ventilation frequency (amount of breaths per minute). The numerical variables displayed on the monitor are: heart rate, blood pressure (systolic, diastolic and mean), oxygen saturation (SpO_2), end tidal CO_2, Positive End Expiratory Pressure, and Peak Airway Pressure. The Philips MP70 monitor is partly adjustable to the personal preference of physicians or nurses. It is possible to rearrange the variables on the screen, alter the colors for the variables and display trend curves for each variable. Many physicians indicated not to be aware of these options during the study, for example by questioning why certain variables were displayed in specific colors. Also, few physicians made use of the possibility to display trend information (this was also found in a study by Drews et al. (2008)). Physicians who indicated they did use the trend information mentioned that they do not use the trend information at the bedside but rather use this information remotely at a later point in time when trying to diagnose a patient.

All monitors that are present on the NICU are closed systems. This means that the data they collect and display are not accessible outside of the system. It is only possible to view and use the data in the way provided by the manufacturers of these monitors, the raw data are not accessible. As a consequence of this, it is not possible to view or analyze the data in a way that is not built in by the manufacturers. Creating a display for cerebral autoregulation requires access to the raw data collected by the several sensors

applied to the patients. As it is not possible to obtain these raw data directly from the monitoring system, a different solution had to be found. The only other registration of the patients' data is the paper logs the nurses keep of the patients' progress. These have a much lower sample rate than the digital monitoring systems. Fortunately, the NICU was running a project at the time of this study called piCare. This project involves a Raspberry Pi computer attached to each patient's bed. This computer reads out the data from the sensors with a sample rate of 1 Hz and stores this information. The goal of piCare is to give researchers access to very specific data to be able to analyze patients over time or compare patients, for example to study the effects of certain treatments on patient outcome. PiCare also provides the opportunity to experiment with ways to visualize the data that are collected from the patients. This is desirable in situations such as this study, where a display is designed for a parameter that is not visualized on the current monitoring tools.

At the time of writing, the NICU was transitioning from a system of paper logs to a more advanced digital system. However, this was not yet in place at the time of the study. Even if it were, this would still have been a closed system without access to the raw data needed to implement the cerebral autoregulation display.

2.5 Implementing a cerebral autoregulation display in the NICU

When designing a cerebral autoregulation monitor it needs to be considered where the monitor will be placed and how the monitor will be used. The most conventional option is to make a bedside monitor that will mainly be used by the nurses. The nurses are in proximity of the monitors continuously due to the way the NICU is organized. They register the progress shown on the monitor and report to the physicians when a parameter moves past certain values set by the physicians. So any kind of display that is at the patients' bedside will be monitored by nurses for most of the time and then interpreted by physicians if anything unusual occurs. A cerebral autoregulation display would be no exception. However, knowledge about the state of cerebral autoregulation is most relevant for physicians when they are deciding on the course of treatment for the patient. This mainly happens during rounds or when discussing the case in the physicians' room. Therefore, it may also be useful for the physicians to be able to use a cerebral autoregulation monitor in the physicians' room. It is also possible to design a bedside monitor that can also be accessed remotely. A subgoal of this study will be to inquire which situation is preferable to the physicians.

Chapter 3

Measuring Cerebral Autoregulation

For the design of the cerebral autoregulation monitor it is necessary to determine which method of constructing a measure for cerebral autoregulation should be used to obtain the values that are displayed to the user. This measure will indicate to what degree cerebral autoregulation is present in the patient. The method of constructing the measure can influence such properties of the monitor as the frequency of presenting new information and the way the measure is visually explained to the user. An overview will be given of different methodologies to construct such a measure for cerebral autoregulation. Before a measure can be constructed it is necessary to establish which physiological data is measured and how. This means comparing methods of measuring cerebral oxygenation and comparing surrogate measures for cerebral perfusion pressure.

As outlined in chapter 1, cerebral autoregulation is the ability to keep cerebral blood flow (CBF) between acceptable margins when there are changes in the cerebral perfusion pressure (CPP). When cerebral autoregulation is absent, CBF will rise and fall to both harmful upper and lower levels during fluctuating CPP. It is difficult to measure cerebral autoregulation for two reasons. First, it is not trivial to determine the strength of the relationship between two signals that represents the state of cerebral autoregulation. Second, the two signals are difficult to measure directly. Measuring CBF directly would involve an invasive procedure in a patient's brain, and measuring CPP would involve a pressure meter in a blood vessel in the patient's neck. Measuring CBF in such a way is extremely dangerous and therefore not an option, measuring CPP as described above is technically impossible. Therefore it is necessary to use surrogate measures to get a representation of the changes in CBF and CPP. How CBF is measured using cerebral

oxygenation exactly will be described in the next section. How CPP is measured will be described after that.

3.1 Measuring Cerebral Blood Flow

Essential to the measurement of cerebral autoregulation is near infrared spectroscopy (NIRS). NIRS measures tissue absorbance of light in the spectral region from 700-1000 nm. This enables determining concentration changes of oxygenated haemoglobin (HbO2), deoxygenated haemoglobin (HbR) and blood volume changes (HbO2 + HbR). Light at the near infrared wavelengths is able to penetrate both biological tissue and bone, allowing NIRS to measure through the intact skull. Villringer et al. (1993) tested whether NIRS performed reliably as a bedside method to measure changes in cerebral haemoglobin concentration and oxygenation, reaching a positive conclusion.

Van Bel et al. (2008) compared different technical implementations of NIRS. The two main compared measures were the cerebral tissue oxygenation index (TOI) and the regional cerebral oxygenation index (rc_{SO_2}). They use the same concept but different algorithms to quantify the measure. Both measure the absorption of NIR light. For the calculation of TOI, absorption is measured at three points very near to each other and then combined using the diffusion equation [1] to obtain one measurement. rc_{SO_2} is calculated by measuring absorption at two optodes 2cm removed from each other and subtracting measurements at the first optode (at 2cm from the light source) from the second optode (at 4cm). By measuring in different places, both of these methods try to account for movement artefacts and issues arising through light scattering. Van Bel et al. (2008) gave an overview of studies in which either or both of these measures for were used and showed that both measures show similar baseline values of cerebral oxygenation in healthy adults. They also showed that both measures react similarly to hyperoxia and hypercapnia, as well as both having a linear correlation with arterial CO_2 values.

Van Bel et al. (2008) therefore concluded that TOI and rc_{SO_2}, despite their underlying technical differences, are both suitable measures for trends in cerebral oxygenation. Wolf and Greisen (2009) came to a similar conclusion.

[1] the diffusion equation is a common equation in physics that describes the behavior of the motion of a group of particles through a material

3.2 Choosing a surrogate for Cerebral Perfusion Pressure

There is a debate in the literature on which measure to use to represent CPP. The measure most commonly used as a surrogate for CPP in adults is mean arterial blood pressure (MABP) (Mitra et al., 2014). However, according to Mitra et al. (2014) there are some drawbacks in using it for preterm infants. First of all, Mitra et al. (2014) argue that spontaneous changes in MABP in preterm infants are less extreme and occur less frequently than in adults. Second, they argue that preterm infants have low systemic flow in the hours after birth, which is not accounted for by MABP-measuring techniques. A third reason, not mentioned by Mitra et al. (2014) but a medical reality, is that MABP can only be continuously measured through invasive means. For these reasons, Mitra et al. (2014) argue that MABP measurements do not accurately account for CPP and resulting CBF. They argue for the use of heart rate (HR) as a surrogate measure for CPP instead. Mitra et al. (2014) state that HR shows greater variability and frequency of fluctuation compared to MABP. They argue that CBF is more directly related to the cardiac output in preterm infants than to MABP and that therefore HR is more suitable as a surrogate measure for CPP in determining cerebral autoregulation. An additional advantage is that HR can be monitored continuously through non-invasive means.

As Mitra et al. (2014) point out, MABP is the preferred surrogate measure for CPP in adults to calculate a measure of cerebral autoregulation. Because most authors also use it for preterm babies (Alderliesten et al., 2013; Eriksen, Hahn, & Greisen, 2015; Gilmore et al., 2011; Kooi et al., 2017; Tsuji et al., 2000; Wong et al., 2008), it is difficult to assess whether using HR indeed yields a more accurate measure of cerebral autoregulation than using MABP. Mitra et al. (2014) attempted to use HR to construct a measure of CA using a moving window correlation. They did not find any significant correlations between MABP and TOI in their study, although they blame this on a technical limitation in data collection. Da Costa et al. (2015) used the same method as Mitra et al. (2014) to calculate a measure of cerebral autoregulation, TOHRx, a moving window correlation between TOI and HR. Da Costa et al. (2015) looked at the correlations of TOHRx with patient outcome in 31 infants, where patient outcome is determined by the *clinical risk index for babies II*, CRIB-II (Parry, Tucker, & Tarnow-Mordi, 2003). They compared the correlation of TOHRx with CRIB-II to the correlation of TOx with CRIB-II. TOx uses the same methodology as TOHRx but takes MABP instead of HR as a surrogate for CPP. They found that TOHRx showed a higher correlation with CRIB-II than TOx, thus favoring heart rate as a CPP surrogate.

Another study that compared HR to MABP was done by Stammwitz, von Siebenthal, Bucher, and Wolf (2016). They compared the coherence between HR and TOI to the coherence between MABP and TOI. Both coherence measures were found to be able to

indicate CA impairment. Stammwitz et al. (2016) found a high correlation between the two coherence measures and thus concluded that they were similarly representative.

Caicedo et al. (2016) suggested a framework for the assessment of cerebral hemodynamics regulation (CHR) using a coherence model which includes both MABP and HR. Caicedo et al. (2016) explicitly mention that MABP would be sufficient to account for just cerebral autoregulation but that they include HR to assess adverse outcomes that are not directly related to cerebral autoregulation per se but to other aspects of cerebral hemodynamics regulation.

This shows there are advocates for the use of HR as a surrogate measure for CPP when determining the state of cerebral autoregulation. However, there are more authors who do not mention this option and use MABP. From the studies that have been done, only two compared HR and MABP directly (da Costa et al., 2015; Stammwitz et al., 2016). Both these studies used a very different methodology to combine HR or MABP with NIRS data (correlation versus coherence) and found conflicting results. Thus, at this point we cannot conclude from the literature that there is an advantage to using HR over MABP.

3.3 Methodologies used in assessing Cerebral Autoregulation

There are two general approaches to determining a measure for cerebral autoregulation in the literature: determining the correlation between a measure for cerebral blood flow and cerebral perfusion pressure, or determining the coherence between those two (Caicedo et al., 2012; Eriksen et al., 2015; Kooi et al., 2017). These two methodologies will be referred to as the time-domain approach (correlation) and the frequency-domain approach (coherence). Within these two approaches, there are differences per study regarding the surrogate measures that are used, and regarding the values of parameters. However, other than these specifics, the methodologies generally fall either within the time-domain approach or the frequency-domain approach. There is more variety in the frequency-domain as the more complicated structure of this approach allows for more choices in the specifics of the implementation.

Both approaches try to assess the strenght of the relationship between CBF and CPP. If this relationship is strong, it means that CBF is not independent from CPP and signals that there is a lack of cerebral autoregulation; this in turn means that a rise or fall in MABP might lead to a too high or low blood pressure in the brain. Therefore both approaches to assessing cerebral autoregulation try to use techniques of determining

the strength between two signals: the measured surrogate for CPP and the measured surrogate for CBF. Both approaches and some example studies will now be discussed in more detail before discussing two studies (Caicedo et al., 2012; Eriksen et al., 2015) that directly compared the approaches.

3.3.1 The time-domain approach

In the time-domain approach to assessing cerebral autoregulation, a moving window correlation of surrogate measures for CPP and CBF is used. A high correlation indicates a strong relationship between the two signals and therefore impaired autoregulation. A moving window correlation between two signals is computed by determining the correlation of the signals over a certain interval, the window. This window is subsequently shifted by a certain amount of time until a series of correlation scores for the whole signal is obtained (or until the correlation score is up–to–date, i.e., when using real time data). This section gives an overview of specific implementations of calculating a moving window correlation.

Alderliesten et al. (2013) specifically looked into the relation between cerebral autoregulation and peri-intraventricular hemorrhages (PIVHs). They used MABP and rc_{SO_2} to assess cerebral autoregulation. The sample rates were not specified, but a 10 minute moving window correlation was used which was updated every minute. A correlation above 0.5 was considered to indicate a lack of cerebral autoregulation. Time periods where oxygen saturation was below 85% were not included in the analysis. Alderliesten et al. (2013) found that their measure showed high correlation values significantly more often before mild and moderate PIVHs and after severe PIVHs. They concluded that therefore this measure may identify patients who are at risk for PIVH.

Mitra et al. (2014) assessed cerebral autoregulation using HR and TOI measurements to calculate the TOHRx. Mitra et al. (2014) do not mention the original sample rates but used 10s average values of HR and TOI to calculate a moving window correlation using a window of 5 minutes (or 300s). These values were found by fitting a range of parameters to the clinical outcome and selecting the ones with the optimal fit. Mitra et al. (2014) found that TOHRx correlated significantly with CRIB-II and that patients with a higher score (indicating impaired autoregulation) required more ventilatory and circulatory support as well as being more likely to show cerebral damage.

Da Costa et al. (2015) calculated TOHRx (which they refer to as TOIHRx) according to the same method as Mitra et al. (2014) but subsequently used these correlation coefficients to calculate $MABP_{OPT}$. This is defined as the value of MABP where autoregulation is strongest in that patient. The idea of $MABP_{OPT}$ stems from the observation

that the degree of cerebral autoregulation varies for different strenghts of the CPP in different patients. When looking at figure 1.1 (see p. 2), this means that where the plateau of the green line lies differs per patient. $MABP_{OPT}$ is determined by dividing MABP recorded in 1 hour periods into 3mm Hg bins and averaging TOHRx within those bins. A fitting method was then used to determine the MABP value associated with the lowest TOHRx value (i.e., the best cerebral autoregulation) to obtain $MABP_{OPT}$. Da Costa et al. (2015) found a significant correlation between divergence from $MABP_{OPT}$ and adverse outcome. Calculating $MABP_{OPT}$ requires 2-4 hours of uninterrupted data however, which in clinical practice might not be possible. This makes it difficult to implement a realtime display, especially when the patient has just arrived. Unfortunately Da Costa et al. (2015) didn't mention any relationship between TOIHRx and adverse outcome.

This overview of studies shows that the methodology used by different authors is similar and that it is possible to link a correlation-based measure of cerebral autoregulation to patient outcome. Because the aim of each study and the evaluation in all papers were slightly different, it is hard to quantify the exact success of using correlation to assess cerebral autoregulation.

3.3.2 The frequency-domain approach

In the frequency-domain approach for assessing cerebral autoregulation, the coherence, i.e., the correlation in the frequency-domain, of MABP or HR and NIRS data is calculated. To calculate correlation in the frequency-domain, the Fourier transform is applied to the signals to find their frequency content and the relative contributions of the frequencies present (the spectral densities). Subsequently the correlation between the frequency spectrum of the two signals is computed. The method involved in calculating the spectral densities is called the Welch method and involves subdividing an epoch in several subwindows. An assumption in this procedure is that the signals are stationary, which in the case of assessing cerebral autoregulation is often not the case. Using a larger amount of subwindows when determining the spectra results in a more accurate estimate of the coherence when the signals contain non-stationarities. Next to the amount of subwindows used, frequency-domain approaches often vary in the frequency bands that are used. Which frequency bands are used determines which part of the frequency content is taken into account when assessing the relation between two signals. As changes in cerebral autoregulation occur relatively slowly, most authors only use low frequencies. This makes sure that a high coherence score is not found due to an unrelated physiological mechanism. Many authors extensively prepare the data by filtering before starting the coherence computations to further avoid spurious results.

Another commonly used measure in the frequency-domain approach is the gain of the transfer function. The transfer function refers to the input – output mapping between the spectral densities of the input and output signal. This function has a so–called magnitude or gain which describes the strength of the relation between the signals. As it largely requires the same steps as calculating coherence and adds extra information, a lot of authors calculate both.

As can clearly be seen, the frequency-domain approach is more difficult to implement than the time-domain approach. To assess what is most suitable for a cerebral autoregulation display, a number of studies will be reviewed that describe specific implementations.

Tsuji et al. (2000) were one of the first to try to determine a bedside measure for cerebral autoregulation. To do so they computed the coherence between MABP and HbD (cerebral intravascular oxygenation, measured using NIRS). They used samples averaged over 5 or 10s, downsampled from a raw signal with a sample rate of 2 Hz. To calculate coherence, the dataset was split into nonoverlapping epochs of 30 minutes, excluding periods where oxygen saturation fluctuated more than 5% in those 30 minutes or where the recording was interrupted. Coherence scores were computed for each 30 minute epoch of continuous data using the following frequency bands: 0 - 0.01 Hz, 0.01 - 0.05 Hz, 0.05 - 0.1 Hz for ultralow, very low, and low frequency ranges, respectively. Tsuji et al. (2000) determined that a coherence score larger than 0.5 in the ultralow frequency bandwith indicates impaired autoregulation. In 15 of the 32 patients, cranial ultranosonic abnormalities were discovered. The authors show that in these patients, the ultralow frequency bandwith coherences show impaired autoregulation, concluding that this method is usable at the bedside to measure impaired autoregulation in relation to adverse outcome.

Wong et al. (2008) assessed cerebral autoregulation by computing the coherence between TOI and MABP. MABP was sampled at 6 Hz and TOI at 1 Hz. Both of these were then downsampled to 1 Hz. Wong et al. (2008) used 20 minute epochs of recordings with stable oxygen saturation for the coherence analysis. Each epoch was divided into 5 segments of 10 minutes wtih 75% overlap. The obtained spectra after Fourier transformation of each of the segments were then averaged over the five segments and coherence was computed. Frequency bands used were 0.003 - 0.02 Hz, 0.02 - 0.05 Hz, and 0.05 0.1 Hz for ultralow, very low, and low frequency, respectively. Wong et al. (2008) found that a high coherence between MABP and TOI in the ultralow frequency range coincided with impaired cerebral autoregulation.

Riera et al. (2014) looked into the difference between assessing cerebral autoregulation using a standard coherence calculation as well as bivariate autoregressive spectral coherence or BiAR-COH. Riera et al. (2014) introduced this more complicated coherence model to more accurately account for time dependencies between the two signals. The two signals used to assess autoregulation were MABP and TOI. These were sampled at 2Hz and divided into 30 minute epochs, provided the change in oxygen saturation was less than 5% in those 30 minutes. Each epoch was then low-pass filtered and resampled at 0.2Hz. Each epoch was checked for covariance stationarity and variation. If an epoch was stationairy and had less than 3% covariance, it was analysed. BiAR-COH was then constructed by creating a bivariate autoregressive model which modeled how each sample of signals can be predicted using the past samples of a given signal and samples from related signals. To get the BiAR-COH, coherence is calculated from the coefficients of the BiAR model. The coherence (both for BiAR-COH and the regular coherence) was obtained by dividing the 30 minute epochs in 10 minute segments with 50% overlap. Both methods were averaged over an ultralow frequency band of 0.003—0.04Hz. Riera et al. (2014) state that BiAR-COH is better at predicting adverse outcome than regular coherence. The interesting thing about BiAR-COH is that it is a combination of time and frequency-domain methods. That seems promising but the drawback of BiAR-COH is that it is reliant on assumptions (stationarity and variation) that required Riera et al. (2014) to exclude up to 35% of their epochs. Not being able to calculate a measure during the excluded epochs makes BiAR-COH unsuitable for bedside monitoring of cerebral autoregulation. Unfortunately BiAR-COH was also not compared to regular time-domain methods, so a comparison of performance cannot be made.

3.3.3 Comparison between time- and frequency-domain approaches

Caicedo et al. (2012) and Eriksen et al. (2015) directly compared different methods of assessing cerebral autoregulation, giving a good indication of their performance under equal circumstances.

Eriksen et al. (2015) compared one time-domain and one frequency-domain approach to assessing cerebral autoregulation. Eriksen et al. (2015) sampled NIRS data (the oxygenation index (OI)) and MABP at 2Hz for both approaches. The data were divided into 10 minute epochs. Periods where oxygen saturation was over 5% or where the patients had undergone some kind of procedure were excluded from the calculations. The time-domain approach that Eriksen et al. (2015) used was COx (like in Gilmore et al. (2011)), a moving window correlation between MABP and OI. To calculate COx, the OI and MABP signals were re-sampled in 10s means. Pearson's R was then computed for 5min windows, sliding every minute. From the six resulting COx values, one mean COx

value for the 10min epoch was computed. This mean COx value was then weighted by the variability in the patient's MABP, assigning a higher weight to those epochs with high MABP variability. The average regression coefficient over the epoch was also calculated and weighted according to MABP variability, to assess the degree of CA impairment. For the frequency-domain approach, Eriksen et al. (2015) used 5min windows with 50% overlap. The coherence was computed for the ultra-low frequency bandwidth ($0.003 - 0.04$Hz). The obtained coherence values were averaged over the 10min epochs. Next to coherence, gain was also computed. Just as in the time-domain approach, the obtained values were weighted with MABP variability.

Eriksen et al. (2015) found that the correlation between the resulting scores of the two methods was weak. The two methods identified different patients as having impaired cerebral autoregulation. They performed an ANOVA to assess relative repeatability of the two methods by comparing the variation within subjects with the variation among subjects. They found that COx, the regression coefficient and gain all had higher (and comparable) repeatability than coherence. The most important finding was however that high coherence and high gain values can arise spuriously when cerebral oxygenation decreases while blood pressure increases due to other mechanisms than cerebral autoregulation. This does not occur with the time-domain approach. Given these spurious results, the added complexity of calculating coherence and gain, and the more stringent assumptions underlying the calculation, Eriksen et al. (2015) conclude that the time-domain approach is more robust and suitable to assess cerebral autoregulation.

Caicedo et al. (2012) also compared several methods of CA assessment. Going further than Eriksen et al. (2015), Caicedo et al. (2012) compared correlation, coherence, alternative coherence and transfer functions gain and phase. Apart from comparing the methods, they also aimed to find the optimal parameters (such as epoch length and window size) for the different methods. All methods used MABP and TOI. The signals were sampled at 6 Hz, low- and high-pass filtered and subsequently downsampled to 0.3Hz. To remove artifacts, a least-squares support vector machine was trained. All segments of the signals that were free of artifacts for 40 minutes or longer were kept for further analysis.

For correlation, the signals were segmented in consecutive overlapping epochs of length T_i, with T_i between 10 and 30 minutes. The overlap between epochs, O_j varied between 10% and 90%. The correlation scores were calculated for each epoch and the mean value of the resulting correlation scores was assigned to each patient, for each T_i and O_j. Finally, a sensitivity analysis was performed to quantify the effect of T_i and O_j on the scores.

The frequency bands used in all the frequency-domain methods were .003 to 0.02 Hz (very low frequency), 0.02 to 0.05 Hz (low frequency), and 0.05 to 0.1 Hz (high frequency). For coherence, the signals were also segmented in consecutive overlapping epochs of length T_i, with T_i between 10 and 30 minutes. Subwindows were fixed to a length of 5 minutes within each epoch. The overlap between subwindows, O_j, was varied between 10% and 90%. Coherence scores were subsequently calculated for each T_i and O_j. The mean values of the coherence scores along the epochs was assigned to the patient, for each T_i and O_j. A sensitivity analysis was performed here as well.

Apart from this regular method of computing coherence, Caicedo et al. (2012) also used another method they named modified coherence (MoCOH). This aims to address the issue that the variation in MABP within an epoch is sometimes too small to get good coherence estimates. To address this issue, in MoCOH every freqeuncy-component of the coherence is weighted by the percentage of MABP power present at that frequency component. Apart from this difference, MoCOH followed the regular coherence method.

Finally, Caicedo et al. (2012) also looked at the use of transfer function analysis to assess cerebral autoregulation. This was done using the exact same method as with coherence apart from calculating gain and phase instead of coherence.

Caicedo et al. (2012) found that for correlation, coherence and MoCOH, epochs between 14 and 22 minutes were optimal, depending on the specific patient. In the frequency-domain methods, overlapping should be larger than 60% and for correlation the overlapping should be larger than 80% for the best results. The latter result is interesting as most authors use non-overlapping windows when calculating the correlation. Correlation and gain were found to be the most robust methods to assess autoregulation. Gain was found to be a little more robust than correlation due to its ability to account for delayed autoregulation dynamics. Coherence was found to be less robust due to its stationarity requirements, which are often not met by real life data. However, when the requirements are met it is better at taking delayed dynamics into account than correlation is. Therefore Caicedo et al. (2012) recommend using a combination of coherence and gain to assess cerebral autoregulation.

3.3.4 Conclusion

Caicedo et al. (2012) and Eriksen et al. (2015) both found the time-domain approach to be more robust than the frequency-domain approach overall, although Caicedo et al. (2012) obtained very good results using transfer function gain, as opposed to Eriksen et al. (2015). It appears that the frequency-domain approach is better suited to calculate values for cerebral autoregulation for research purposes after all the data has been

collected, while the time-domain approach is more suited for calculating a measure for cerebral autoregulation in real time. Given the greater complexity of calculating coherence over correlation and the apparent lower robustness of coherence-based measures, the time-domain method will be used in the design of the cerebral autoregulation monitor. Finding the most accurate way to assess cerebral autoregulation is outside the scope of this study and will, judging by the state of the literature, still take several years. It appears that the time-domain method does not have much room for improvement over the current standard. Thus, finding a more suitable approach to autoregulation assessment will most likely depend on improvements in the robustness of the frequency-domain approach or to new measuring techniques.

Chapter 4

Situation Awareness and Clinical Reasoning

The main aim of this study was to investigate whether a cerebral autoregulation monitor increases the situation awareness of physicians at the NICU. Determining the most suitable way to assess cerebral autoregulation is only the first step in that process. To design a display that enhances situation awareness, it is necessary to be aware of how situation awareness relates to displays and how it ties in with clinical reasoning and decision making.

4.1 Using a patient monitor to achieve situation awareness

Drews and Westenskow (2006) provide a review of data displays for patient monitoring in anesthesia. They approach data displays from the goal of increasing clinicians' performance and patient safety. The authors argue that data displays play a role in this by enhancing situation awareness.

Situation awareness can be described as a general understanding of the task environment, how the current state of that environment arises from the past and the ability to make predictions about the future of the environment (Endsley, 1995). Endsley (1995) describes situation awareness as consisting of three levels which can be achieved by an operator. Level 1 (detection) is the operator receiving information and detecting changes in the environment. Level 2 (diagnosis) is the operator being aware of how the information can be integrated and how it relates to the present goal; it is this level where understanding of the information is achieved. Level 3 (prediction) is the operator being

able to predict the future states of the system (in this case, the patient) and evaluate the possible outcomes of different courses of action.

In anesthesia, systems that increase situation awareness diminish the time between the occurrence of a critical event and a suitable action to correct that event. Drews and Westenskow (2006) equate detection of a critical event to having level 1 situation awareness, diagnosis of the event to level 2 situation awareness and being able to come up with a plan for treatment to level 3 situation awareness. For level 1 and level 2, Drews and Westenskow (2006) provide display design recommendations to enhance that level of situation awareness.

To enhance level 1 situation awareness in a critical care context, Drews and Westenskow (2006) recommend that graphical displays should implement design principles such as symmetry (for example, a display that is symmetrical when the situation is normal and asymmetrical when it is not) and emergent features (such as a bar graph where all bars form one horizontal line in a normal situation) so that changes can easily be detected by seeing that a specific data point deviates from the norm. This design recommendation is based on decades of information visualization research (for a summary see Liu, Cui, Wu, and Liu (2014); North (2012)) and human factors literature. For an example of the latter, see Wickens and Hollands (2000), who give similar recommendations when discussing displays that facilitate detecting changes in the domains of aviation and nuclear powerplants. Liu et al. (2014) show how salient different types of properties of visual glyphs are. A visual glyph is a visual entity that corresponds to a data set entity. Examples of this are points, lines and regions. The properties of these glyphs, such as position, size and color determine how easy it is to distinguish between two different glyphs. In the case of these examples. the most salient property is position and the least salient property is color. This information can also help to facilitate level 1 situation awareness by choosing the properties and glyphs representing the data set in such a manner that new or deviating glyphs are easily detected.

Klein (2008) describes rapid decision-making by experts as being guided, where possible, by recognizing patterns in the situation and choosing a course of action that is appropriate given the expert's past experiences. When it isn't possible for the expert to match a pattern, they revert to using an analytical pocess. Physicians, being medical experts, make decisions in this way as well (Cnossen, 2015; Drews & Westenskow, 2006). Drews and Westenskow (2006) therefore recommend that displays make use of distinctive and easy to memorize patterns to enhance level 2 situation awareness. In addition to the use of the described patterns that can also enhance situation awareness at level 1, this refers to specific visual patterns that are formed when a specific situation arises. When the same or a similar pattern occurs, a physician can match the visual pattern to the

memory of a situation that occurred in the past and apply the experience gained in the past to the current situation. This diminishes or even removes the need to analytically determine the significance of what is displayed, reducing the time needed to make a decision. So to enhance level 2 situation awareness, a display design should facilitate the recognition of unique patterns in unique situations. However, the design should also take into account that a situation may occur which is not recognized by the user of the display. If that happens, the design should be as clear as possible to support an analytical assessment of the situation using the display.

Enhancing level 3 situation awareness can be achieved by showing trend information, so that future states of the patient can be anticipated (Drews et al., 2008). In addition, it is important to be able to compare the current state of the patient with the desired state of the patient to determine if enough progress is being made and to indicate in which direction the progress should be made (Drews & Westenskow, 2006). Therefore, to enhance level 3 situation awareness, it should be clear from the display design what the desired state of the patient is and how the current situation relates to that desired state.

Agutter et al. (2003) describe the design and evaluation of a display for monitoring cardiovascular changes during anesthesia that reflect the aforementioned design principles (see figure 4.1). The authors chose to design a metaphorical display that was intended to look like the represented physiological system and showed emergent features with changing variables. They found that using the metaphorical display as an addition to the traditional monitor was beneficial in detecting, diagnosing and treating critical events in the cardiovascular system during anesthesia in a simulated context compared to using the traditional monitor in combination with a display that showed only the numerical values also displayed in the metaphorical display. Van Amsterdam, Cnossen, Ballast, and Struys (2013) also designed a metaphorical monitor for anesthesia. They did not find any significant improvements in decision time or accuracy in a simulated setting when using the metaphorical monitor over using the standard monitor. These studies show that it is possible to design a monitor using the recommendations outlined above to create a monitor that leads to performance equal to a traditional monitor in a clinical setting and potentially even to better performance.

The recommendations put forward by Drews and Westenskow (2006) and Drews et al. (2008) are very valuable in providing design recommendations for enhancing situation awareness. It is however important to consider that these recommendations are for an anesthesia context, which is slightly different from the context of the NICU. The monitors that are used in both contexts are the same, but the task environments are different. While in anesthesia the monitors are (theoretically) monitored constantly, the

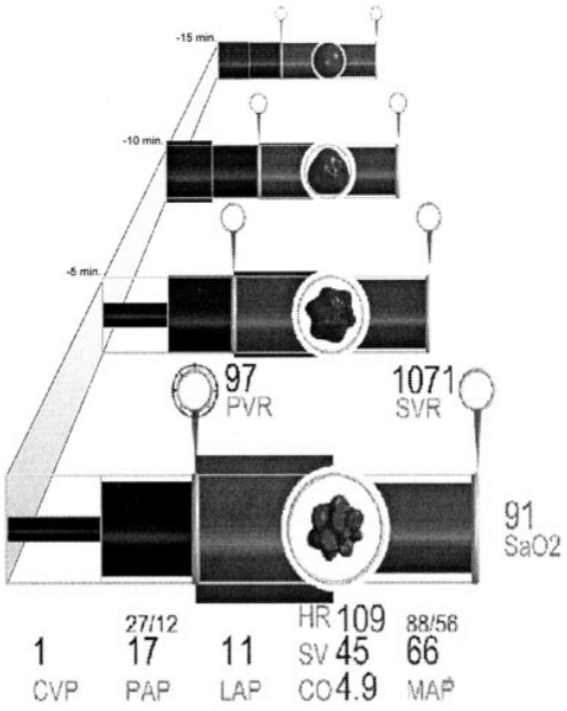

FIGURE 4.1: A metaphorical display for monitoring cardiovascular changes during anesthesia, taken from Agutter et al. (2003). The display shows the relationship between variables such as pulmonary arterial pressure, left arterial pressure, cardiac output, blood pressure and more at intervals of 5 minutes.

monitors are checked periodically in the NICU (Drews et al., 2008). This makes detecting changes more cognitively demanding as it occurs across several different patients and requires maintenance of several mental representations at once. Drews et al. (2008) found in a study that nurses often have trouble integrating the different variables on a patient monitor. He also found that trend information is very important to the nurses but hard to use in the current patient monitors. He therefore argues that in the context of the NICU it is even more important to have displays that facilitate recognition of patterns and to show easy to interpret trend information. Drews and Doig (2014) designed a display (see figure 4.2 for monitoring vital signs in an ICU based on the findings from Drews et al. (2008)) and found a significant reduction of time needed for nurses to integrate different variables with their new design. This finding underlines the importance of adding trend information to enhance level 3 situation awareness in the design of a monitor for the NICU.

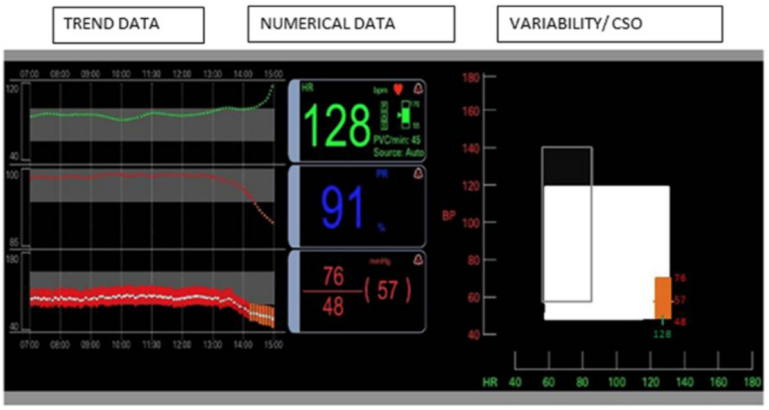

FIGURE 4.2: A display for monitoring monitoring vital signs in an ICU, taken from Drews and Doig (2014). The display shows long trend information and numerical data for vital parameters on the left and middle panels. The right side shows the current state object (CSO), an alternative visual representation of the middle panel.

4.2 Making sense of the environment

While achieving good situation awareness does not automatically lead to good decision making (Endsley, 1995), good situation awareness is clearly vital in reasoning about a patient and making decisions about that patient. Research into clinical reasoning is approached from two main angles. Much of the literature focuses on the *dual-process model* of reasoning (Croskerry, 2009; Kahneman, 2011; Norman, 2005). The other main angle focuses on *sensemaking models* of reasoning (Klein, Moon, & Hoffman, 2006a). Both approaches will be explained in detail below.

The dual-process model of reasoning stems from the so-called heuristics and biases approach to reasoning. Its origins lie in Tversky and Kahneman's seminal paper which demonstrated that there are certain cognitive biases and heuristics that people use when making decisions (Tversky & Kahneman, 1975). Even experts in their field are susceptible to these biases, which can lead to errors in decision making despite experts' many years of training. The dual-process model of reasoning explains how this can happen by dividing reasoning into two systems. System 1 is responsible for unconscious, non-analytic reasoning and system 2 is responsible for analytic reasoning. The dual-process model describes that when a fast, intuitive decision is made, this is done by system 1 through a process akin to naturalistic decision making (Klein, 2008). This type of fast, intuitive decisions occur when the situation seems familiar enough to rely on intuition or when circumstances prevent analytic reasoning. When pattern recognition doesn't lead to a satisfactory solution or the circumstances allow for more deliberation, system 2 comes into action and tries to solve the problem by a conscious, logical, analytic process where hypotheses are formed and data are gathered to prove or disprove the hypothesis (Kahneman, 2011; Norman & Eva, 2010).

Some of the cognitive heuristics and biases that commonly occur in clinical reasoning are described by Norman and Eva (2010). They describe several of the heuristics and biases that were discovered by Tversky and Kahneman (1975) and how they can affect clinical reasoning. An example is the *availability heuristic*. This affects the probability estimate a physcian makes about a diagnosis being correct. It results in a higher likelihood being attributed to the first diagnosis that comes to mind than is statistically valid. Another example is *cue primacy*, where a physician assigns more weights to the first few cues of a series of cues that come in. This can lead to *anchoring* in which a hypthosis is mostly built on the first few cues that come in, ignoring or attributing too little weight to newer information. These are only two of many more examples of similar cognitive biases.

While the dual-process model can explain a lot of the traits of expert reasoning, there is also a group of researchers that focuses on sensemaking theories. This group has grown out of the naturalistic decision making field. Where the dual-process model literature often focuses on the shortcomings of intuitive reasoning, the sensemaking field focuses on the benefits (Kahneman & Klein, 2009). This alternative approach entails a greater focus on the process of reasoning than on flaws in reasoning. Sensemaking theories can therefore function as a guide to reasoning by describing the reasoning process in its entirety and describing how a physician interacts with a data set to reason about it. Sensemaking theories function as a guide to reasoning not only when there are flaws in the reasoning process but in the majority of cases, where the physician's approach to reasoning leads to the desired result. Sensemaking theories and the dual-process model can complement each other; sensemaking theories can show how reasoning normally occurs and the dual-process model can show where the pitfalls in reasoning are.

At the base of many sensemaking theories is the Data/Frame Theory proposed by Klein et al. (2006a). Klein et al. (2006a) propose a framework for sensemaking that can be seen as a continuation of naturalistic decision making work. Klein et al. (2006a) describe sensemaking as complementary to situation awareness and see it as "a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively." (Klein et al., 2006a, p. 71).

In the proposed framework the key concept is a *frame*. This can be described as one person's perspective or interpretation of a collection of data. A frame can also be seen as a hypothesis. Frames determine which data are relevant but data can also influence and change the frames. Figure 4.3 shows how different sensemaking activities occur using this basic building block. The elaboration cycle, in the left of figure 4.3, describes the elaborating of a frame when new data arise or when the frame is questioned and new data are required to address the issue. In the elaboration cycle, the frame is preserved

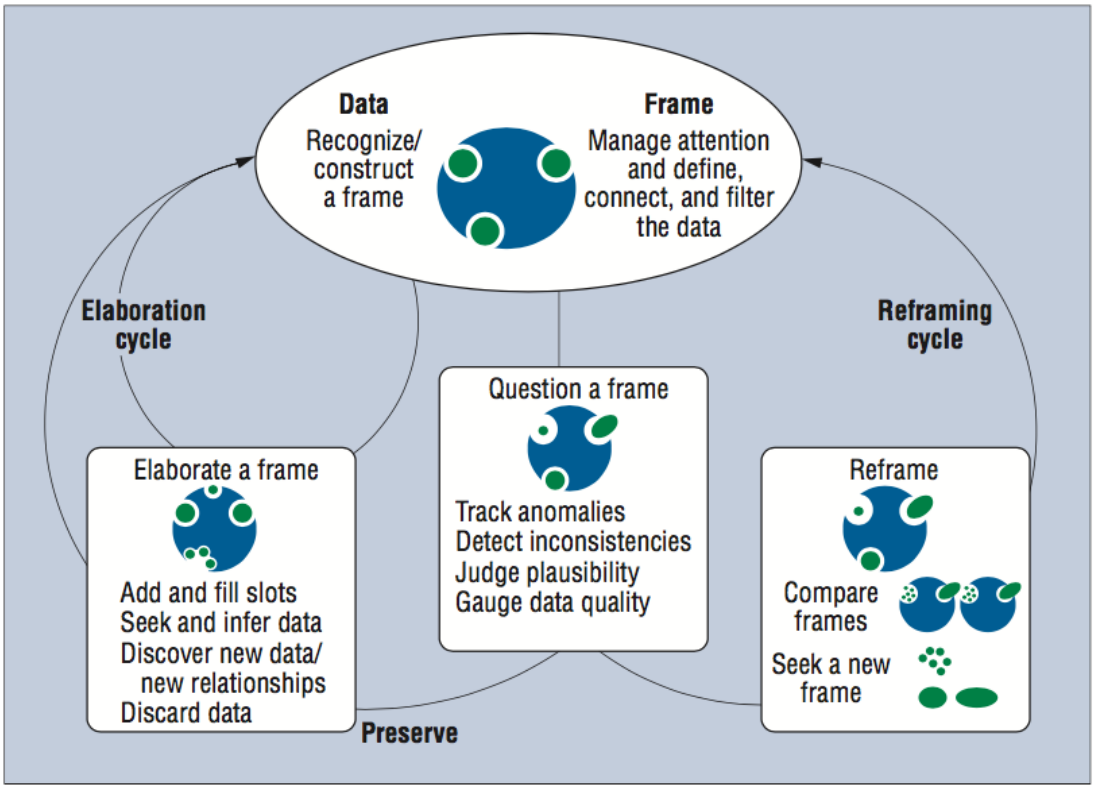

FIGURE 4.3: The Data/Frame Theory of sensemaking, taken from Klein et al. (2006b)

but adjusted. Alternatively, the reframing cycle, as displayed in the right of figure 4.3, can occur. In the reframing cycle, a frame is discarded and replaced by another frame after the original frame has been questioned. Whether a frame is elaborated or reframed depends on the discrepancy between the explanation provided by the frame and the data. When this discrepancy is small, the frame can be elaborated. When the discrepancy becomes larger, the explanation provided by the frame becomes untenable and the frame needs to be reframed. Whether a frame will be elaborated or reframed depends on the person engaged in sensemaking. The two cycles occur within closed loops of mental model formation (looking backwards to explain data) and mental simulation (anticipating the future). Together these two loops form the continuous process of sensemaking.

To illustrate how the sensemaking framework can help in interface design in a medical context, Aselmaa et al. (2017) used a contextualized sensemaking model in the interaction design for a tool that helps radiologists. Several concepts are added to the sensemaking terminology that was already discussed by Aselmaa et al. (2017). A *goal* is the desired outcome of a sensemaking process and a *gap* is a discrepancy between data and frame or two different frames. Gaps are seen as the trigger for sensemaking activities. The task that Aselmaa et al. (2017) studied was tumor contouring. This is the process of outlining a tumor on medical images (contouring needs to be done on multiple

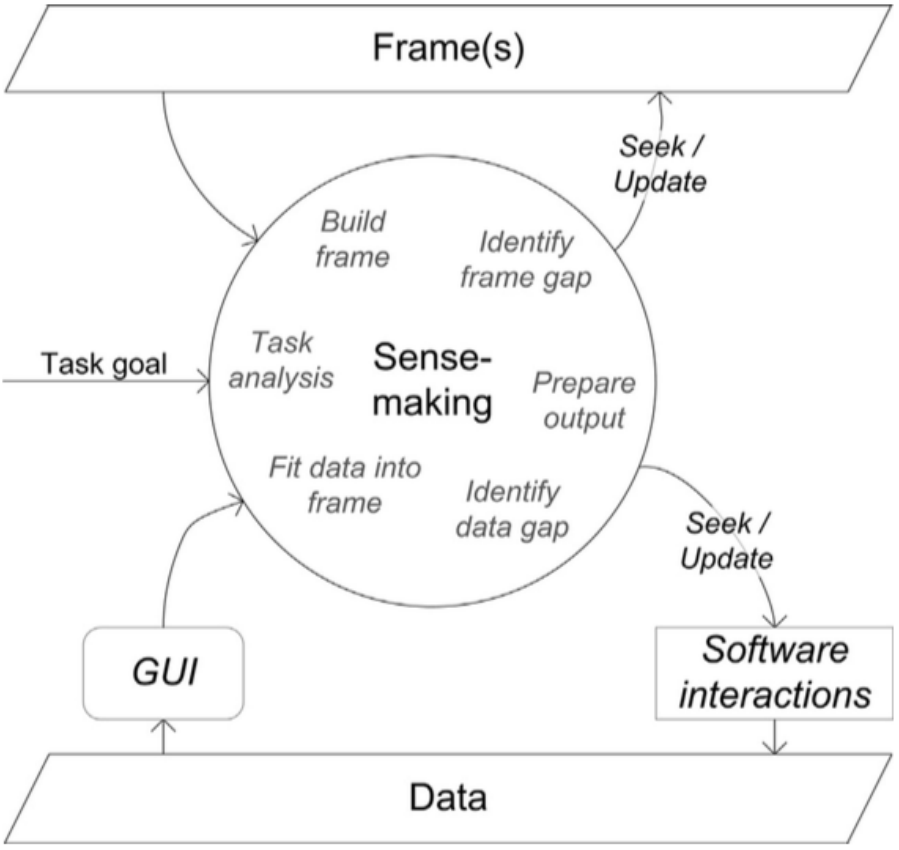

FIGURE 4.4: A sensemaking model in the context of software use, taken from Aselmaa et al. (2017).

images as a tumor is 3D and the images are 2D, requiring the tumor to be contoured on a set of images to get its volume) as a preparation for radiotherapy. Aselmaa et al. (2017) made a contextualized sensemaking model of this task based on a general sensemaking model described by Zhang and Soergel (2014). There are seven sensemaking activities in this general model: task analysis, identification of gaps, information seeking, building frames, fitting data into frames, updating frames and preparing task output, as can be seen in figure 4.4.

Identification of gaps (in data or frame) is the driving force behind sensemaking in this model. When a gap is identified, information seeking activities occur in order to find data or a frame to fit the gap. Building a frame occurs in symbiosis with fitting data into a frame and together this is referred to as *gap bridging* in the model. Apart from identifying gaps, seeking information and bridging gaps, updating the frame and preparing task output are continuously happening. In the context of using software, information seeking and preparing task output is facilitated through the software. All the data are presented using the graphical user interface (GUI), which therefore plays an integral part in the sensemaking process. The general model can be seen in figure

4.4. This model was further developed into a contextualized model, as can be seen in figure 4.5. This contextualized model was then used to inform interaction design choices for a software tool for tumor contouring, as well as explaining behaviour of participants while testing the software.

The contextualized sensemaking model is based on three phases of the tumor contouring task that Aselmaa et al. (2017) observed in physicians performing this task. These phases are: familiarization, action, and evaluation. Figure 4.5 shows how sensemaking occurs in these three phases from left to right. In the familiarization phase, the physician becomes acquainted with the task and the data and identifies where gaps exist between data and frames. The physician starts from a *general tumor frame* (as seen at the top of figure 4.5), which is formed by the physician's experience, knowledge and expectations. In the familiarization phase, this general tumor frame is updated through iterative sensemaking cycles towards an initial frame of the tumor to be contoured. The second phase is the action phase, where the actual contouring is performed as well as navigating the contouring software. During this phase, sensemaking activities occur that lead to updating the initial frame and gradually changing it to a specific frame for the specific tumor. Finally, in the evaluation phase, the physician compares the outcome of the contouring process to their clinical experience and other medical information about the patient. If gaps between data and frame are identified during this phase, the physician returns to the action phase.

The general model of sensemaking using software shows that a GUI presents data to a person engaged in sensemaking activities to allow them to fit that data into a frame. Thus, how these data were presented determines how the data are perceived and fitted into a frame. The contextualized sensemaking model is an example of how this can be applied to a medical task. In the tumor contouring task, the GUI not only functions to show the data as they are presented at the start of the task but also to allow the physicians to observe and reflect on the actions they perform during the task.

4.3 Conclusion

The design of a cerebral autoregulation display should make use of the recommendations that have been described in this chapter to enchance situation awareness. For enhancing situation awareness at level 1, the recommendation is to make use of design principles such as symmetry and emergent features, as well as knowledge about visual glyphs and their properties to make new or deviating data stand out from the rest (Drews & Westenskow, 2006; Liu et al., 2014). To enhance situation awareness at level 2, the design

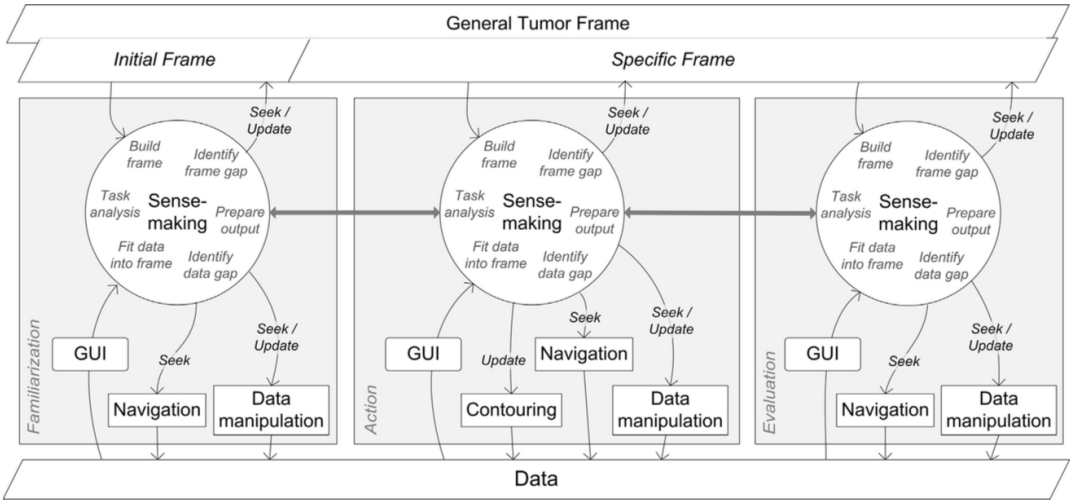

FIGURE 4.5: A contextualized sensemaking model for a tumor contouring task, taken from Aselmaa et al. (2017).

should facilitate memorizing and recognising visual patterns that occur based on patterns in the data (Drews & Westenskow, 2006). To enhance situation awareness at level 3, trend information should be provided so that future states of the patient can be anticipated (Drews et al., 2008). In addition to these recommendations that directly impact the visual design, it is important to attempt to avoid cognitive biases such as the availability heuristic or cue primacy that can occur in clinical reasoning. The sensemaking model of clinical reasoning provides a framework of how physicians could interact with the designed monitor, which can be used when evaluating a preliminary design. The sensemaking and dual-process models of reasoning also provide frameworks that can be used when analyzing how physicians use the designed monitor while reasoning.

Chapter 5

Design and Implementation of the Cerebral Autoregulation Monitor

This chapter will describe the design and technical implementation of the cerebral autoregulation monitor. The monitor was designed and implemented in such a way that it would be possible to use it in practice at the NICU. The implementation described here was intended as a working prototype. The process of designing the monitor was based on the usability engineering lifecycle as described by Nielsen (1992). The usability engineering lifecycle focuses on iterative design and being in constant contact with the users and the environment for which a monitor is being designed. This chapter will first describe the design process, followed by the technical implementation of the cerebral autoregulation calculations, and finally, the design of the monitor.

5.1 Design process

The design process can roughly be described as follows: the predesign phase, which had the goal of getting to know the users and their tasks, started off with research into cerebral autoregulation measures, information visualization, situation awareness and clinical reasoning. In addition to this, two days were spent observing physicians in the NICU and one day observing the nursing staff. Based on the literature study and discussing cerebral autoregulation with physicians, a method of assessing cerebral autoregulation was determined. This concluded the predesign stage. Subsequently, the design phase started, which had the goal of achieving a finalized interface that serves the users' need. In order to achieve a finalized interface, the usability-engineering lifecycle requires the creation of a prototype that is empirically evaluated to determine if it meets the users' needs (Nielsen, 1992). The end-goal of this study was a final,

evaluated prototype of the cerebral autoregulation monitor. Achieving a completely finalized interface according to the usability-engineering lifecycle would require making changes to the evaluated prototype according to the findings of the evaluation study and subsequently evaluating how well the interface serves its purpose over time after having been implemented in its environment. This was beyond the scope of this study but may be a topic for future work.

The next step in the design process was to create mock-ups of the cerebral autoregulation monitor. These mock-ups were evaluated informally by both intended users and by students of and experts on human-machine interaction. This group consisted of students of the human-machine communication program at the university of Groningen, an assistent professor in cognitive engineering, a data-scientist at the UMCG and two neonatologists. The evaluation consisted of explaining the intended use of the monitor, asking whether it was clear what the monitor displayed and how it displayed it, and what the evaluator's opinions on specific design decisions were. After collecting impressions and preferences, a final mock-up was made that was again evaluated by the same group. Finally, a working prototype was implemented. The prototype was evaluated once more before reaching the stage of final design to be used in the evaluation study. An overview of the design and the considerations that went into it will follow in the next sections.

5.2 Calculation of cerebral autoregulation

After reviewing the options (see chapter 3), a moving window correlation was chosen as the measure of cerebral autoregulation in the monitor. As described in chapter 3, correlation methods are currently more robust than coherence-based methods, they are less likely to show spurious results, and they rely on less assumptions which are prone to be violated in real-time monitoring settings. The signals used in the correlation were MABP and rc_{SO_2}. The choice for MABP was made as it is the most widely used surrogate for CPP in cerebral autoregulation literature and therefore already consistent with the physicians' concept of cerebral autoregulation. Apart from that, there is no clear indication that using HR instead of MABP would lead to a more valid measure. There was no choice in using rc_{SO_2} as this is the form of NIRS data that the available machines at the NICU support. The raw signals for MABP and NIRS were obtained using the piCare project (see chapter 2).

Specific parameter values were chosen based on Caicedo et al. (2012)'s comparison of moving-window correlation implementations. It was decided to downsample the rc_{SO_2} and MABP signals to 0.1Hz. The absolute value of Pearson's R was calculated for windows of 300s with a moving constant of 60s, which means that one new autoregulation

value was generated per minute, this corresponds to an overlapping percentage of 80% between windows. Each score for cerebral autoregulation was therefore calculated as:

$$CA_{score} = |R(MABP, rc_{SO_2})| = \frac{\epsilon[(MABP - \mu_{MABP})(rc_{SO_2} - \mu_{rc_{SO_2}})]}{\sigma_{MABP} * \sigma_{rc_{SO_2}}}$$

where $|R(MABP, rc_{SO_2})|$ is the absolute correlation coefficient for the current window, $\epsilon[]$ respresents the mathematical expectation operator, μ_{MABP}, $\mu_{rc_{SO_2}}$, σ_{MABP}, $\sigma_{rc_{SO_2}}$ represent the mean value and standard deviation of the part of the MABP and rc_{SO_2} signals that fall within the current window.

The calculation was implemented as a python script that is called upon every minute to retrieve up-to-date cerebral autoregulation data and send it to the monitor.

5.3 The final design

Figure 5.1 shows the final prototype of the cerebral autoregulation monitor. The interface was implemented using HTML, CSS and JavaScript. The graphs were made using D3 (Bostock, Ogievetsky, & Heer, 2011), an advanced graphing library with tools to make dynamic, real-time updating graphs. The reasons for implementing NeoCAM as a web interface were twofold. First, it allows easy integration with piCare as that is also a web-based system. Second, the implementation of NeoCAM as a web interface makes it convenient to use in different capacities. As a web interface, NeoCAM can be used as a bedside monitor, a monitor from a distance and even as a mobile monitor if desired.

The design of the interface was based on a set of requirements that are based on the recommendations that follow from chapter 4. The monitor should:

1. *Show the state of cerebral autoregulation every minute*: based on the calculation of cerebral autoregulation the monitor will have to show a value for cerebral autoregulation every minute.

2. *Facilitate detection of changes in the state*: to enhance situation awareness at level 1, the designed monitor should make it easy to detect new or deviating data.

3. *Aid the recognition of patterns*: to enhance situation awareness at level 2, the designed monitor should present the data in such a way that visual patterns can be memorized and recognized at a future time if a similar pattern occurs.

4. *Show trend information*: to enhance situation awareness at level 3, the designed monitor should show trend information over a longer period of time than the standard vital parameter monitors in the NICU.

First, a general description of the monitor will be given and subsequently it will be explained how the design choices were informed by the requirements.

The cerebral autoregulation monitor shows 30 minutes of data (the bars), the most recent value (the number on the right), and a trend arrow. There is an information popover explaining the calculation of the cerebral autoregulation values that can be opened by pressing the button above the number showing the most recent value. The monitor is a bar-chart where one bar is added per minute. When a bar is added, it appears on the right side of the chart (corresponding to the current time), at the same time the leftmost bar (the oldest value) is removed. This means that the number of bars in the chart is always constant, unless measuring has started less than 30 minutes ago. Every minute Pearson's R is calculated between the values of the last five minutes of the MABP and rc_{SO_2} signals. Each bar therefore represents the degree to which cerebral autoregulation was present in the last five minutes. This explanation can also be found in the information popover and the heading of the monitor serves as a short reminder. The height and color of each bar are determined by the cerebral autoregulation value: the legend on the left of the graph shows how the colors relate to the cerebral autoregulation values; low cerebral autoregulation values correspond to green, medium values to yellow and high values to red. Between these three main colors are gradients mapping each value to a specific hue. This allows for immediate recognition of a specific value by recognizing the hue of the bar in the graph.

Next the motivation behind these design choices will be discussed based on the requirements stated above.

Show the state of cerebral autoregulation every minute

The first requirement stems from the implementation of the cerebral autoregulation calculation. The chosen method of calculation yields one cerebral autoregulation value per minute. These values are displayed in the bar chart where each bar represents one cerebral autoregulation value. To allow the physicians to monitor the state of cerebral autoregulation accurately, every value is displayed in two places: in the bars of the chart and in the number on the right of the chart. The numerical correlation value is added for higher accuracy as it is difficult to read the exact value in two decimal points accuracy from the bar chart.

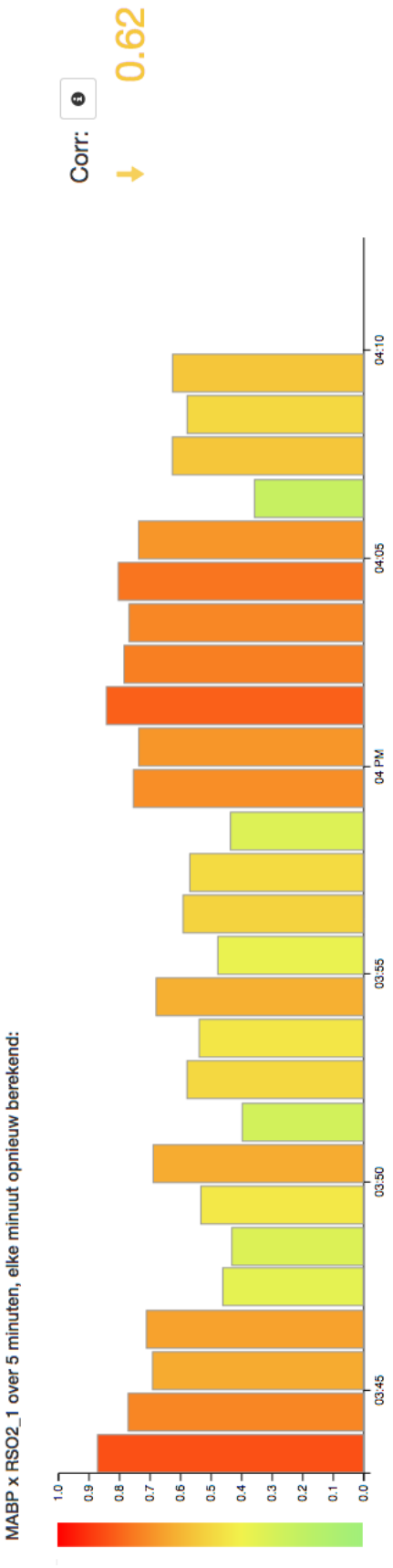

FIGURE 5.1: The cerebral autoregulation monitor.

Facilitate detection of changes in the state

The choice was made to display the timeline as a bar-chart and not as a line-chart. This choice was made because it is possible to add more visual properties to bars that help distinguishing between two data-points. According to information visualization theory, the three most effective properties to distinguish between two visual glyphs are in that order: position, size and color (North, 2012). Position and color could be applied equally easy to points (which would be the glyphs representing a data point in the case of a line-chart) as to bars but it is easier to apply the size property when using bars as glyphs. When making use of size as a property, the use of color becomes more effective as there is a larger surface for presenting the color. There is redundancy in the use of size and color in the design of the cerebral autoregulation monitor. Both these properties are mapped to the cerebral autoregulation value. As the value increases, the height (and therefore the size) of a bar increases and the color shifts along a spectrum from green to yellow to red. These colors were chosen as they represent the familiar use of green for safety, red for danger and yellow for in between (Wickens & Hollands, 2000). The color scheme was tested for color-blindness and it is still possible to distinguish between the different colors in grayscale (with low cerebral autoregulation values corresponding to light bars and high values corresponding to dark bars).

The property of position is used to indicate the point in time that the cerebral autoregulation value represents. New values always appear on the right of the chart, making the detection of new information easy and predictable. The use of size and color makes seeing what value a bar represents and how much it deviates from other bars around it easy. To add redundancy, the number on the right of the chart showing the latest cerebral autoregulation value is presented in the same color as the bar it represents.

Aid the recognition of patterns

To aid the recognition of patterns in the data, the choice was made to consistently show 30 minutes of data and to always have the same number of bars presented in the monitor. By keeping the interval and the number of bars constant the chart gives rise to an emergent feature in the form of the shape that the collection of bars takes. This shape, or visual pattern, always represents 30 minutes of changes in the state of cerebral autoregulation. This allows physicians to familiarize themselves with the way the state of cerebral autoregulation can progress over time by remembering the shapes. This can aid them in naturalistic decision making processes by providing easy to recognize situations that can be linked to courses of action that have proven effective in the past.

Show trend information

To allow physicians to see how cerebral autoregulation has developed over time and help them anticipate the future state of cerebral autoregulation, the monitor shows trend information. The choice was made to show trend information in two ways, as a timeline of cerebral autoregulation values and as an arrow summarizing the trend of cerebral autoregulation values over the last five minutes.

The timeline allows physician to track the progression of the state of cerebral autoregulation in the patient. The interval the timeline shows is always 30 minutes. This value was based on conversations with physicians. The 30 minute interval is long enough to capture more than one state of cerebral autoregulation while preserving a lot of visual detail in the bars shown in the graph.

Using the trend arrow, a physician can get an immediate impression of the recent trend without having to inspect the timeline. The trend arrow points upwards to indicate increasing cerebral autoregulation values (and therefore decreasing cerebral autoregulation) or downwards to indicate decreasing cerebral autoregulation values. When the cerebral autoregulation values are constant, the arrow disappears. The direction of the arrow is based on the slope of the linear regression equation calculated over the last five values.

Chapter 6

Evaluation Study

6.1 Introduction

The evaluation study aimed to achieve two goals. The first goal was to find the answer to the question: does the addition of a cerebral autoregulation monitor to the standard monitors at the NICU enhance situation awareness for physicians? As part of answering this question, the study sought to establish how often physicians reason about cerebral autoregulation when it can be monitored compared to when it can't. Additionally, the study aimed to find how the presence of the cerebral autoregulation monitor influences the physicians' reasoning process and the diagnosis and treatment plans that are the outcome of the reasoning process, as well as how the physicians experienced working with a cerebral autoregulation monitor, what their attitude was towards implementing a cerebral autoregulation monitor in the NICU, and how they currently reason about cerebral autoregulation when assessing a patient.

The second goal of the study was to evaluate the design of the cerebral autoregulation monitor in terms of usability. This entailed evaluating if the visual design of the monitor was clear and pleasant for the physicians, whether the physicans trusted the information displayed by the monitor, finding specific usability issues the physicians encountered while working with the monitor and evaluating the physicians' willingness to adopt the monitor in their workflow.

In the study, eight neonatologists were provided with a patient simulation task in which they were asked to make a diagnosis and treatment plan for eight cases while thinking out loud. The participants determined themselves how much time they spent on each case but a maximum case length was set to make sure all cases could be completed during the experimental session. The task consisted of two blocks of four cases each. In the

first block, the cases were displayed on an interface that resembled the standard vital parameter monitors in the NICU. In the second block, the cerebral autoregulation monitor was added to this interface. The block without the cerebral autoregulation monitor was always presented first. Presenting the physicians with a sequence of cases was intended to simulate the physician's rounds, where they visit all patients in a ward, assess their situation and discuss the actions to take during the shift. The cases were divided in two categories: stable cases, where the patients did not require intervention, and unstable cases, where the patients had impaired cerebral autoregulation and required some form of intervention. Each case consisted of a simulation of the vital parameters of a patient during at most 30 (in the stable cases) or 50 (in the unstable cases) minutes, a description of other clinical information, and the patient history. The difference in maximum simulation length between the stable and unstable cases was to prevent participants spending too much time on the stable cases due to the limited length of the experimental session. In the experiment none of the participants exceeded the maximum simulation length set for stable cases, both in the simulations with stable and the simulations with unstable cases. The simulations were played at 2x normal speed by default to allow improvement or deterioration in the patient state to happen within a realistic time-frame while being able to fit more cases in one experimental session. This was necessary as the length of the experimental sessions was constrained due to the full schedule of the physicians. After each block of trials, the participants filled out a usability questionnaire about the monitor interface. At the end of the experimental session a semi-structured interview with the participant was conducted. The recordings of the experimental sessions were transcribed and analyzed qualitatively for reasoning patterns, frequency of reasoning about cerebral autoregulation, and the outcome of the reasoning process per monitor condition. The semi-structured interviews were also analyzed qualitatively to gain insight into the participants' attitude towards using cerebral autoregulation when reasoning about patients, their opinion on the addition of a CA monitor, and their opinion on the design of the monitor. The time needed for a physician to make a diagnosis and treatment plan was measured and compared between the two different monitor conditions. The usability questionnaire scores were also compared between the two monitor conditions.

It was hypothesized that:

- The addition of a cerebral autoregulation monitor to the standard monitors would enhance situation awareness for the physicians.
- The physicians would reason more frequently about cerebral autoregulation when the CA monitor was present than when it was absent.

- The presence of the CA monitor would increase the time needed to complete a case.
- The presence of the CA monitor would directly inform diagnoses and treatment plans, leading to different diagnoses and treatment plans when the CA monitor is present compared to when it is absent.

This chapter will describe the methodology of the evaluation study in more detail.

6.2 Participants

Eight neonatologists (out of fourteen employed at the time of the experiment) of the neonatology department of the University Medical Center of Groningen participated in the experiment. Six of them had over ten years experience as attending neonatologist. Two participants had between 0 and 2 years of experience as attending neonatologist. One of the participants had done previous research into cerebral autoregulation and could be described as an expert. The rest of the participants were familiar with the concept of cerebral autoregulation having learned it as part of their training to become a neonatologist.

6.3 Materials

6.3.1 Experimental Interface

Simulated patients were used in the evaluation study. Each case consisted of a simulation of the changing vital parameters and state of cerebral autoregulation of a patient during 30 (in the stable cases) or 50 (in the unstable cases) minutes. The vital parameters were shown at a frequency of 1 Hz and the cerebral autoregulation at one value per minute. In addition each case consisted of a description of additional clinical information and the patient history. As there was no access to a simulated bedside environment with a patient simulator in which to place the monitor, it was decided to provide all relevant information in the experimental interface which can be seen in figure 6.1. The information about the patient is divided into two separate sections. The first is the clinical information section, which features the patient's name, gender and age and all other current information relevant to the specific case. This includes variables not shown in the monitor (such as temperature or respiratory support system settings), results of lab tests and information about the physical appearance of the patient.

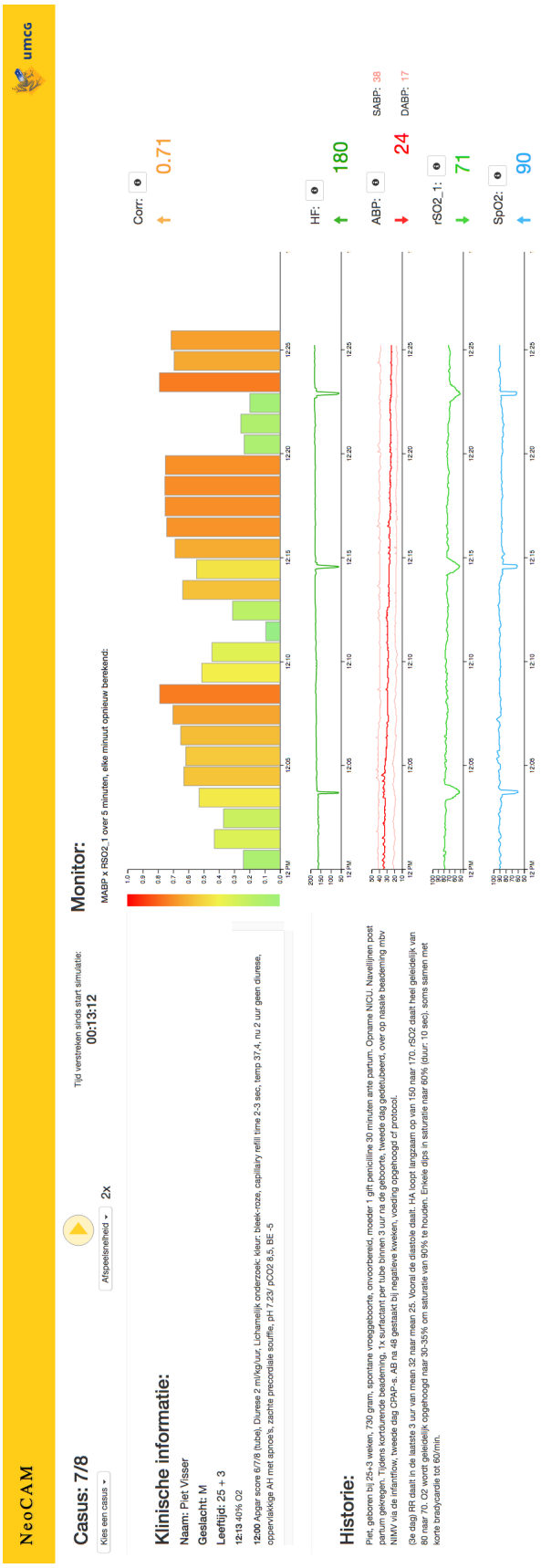

FIGURE 6.1: The interface as used in the evaluation experiment.

FIGURE 6.2: The experimental controls, clinical information and patient history.

The second section is the patient history. This is an overview of the medical history of the patient since birth (sometimes including the history of the patient's mother before birth, if relevant). This section ends with stating how long it has been since birth. Both these sections are written in the style of the handover documents the physicians are used to seeing and writing every day. Designing an alternative interface to display this information lay outside of the scope of this study.

At the top of figure 6.2 are the buttons which control the experimental session. The top left is a statistic showing the current case number and the total number of cases. Next to that is a menu to select the cases. Next to that is a pause and play button which is used to start and stop the simulation. This was included for the eventuality of interruptions during the experiment, as physicians could potentially be called away at any moment. Below the play-button is a menu from which to select the play speed of the case. This was set to 2x normal speed by default to allow for improvement or deterioration in the patient state to happen within a realistic time-frame while being able to include more trials in the experimental session. The playback speed is adjustable in case a physician expressed the desire to monitor the case at 1x speed to see greater detail. Finally, to the right of the controls is a timer showing how much time has elapsed in simulation time, i.e., this runs slower if the playback-speed is decreased. The buttons controlling the experiment were only used by the experimenter.

The monitor contains the cerebral autoregulation display (depending on condition) and a graph for heart rate, mean arterial blood pressure, cerebral NIRS, and oxygen saturation (see figure 6.3). These graphs were intended to replicate the basic functionality of the normal Philips MP70 monitors at the NICU. Each of these graphs shows the trendline of the past half hour. This is a divergence from the standard Philips monitor that was chosen to make the displays consistent with the cerebral autoregulation display and to allow the physicians to compare MABP and NIRS trends for themselves. The color used for each variable corresponds to the default colors used for that variable on the Philips monitors at the NICU.

Each of the traditional vital parameter monitor graphs consists of a line-chart that shows the development of the variable over time. The time domain on the horizontal axis is always 30 minutes with the variable plotted at a frequency of 1 Hz. Each graph's vertical axis range boundaries are set to values relevant to the specific variable. These values can be found in table 6.1. One of the graphs contains multiple lines. This is the arterial blood pressure graph. This not only shows mean arterial blood pressure but also lines for systolic and diastolic blood pressure, see figure 6.4.

Next to the time graphs, each display in the monitor contains a number showing the current value of that variable (or variables in case of the blood pressure display). These

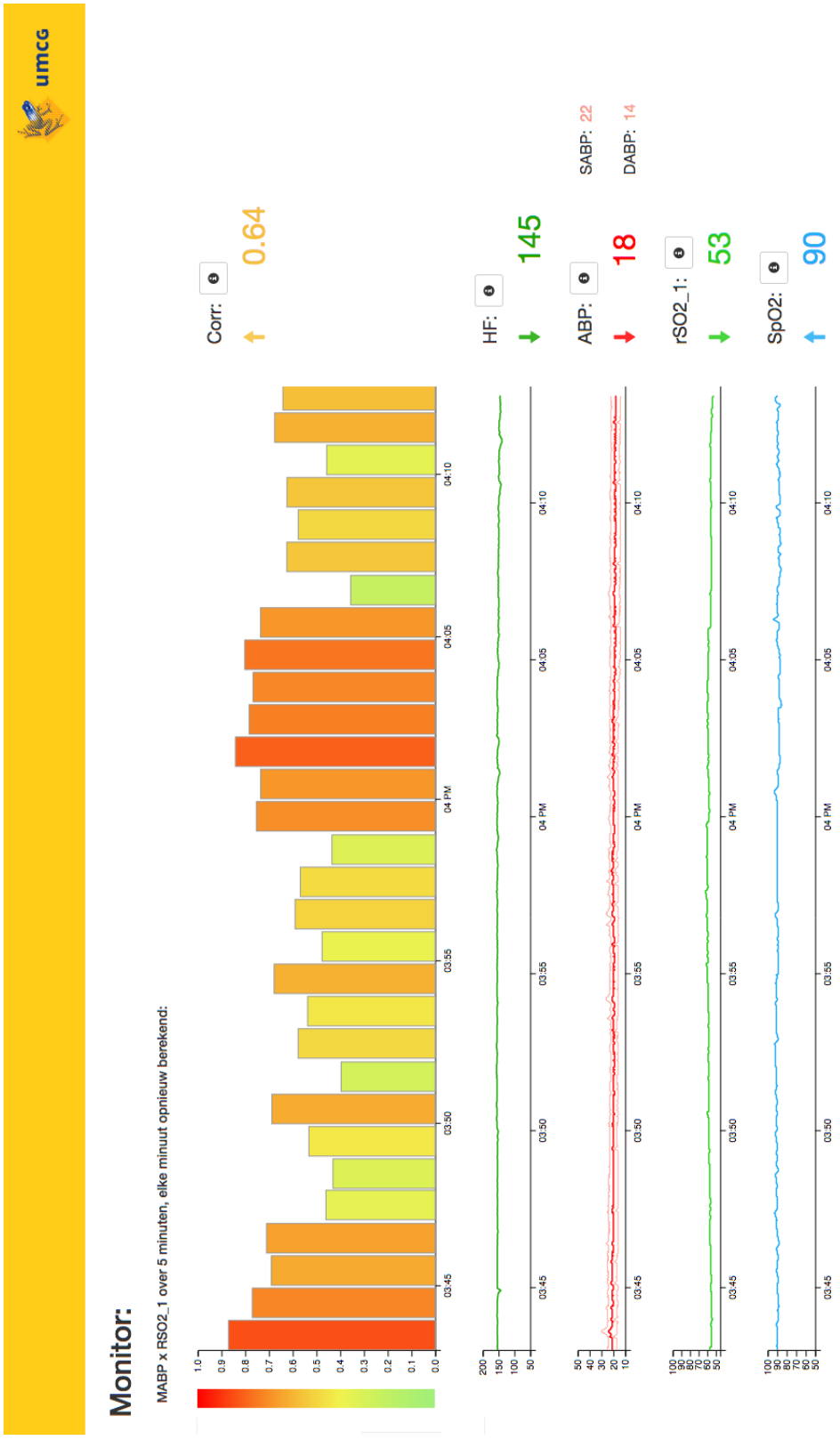

FIGURE 6.3: The cerebral autoregulation display and graphs for the traditional monitor variables.

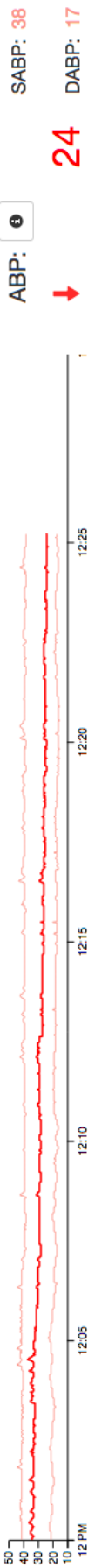

FIGURE 6.4: The arterial blood pressure graph. The dark-red line in the middle of the graph shows mean arterial blood pressure. The other two lines show systolic and diastolic blood pressure respectively.

TABLE 6.1: The min and max values of the y-axis per variable in the monitor

Variable	y_{min}	y_{max}
Heart Rate	50	200
Arterial Blood Pressure	10	50
rc_{SO_2}	40	100
Oxygen Saturation	50	100

numbers blink to show that they are updating even when the value stays the same. Next to each number is an arrow indicating the trend of the variable over the past five minutes. This trend is obtained by calculating the slope of the linear regression function of the last five minutes of data per variable. If the slope is positive, the trend arrow points upwards, if it is negative it points downwards. When the slope is zero the arrow is not shown. If there are less than five minutes of data, the trend arrows are not shown. Finally there is an information button next to the heading of each variable, which when clicked opens a popover window with a short explanation of the variable(s) and the calculation of the trend arrow.

During the experiment the interface was presented on a 24" Iiyama E2483HS monitor powered by a 13" 2015 Apple Macbook Pro (base model). The participants were also provided with a mouse to be able to toggle the info dialogs in the display. All sessions were recorded using the Macbook's internal microphone and "Audacity" recording software.

6.3.2 Simulation cases

The evaluation study consisted of eight trials based on a simulated patient case. Each case represented a different simulated patient. These cases fell into two categories. Half of the cases were 'stable' cases, which were designed to be used to familiarize the participants with the displays. These cases were designed to serve as control cases by showing a patient in a relatively normal and healthy state, not requiring any intervention from the participants during the simulation. The remaining four cases were 'unstable' cases. These cases were designed to require intervention by the participants during the simulation. Each of the hard cases showed patients with impaired cerebral autoregulation along with varying other symptoms. As there is no protocol in place at the UMCG for the treatment of cerebral autoregulation issues, these cases were not designed with a desired plan of action in mind. All unstable cases were designed by one neonatologist who did not participate in the experiment. All stable cases were designed by the experimenter based on guidelines about stable patients provided by the same neonatologist. All materials were presented in Dutch. The cases were counter-balanced across conditions so

that each case was presented equally often with the cerebral autoregulation monitor as without the cerebral autoregulation monitor.

Each case consisted of a 'real-time' simulation of vital parameter (heart frequency, blood pressure, oxygen saturation and cerebral oxygenation as measured by NIRS) and cerebral autoregulation values that showed the patient's progression over a period of 30 (stable cases) or 50 (unstable cases) minutes. As the simulations were played at 2x playback speed this meant the cases covered at most 15 or 25 minutes of actual time. The progression of the vital parameter and cerebral autoregulation values in the simulations was defined roughly by the neonatologist designing the cases. This meant that an indication of a mean and of the variance were given per value per specific time interval in minutes. The specific values per second (in case of the vital parameters) were then derived from real patient data to induce realistic variability into the simulations. The cerebral autoregulation values were manufactured to fit the design of the case and where necessary to fit with the trends shown in blood pressure and NIRS values, to accurately simulate a strong or weak correlation between the two signals.

As stated above, each case contained a collection of clinical information and an overview of the patient's history. The clinical information consisted of the name, gender and age of the patient as well as all information that a physician would normally see or have access to at the bedside of the patient. This includes settings of respiratory support systems, physical appearance of the patient, variables not shown on the monitor such as temperature and urination data, and lab results. This information was presented as a block of text similar to what the physicians are used to seeing in their handover sheets. The history of the patient was presented in a separate block of text. This history consisted of an overview of how the clinical information changed over time, any procedures the patient underwent and information about the medical state of the mother and about the birth process, if relevant. The specifics of all cases used in the study can be found in appendix A. The considerations that went into designing the cases will be discussed next.

Stable cases

The stable cases served as controls and to acquaint the participants with the experimental interface and cerebral autoregulation monitor. These cases consist of patients who are admitted to the NICU because they were born prematurely but who are otherwise stable. This means that the patients' vital parameters and cerebral autoregulation values are all within normal limits during the simulation. These patients represent the patients who are normally on the ward who do not require acute medical attention and

as such add realism to the concept that a block of trials during the evaluation study simulates rounds in the ward.

Unstable cases

The unstable cases were designed to require intervention by the physician. These cases consisted of patients who show a range of possible symptoms and vital parameter patterns that occur commonly in sick patients in the NICU and that require acute medical attention. Each patient suffered from impaired cerebral autoregulation. The rest of the symptoms were chosen to be consistent with, but not exclusive to patients with cerebral autoregulation dysfunction. The cases were designed to allow treatment according to normal protocol in the condition without the cerebral autoregulation monitor present and to see whether the presence of a cerebral autoregulation monitor led to a deviation from the normal course of action when determining a treatment plan. Additionally the design of the cases was intended to stimulate the physicians to reason about cerebral autoregulation without the monitor present by ensuring the blood pressure and cerebral oxygenation data showed similar patterns and by choosing symptoms that would commonly require decisions to be made about blood pressure treatment, which would be affected by knowledge about the state of cerebral autoregulation.

6.3.3 Questionnaire

To be able to gain an insight into the usability of the cerebral autoregulation monitor, a questionnaire was presented to the participants, once after completing the first and once after the second block of trials. This allowed a comparison between the usability scores for the standard monitor and for the standard and cerebral autoregulation monitor together. The items in the questionnaire were based on eight items from the standard usability scale (SUS) (Brooke, 1996) plus three additional questions. All items used in the questionnaire can be seen below. Each item in the questionnaire consisted of a statement about the system that was rated by the participant on a five point Likert scale that ranged from “Strongly disagree” to “Strongly agree”. The score of each individual question was obtained by taking the position on the Likert scale as 0 - 4. Two scores were calculated for the questionnaire: the overall score on the questionnaire and the SUS score, made up of only the questions that were taken from the SUS. The score on the entire questionnaire was calculated as the sum of the scores on the individual questions. Obtaining the SUS score is achieved by taking the question score minus one for questions 1,3,5,7 and five minus the question score for questions 2,4,6,8; then multiplying the sum

of these scores with 2.5. These scores can range from 0 to 100. As two questions were omitted, the maximum attainable SUS score on this questionnaire is 80.

The SUS was chosen as the source for most items in the questionnaire as it is a validated questionnaire that has been used to assess usability in a great number of different domains and scenarios. Therefore using the SUS makes it possible to place the usability of the cerebral autoregulation monitor in a broader context. As the SUS is intended for use with an interactive system and the interface used in this experiment only presents information to the user, two questionnaire items pertaining to the interaction with the system were ommitted. These were items 5 and 6 from the standard SUS: "I found the various functions in the system were well integrated" and "I thought there was too much inconsistency in this system". The items taken from the SUS used in this study were:

1. I think that I would like to use this monitor frequently (NL: ik denk dat ik het prettig zou vinden om deze monitor vaker te gebruiken)
2. I found the monitor unnecessarily complex (NL: Ik vond de monitor nodeloos ingewikkeld)
3. I thought the monitor was easy to use (NL: Ik vond de monitor makkelijk in gebruik)
4. I think that I would need the support of a technical person to be able to use this monitor (NL: Ik denk dat ik hulp nodig heb van een technisch persoon om van de monitor gebruik te kunnen maken)
5. I would imagine that most people would learn to use this monitor very quickly (NL: Ik kan mij voorstellen dat de meeste mensen makkelijk leren om deze monitor te gebruiken)
6. I found the monitor very cumbersome to use (NL: Ik vond de monitor moeilijk te gebruiken)
7. I felt very confident using the monitor (NL: Ik voelde me heel zelfverzekerd in het gebruik van de monitor)
8. I needed to learn a lot of things to get going with this monitor (NL: Het duurde lang voor ik de monitor goed genoeg door had om er echt mee aan de slag te kunnen)

Three additional items were added to the questionnaire to gain more insight into whether the participants like the cerebral autoregulation monitor enough to recommend it to other physicians, whether the monitor inspired a sense of confidence in the physicians and

whether the monitor provided all the necessary information. These additional questions were:

9 I would recommend another physician to use this monitor (NL: Ik zou een andere arts aanraden om van deze monitor gebruik te maken).

10 I would dare to trust on information presented in this monitor (NL: Ik zou op informatie van deze monitor durven vertrouwen).

11 It was easy to find the information I was looking for in the monitor (NL: Ik vond het makkelijk om de informatie die ik zocht in de monitor the vinden).

6.3.4 Semi-structured interview

At the end of the experimental session a semi-structured interview was conducted with the physician. This interview was conducted with two main goals. The first goal was to gain further insight from the physicians into how they experienced working with a cerebral autoregulation monitor, what their attitude was towards using a cerebral autoregulation monitor in the NICU, how they currently reason about cerebral autoregulation when assessing a patient, and the influence the physicians themselves think a monitor will have on their reasoning about cerebral autoregulation.

The second goal was to obtain feedback from the physicians on the design of the cerebral autoregulation monitor. More specifically, feedback was sought on the visual representation of cerebral autoregulation, the measure for cerebral autoregulation used and on anything that was missing or unpleasant in the display.

The interview did not adhere to a fixed structure but the following list of questions served as starting point and guideline during the interview:

1. Do you think you would use the cerebral autoregulation display in daily practice?
    - In what situation would you use the display?
    - Would you use the display at the bedside or remotely?
    - To what degree do you take cerebral autoregulation into account when reasoning about patients?
        - Do you think the display would change that?
2. Did you think the cerebral autoregulation display was informative?
    - Was there information you were missing in the display?

- What did you think of using correlation as a measure?
- What did you think of the use of colors?
- What part of the display took the most getting used to?

The participants were invited to ask any further questions or make remarks if they had any at the end of the session.

6.4 Design

The study featured 2 x 2 conditions: monitor without the cerebral autoregulation display showing stable or unstable cases, and monitor with the cerebral autoregulation display showing stable or unstable cases. In a within-subjects design, participants were first presented with the monitor without the cerebral autoregulation display and then with the monitor with the cerebral autoregulation display. The order of presentation of cases was balanced across subjects and monitor conditions. Each participant was presented with a different and unique order of cases and each case was presented equally often in the condition without the cerebral autoregulation display and the condition with the cerebral autoregulation display.

The nature of the study is qualitative. The dependant variables which we are most interested in will be the outcome of the reasoning process across conditions. Below, we will also look at three quantitative variables, the amount of time (in seconds) spent on a case and the overall and SUS scores on the usability questionnaire. Finally, there will be qualitative data in the form of the semi-structured interview. This is independent of conditions.

6.5 Procedure

At the start of an experimental session participants were told that they were going to evaluate a new custom monitor for the piCare project. They already knew beforehand due to communication within the department that part of the study was related to cerebral autoregulation. This was not repeated at the beginning of the experimental session although several participants asked about it. Participants were briefed that they were going to see two variants of a monitor and that for each variant there would be a block of four trials consisting of assessing a simulated case. It was explained to the participants that a block of trials represented doing rounds in the hospital ward. The participants were instructed to look at the case presented on the monitor, assess the

situation of the patient presented on the monitor and describe what course of action they would take in this situation. It was explicitly mentioned that the participants were free to take any course of action, including deciding not to undertake an action or to ask for further information. The participants were instructed to think out loud during the task.

The participants were informed that a usability questionnaire would be presented after each variation of the display and that an interview would be conducted at the end of the session. The participants were asked to give their informed consent to participate in the study and for the session's audio to be recorded.

After the instructions, the participants were presented with the first block of trials using the monitor without the cerebral autoregulation display. They were presented with two trials of stable cases to learn to read the monitor and to serve as control trials. For each case the participants were instructed to indicate when they were satisfied with their assessment of the patient and the treatment plan. The experimenter stopped the trial at that time and started the next trial. After two stable cases, the participants performed two trials with unstable cases.

After finishing the first block of trials, the participants filled out the usability questionnaire for the standard monitor condition. This was followed by a brief instruction explaining the cerebral autoregulation display. This instruction entailed describing the relationship between mean arterial blood pressure and NIRS and explaining that each bar in the display represents the correlation value of these two signals over the last five minutes. The purpose of this explanation was to make clear how the interface of the cerebral autoregulation display is related to the concept of cerebral autoregulation, which the physicians were already familiar with.

After this explanation, the second block of trials started. This also consisted of two trials with stable cases, followed by two trials of unstable cases. Afterwards, the participants filled out out the usability questionnaire for the cerebral autoregulation monitor condition. The experimental session ended with conducting the semi-structured interview.

6.6 Analysis

The small sample size meant that this study was not suitable for statistical analysis. The difference between the quantitative variables, mean overall questionnaire score, mean SUS score and mean time required to complete a case, were therefore examined exploratively.

All experimental sessions were transcribed from the audio recordings. From these transcriptions, a list of diagnoses was made for each case per condition and participant. These were compared per monitor condition to see if the presence of the cerebral autoregulation monitor leads to a different diagnosis. To assess the impact of the cerebral autoregulation monitor on the reasoning process, a similar list of treatment plans was made. Each entry in this list consists of a list of steps that the participant wants to undertake to treat this specific case and the motivation for these steps, if stated by the participant. In this context a step is defined as an action or series of actions that serves one medical purpose, e.g., "taking a blood gas test" or "increasing respiratory support" are seen as one step even though they may consist of multiple physical acts. These plans were compared on what specific medical actions were included, number of steps, the amount of reasoning about cerebral autoregulation and the influence of the cerebral autoregulation monitor on the reasoning process. To be able to assess the whole reasoning process and not only the final plan, a list was also compiled for each case and participant of plans or diagnoses that were considered but abandoned and the reason of abandonment. Finally to gain a qualitative insight into the user experience, all remarks about usability or the design of the monitor found during the trials were compiled in a general list, roughly divided into positive and negative remarks.

The answers and remarks made during the semi-structured interview were first grouped per question (see section 6.3.4) where possible. These answers were compiled into a list per question and their contents were then examined for similar and contrasting points of view. All remarks that did not fit a specific pre-defined question about usability were added to the general list of usability remarks. All other remarks and questions relevant to the study were added to a miscellaneous list and its content was examined to find if there were common themes or patterns. If so, they were grouped, else they were marked as individual opinions.

Chapter 7

Results

The evaluation study was designed to test how the presence of the designed cerebral autoregulation monitor influenced the situation awareness of physicians while assessing and treating a patient. The amount of time needed to assess a patient and come up with a plan for treatment was compared between conditions, as were the usability questionnaire scores. In addition to this, a qualitative analysis was carried out on the transcriptions of the experimental session to assess the presence of the cerebral autoregulation monitor on the clinical reasoning process and its outcome. The transcription data were also analyzed for usability feedback and the physicians' perspective on reasoning about cerebral autoregulation and the possible introduction of a cerebral autoregulation monitor in the NICU.

The next sections describe the results of the evaluation study, starting with the quantative data on task duration and usability scores, followed by the qualitative data on clinical reasoning, usability and the interview.

7.1 Amount of time needed to complete the task

The mean amount of time needed to complete the task in seconds per monitor condition and case category for each case can be seen in figure 7.1. The mean amount of time per stable case without the CA monitor was 81.4 seconds (SD = 34.5). The mean amount of time per stable case with the CA monitor was 75.4 seconds (SD = 54.9). The mean amount of time per unstable case without the CA monitor was 194.3 seconds (SD = 136.8). The mean amount of time per unstable case with the CA monitor was 193.6 seconds (SD = 158.3).

Overall, the task took longer to complete for the unstable cases than for the stable cases.

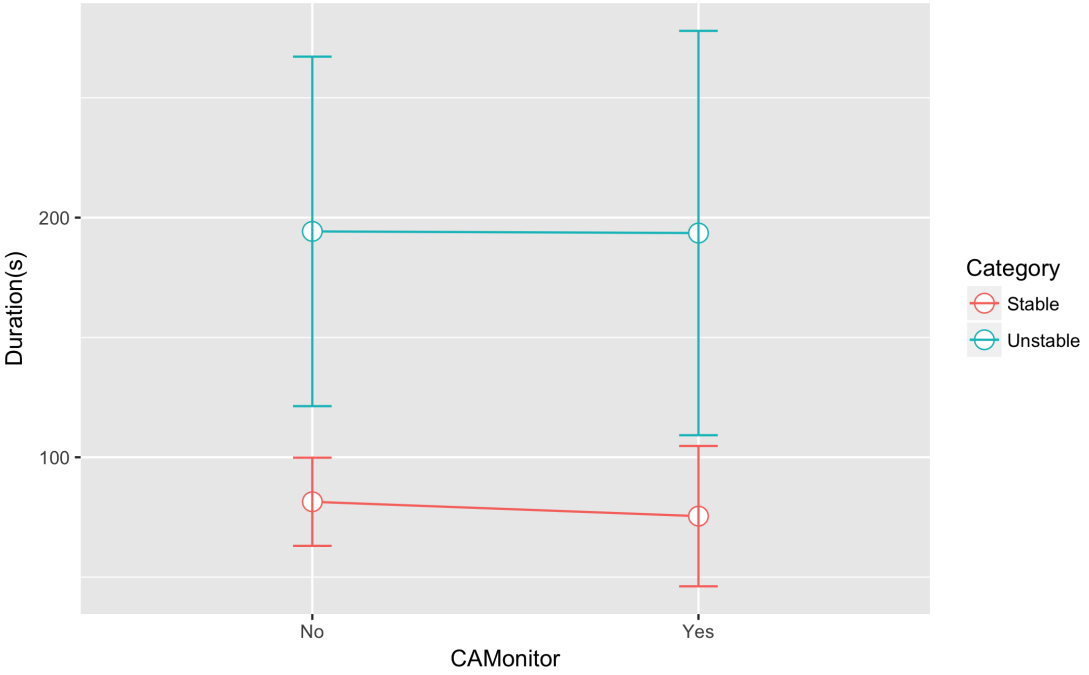

FIGURE 7.1: The duration in seconds of cases per category and monitor condition. Error bars represent 95% confidence intervals.

Overall there does not appear to be a meaningful difference between monitor conditions in how long it took participants to complete the task.

7.2 Usability questionnaire

Figure 7.2 shows the mean usability score of the total questionnaire per monitor condition. The mean usability score of the monitor without the CA monitor was 34.8 (SD = 4.6). The mean usability score with the CA monitor was 31.1 (SD = 7.6).

The mean SUS score of the interface without the CA monitor was 65.6 (SD = 11.1). The mean SUS score of the interface with the CA monitor was 58.1 (SD = 14.7) (See figure 7.3).

The CA monitor scored consistently lower on the usability questionnaire when looking at the overall score and SUS score. Looking at the mean scores for each individual item (Figure 7.4), there are items where both monitors scored equal or the CA monitor scored higher.

The CA monitor scored an average of 3 on item 1 compared to 2.87 for the standard monitor. The CA monitor scored an average of 2.50 on item 9 compared to 2.38 for the standard monitor. Mean scores were equal on item 2 (3.50) and item 10 (3.13).

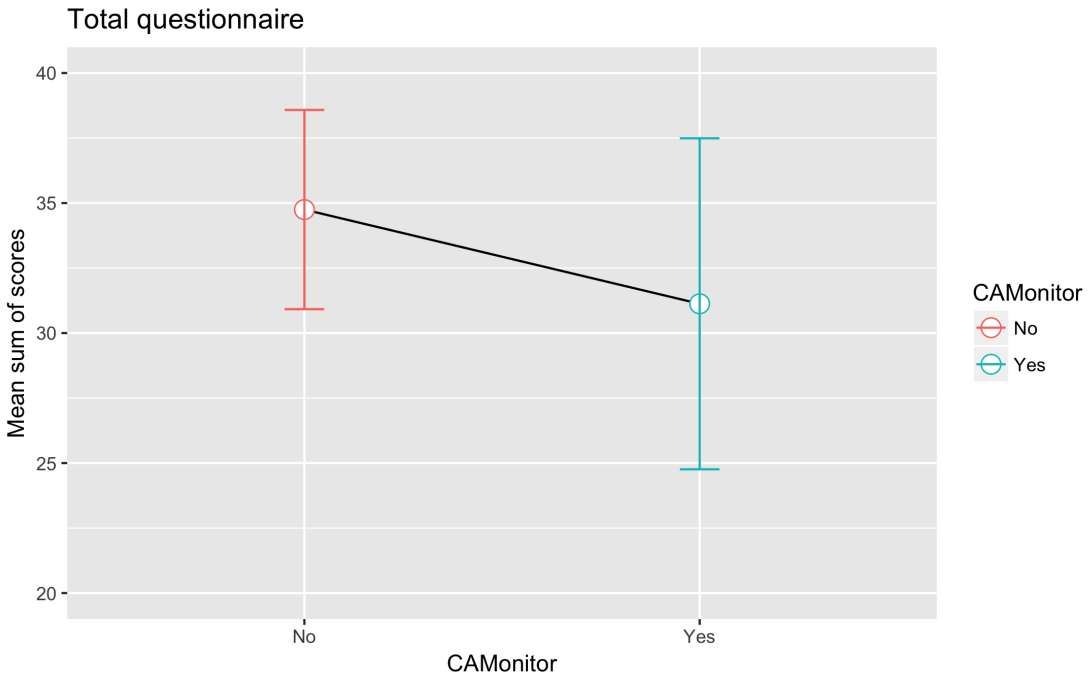

FIGURE 7.2: Mean usability questionnaire scores per condition. Error bars represent 95% confidence intervals.

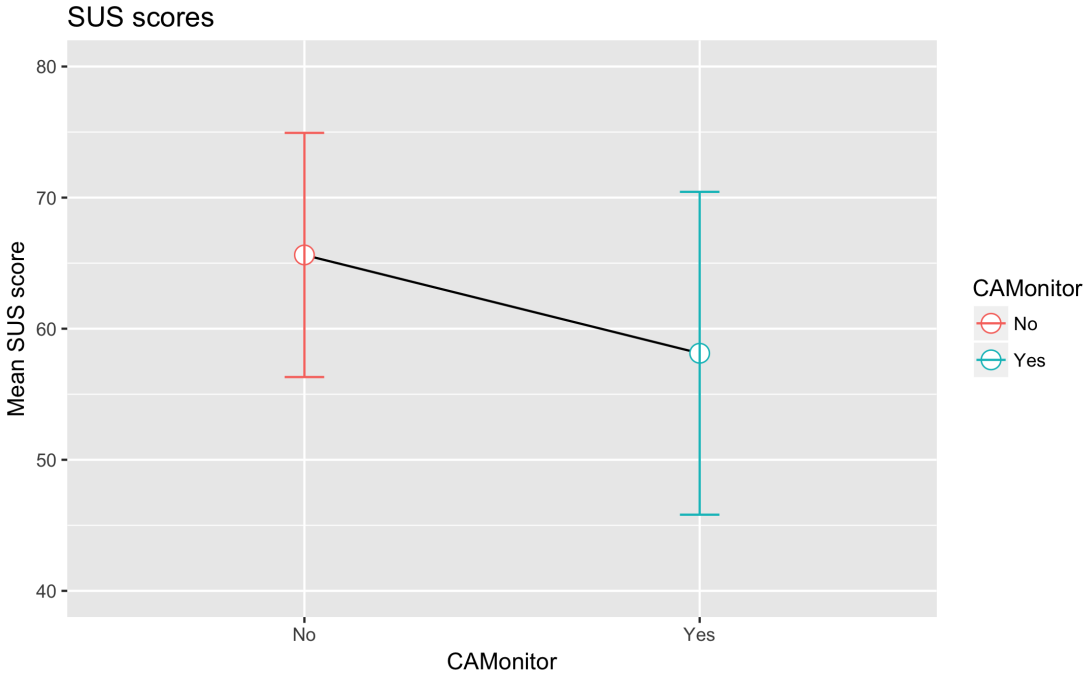

FIGURE 7.3: The score on the part of the questionnaire taken from the SUS per condition. Error bars represent 95% confidence intervals.

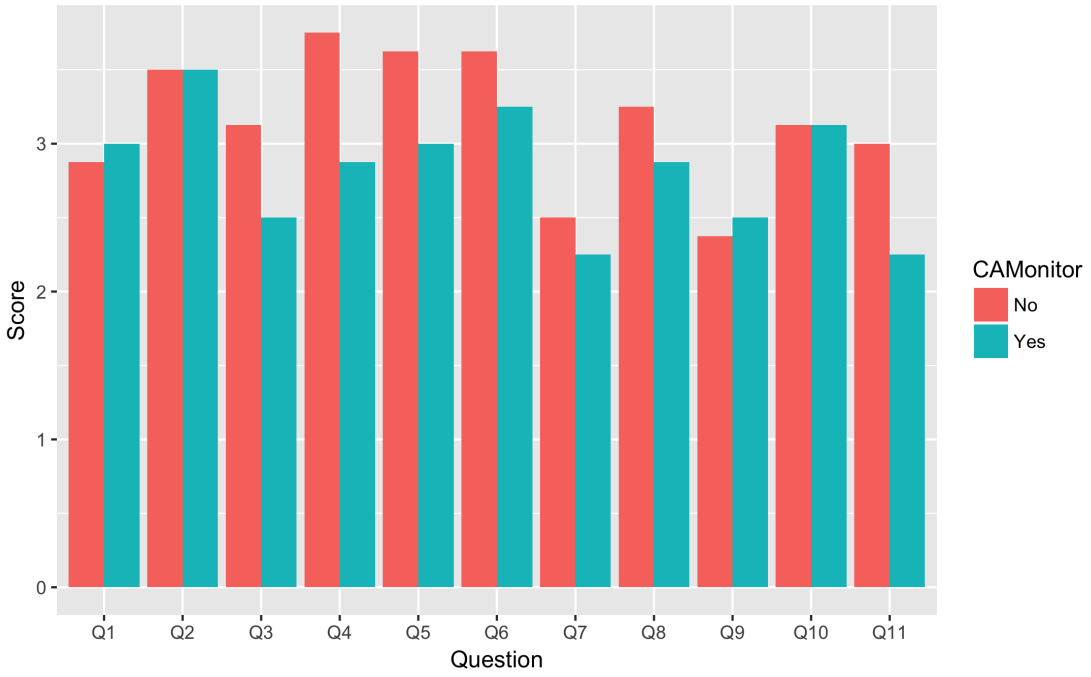

FIGURE 7.4: Mean score per questionnaire item per condition.

7.3 Clinical reasoning

First, the clinical reasoning performance of the physicians for the stable cases will be discussed per monitor condition. After that, the clinical reasoning performance of the physicians for the unstable cases will be discussed per monitor condition.

In 28 of 32 (8 participants $\times$ 4 cases) cases, the participants correctly identified the patient as stable and decided not to take any action. Of the four cases where the reasoning process lead to a different outcome, two were in the no CA monitor condition and two were in the CA monitor condition. In two separate cases in the no CA monitor condition, the participant requested more information about circulatory support. In the CA monitor condition there was one case where the participant decided the patient needed more respiratory support. In the other case in the CA monitor condition the participant indicated they would check up on the patient more regularly due to worries about the values shown on the cerebral autoregulation monitor. As the stable cases were designed to show values of cerebral autoregulation that did not require intervention this was unjustified.

In all 32 unstable cases, the participants correctly identified the patient as unstable. In four cases, the treatment plan was directly informed by the values on the CA monitor. In two separate cases, the participant decided to make an echo of the brain. In one case a treatment option (increasing the respiratory support) was rejected due to the absence of CA shown on the monitor. In one case the participant did not know what action to take

given the values on the CA monitor and decided that their course of action would be to ask help from another physician. In the cases where the course of action was not directly informed by the CA monitor, the outcome of the clinical reasoning process was similar in both monitor conditions. This means that the diagnosis and general treatment plan for a patient that the participants made was often the same, regardless of participant and monitor condition. There was no indication of extra doubt or enthusiasm for certain treatments in the CA monitor condition compared to the No CA monitor condition. However, there were notable differences in the amount of reasoning about CA between the monitor conditions as well as in the contents of the treatment plans.

Reasoning about cerebral autoregulation was a lot more frequent in the CA monitor condition. In only 1 out of 16 (4 participants $\times$ 4 cases) cases in the no CA monitor condition the participant explicitly mentioned cerebral autoregulation when reasoning about the patient. In the remaining 15 cases there were no indications that the participants were reasoning about cerebral autoregulation. In the CA monitor condition all participants explicitly mentioned cerebral autoregulation in all 16 cases. This means that the physicians explicitly stated they took note of the values on the CA monitor and that they were aware of the state of CA in the patient, not necessarily that CA directly informed the decisions made for the treatment plan.

Interestingly, the treatment plans in the CA monitor conditions often contained more steps while trying to achieve the same goal as in the no CA monitor condition. The mean number of steps per case and per condition is shown in table 7.1. Only in case *unstable 1* was the mean number or steps greater in the no CA monitor condition than in the CA monitor condition. In case *unstable 4*, the mean number of steps was equal in both conditions. In case *unstable 2* and *unstable 3* the mean number of steps was larger in the CA monitor condition than in the no CA monitor condition. The mean number of steps in the treatment plans for these two cases is noticeably larger than in case *unstable 1* and *unstable 4*, indicating that case *unstable 2* and *unstable 3* stimulated the physicians to make more complex plans. It is in these more elaborate plans where the difference in the reasoning process per condition can be seen best.

An interesting case to study in more detail is case *unstable 3*, here the difference in mean number of steps is the greatest and the diagnosis the physicians made in 7 of 8 cases (so regardless of condition) was "blood pressure too low" (the diverging diagnosis was a pneumothorax, in the no CA monitor condition), which is something that needs to be treated acutely and can require a different course of action depending on the presence of cerebral autoregulation (although no physician took a different course of action based on the cerebral autoregulation values). The treatment plans in the condition with no CA monitor were:

- In two cases, the treatment plan consisted only of one step: treating with volume expansion.
- In one case the treatment plan consisted of two steps: taking a blood gas test and administering antibiotics.
- In the case with the pneumothorax diagnosis, the treatment plan consisted of treating with volume expansion, increasing respiratory support and making a thorax x-ray.

The treatment plans in the condition with CA monitor were:

- In one case the treatment plan consisted of one step: treating with volume expansion.
- In one case the treatment plan consisted of three steps: treating with volume expansion, preparing cardiotonica and increasing respiratory support.
- In one case the treatment plan consisted of four steps: treating with volume expansion, increasing respiratory support, increasing the body temperature and taking a blood gas test.
- One treatment plan consisted of five steps: treating with volume expansion, preparing cardiotonica, starting to measure the pre- and post-ductal saturation difference, increasing body temperature and making an echo of the head.

As can be seen, nearly all of the treatment plans for case unstable 3 focus on elevating blood pressure by treating with volume expansion. The plan in the CA monitor condition was more elaborate in each case apart from one, adding additional treatments or preparing medicine for eventualities, whereas in the no CA monitor condition the participants mostly seem satisfied with only trying to raise the blood pressure by treating with volume expansion. A similar pattern can be seen in case *unstable 2*, where the diagnosis in 7 of 8 cases was pulmonary hypertension (with shunting being the diverging diagnosis). The treatment plans here also largely agreed between the conditions, with extra steps added in the CA monitor condition. A notable difference between the conditions that is related to the increased number of treatment steps is the greater frequency of conditionals in the treatment plans. In the no CA monitor condition it was never mentioned that a certain step would only be taken depending on the outcome of a previous step in the treatment plan. However, in the CA monitor condition some of the treatment plans contained multiple conditional steps.

TABLE 7.1: The mean number of steps in the treatment plan per case and condition (n = 4 for each case in each condition)

Case	No CA Monitor	CA Monitor
Unstable 1	2.00	1.50
Unstable 2	3.75	4.75
Unstable 3	1.75	3.25
Unstable 4	1.25	1.25

It can be seen from these examples that even though cerebral autoregulation was never mentioned as a motivation, the treatment plans in the condition with CA monitor consisted of more and different steps compared to the treatment plans in the no CA monitor condition.

7.4 General remarks and interview

In general the participants found the monitor clean, clear and pleasant looking. They mentioned that they liked the amount of overview the monitor gave them. Participants mentioned that they preferred the 30 minute trends in the monitor over the short trends in the Philips monitor they normally use. The presence of color and the choice of colors in the CA monitor was also received favorably by the participants. Several participants mentioned that they did not feel the need to look at the scale to see what value a bar represented because it was already clear from the color and size of the bar what was happening with the patient. None of the participants found the correlation scale or the underlying calculation confusing.

Multiple participants reported that they missed the fractional tissue oxygen extraction (FTOE) in the monitor. This value is obtained by the formula FTOE = (SpO_2 - rc_{SO_2})/SpO_2. Since SpO_2 and rc_{SO_2} are displayed on the monitor, it was possible for the participants to calculate the value for FTOE. The physicians that mentioned missing FTOE in the monitor also calculated it during the task. These participants also used FTOE in reasoning about cerebral autoregulation, trying to relate the FTOE score to the values shown in the CA monitor to see if the familiar FTOE scores matched the unfamiliar CA values in the monitor in terms of what they indicated about the patient's state.

Some participants found the monitor graphs too small and the CA graph too large. There were also participants who found the trendlines hard to read and suggested that ticklines or a grid could be added to the graphs to make it easier to read and compare

specific values. The label of the rc_{SO_2} (NIRS) graph was not clear enough for all participants. Some participants initially thought it represented renal instead of cerebral NIRS measurements and had to be corrected by the experimenter. It was not clear to all participants what the trend arrows were based on. Some reported that they did not find them intuitive or helpful and that the trendlines provided clear enough trend information. One participant did not find the blood pressure graph clear and thought the systolic and diastolic blood pressure lines showed alarm boundaries. This issue arose during one of the first stable cases where the lines were very stable.

In the interview, all participants indicated that they could see some use in the CA monitor in their work. Five participants indicated that they currently did not use cerebral autoregulation when reasoning about patients but that they did use NIRS values and FTOE scores to reason about potential cerebral damage. The physicians who stated that they did use cerebral autoregulation when reasoning about patients indicated that they thought a cerebral autoregulation monitor would be useful. The participants who indicated not to reason about cerebral autoregulation stated that they did not see the immediate need for a cerebral autoregulation monitor, as they had never missed cerebral autoregulation values before. However, after having used the cerebral autoregulation monitor these participants agreed that it might be beneficial to reason about cerebral autoregulation more regularly, and that the CA monitor would stimulate that. All participants thought that having the CA monitor would increase how often they would take cerebral autoregulation into account when treating a patient. Since a monitor like the one in this study does not currently exist, nor does another way of accessing the information that it displays, participants expressed uncertainty about how to use the cerebral autoregulation values when making decisions.

The most useful location for the CA monitor was universally agreed to be at the bedside. Some participants would like to see it as a bedside device with the possibility of accessing the data remotely as well. Some participants voiced concerns that if the CA monitor would be used as a bedside device that it might worry visiting parents to see a monitor filled with red bars when there is a lack of cerebral autoregulation. Since cerebral autoregulation cannot be restored immediately by the physicians, it is possible that the bars stay red even if adequate action has already been taken. Some participants also voiced concerns that having a monitor at the bedside with unchanging red bars might influence their own decisions negatively by erring on the side of caution, for example by not administering treatments to patients that cause large changes in blood pressure out of fear for cerebral damage even though withholding the treatment actually threatens the patient's health more than administering the treatment could.

Chapter 8

Discussion

In this study, a monitor was developed that shows the state of cerebral autoregulation in sick and premature babies. This monitor was used in a study that aimed to establish how the presence of a cerebral autoregulation affects the situation awareness and clinical reasoning process of physicians in the NICU as well as the physicians' attitudes towards reasoning about cerebral autoregulation and towards introducing a cerebral autoregulation monitor in the NICU. The study's secondary goal was to evaluate the designed monitor in terms of usability.

This final chapter will discuss the conclusions that can be drawn from the evaluation study and areas of focus for future research.

8.1 Conclusions

8.1.1 Time needed to complete a case

It was hypothesized that adding a cerebral autoregulation monitor to the standard monitors in the NICU would increase the amount of time needed to diagnose a patient and make a treatment plan. However, even though there was too little data to perform a meaningful statistical analysis, the time needed to diagnose a patient and come up with a treatment plan does not appear to differ between monitor conditions. This means that the addition of a cerebral autoregulation monitor does not appear to make the clinical reasoning process proceed more or less rapid.

8.1.2 Usability questionnaire

The overall score on the usability questionnaire and the SUS score were both lower on the monitor with cerebral autoregulation. This indicates that physicians find it easier and more comfortable to only use the monitor they are experienced and familiar with compared to using the cerebral autoregulation monitor in addition to this standard monitor.

On the SUS part of the questionnaire the new monitor scores 58.1 on average. Interfaces with SUS scores above 70 are generally seen as acceptable in terms of usability and interfaces with scores under 50 as having serious usability flaws (Bangor, Kortum, & Miller, 2008). It is impossible to know the score the monitor would have attained using the full SUS questionnaire but achieving a marginally acceptable score using the partial score is promising. To put the partial SUS score in perspective, it can be expressed as a percentage. The score on the full SUS is also a percentage of the maximum attainable score (Brooke, 1996). When doing this, the CA monitor attains a score of $(58.1/80) \times 100 = 72.6\%$, which translates to an acceptable level of usability. It does have to be noted that when using the SUS with a relatively small sample size such as in this study, it is advised to be cautious about interpreting the results (Bangor et al., 2008).

When looking at individual questions, the new monitor scored higher than the old monitor on the questions 'I think that I would like to use this system frequently" and"I would recommend another physician to use this monitor". The new monitor scored equal to the old monitor on "I would dare to trust information presented in this monitor". This indicates that even though the physicians found the new monitor less easy to use they are positive about the monitor being adopted in the department and that they trust the information the monitor provides.

8.1.3 Clinical Reasoning

As hypothesized, the physicians reasoned more frequently about cerebral autoregulation when the cerebral autoregulation monitor was present than when it was absent. The presence of the cerebral autoregulation monitor also influenced the outcome of the clinical reasoning process but not as hypothesized. The presence of the cerebral autoregulation monitor did not lead to different diagnoses and only directly influenced the chosen treatments in a few cases. Even though the focus of the treatment plans is very similar in both monitor conditions, the treatment plans in the cerebral autoregulation monitor condition contained a greater number of steps and conditionals. This indicates that the physicians may have experienced a greater level of uncertainty about their treatment plans in the condition with cerebral autoregulation present.

The finding that the treatment plans contain more steps and conditions when the cerebral autoregulation monitor is present may be explained by a lack of knowledge about relating cerebral autoregulation values to patient outcome. The cerebral autoregulation information provides uncertainty about which decision to make as its consequences are unfamiliar. Many participants also indicated they experienced uncertainty about what to do in the cerebral autoregulation monitor condition. In terms of naturalistic decision-making (Klein, 2008) this can be explained as follows: it becomes harder to make a decision based on the recognition of a pattern, so a more analytical process follows. Or in the terms of the dual-process model Kahneman (2011) is could be axplained as: system 1 reasoning is not possible and therefore the participants switch to system 2. This may explain why the participants mention a greater number of steps in the treatment plan: they rely on formal thinking rather than heuristics and therefore deliberately consider and mention each step. When the cerebral autoregulation monitor is not present, the recognition of the situation leads to the selection of a more automated plan that the physician is very familiar with. In this case, the physicians only explicitly mentions one step while more steps may be implied but not mentioned because they are obvious to the physician. When quick decision-making is disrupted due to the presence of the cerebral autoregulation monitor, the more analytical reasoning process that is necessary to make sense of the situation may cause more steps and conditionals to be mentioned in the treatment plan.

Another possibility is that the disruption in recognizing patterns caused by the unfamiliar information does not interfere with quick decision making but does decrease confidence in the decision. This may lead the physician to consider the selected plan more carefully, reasoning about if and how the cerebral autoregulation values interfere with this plan, which does not account for them. In terms of the Data/Frame theory Klein et al. (2006b): the information provided by the cerebral autoregulation values may cause the initial frame (the decision the physician would normally take without the unfamiliar cerebral autoregulation values) to be elaborated, leading to a more specific frame and thus a more elaborate plan.

Somewhat at odds with these explanations is that the time it takes to form a plan appears to be similar in both conditions, while it would be expected that these cases would take longer if either of the previous explanations would hold. The lack of difference in the amount time needed to complete a case could in turn be explained by a learning effect caused by the fixed order of presentation of the monitor conditions in the study. As the participants were always first presented with the no CA monitor condition before being presented with the CA monitor condition, the participants were much more familiar with most of the interface in the CA monitor condition, which may have compensated for a longer reasoning process in this condition.

Finally it has to be noted that obtaining a treatment plan containing more steps and conditions can be interpreted as a positive consequence of the addition of the cerebral autoregulation monitor. The greater number of steps indicates that treating a patient using the cerebral autoregulation monitor is more challenging for physicians but it might lead to a better outcome for the patients due to a more detailed, individual treatment approach compared to a quicker approach that may overlook some of the details of the medical state of each specific patient.

8.1.4 General remarks and interview

The users were generally satisfied in terms of usability. There were some issues in the experimental interface, such as a lack of gridlines and the relative sizes of the different charts. However, all participants were satisfied with the way the cerebral autoregulation monitor was designed.

The participants were all positive about the idea of using a monitor like this in daily practice at the NICU. The participants all indicated that they thought they would reason about cerebral autoregulation more often if the monitor was used in daily practice. This thought is supported by the observed behaviour in the study. The participants indicated that their preferred implementation of the cerebral autoregulation monitor would be as a bedside monitor with the possibility of also viewing the monitor remotely (which is already possible using the current design). The participants did express uncertainty about how to use the information provided by the cerebral autoregulation monitor. This is also supported by the observed behaviour in the study. The expressed uncertainty was based on two causes. On the one hand, the physicians were not sure how to use the cerebral autoregulation values when determining a treatment plan. On the other hand they all had different ideas on how to implement the monitor organizationally. The most conventional way of implementing the monitor organizationally would be to have the nursing staff read the monitor's values periodically (at a currently unspecified interval) and adding this to the patient charts that the physicians later receive. As a further potential improvement, an alarm could be added for when the cerebral autoregulation values rise above a threshold defined by the physicians. Finding more inventive ways of deploying the cerebral autoregulation monitor in the NICU could be a topic for further study.

8.1.5 Overall

It can be concluded from this study that physicians at the NICU like the way the cerebral autoregulation monitor was designed, but are still unsure on how to use it. The monitor made it possible to observe the state of cerebral autoregulation and detect changes. And this in turn has made the physicians relate this information to the current state of the patient and to the goal of treating the patient. In terms of situation awareness as described by Endsley (1995), this equates to increased level 1 and level 2 situation awareness. This also fits with how Drews and Westenskow (2006) relate situation awareness to patient monitors in anesthesia.

The novelty of the possibility to use cerebral autoregulation in reasoning and the limited knowledge about how to exactly use it to inform clinical decisions means that at this point it cannot be said whether the monitor also enhances level 3 situation awareness. This knowledge gap prevents the physicians from predicting the future state of the patient accurately while taking the cerebral autoregulation values into account. Further study that relates cerebral autoregulation values to treatment and patient outcome is required to bridge this knowledge gap. It then remains to be seen whether the current design of the cerebral autoregulation monitor is the one most suited to enhancing level 3 situation awareness. It will however be possible to use it as a tool to study how the knowledge needed to achieve level 3 situation awareness can be attained.

The results of this study indicate that at the current level of knowledge about cerebral autoregulation in relation to patient outcome, the addition of a cerebral autoregulation monitor to the standard monitor at the NICU mainly leads to more uncertainty in the clinical reasoning and decision making process. The addition of the cerebral autoregulation monitor did not appear to have a great effect on which treatments are administered and withheld. Therefore, using the cerebral autoregulation monitor in the NICU may cautiously be encouraged. Even though it adds uncertainty, the addition of the cerebral autoregulation display did not slow down the physicians' clinical reasoning process or cause any other adverse effects. The addition of the cerebral autoregulation monitor leads physicians to think more explicitly about the consequences of certain treatments and to make a more elaborate and explicit treatment plan. At the very least, the cerebral autoregulation monitor can be used as a tool to study the use of cerebral autoregulation in clinical reasoning by providing a way for the physicians to use the cerebral autoregulation values and by providing a platform to study how certain treatments and circumstances affect the cerebral autoregulation values. This should however happen in an experimental context first. The current state of knowledge about the relation between cerebral autoregulation, treatment and patient outcome is not yet sufficient to adopt the use of the cerebral autoregulation monitor in daily clinical practice. However, as soon

as this knowledge is more advanced, the monitor can successfully be implemented at the NICU.

8.2 Further Research

There are several directions in research that can be taken to increase the chances of a successful implementation of a cerebral autoregulation monitor at the NICU. First af all, this study was limited to few participants and only one NICU. While this is enough to assess the cerebral autoregulation monitor's usefulness for the population of physicians in this NICU, this study should be followed up by studies across different NICUs with larger sample sizes to see if its findings hold. Additionally a follow-up study should be situated in a more realistically simulated environment. The way cases were presented in this study was not very realistic. Several participants in the study expressed the desire to see the patient when making decisions about treatment. This means that ideally a similar study should take place in a simulated environment that more closely resembles the reality of the department or at the department itself. It would also be interesting to do a similar study using data of actual patients with impaired autoregulation instead of simulated data.

There were some participants who questioned the reliability and validity of the calculation underlying the visualization of the monitor. As discussed extensively in chapter 3, there is no definitive way yet to construct a measure of cerebral autoregulation. While the method that was chosen in this study was judged to be the most reliable at this point in time, it is not possible to show the exact ecological validity of the measure. With the relatively simple correlation measure it is still possible that the cerebral autoregulation monitor shows false positives or negatives. More research is needed to establish a definitive method to accurately assess the state of cerebral autoregulation in the patient.

Research that studies patient outcome in relation to cerebral autoregulation values and administered treatments should be expanded. The current studies use different methodologies to construct cerebral autoregulation values and look at different measures of patient outcome. This leads to difficulty in designing protocols for treatment based on cerebral autoregulation values. A cerebral autoregulation monitor can only be used to its fullest potential when it is possible to determine how the data it displays should inform actions taken by physicians. At that point in time the monitor will be able to enhance situation awareness at level 3.

The monitor designed in this study can function as a tool for studying the relationship between cerebral autoregulation values, patient outcome and administered treatments

by allowing researchers and researching physicians to use the cerebral autoregulation monitor to observe the progression of the state of cerebral autoregulation and patient outcome. The monitor could be used in conducting controlled trials and observing how the values on the monitor respond to certain treatments being administered in patients who show impaired cerebral autoregulation and certain other symptoms. This type of research would allow the physicians to develop an intuition for how cerebral autoregulation behaves over time in patients and come up with appropriate strategies for treatments that take cerebral autoregulation into account.

References

Agutter, J., Drews, F., Syroid, N., Westenskow, D., Albert, R., Strayer, D., ... Weinger, M. B. (2003). Evaluation of Graphic Cardiovascular Display in a High-Fidelity Simulator:. *Anesthesia & Analgesia*, *97*, 1403–1413.

Alderliesten, T., Lemmers, P. M., Smarius, J. J., van de Vosse, R. E., Baerts, W., & van Bel, F. (2013). Cerebral Oxygenation, Extraction, and Autoregulation in Very Preterm Infants Who Develop Peri-Intraventricular Hemorrhage. *The Journal of Pediatrics*, *162*(4), 698–704.e2.

Aselmaa, A., van Herk, M., Laprie, A., Nestle, U., Gtz, I., Wiedenmann, N., ... Goossens, R. H. (2017). Using a contextualized sensemaking model for interaction design: A case study of tumor contouring. *Journal of Biomedical Informatics*, *65*, 145–158.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., ... Lawn, J. E. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet*, *379*(9832), 2162–2172.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309.

Brooke, J. (1996). SUS–A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Caicedo, A., Alderliesten, T., Naulaers, G., Lemmers, P., van Bel, F., & Van Huffel, S. (2016). A New Framework for the Assessment of Cerebral Hemodynamics Regulation in Neonates Using NIRS. In C. E. Elwell, T. S. Leung, & D. K. Harrison (Eds.), *Oxygen Transport to Tissue XXXVII* (Vol. 876, pp. 501–509). New York, NY: Springer.

Caicedo, A., Naulaers, G., Lemmers, P., van Bel, F., Wolf, M., & Van Huffel, S. (2012). Detection of cerebral autoregulation by near-infrared spectroscopy in neonates:

performance analysis of measurement methods. *Journal of Biomedical Optics*, *17*(11), 117003–117003.

Cnossen, F. (2015). Cognitive Skills in Medicine: An Introduction. In P. Lanzer (Ed.), *PanVascular Medicine* (pp. 4719–4752). Berlin, Heidelberg: Springer.

Croskerry, P. (2009). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education*, *14*(S1), 27–35.

da Costa, C. S., Czosnyka, M., Smielewski, P., Mitra, S., Stevenson, G. N., & Austin, T. (2015). Monitoring of Cerebrovascular Reactivity for Determination of Optimal Blood Pressure in Preterm Infants. *The Journal of Pediatrics*, *167*(1), 86–91.

Drews, F. A. (2013). Human Factors in Critical Care Medical Environments. *Reviews of Human Factors and Ergonomics*, *8*(1), 103–148.

Drews, F. A., & Doig, A. (2014). Evaluation of a Configural Vital Signs Display for Intensive Care Unit Nurses. *Human Factors*, *56*(3), 569–580.

Drews, F. A., Musters, A., & Samore, M. H. (2008). Error producing conditions in the intensive care unit. In H. K. Battles, J. Battles, & M. Keyes (Eds.), *Advances in patient safety: New directions and alternative approaches (vol. 3: Performance and tools).* Rockville (MD): Agency for Healthcare Research and Quality (US).

Drews, F. A., & Westenskow, D. R. (2006). The right picture is worth a thousand numbers: data displays in anesthesia. *Human Factors*, *48*(1), 59–71.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32–64.

Eriksen, V. R., Hahn, G. H., & Greisen, G. (2015). Cerebral autoregulation in the preterm newborn using near-infrared spectroscopy: a comparison of time-domain and frequency-domain analyses. *Journal of Biomedical Optics*, *20*(3), 037009–037009.

Gilmore, M. M., Stone, B. S., Shepard, J. A., Czosnyka, M., Easley, R. B., & Brady, K. M. (2011). Relationship between cerebrovascular dysautoregulation and arterial blood pressure in the premature infant. *Journal of Perinatology*, *31*(11), 722–729.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Strauss and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors*, *50*(3), 456–460.

Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, *21*(4), 70–73.

Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems*, *21*(5), 88–92.

Kooi, E. M. W., Verhagen, E. A., Elting, J. W. J., Czosnyka, M., Austin, T., Wong, F. Y., & Aries, M. J. (2017). Measuring cerebrovascular autoregulation in preterm

infants using near-infrared spectroscopy: an overview of the literature. *Expert Review of Neurotherapeutics*, *17*(8), 801–818.

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer*, *30*(12), 1373–1393.

Mitra, S., Czosnyka, M., Smielewski, P., O'Reilly, H., Brady, K., & Austin, T. (2014). Heart rate passivity of cerebral tissue oxygenation is associated with predictors of poor outcome in preterm infants. *Acta Paediatrica*, *103*(9), e374–e382.

Nielsen, J. (1992). The usability engineering life cycle. *Computer*, *25*(3), 12–22.

Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Medical Education*, *39*(4), 418–427.

Norman, G., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education*, *44*(1), 94–100.

North, C. (2012). Information Visualization. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics, Fourth Edition* (4th ed., pp. 1209 – 1236). John Wiley & Sons.

Parry, G., Tucker, J., & Tarnow-Mordi, W. (2003). CRIB II: an update of the clinical risk index for babies score. *The Lancet*, *361*(9371), 1789–1791.

Riera, J., Cabaas, F., Serrano, J. J., Bravo, M. C., Lpez-Ortego, P., Snchez, L., ... Pellicer, A. (2014). New Time-Frequency Method for Cerebral Autoregulation in Newborns: Predictive Capacity for Clinical Outcomes. *The Journal of Pediatrics*, *165*(5), 897–902.

Stammwitz, A., von Siebenthal, K., Bucher, H. U., & Wolf, M. (2016). Can the Assessment of Spontaneous Oscillations by Near Infrared Spectrophotometry Predict Neurological Outcome of Preterm Infants? In C. E. Elwell, T. S. Leung, & D. K. Harrison (Eds.), *Oxygen Transport to Tissue XXXVII* (Vol. 876, pp. 521–531). New York, NY: Springer.

Tsuji, M., Saul, J. P., du Plessis, A., Eichenwald, E., Sobh, J., Crocker, R., & Volpe, J. J. (2000). Cerebral intravascular oxygenation correlates with mean arterial pressure in critically ill premature infants. *Pediatrics*, *106*(4), 625–632.

Tversky, A., & Kahneman, D. (1975). Judgment under Uncertainty: Heuristics and Biases. In *Utility, Probability, and Human Decision Making* (pp. 141–162). Springer, Dordrecht.

van Amsterdam, K., Cnossen, F., Ballast, A., & Struys, M. M. R. F. (2013). Visual metaphors on anaesthesia monitors do not improve anaesthetists' performance in the operating theatre. *British Journal of Anaesthesia*, *110*(5), 816–822.

van Bel, F., Lemmers, P., & Naulaers, G. (2008). Monitoring Neonatal Regional Cerebral Oxygen Saturation in Clinical Practice: Value and Pitfalls. *Neonatology*, *94*(4), 237–244.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., & Dirnagl, U. (1993). Near infrared

spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience letters*, *154*(1), 101–104.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed ed.). Upper Saddle River, NJ: Prentice Hall.

Wolf, M., & Greisen, G. (2009). Advances in Near-Infrared Spectroscopy to Study the Brain of the Preterm and Term Neonate. *Clinics in Perinatology*, *36*(4), 807–834.

Wong, F. Y., Leung, T. S., Austin, T., Wilkinson, M., Meek, J. H., Wyatt, J. S., & Walker, A. M. (2008). Impaired Autoregulation in Preterm Infants Identified by Using Spatially Resolved Spectroscopy. *Pediatrics*, *121*(3), e604–e611.

Zhang, P., & Soergel, D. (2014). Towards a Comprehensive Model of the Cognitive Process and Mechanisms of Individual Sensemaking. *Journal of the Association for Information Science and Technology*, *65*(9), 1733–1756.

Appendix A

Simulation cases

This appendix shows the details of every case used in the evaluation study. The clinical information and history are exactly as shown to the participants (except translated to English from Dutch). The value progression was simulated to occur in 2x real time on the various monitors in the experimental interface. At the of the appendix a glossary of abbreviations can be found.

A.1 Stable cases

1. Name: Sophie de Vries, **Sex**: Female, **Age**: 25 weeks + 3 days

Clinical Information: Peeing 2-3 ml/kg/hour, pink complexion, mobile, warm extremities, well circulated. CPAP-S: 5 cm, 21% O2, no heart murmur
History: Sophie, born at 25+3 weeks, 780 grams, spontaneous preterm birth, unprepared, mother received 1 dose penicillin 30 minutes ante partum. Admitted to the NICU. Received umbilical venous catheters post partum.
Value progression: **HF**: stable at 150 +/ 10. **MABP**: stable at 35 +/- 5, **SABP**: stable at 45 +/- 5, **DABP**: stable at 20 +/- 5. rc_{SO_2}: stable at 80 +/- 5. SpO_2: stable at 90 +/- 5. **CA**: stable around 0.3, no fixed variation.

2. Name: Michael Jansen, **Sex**: Male, **Age**: 29 weeks

Clinical Information: Pink complexion, mobile, well circulated, CPAP support, 23% O2. No arterial lines
History: Michael, born at 29 weeks, unprepared, spontaneous preterm birth, parents' first child. Admitted to NICU due to O2 demand. No umbilical venous catheters.

Value progression: **HF**: stable at 153 +/ 5. **MABP**: stable at 35 +/- 5, **SABP**: stable at 45 +/- 5, **DABP**: stable at 25 +/- 5. textbfrc_{SO_2}: stable at 80 +/- 5. SpO_2: stable at 95 +/- 5. **CA**: Stable around 0.2 +/- 0.1 with some peaks at 0.4.

3. Name: Jamilla Kompany, **Sex**: Female, **Age**: 27 weeks

Clinical Information: Peeing 2 ml/kg/hour, pink complexion, mobile, well circulated. 25% O2.
History: Jamilla, born at 27 weeks, spontaneous preterm birth, unprepared, 950 grams. Started out well, Admitted to the NICU due to O2 demand. No umbilical venous catheters.
Value progression: **HF**: stable at 160 +/ 10. **MABP**: stable at 35 +/- 5, **SABP**: stable at 50 +/- 5, **DABP**: stable at 20 +/- 5. rc_{SO_2}: stable at 80 +/- 5. SpO_2: stable at 90 +/- 5. **CA**: stable around 0.2, no fixed variation. **CA**: Varying in the range of 0.1-0.4.

4. Name: Ronald Maassen, **Sex**: Male, **Age**: 30 weeks

Clinical Information: Pink complexion, peeing 1 ml/kg/hour, CPAP support. 21% O2.
History: Ronald, born at 30 weeks, parents' second child, unprepared, spontaneous preterm birth. 1320 grams. No umbilical venous catheters.
Value progression: **HF**: stable at 151 +/ 5. **MABP**: stable at 30 +/- 5, **SABP**: stable at 40 +/- 5, **DABP**: stable at 20 +/- 5. SpO_2: stable at 85 +/- 5. SpO_2: stable at 90 +/- 5. **CA**: stable around 0.2, no fixed variation.

A.2 Unstable cases

1. Name: Piet Visser, **Sex**: Male, **Age**: 25 + 3/7 weeks

Clinical Information: Apgar score 6/7/8 (tube), Diuresis 2 ml/kg/hour, Physical examilantion: complexion: pale pink, capillairy refill time 2-3 sec, temp 37,4, no diuresis in last 2 hours, superficial respiration with apnea, soft precordial heart murmur, pH 7.23/ pCO2 8,5, BE -5
History: Piet, born at 25+3 weken, 730 grams, spontaneous premature birth, unprepared, mother received 1 dose of penicillin 30 minutes ante partum. Admitted to the NICU. Recieved umbilical venous catheters post partum. During short mechanical ventilation, 1x surfactant per tube within 3 hours of birth, Detubed on second day, switch to nasal ventilation using NIMV via infantflow, second day CPAP-s. AB stopped after 48 hours because of negative cultures. Nutrition heightened following the protocol.

(3th day:) ABP decreases in the last three hours from mean 32 to mean 25. DABP is decreasing in particular. HF is rising slowly from 150 to 170. rc_{SO_2} decreases very slowly from around 80 to around 70. Added 02 is gradually heightened to 30-35% to keep oxygen saturation of 90%. Some dips in oxygen saturation to 60% (duration: 10 sec). Sometimes together short bradycardia to 60/min.

Value progression:
CA is cycling between low values around 0.2-0.4 and high values around 0.6-0.8 every five minutes. These cycles roughly coincide with the bradycardias.

10:00 start of data. **MABP**: 32, **SABP**: 41 **DABP**: 19, **HF**: 150 +/- 10, rc_{SO_2}: 74, SpO_2: 90.

10:30: **MABP** decreasing to a mean around 25. **DABP** decreasing in particular, increasing the difference between **SABP** and **DABP** to 35/17. **HF** is slowly rising to 170 +/- 10. rc_{SO_2} is decreasing gradually to 70 +/- 5. SpO_2 stays around 90% with some values more towards 80-85%. There are more severe dips in SpO_2 towards 60% together with bradycardia to 60/min, these last for 5 seconds.

10:50: **MABP** decreases further to a mean around 22, **HF** rises to a mean around 185, rc_{SO_2} decreases towards 55.

2.Name: Maarje Kersenbos, **Sex**: Female, **Age**: 31 weeks

Clinical Information: Apgar score 4/6/8(tube) at mediocrily respirating and moaning child with FiO2 of 80%. Admitted to the NICU, umbilical venous catheters, started amoxicillin/gentamycin. First blood pressure 60/40 (50). HF 170/min. Started pre- and post-ductal saturation measurements (re arm and foot). 60% O2.
History: Maartje, born at 31 weeks, born after long premature rupture of the membranes (4 days). Mother received celestone and antibiotics ante partum. Prevention of premature birth stopped. It turns out mother has a urinary tracht infection. Childbirth proceeded, eventually tachycardic CTG, smelling child.

3 hours post partum (30 minutes ago now): Sat 90% with 60% O2. Pre-postductal saturation difference at times 10% in the disadvantage for postductal saturation.

Value progression:
13:33: start of data. **MABP**: 32, **SABP**: 45, **DADP**: 30. **HF**: 175, rc_{SO_2}: 75, SpO_2: 90%. **CA**: 0.3

14:00: **MABP** is decreasing towards a mean around 30, **DABP** and **SABP** are decreasing with it. **HF** is increasing towards a mean around 183. rc_{SO_2} is decreading to a mean around 70. **CA** is increasing and stabilising at 0.7 +/- 0.1.

14:20: **MABP, DABP and SABP** are decreasing further towards a mean of 28. **HF** has risen further towards a mean around 190. rc_{SO_2} has decreased to a mean around 65.

3. Name: Mohamed Azriouil, **Sex**: Male, **Age**: 24 + 3/7 weeks.

Clinical Information: Temp 36,1. Complexion: grey-pink. First pH 2 uur post partum (now): 7.17/ 6,5/15/-12. Hb 12,5. Ventilated using HFO, MAP 9, amplitude 25, 10 Hz. 25% O2.
History: Mohammed, born after pregnancy of 24+3/7 weeks. Mother was G5P0 (repeated spontaneous miscarriages in VG). Braxton Hicks contractions and shortening of the cervix lenght since 22 weeks. Celestone at 23 + 5/7 weeks. Childbirth proceeded after stopping of tocolysis. Parents' first child. Due to agreed upon policy of active intervention Mohamed received a tube shortly after birth and received endotracheal surfactant after admission to NICU.

Currently we are at 2 hours post partum.

Value progression:
15:43: start of data **MABP**: 20, **SABP**: 25, **DABP**: 17, **HF**: 150 +/- 10, rc_{SO_2}: 50-60, SpO_2: 88-90, **CA**: 0.7 +/- 0.2 (this remains for the duration of the simulation)

16:33: gradually during the 50 minutes the **MABP, SABP and DABP** rise to a mean of 24-28. **HF** decreases to a mean of 140 +/- 10.

4. Name: Jan van der Vliet, **Sex**: Male, **Age**: 29 weeks

Clinical Information: Apgar score 7/8/8, Complexion: pink, Little urination (1 ml/kg/hour), capillairy refill time 3 sec. O2 demand reducing to 21% in the hours after increasing CPAP pressure.
History: Jan, born at 29 weeks, unprepared, parents' first child, spontaneous premature birth, 1290 grams. Started out well (apgar score 7/8/8), CPAP support. Admitted to the NICU due to O2 demand. CPAP increased in first hours to 11cm H2O. Afterwards 21-25% O2 with saturations of 90-93%. No umbilical venous catheters.

First day after birth.

Value progression:
15:33: start of data. **MABP**: 28, **SABP**: 41, **DABP**: 19, **HF**: 155 +/- 10, rc_{SO_2}: 88, SpO_2: 90-94, **CA**: 0.5, gradually rising.

16:03: **HF** is gradually rising towards 165 +/- 10, rc_{SO_2} is decreasing towards a mean of 78, **CA** has risen to 0.8 +/- 0.1.

16:23: **HF** has risen to a mean of 175 +/- 10, rc_{SO_2} is decreasing towards a mean of 78, **CA** has slightly fallen to 0.7 +/- 1.

A.3 Glossary

HF	Heart Frequency
MABP	Mean Arterial Blood Pressure
SABP	Systolic Arterial Blood Pressure
DABP	Diastolic Arterial Bloood Pressure
rc_{SO_2}	cerebral oxygenation as measured by NIRS
SpO_2	oxygen saturation of the lungs
CA	Cerebral Autoregulation
02	Oxygen
CPAP	Continuous positive airway pressure
FiO2	Fraction of inspired oxygen