

Early Recognition Of The Deteriorating Surgical Patient Using HealthPatch MD, A Wireless And Wearable Vital Signs Monitor

An Early Clinical Feasibility Study

ERIK HUIZINGA, BSc.

15-4-2017

A thesis submitted in fulfilment of the requirements for the degree

MASTER OF SCIENCE

in

TECHNICAL MEDICINE – MEDICAL SENSING & STIMULATION

Supervisors

Prof. dr. ir. H.J. Hermens

Drs. N.S. Cramer Bornemann

E. Groot Jebbink, MSc.

Prof. dr. C.J. Kalkman

Dr. T.J. Blokhuis

M.J.M. Breteler, MSc.

Table of contents

| | |
|---|-------|
| Chapter 1. Introduction | 1 |
| Chapter 2. HealthPatch MD by Vital Connect | 5 |
| Chapter 3. Simulation experiment with HealthPatch MD | 13 |
| Chapter 4. Measurements with HealthPatch MD on a clinical ward | 25 |
| Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPRESSON | 49 |
| Chapter 6. Predicting risk of adverse events | 65 |
| Chapter 7. Discussion and conclusion | 77 |
| Appendix A. Bland-Altman Analysis in a MATLAB implementation | I |
| Appendix B. Survival analysis of HealthPatch measurements | III |
| Appendix C. Plots of all participations' vital signs as measured by HealthPatch MD and Spacelabs XPRESSON | V |
| Appendix D. Tables | XIX |
| Appendix E. Honeycomb plots in a MATLAB implementation | XXIII |

Full table of contents

| | |
|--|----|
| Chapter 1. Introduction | 1 |
| §1.1 Study rationale | 1 |
| §1.2 Research questions and goals | 2 |
| Chapter 2. HealthPatch MD by Vital Connect | 5 |
| §2.1 Physical features | 5 |
| §2.2 HealthPatch MD signals | 5 |
| §2.2.1 Electrocardiogram | 5 |
| §2.2.2 Heart rate | 6 |
| §2.2.3 Acceleration | 6 |
| §2.2.4 Respiratory rate | 6 |
| §2.2.5 Other signals | 7 |
| §2.3 Battery life | 8 |
| §2.4 Memory | 8 |
| §2.5 The Vital Connect Platform | 8 |
| §2.5.1 Data flow from HealthPatch to relay | 8 |
| §2.5.2 Storage on the relay | 8 |
| §2.5.3 Transmission from relay to a remote data server | 9 |
| §2.5.4 Access on the remote data server | 10 |
| §2.6 Practical use of HealthPatch MD | 10 |
| §2.7 Validation status of HealthPatch MD | 10 |
| §2.8 VitalPatch | 11 |
| §2.9 Discussion | 11 |
| Chapter 3. Simulation experiment with HealthPatch MD | 13 |
| §3.1 Introduction | 13 |
| §3.2 Background | 13 |
| §3.2.1 HealthPatch measurement methods in comparison with typical bedside monitors | 13 |
| §3.3 Materials and methods | 15 |
| §3.3.1 Disassembly of HealthPatch MD | 16 |
| §3.3.2 Vital signs simulator | 16 |
| §3.3.3 Experimental setup | 17 |
| §3.3.4 Measurement protocol | 17 |
| §3.3.5 Analysis | 18 |
| §3.4 Results | 21 |
| §3.4.1 Measured signals | 21 |
| §3.4.2 Behaviour of the relay software and remote server | 22 |
| §3.5 Discussion | 23 |
| Chapter 4. Measurements with HealthPatch MD on a clinical ward | 25 |
| §4.1 Measurements on the surgical medium care ward | 25 |
| §4.1.1 Patients and setting | 25 |
| §4.1.2 Measurement protocol | 26 |
| §4.2 Data acquisition and storage | 26 |

| | |
|---|-------|
| §4.2.1 Data acquisition _____ | 26 |
| §4.2.2 Data storage _____ | 27 |
| §4.3 Data pre-processing _____ | 31 |
| §4.3.1 Reading raw data _____ | 32 |
| §4.3.2 Resampling and filtering _____ | 33 |
| §4.3.3 Synchronising HealthPatch and Spacelabs _____ | 34 |
| §4.3.4 Graphing participations for exploratory analysis _____ | 40 |
| §4.4 Analysis of validity of measurements _____ | 40 |
| §4.4.1 Agreement analysis _____ | 40 |
| §4.4.2 Agreement analyses of HealthPatch and Spacelabs _____ | 44 |
| §4.5 Discussion _____ | 45 |
| §4.5.1 Experiences with HealthPatch MD in practical measurements _____ | 45 |
| §4.5.2 The battery phenomenon _____ | 46 |
| §4.5.3 Annotating plots of participations or measurements _____ | 47 |
| §4.5.4 Bland-Altman analysis _____ | 47 |
| Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPREZZON _____ | 49 |
| §5.1 Measurement results _____ | 49 |
| §5.1.1 Results of the patient sample _____ | 49 |
| §5.2 Agreement analysis results _____ | 52 |
| §5.2.1 Agreement analysis of heart rate _____ | 53 |
| §5.2.2 Agreement analysis of respiratory rate _____ | 58 |
| §5.3 Discussion _____ | 59 |
| §5.3.1 Case study: exploration of vital signs of a deteriorating patient _____ | 59 |
| §5.3.2 Agreement analysis _____ | 60 |
| Chapter 6. Predicting risk of adverse events _____ | 65 |
| §6.1 Early Warning Scores _____ | 66 |
| §6.2 Methods to predict risk of deterioration using repeated measurements of vital signs _____ | 68 |
| §6.2.1 Matching the number of inputs and outputs _____ | 68 |
| §6.2.2 Feature extraction and exploration _____ | 68 |
| §6.2.3 Predictive modelling for time series _____ | 69 |
| §6.3 Discussion _____ | 75 |
| Chapter 7. Discussion and conclusion _____ | 77 |
| §7.1 Discussion _____ | 78 |
| §7.1.1 Wearable vital signs monitoring technology in literature _____ | 78 |
| §7.1.2 Strengths and limitations _____ | 78 |
| §7.1.3 Early Warning Scores may be used to estimate sample sizes for future studies _____ | 80 |
| §7.2 Conclusion _____ | 81 |
| Appendix A. Bland-Altman Analysis in a MATLAB implementation _____ | I |
| Appendix B. Survival analysis of HealthPatch measurements _____ | III |
| Appendix C. Plots of all participations' vital signs as measured by HealthPatch MD and Spacelabs XPREZZON _____ | V |
| Appendix D. Tables _____ | XIX |
| Appendix E. Honeycomb plots in a MATLAB implementation _____ | XXIII |

Chapter 1. Introduction

*Change is the law of life
and those who look only to the past or present
are certain to miss the future*
John Fitzgerald Kennedy

§1.1 Study rationale

Monitoring of vital signs is paramount for the recognition of deterioration of hospitalised patients and clinical decision support. Changes in vital signs often precede acute physiological deterioration due to complications [1–8], particularly after major surgery [9–13]. Unrecognised complications cause patient decline and if vital instability is not adequately managed patients may die (failure to rescue) [14]. Such critical adverse events are often measured with certain outcomes, e.g. activation of a medical emergency team (MET) or Rapid Response Team (RRT), unplanned readmission to an intensive care unit (ICU), unplanned reoperation or specific complications, for example sepsis. For industrialised countries, the World Health Organization (WHO) reports 3–22% of inpatients undergoing surgery suffer from major complications and failure to rescue occurs in up to 0.8% [15]. About half [15] of these postoperative adverse events are considered preventable [12,16,17]. The resulting prolonged admissions and increased number of readmissions and interventions aggravate the burden of healthcare for patients, caregivers, hospitals and society [18,19].

Vital signs monitoring is key for the recognition of physiological changes, and thus for their adequate management. For this and the aforementioned reasons, ICUs and other high care wards continuously monitor patients' vital signs, such as heart rate and respiratory rate. In contrast, regular hospital wards do not monitor vital signs as frequently. In general, monitoring on these wards is performed manually by nurses, often no more than once per work shift (i.e., every 8 hours). This may lead to missed or delayed recognition of critical changes in vital signs [9].

One reason for the lower monitoring frequency is that the risk of deterioration is considered to be lower in regular ward patients. This may not always be the case [7,20]. For example, Barwise et al. report that RRT activation was required in about 40–60 per 1000 discharges in 2012 in the Mayo Clinic [7]. Another study found that in cancer patients the four most common ICU admission characteristics were sepsis, cardiac insufficiency, respiratory insufficiency and cardiopulmonary resuscitation (CPR) [20]. For all these complications, it has been shown they may be avoidable if recognised in an early stage with monitoring of vital signs.

Other reasons for lower frequency and quality of vital signs monitoring are the increasing pressure to reduce admission times, growing health expenditure, trends to reduce nurse and doctor staffing and the increasing elderly population [17,21–24]. These factors contribute negatively to both medical and surgical patient safety. At night time patient safety reduces even further, because the number of nurses and doctors per patient typically is lower during night shifts compared to day and evening hours [23].

In other words, a contrasting situation exists in which patient safety can be enhanced through frequent monitoring of vital signs, but monitoring of deterioration and its management need to be improved. This is especially the case on regular wards, where resources are fewer for a larger number of patients. This decrease in ability to timely recognise deterioration from ICU to regular ward continues after hospital discharge. For example, at home or in nursing homes frequent patient monitoring is virtually non-existent. For all the aforementioned reasons, *continuous monitoring of vital signs for timely recognition of physiological deterioration of the surgical patient is of current interest.*

The recent trend of wearable and wireless smart technologies has resulted in unobtrusive devices that provide opportunities to monitor vital signs without the limitations of bulky, wired, bed-bound and manual monitoring. A wireless device for continuous and remote monitoring of vital signs has the potential to monitor patients during their hospital stay and after discharge. Patients with a wearable monitor can move freely while they recover, without the limitations of wired monitoring. Such a wearable system with embedded smart software may be able to (automatically) recognise patients at elevated risk of critical deterioration in an early stage. In general, any system to detect and react to physiological deterioration is called a physiological track-and-trigger system [25,26]. Currently, continuous monitoring of vital signs requires a caregiver to be physically nearby at all times. This paradigm will shift if a wearable physiological track-and-trigger system proves capable of early recognition of the deteriorating patient.

To implement wearable vital signs monitors their performance must be evaluated. Due to the underlying wireless technology, nearly continuous data acquisition and transmission become a possibility, independent of space and time. Various challenges arise from this situation. Signal continuity and validity must meet requirements for use in clinical practice. For example, one challenge is the reliability of the data transmission, which might be compromised because wireless connectivity is more fragile than classical wired connections. Another challenge is that both normal and abnormal physiology must be presented accurately without a high false alarm rate, as is common in current state-of-the-art bedside monitors [27,28]. A high false alarm rate causes alarm fatigue, which in turn can reduce patient safety.

Currently no standardised validation protocol exists for this type of monitoring technology. This is reflected by the vastly varying quality of assessments found in (grey) literature [29]. Often the conclusions are too optimistic; developments show promising results but are too technical and not (yet) feasible for clinical practice; validation assessments are performed in small and unrepresentative population samples; or quality of physiological measurements is low [30–32].

A number of medical-grade, CE marked and FDA approved wearable vital signs monitoring devices are currently available [31]. Despite approval for medical use, very little evidence exists on their added clinical value and whether the devices perform as specified by their manufacturers. The added clinical value, i.e. enhancement of patient safety and decrease of the burden of hospitalisation, has not been studied in well-defined clinical validation studies. This makes identification of wearables suitable for monitoring of vital signs a difficult task. Their added clinical value for the postoperative population and reimbursement are as of yet unclear.

The goals of this thesis research were to evaluate one wearable vital signs monitor through technical and clinical assessments: HealthPatch® MD (Vital Connect, Campbell, CA USA). This is a patch with integrated sensor technologies for the measurement of vital signs and wireless connectivity for the transmission of data. The research presented in this thesis aims to perform an early clinical feasibility study of this wearable. This is done in five following chapters, each of which having different goals and answering different research questions. After these chapters, the last chapter discusses and concludes this thesis.

§1.2 Research questions and goals

The main question being investigated is:

Can HealthPatch MD be used in clinical practice for early recognition of the deteriorating surgical patient?

This is described in multiple chapters, each with their own research questions and goals listed in Table 1.

Table 1 – This table lists the research questions and goals of this thesis and in which chapter they are described.

| Research questions | Goals | Chapter |
|---|--|---|
| What is HealthPatch MD, what does it do and how does it work? | To specify technically what HealthPatch MD features, functionality and specifications are as advertised | Chapter 2. HealthPatch MD by Vital Connect |
| How well does HealthPatch MD perform in simulated patients? | To obtain better understanding of the technical performance of HealthPatch MD through a simulation experiment | Chapter 3. Simulation experiment with HealthPatch MD |
| How are measurements with HealthPatch MD performed in clinical practice? | To obtain measurements with HealthPatch MD in a patient population at high risk of adverse events: postoperative patients on the surgical medium care ward in the UMCU | Chapter 4. Measurements with HealthPatch MD |
| How can data from HealthPatch MD be acquired? | To acquire raw measurement data systematically after performing measurements | Chapter 4. Measurements with HealthPatch MD |
| How can data from HealthPatch MD be stored and pre-processed for further analysis? | To store and pre-process the raw data while facilitating further analyses | Chapter 4. Measurements with HealthPatch MD on a clinical ward |
| What methods are available to measure agreement between measurement methods of time series? | To review and discuss various methods for agreement analysis between measurement methods of multiple time series of physiological data | Chapter 4. Measurements with HealthPatch MD on a clinical ward |
| What do heart rate and respiratory rate measurements by HealthPatch MD look like? | To describe a measurement with the corresponding patient's case. | Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPRESSON |
| How well do HealthPatch MD signals agree with the bedside reference standard? | To perform agreement analysis on HealthPatch MD data with the bedside reference monitor used in the UMCU | Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPRESSON |
| How is risk of deterioration in the surgical patient measured? | To summarise the methods used to objectively assess risk of physiological deterioration on the surgical medium care ward of the UMCU | Chapter 6. Predicting risk of adverse events |
| How can risk time series be predicted based on other time series? | To review and discuss prediction modelling methods for multiple physiological time series | Chapter 6. Predicting risk of adverse events |

Chapter 2. HealthPatch MD by Vital Connect

*If we don't know life
how can we know death?*
Confucius

This chapter describes the HealthPatch® MD (Vital Connect, Campbell, CA USA) system. The system's advertised specifications, features and functionality relevant to this thesis are described in the following section. The details in this chapter are used for reference in upcoming chapters.

§2.1 Physical features

HealthPatch MD is a wireless wearable vital signs sensor system in the form of an adhesive patch and a sensor module which is inserted into the patch upon activation. Figure 1 shows these two elements. Together they form the HealthPatch.

The adhesive patch has dimensions 115×40×7 mm³ and weighs 11 grams with the sensor module inserted [33]. The patch is available in two grades: gentle and active. The active grade has a stronger adhesive compared to the gentle grade, hence it is better suited for high-activity measurements (e.g. sports) and possibly sweating patients. However, the gentle grade has the advantage that it can be reapplied after temporary removal of the patch, whereas the active grade cannot be reapplied. The gentle grade patch may therefore be more practical in situations where it needs to be removed temporarily, e.g. during a medical intervention or radiological chest exams. On the other hand, the active grade patch is better suited for use in very wet conditions, such as showering. The adhesive patch consists of a zinc-air cell battery (the white circle in the middle part of the patch in the figure), some flexible electronics with a connection for the module, a thermistor and two electrocardiography (ECG) electrodes. The patch has a second grading system: medical and non-medical. The medical variant name has the MD suffix; the non-medical grade patch does not have any suffix. The MD grade adhesive patch supports the MD grade sensor module, which supports transmission of the ECG signal.

The sensor module is the second element in a HealthPatch. It contains more advanced circuitry, an embedded processor and additional sensor technology, along with a Bluetooth radio transceiver for wireless connection with a relay device (described later). The module is also available in an MD and non-MD variant and must be matched to the medical grade of the adhesive patch.

§2.2 HealthPatch MD signals

The HealthPatch measures the following physiological signals using the integrated sensors. Only those signals relevant in this research are described, all others are mentioned in one section below. The relevant sensing methods are illustrated in Figure 2. The values and algorithms in the following sections are as specified, advertised and researched by Vital Connect, but not as measured in this research [34–38].

§2.2.1 Electrocardiogram

A single-lead ECG is measured in μV using the two skin electrodes in the adhesive patch. It is sampled at 125 Hz. Note that ECG is measured independent of the medical grade of the HealthPatch. Both the medical and non-medical grade variants measure the ECG, but only the medical grade can transmit the ECG data to the relay device. ECG values range from -10 mV to +10 mV.



Figure 1 – The HealthPatch MD wearable studied in this thesis is a sensor system consisting of an adhesive patch (left) and a sensor module (right), which is inserted into the patch upon activation.

§2.2.2 Heart rate

Heart rate in beats per minute (bpm) is measured using the single-lead ECG. The algorithm is based on the R-R interval, which is the time between individual R peaks in the QRS complex. The reciprocal of the R-R interval signal is passed through a moving average filter of 10 heart beats to calculate a heart rate sample. Heart rate is sampled once per four seconds, i.e. 0.25 Hz. Measured heart rate values lie in the range 30–200 bpm with an accuracy of ± 5 bpm or 10%, whichever is greater.

§2.2.3 Acceleration

Tri-axial acceleration is measured using a MEMS¹, an accelerometer. It needs to be calibrated after initiation of a measurement. This can be done automatically by walking around for about a minute after application of the patch to the skin. A manual calibration can be performed by positioning the subject in an upright position, e.g. standing or sitting, and then forcing recalibration in that vertical position using the relay device's software.

§2.2.4 Respiratory rate

Respiratory rate in breaths per minute (brpm) is measured based on a weighted average of three individual algorithms to calculate respiratory rate. These algorithms are based on two of the aforementioned signals. One algorithm is based on the accelerometer signal, the other two are ECG-based:

1. Acceleration of the chest is measured using the accelerometer, from which the respiration signal can be obtained. The algorithm is based on the change in chest angle during respiration with respect to the direction of gravity.
2. Two ECG components are used:
 - a. Due to various physiological changes during respiration, the heart rate changes with inspiration and expiration. This phenomenon is called respiratory sinus ar-

¹ Microelectromechanical system

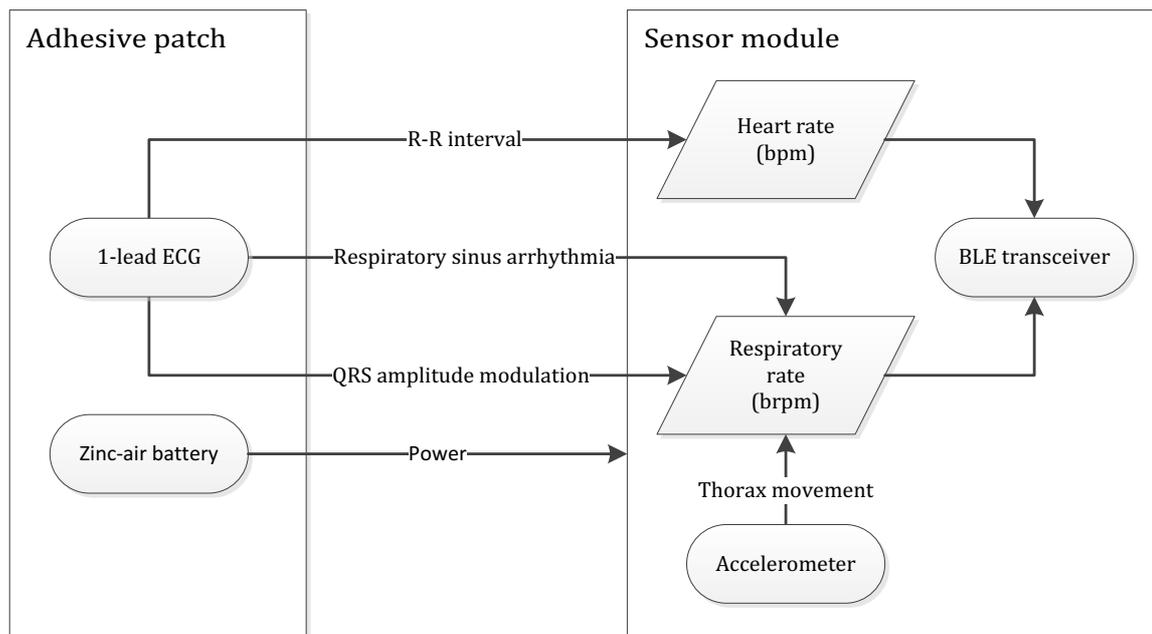


Figure 2 – Heart rate and respiratory rate are calculated in the HealthPatch sensor module based on two physical measurements performed by sensors within the adhesive patch and the sensor module. Bpm: beats per minute, brpm: breaths per minute, BLE: Bluetooth Low Energy.

rhythmia (RSA) [39]. During inspiration, heart rate increases; during expiration, heart rate decreases. With this variability, the R-R interval varies as well. The R-R interval signal over time can therefore be used to estimate respiratory rate.

- b. Due to the changes in the cardiac axis with respect to the chest wall during a respiration cycle, the ECG QRS complex amplitude varies (modulation). This modulation signal contains the respiratory rate.

In the respiratory rate algorithms of the HealthPatch it is assumed the rate lies in the ranges 6–42 brpm. A custom 45-second moving average window is used to obtain three respiratory rate signals. The three signals are then averaged using a quality measure to ensure the best signals are weighted more heavily towards the final respiratory rate signal. In this weighting, the accelerometer signal is the only component used if its quality exceeds a certain threshold and if the respiratory rate is greater than half the heart rate. This is because the ECG-derived respiratory rate signals are sampled by every heart beat and thus they can only measure respiratory rate up to half the heart rate (due to the Nyquist-Shannon sampling theorem [40]), otherwise cardiac aliasing occurs.

The final respiratory rate signal is sampled once per four seconds, i.e. 0.25 Hz. Values lie in the range 4–42 brpm with an accuracy of ± 5 brpm.

§2.2.5 Other signals

Various other signals—both physiological and technical—are measured by HealthPatch, but are not of interest in this research. Available physical/physiological signals are skin temperature, heart rate variability, fall detection, step count, posture (laying down, sitting, standing, walking, running, leaning back), R-R interval, stress level, energy expenditure (both total and rate), activity (both total and rate). Available technical signals are patch skin application status, contact impedance, battery level and sensor memory level.

§2.3 Battery life

As specified by Vital Connect, HealthPatch has an embedded zinc-air cell battery. The battery life is advertised to be at least 72 hours (3 days) when ECG transmission is enabled and 96 hours (4 days) when ECG transmission is disabled. Transmission can be switched on or off using the relay device's software.

§2.4 Memory

Because any wireless connection is fragile by nature, transmission of signals from HealthPatch to relay may be interrupted. During an interruption HealthPatch continues to measure and stores the acquired data in memory. This memory is limited depending on the number of signals requested for transmission, e.g. transmission of the default signals plus ECG data requires more memory than just the default signals. The default signals are heart rate, R-R interval, respiratory rate, skin temperature, step count, posture and fall detection.

Vital Connect specifies no less than 10 hours of data can be stored in a HealthPatch, but typically 18 hours with ECG data and 40 hours without can be stored. After the connection with the relay device has been re-established, the stored data is transmitted piecewise to the relay device.

§2.5 The Vital Connect Platform

The Vital Connect Platform is what Vital Connect calls the network formed by a HealthPatch, a relay device and an optional data server. This section describes these network elements and the connections between them. An overview of the platform is illustrated in a data flow diagram in Figure 3.

§2.5.1 Data flow from HealthPatch to relay

The HealthPatch collects and stores data for transmission. This is done via the Bluetooth Low Energy (BLE) protocol. BLE is relatively energy-efficient and nowadays generally available in wireless computers and other wireless connected devices, e.g. smartphones and tablets. The typical connection range of BLE devices is up to 10 metres within line of sight.

§2.5.2 Storage on the relay

The receiving BLE device is called the relay device, because it is a transmission hub in the Vital Connect Platform. The relay device uses its BLE transceiver to communicate with a HealthPatch. For this to work it needs software that can work with Vital Connect's specific protocols and data

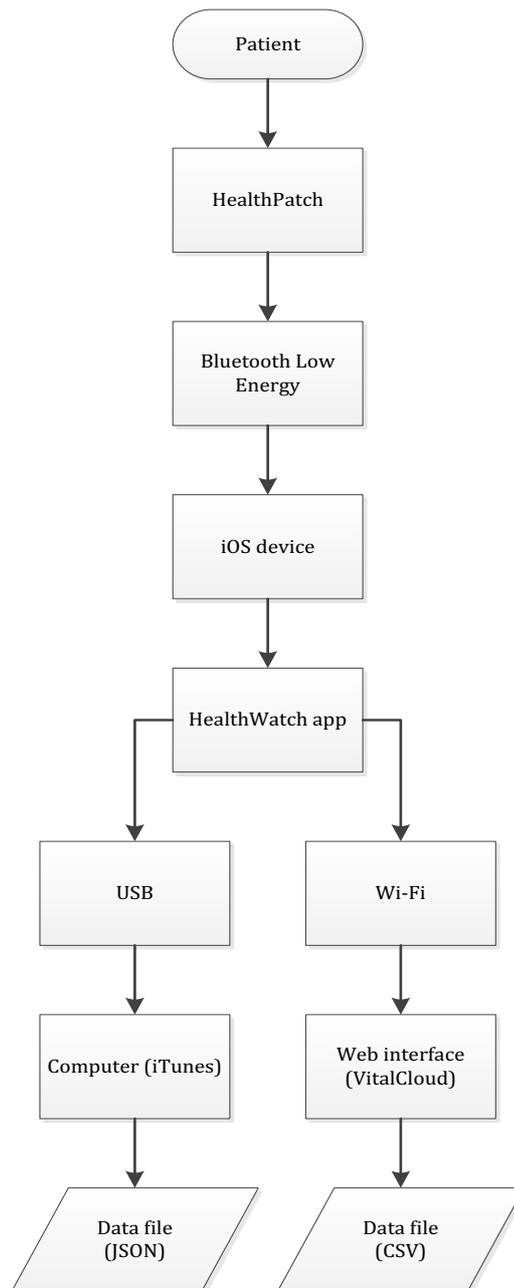


Figure 3 – The data flow diagram from patient through HealthPatch MD results in either a JSON or a CSV file of raw vital signs data.

formats. Vital Connect provides an application that does this for iOS (Apple Inc., Cupertino, CA USA) called HealthWatch. They are currently developing an application for Android (Google Inc., Mountain View, CA USA) and they provide an SDK² to develop custom relay software for iOS, Android or embedded devices.

The HealthWatch iOS app provides a monitoring interface where current and historical data can be watched. Some basic monitoring functionality is available, such as warnings (e.g. bad skin contact) and notifications (e.g. threshold alarms).

The relay device stores, processes and optionally transmits the data to a remote data server. Storage on the relay device is managed by the HealthWatch application. It does this in the JSON³ data language. This data format knows widespread use in programming and data interfaces. It allows to store multiple so-called objects or arrays with one or multiple values, which can be objects or arrays themselves. For example, a HealthPatch JSON file could contain a heart rate object consisting of an array of multiple values, being the bpm samples measured by the patch. For details on the actual JSON file format as stored by the HealthWatch application for iOS, the reader is referred to the Vital Connect developer resources [34] and the HealthPatch Clinical Resource Guide [35].

Transmission of data from relay to a remote data server is optional, because the relay device itself can be used as a local (as in: being near the patient) database. All the measurement data is stored on the relay device, so a remote (as in: not necessarily being near the patient) data server is not required. Using the relay device as local data storage without transmission to a remote data server can be preferred for various reasons, e.g. security of the remote connection. Vital Connect calls this 'serverless mode'.

To acquire the locally stored data the caregiver must be physically near the relay device. The HealthWatch JSON files can be obtained from the iOS relay device using Apple iTunes, which is computer software to connect with iOS devices.

§2.5.3 Transmission from relay to a remote data server

In contrast to serverless mode, as described in the previous section, the relay device can be configured to transmit data to a remote server. This functionality is easily used if the relay device, e.g. a smartphone, has other network connections available to the internet or an intranet. This allows for the local data to be transmitted easily to a remote server. Transmission to a remote server allows for a HealthPatch measurement to take place without the requirement of a caregiver being physically present near the relay device. If the caregiver has access to the remote server, the patient may roam freely as long as the HealthPatch is connected with the relay device, which in turn is connected with the remote data server. Such a network completely different from the continuous monitoring network paradigm as currently found in acute care settings in hospitals. Hence, the Vital Connect platform (and similar) has the potential to become a disruptive innovation when implemented in clinical practice.

Transmission to the remote data server takes place in an almost real-time fashion. A data packet is typically sent every few seconds and transmission normally another few. The HealthWatch iOS application provided by Vital Connect can be configured to transmit data to the remote data server as soon as it has been received from a patch, but regular intervals can be configured to reduce the power consumption of the relay device software. This would be beneficial for a relay device depending on a battery, e.g. a smartphone, but would increase the transmission time of the packet.

² Software Development Kit

³ JavaScript Object Notation, www.json.org

§2.5.4 Access on the remote data server

The remote data server can for example be a web interface accessed through a web browser. This is exactly what Vital Connect provides. For demonstration purposes, they have developed a web interface called HealthWatch⁴ (like the iOS application), where a live connection (notwithstanding connection latency) can be made remotely with a relay device while performing measurements. HealthWatch also provides heart rate variability (HRV) details, which are calculated on the remote server and unavailable on the relay device. The web interface also provides a historical data overview. Historical data can be downloaded in comma-separated-values (CSV) format, which is a less powerful but easier to understand data file format compared to JSON. The HealthWatch CSV files contain less technical information about the measurements, but enough for this research: heart rate and respiratory rate are available in these files.

§2.6 Practical use of HealthPatch MD

To use HealthPatch MD, one needs an adhesive patch (medical or non-medical), a compatible (medical or non-medical) sensor module and a relay device. The HealthPatch is activated upon insertion of the module into the patch. One of three available application sites can be chosen based on researcher, caregiver or subject preference. The three locations are:

1. at a 45° angle on the left m. pectoralis major
2. vertically on the sternum
3. over the left midclavicular line near the 6th and 7th intercostal spaces

The location of preference is 1, because of the easiest accessibility and comfort for the subject. Vital Connect suggests women to use locations 1 or 3 and obese people to use locations 1 or 2. Excessive chest hair can be removed and the skin can be cleaned with alcohol prior to application for enhanced skin contact and reduced long-term irritation of the skin.

Upon removal of a HealthPatch, the sensor module is removed from the adhesive patch. The patch can be recycled. The sensor module can be cleaned with an alcohol wipe and stored for reuse in another patch.

§2.7 Validation status of HealthPatch MD

Small validation studies in limited populations and settings have been performed with HealthPatch. Both PubMed⁵ and Scopus⁶ list the same two results for 'HealthPatch' performed by Vital Connect [41,42] (date accessed: 27 September 2016). Other studies have been conducted by Vital Connect [36–38]. These five studies have in common that the included subjects are not very representative of a hospital population and the observations were done in controlled settings instead of spontaneous.

The first study [36] included 15 subjects aged 63–79 without specified medical conditions. They were asked to breathe spontaneously for four minutes and perform metronome breathing to simulate various respiratory rates (12–24 brpm), each rate for three minutes. The quality outcome was measured as mean absolute error (MAE) compared to nasal cannula capnography.

The second study [37] again included 15 subjects aged 63–79 without specified medical conditions. It is mentioned this is the same sample as used in the first study. Furthermore, 10 subjects aged 18–29 were included. In this study the outcomes of interest were heart rate, HRV, posture detection, step counting and fall detection. The younger subjects performed the simulated activities of daily living (ADL), the older subjects performed the protocol of the first study, which may possibly be the exact same experiments. Again, the measure of quality was MAE and

⁴ accessible at <https://healthwatch.vitalconnect.com/>

⁵ <https://www.ncbi.nlm.nih.gov/pubmed/>

⁶ <https://www.scopus.com/>

furthermore root-mean-square error (RMSE) with respect to an unspecified Actiheart device and nasal cannula capnography.

The third study [38] included 76 subjects aged 59–85 with various medical histories, but who were otherwise healthy elderly volunteers. They wore HealthPatches for 50 days. Various practical usability measures were assessed during the study. However, only on days 1 and 4 the validation measurements were done similar to the previous studies: simulated ADL and metronome breathing were used to compare heart rate and respiratory rate with reference devices using MAE as the performance metric.

The fourth and fifth studies [41,42] evaluate specific applications of the HealthPatch. These studies are demonstrations of the patch applied for sleep apnoea screening and psychological acute stress monitoring. Since they do not validate heart rate or respiratory rate, they are not further evaluated here.

The FDA 510(k) clearance for marketing as a medical device for HealthPatch MD can be found in three FDA documents [43–45]. The first FDA document does not specify on which study the validation was based. The second document mentions a study with 76 participants who wore the patch over 50 days, which may correspond to the study mentioned above with the same number of participants and methods [38]. The third FDA document mentions the same previous experiments were deemed sufficient. It is noteworthy the VitalPatch, a new device by Vital Connect (see also §2.8) similar to HealthPatch MD, is mentioned in the last FDA document.

In general, the validation measure used in Vital Connect’s studies was MAE. It provides insight in how well the mean error between two measurement methods is, but does not tell how large the individual errors are. However, the latter is of interest in validation of continuous heart rate and respiratory rate monitoring, because individual observations must not deviate much from the true values (or an accepted reference method). Even if the MAE is low, individual errors may be large. The MAEs found in the studies were less than 3 bpm for heart rate and less than 3 brpm for respiratory rate. This number is acceptable for heart rate, but it is not for respiratory rate if a clinician wants to distinguish a tachy- or bradypnoea from a normal breathing rate. Nonetheless, Vital Connect advertises heart rate, respiratory rate and other signals measured by HealthPatch as validated [35]. More specifically, they mention respiratory rate accuracy to be ± 3 brpm in the range of 10–30 brpm, validated by clinical studies. The accuracy is ± 1.5 brpm for the entire range of 4–42 brpm, validated by simulation studies. This is more accurate than the ± 5 brpm mentioned in §2.2.4, but the numbers here are officially only validated for use in the USA and its territories, which refers to the FDA-clearance.

§2.8 VitalPatch

A successor to the HealthPatch MD has been introduced in March of 2016: VitalPatch [46]. It is the next iteration of the HealthPatch based on a collaboration with Philips. The hardware of the VitalPatch has an integrated disposable sensor module instead of a separate reusable one. At this time it is unclear to what extent the VitalPatch differs from HealthPatch MD, other than the aforementioned integrated sensor module, because all other specifications are identical [33,47]. Vital Connect may have plans for further development of their software platform, as they now collaborate with a ‘Digital Health Platform’ called BePatient⁷ and plan to further develop their mobile applications (announced through communication with their technical support).

§2.9 Discussion

This chapter discussed various features and specifications of HealthPatch MD, as marketed by Vital Connect. This is important information to have for the methods and analyses in the upcoming

⁷ <https://www.bepatient.com/>

ing chapters. Sections §2.1–§2.6 described the theoretical and practical use of the patch. Sections §2.7 and §2.8 reviewed the current status and future developments of the patch.

It is noteworthy that the validation studies performed to obtain FDA 510(k) clearance for marketing as a medical device were performed with few studies of limited methodological value. The patient samples, the experimental designs and analysis methods used in those studies may be improved, which may result in higher quality evidence about the validity of measurements by HealthPatch MD. Evidence for being acceptable for continuous and reliable monitoring of heart rate and clinical decision support has not yet been demonstrated.

Chapter 3. Simulation experiment with HealthPatch MD

Intelligence is the ability to adapt to change

Stephen Hawking

§3.1 Introduction

To obtain a better understanding of the technical properties, features and reliability of the HealthPatch MD we conducted a simulation experiment before any other experiments with data from measurements performed on patients. This chapter describes the steps taken in this experiment. The goal was to obtain an in-depth understanding of the behaviour of the HealthPatch signals and data when various normal and abnormal heart rate and respiratory rate patterns are presented to the device. It is useful to know this simulated behaviour before patient measurements are performed in clinical practice. After all, some heart rate and respiratory rate patterns are relatively rare in clinical practice, but can be easily simulated, e.g. asystole or ventricular fibrillation (VF). Furthermore, the physical principles on which wearable vital signs monitors base their measurements can differ substantially [31]. A simulation experiment can shed light on the inner workings of these devices.

This chapter presents the physical and physiological background, methods and results of the simulation experiment with HealthPatch MD. The methods consist of a disassembly of the patch, connection of the inner electronics to a vital signs simulator, performing simulations of various heart and respiratory rhythms and the analysis of the results.

§3.2 Background

To connect a HealthPatch to a patient simulator, which is ordinarily used to test bedside vital signs monitors, it is essential to study its internal electronics. The HealthPatch does not measure all vital signs based on the same physical and physiological principles as a bedside monitor. This section describes the HealthPatch electronics and their use.

§3.2.1 HealthPatch measurement methods in comparison with typical bedside monitors

ECG, heart rhythm and heart rate

Regular bedside monitors measure various ECG leads derived from various available electrode locations. Commonly used locations are the extremity electrodes, which are right arm (RA), left arm (LA), left leg (LL, also called foot (F)) and right leg (RL, also called neutral (N)). Bedside monitors use three (RA+LA+LL) or more electrodes. In contrast, HealthPatch uses only two electrodes. This raises a question: can HealthPatch measure a signal from a simulator that assumes three or more available leads? Also, how does HealthPatch cope with the absent reference electrode (RL)? In theory, the simulator could apply too much current to the HealthPatch if the reference potential is too different from the HealthPatch 'ground' or zero potential (the average between its two electrodes). This is not expected to be the case, however, because the simulator generates signals of physiological order of magnitude, which is precisely what HealthPatch is designed for.

Another concern is the amplitude of the simulated signals and the amplitude of the measured signals, which depends on electrode separation distance. The HealthPatch electrodes are about 10 cm apart, whereas regular bedside monitors electrodes are much further apart. For example, the RA and LL electrodes placed on shoulder and hip can exceed 80 cm in distance. A

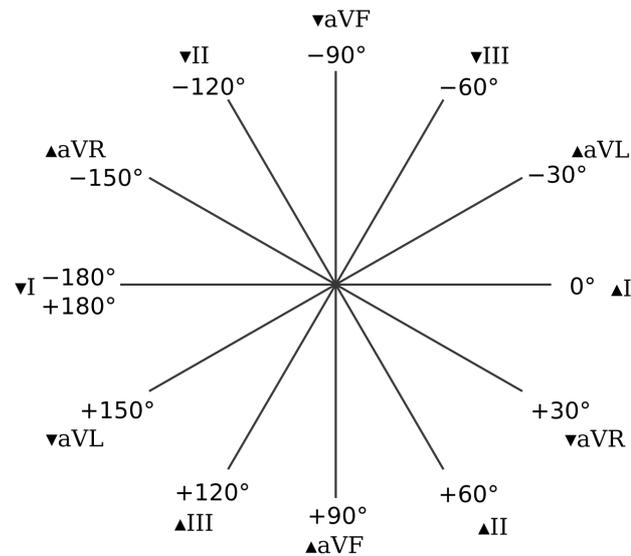


Figure 4 – The hexaxial reference system indicates the angular direction of ECG leads I, II, III, aVR, aVL and aVF. Note that the leads' positive (▲) and negative (▼) directions are indicated, hence the twelve directions of the six leads. Image reused under license from Wikipedia Commons [95].

relatively shorter distance between electrodes means a relatively smaller potential difference is measured [48]. This is easily understood if we take the limit of the distance approaching zero: if it were zero, no potential difference would be measured as both electrodes measure the exact same signal. If it were very close to zero, the potential difference would increase, but still be small. Thus, a larger distance creates the ability to measure greater potential differences (although too large distances would measure too much noise instead of the signal of interest). The concern in this is that the simulator generates signals, assuming they are measured on the extremities. Because HealthPatch measures them much more centrally, nearly on top of the heart, its electrodes measure different signals in practice. It is unknown how this affects the HealthPatch measurements.

The shorter distance between the HealthPatch electrodes also has an advantage. Noise is to be rejected from the ECG signal, which can be partially achieved if both electrodes measure the same noise (common mode). The differential signal then rejects the common mode (differential noise reduction). To achieve perfect differential noise reduction, both electrodes must measure the exact same noise. This can be achieved by placing the electrodes close to each other. In the same reasoning as above, if the distance were zero between the electrodes, they would measure the exact same noise, i.e. the common mode contains all the noise. The differential then would perfectly reject the noise. Trivially, the common mode then also contains all the signal of interest, which results in total rejection of the ECG. With increasing distance between the electrodes, the measured noise changes, hence the common mode rejection does not remove all the noise from the differential signal. If the distance is not too large, a large fraction of the noise will still be rejected by the differential signal.

HealthPatch achieves most of its noise reduction likely through common mode rejection, because the electrodes are located quite closely. In contrast, a regular bedside monitor can use a reference electrode to enhance its common mode rejection ratio. The measurement electrodes then need not be closely located.

The ECG leads measured by bedside monitors are calculated from the electrode locations on the skin. They indicate the electric potential direction of the heart vector during the cardiac cycle. The extremity ECG—three or four electrodes on positions RA, LA, RL, and LL—can measure six leads. They are I, II, III, aVR, aVL and aVF. The angular direction of these leads is illustrated in

Figure 4. Both the positive and negative directions are drawn, hence twelve instead of six directions. This is called the hexaxial reference system, wherein the centre of the system is located in the heart. Taking different combinations of differential potentials from the electrodes results in the hexaxial system leads. For example, differencing LL and RA results in lead II. Leads II and minus aVR point approximately in the same direction as the HealthPatch applied on position 1 (at a 45° angle on the left m. pectoralis major, see also §2.6). Hence, HealthPatch approximately measures the heart vector in the direction of leads II and minus aVR.

It is expected HealthPatch can accurately measure the simulated heart rate based on its single lead ECG. The HealthWatch iOS application and web interface do not provide automatic heart rhythm pattern classification, but simulated heart rhythm morphologies are expected to be recognised manually.

Respiration and respiratory rate

Typical bedside vital signs monitors offer a patient-friendly respiration measurement method based on chest impedance variability. During the respiratory cycle, inspiration causes the chest to expand, which changes the volume and shape of the conducting tissue between ECG electrodes. Likewise, expiration causes the opposite changes. The result is a changing distance between the ECG electrodes, which has an effect on the impedance between them. Over time, the respiratory cycle generates a signal in the chest impedance from which the respiratory rate can be estimated.

However, HealthPatch does not estimate respiratory rate based on chest impedance variability. Its methods, based on accelerometry and ECG features, are described in §2.2.4. This might pose a challenge in this simulation experiment, because the respiratory rate simulator is made for monitors that measure chest impedance variations. If the simulator would realistically generate an ECG containing QRS complex amplitude modulation and RSA, HealthPatch should be able to estimate the respiratory rate. Otherwise it is not expected the patch is able to measure the respiratory rate accurately, if at all.

§3.3 Materials and methods

One HealthPatch MD was used for measurement of simulated heart and respiration signals. It was opened and connected to a simulator. Measurements were performed in the course of nearly 1½ hour, in which various heart and respiration rhythms were generated by the simulator. The results were analysed and are presented after this section.

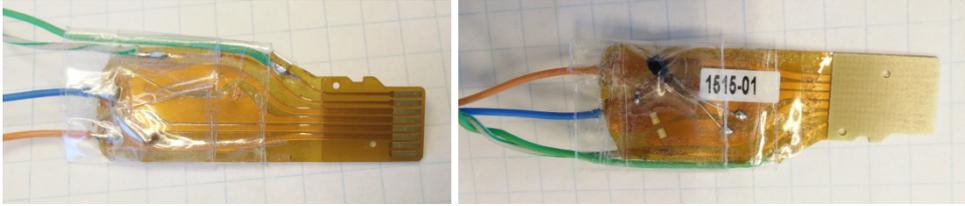


Figure 5 – The internal electronics of a HealthPatch MD are shown after removal of the surrounding materials, ECG electrodes and battery. Additional cables were applied: the green and green-white (upper two in left panel) wires connect to the ECG electrodes; the blue (middle) and red (bottom in left panel) wires connect to the power source. Blue is negative, red is positive. To the right end of the strip a connection similar to a micro-SD card can be seen. The left panel shows the front side of the circuitry, the right panel shows the backside, i.e. the side faced towards the patient’s skin. In the right panel, the thermistor is visible as a black component soldered to two electrodes. The photograph was taken on 5×5mm² grid paper for reference.

§3.3.1 Disassembly of HealthPatch MD

The HealthPatch was opened to reveal its electronics, it was studied and connected to the simulator (described in the next section). The disassembly revealed a small, flexible strip with some electronic connections to the two ECG electrodes and the zinc-air cell battery. These connections are on one side of the strip, on the other side the connection for the sensor module was found. It looks very similar to a micro-SD card connection in shape and positioning of the connections. Figure 5 shows the strip’s front and back. The back shows the thermistor for skin temperature measurements. Wires were soldered where the battery and ECG electrodes previously connected. Through these wires both power and signal could be delivered to the HealthPatch sensor module. On the sensor module side (the micro-SD-alike connection), the eight connections have the following functions:

1. Connects to the RA skin electrode.
2. Connects to the LL, RL or LA skin electrode, depending on orientation (see also §2.6).
3. Connects to the thermistor positive pole.
4. Connects to the negative pole of the power source, i.e. reference. It also connects to the thermistor negative pole.
5. Not connected.
6. Connects to the positive pole of the power source.
7. Not connected.
8. Not connected.

§3.3.2 Vital signs simulator

Simulation of heart and respiration signals was done with a ProSim 4 by Fluke (Fluke corp., Everett, WA USA) [49]. This simulator can generate various physiological signals. Of these, the following vital signs and specifications are relevant to this simulation experiment.

A three or more (up to twelve) lead ECG can be connected. Sinus rhythm (SR) can be generated at various frequencies in the range 30–320 bpm with an accuracy of 1% of the rate. Arrhythmias that can be simulated are atrial fibrillation (AFib, coarse and fine), premature ventricular contraction (PVC, left ventricular), ventricular tachycardia (VT, 160 bpm or 200 bpm), ventricular fibrillation (VFib, coarse and fine), second or third degree atrioventricular (AV) block and asystole. Some artificial signals, such as a square wave, can be generated as well. The amplitude of the ECG signal can be set. Respiration can be simulated at 0 brpm, or 10–100 brpm in 10 brpm increments. Respiration is simulated as variation in chest impedance.



Figure 6 – The experimental setup overview shows the Fluke ProSim 4 patient simulator (front left) the disassembled HealthPatch (centre) and the iPad mini (right).

§3.3.3 Experimental setup

The connection to the simulator was realised by attaching regular ECG electrode clamps to the wires connected to the HealthPatch. These clamps can readily be applied to the ProSim 4 simulator. The connection to a power source was realised via general wire clamps. The power source was set to generate a 1.4VDC potential, which is the nominal voltage of the zinc-air cell battery normally incorporated in a HealthPatch.

At the start of the measurements, the HealthPatch MD was activated by inserting a corresponding sensor module. A BLE connection was established with an Apple iPad mini™ (model A1455) as the relay, running the HealthWatch application. The iPad was connected to the HealthWatch remote data server through Wi-Fi. A computer showed the HealthWatch web interface during the course of the measurements. The ProSim 4 patient simulator was set to generate SR at 80 bpm and respiration at 20 brpm. The HealthPatch ECG electrodes were connected to the simulator's RA and RL electrodes. The setup was successful in one try: vital signs were seen immediately after the measurement commenced. Figure 6 shows an overview of the experimental setup.

The relay software was set to upload every 5 seconds to the HealthWatch web interface, which enabled nearly live online monitoring of the measurements.

§3.3.4 Measurement protocol

Every few minutes, the ProSim 4 was set to simulate different heart rhythms and respiration signals. The measurement protocol is listed in Table 2 for every time a change was made. Initially, measurements were performed on lead RL–RA (vertically oriented, i.e. $+90^\circ$ in the hexaxial reference system), but because HealthPatch MD approximately measures lead II, a switch to LL–RA was made on minute 28 of the measurement. A switch to another patient simulator with similar technical specifications was made on minute 65, because this allowed for easier amplitude changes. These changes were made to test the ECG sensitivity of the HealthPatch. This was done by reducing the generated signal amplitude with stepwise decrements until the HealthPatch ceased to measure a signal.

Furthermore, the iOS HealthWatch application was tested by using some of its functions. A number of warning and alarm thresholds was set to test the reactions of the monitor interfaces (both application and web interface) to the various rhythms.

§3.3.5 Analysis

Data was analysed by exploration. A manual comparison between simulated and measured heart rate and respiratory rate was done, because the true values are only coarsely known. For example, the true heart rate during coarse or fine AFib as generated by the simulator is unknown, hence no comparison can be made with the HealthPatch. For such a comparison, a Bland-Altman analysis could be carried out, for which the simulator heart rate and respiratory rate time series would need to be available. All analyses were done with raw data obtained from the iPad without any pre-processing.

Table 2 – The simulation measurement protocol is tabulated for every point in time where a different heart rhythm or respiratory rate was set. SR: sinus rhythm. Bpm: beats per minute. Brpm: breaths per minute. PVC: premature ventricular complex. VT: ventricular tachycardia. RL: right leg. LL: left leg. AFib: atrial fibrillation. VFib: ventricular fibrillation. AV: atrioventricular.

| Minute | Heart rhythm | Respiratory rate |
|--------|--------------------------------------|--------------------|
| 0–13 | SR 80 bpm | – |
| 0–31 | – | 20 brpm |
| 14–18 | Asystole | – |
| 19–20 | SR 80 bpm | – |
| 20–23 | PVC | – |
| 24–26 | SR 80 bpm | – |
| 27–29 | VT 200 bpm | – |
| 28 | Electrode repositioned from RL to LL | – |
| 30–35 | Bradycardia 30 bpm | – |
| 32 | – | Bradypnoea 10 brpm |
| 33–end | – | Tachypnoea 30 brpm |
| 36–37 | Coarse AFib | – |
| 38–40 | Fine AFib | – |
| 41–42 | Asystole | – |
| 43 | SR 60 bpm | – |
| 44–47 | Bradycardia 30 bpm | – |
| 48–52 | Coarse VFib | – |
| 53–56 | Third degree AV Block | – |
| 57–59 | Second degree AV Block | – |
| 60–64 | SR 80 bpm | – |
| 65 | Patient simulator change | – |
| 66–74 | Signal amplitude tests | – |
| 69 | VT 180 bpm | – |
| 72–end | SR 80 bpm | – |

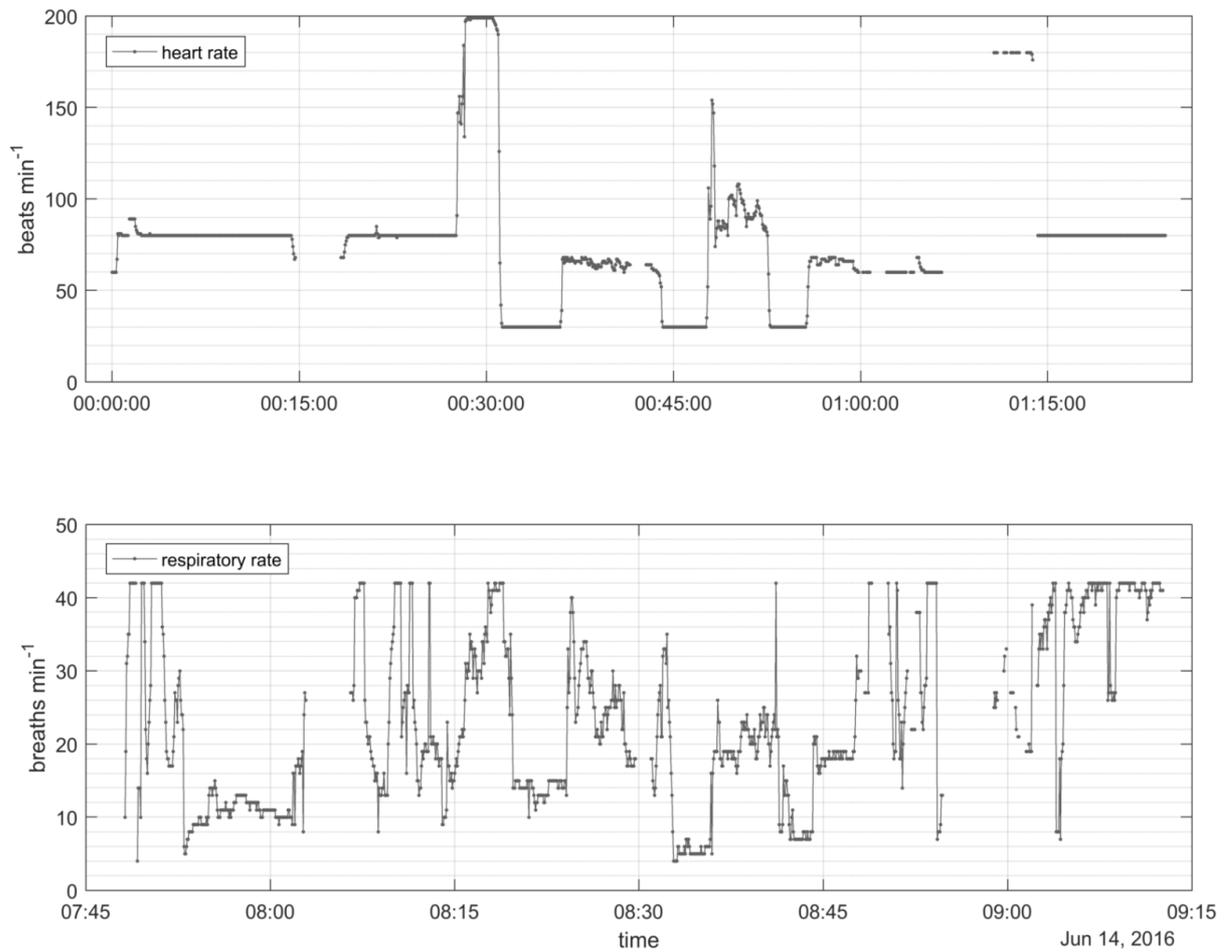


Figure 7 – The two graphs in this figure show an overview of the HealthPatch data throughout the simulation experiment. The upper panel shows heart rate in beats per minute vs. time, the lower panel shows respiratory rate in breaths per minute vs. time.

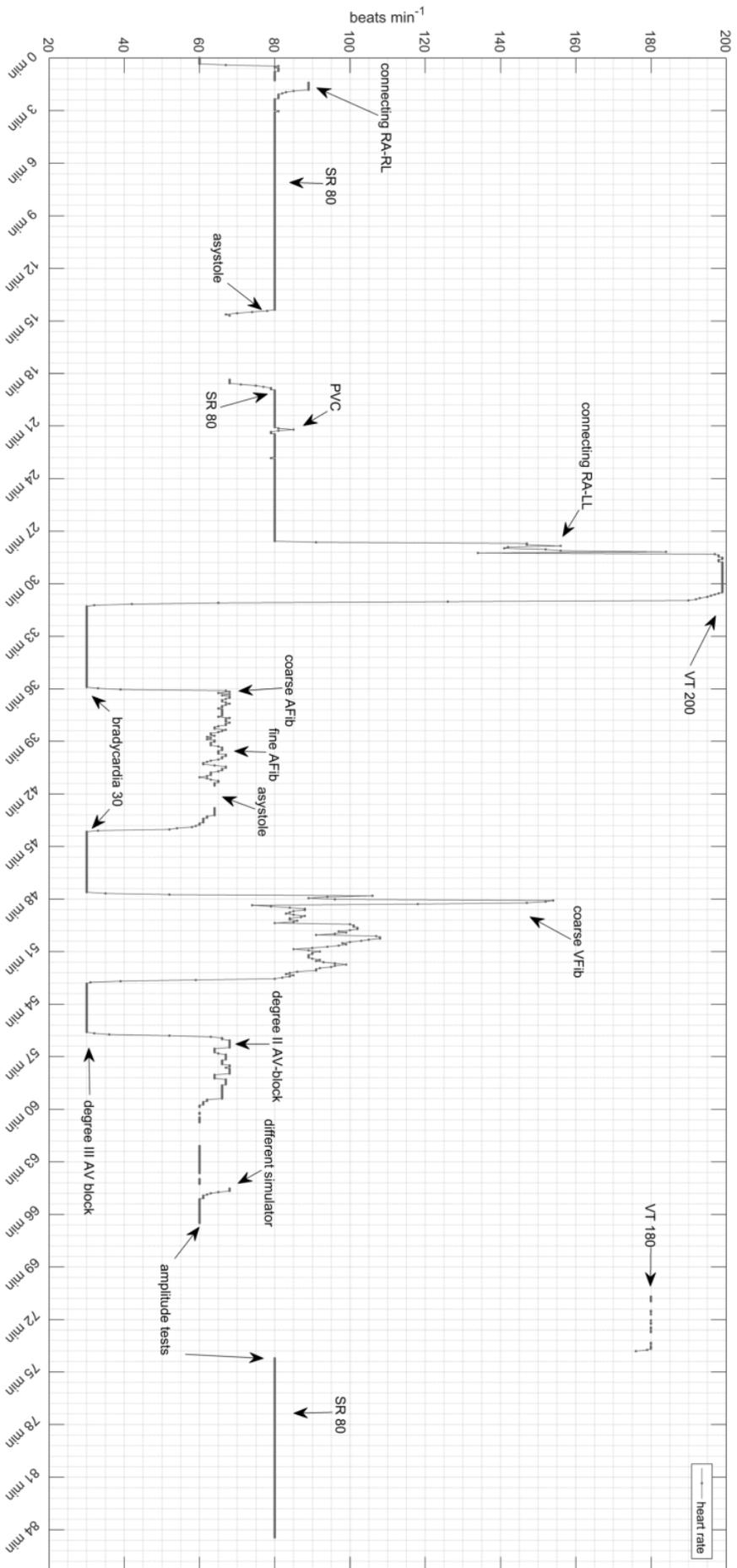


Figure 8 – The HealthPatch heart rate in this graph is annotated with the measurement protocol found in Table 2.

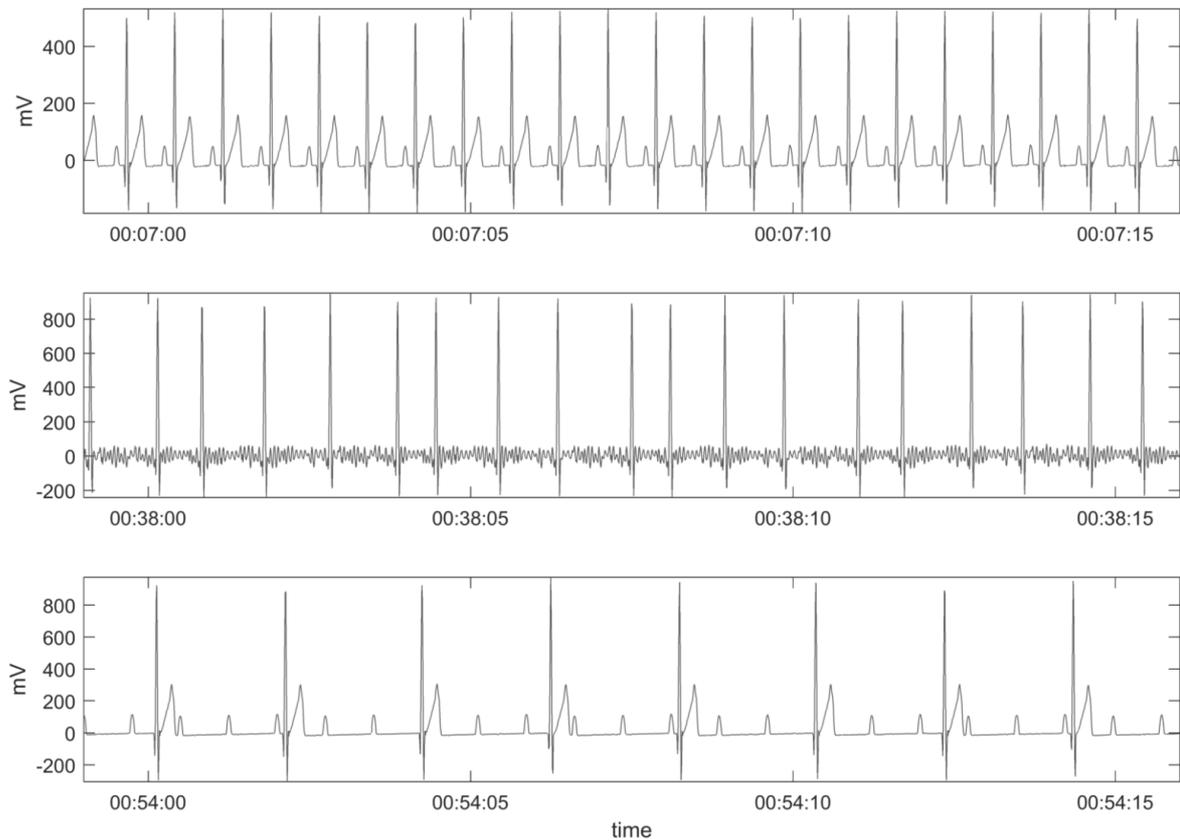


Figure 9 – Various simulated ECG recordings by HealthPatch MD are shown in the three panels. The upper panel shows a sinus rhythm, the middle panel shows fine atrial fibrillation, the lower panel a third-degree atrioventricular block.

§3.4 Results

§3.4.1 Measured signals

Heart rate and respiratory rate as measured by HealthPatch MD throughout the simulation protocol are graphed in Figure 7.

ECG, heart rhythm and heart rate

It was found the HealthPatch ECG signal uses adaptive filtering algorithms to obtain a maximally noise-free signal, which was seen on the HealthWatch web interface. This was seen in the first minute of the measurements, directly after connecting the patient simulator.

Heart rate and a number of arrhythmias are measured accurately by the HealthPatch. Figure 8 shows the heart rate measurement with the protocol annotated in the graph. Some abnormal heart rhythm patterns are easy to see and recognise in the HealthPatch ECG, but do not result in a characteristic change in measured heart rate. For example, a third-degree AV block leads to a ventricular bradycardia, but is not distinguishable from a regular bradycardia or another conductivity cardiopathy based on heart rate alone. Both simulated tachycardia and bradycardia are precisely followed by the HealthPatch: 200 bpm is measured as 199 bpm. This could be a measurement error by HealthPatch or a simulation error by the ProSim 4, which has an accuracy of 1%, i.e. 2 bpm. A simulated bradycardia at a constant 30 bpm is measured as a constant 30 bpm. The patch seems to be insensitive to PVCs: heart rate barely changes during that simulation period. This could be explained by the moving average filter over 10 heart beats used by Health-

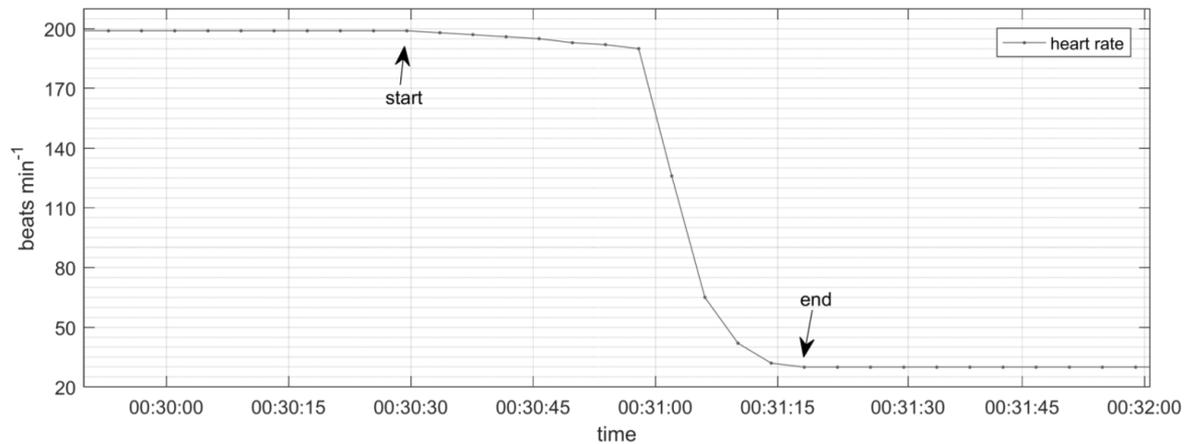


Figure 10 – A detail of the HealthPatch heart rate measurement is shown. Simulated heart rate drops instantaneously from VT at 200 bpm to bradycardia at 30 bpm. The HealthPatch heart rate drops slowly due to its 10 heart beat moving average filter. The filter effect can be seen in the figure, indicated by the arrows.

Patch (see also §2.2.2). A few short sections of the ECG measurement are depicted in Figure 9, in which three rhythms can be clearly recognised.

The effect of the moving average filter is clearly visible in Figure 10: the measured VT slowly falls before reaching the actual bradycardia at 30 bpm. The fall takes about 45 seconds, while it actually happened instantaneously in the simulation (hence the fall is a negative step response).

Depending on the generated amplitude, which was varied near the end of the measurement, HealthPatch was or was not able to measure anymore. At a generated ECG amplitude of 0.3 mV all signals were continuously measured. At 0.25 mV or lower however, the signal starts to drop due to uncertain calculations or is completely absent.

Respiration and respiratory rate

As expected, HealthPatch cannot measure respiratory rate based on chest impedance variability. Because the ProSim 4 generates its respiratory rate signals based on this principle, the result is that HealthPatch respiratory rate does not measure the simulated rate. The respiratory rate seen in Figure 7 varies greatly and is not near the simulated respiratory rates, which were set to constant values throughout large parts of the measurement protocol. No correlation between simulated and measured respiratory rate was seen. However, positive correlation between HealthPatch heart rate and respiratory rate was found: when heart rate was very high or very low, respiratory rate often was too. In these parts of the measurement, the respiratory rate signal seems to vary about a constant average value that correlates with heart rate. For example, between minutes 44 and 48, respiratory rate varies closely around a constant average of about 5 brpm. It is remarkable that the measurement limits of 4 brpm and 42 brpm are reached.

§3.4.2 Behaviour of the relay software and remote server

The relay software on the iPad mini, i.e. the HealthWatch application, and the HealthWatch web interface generally reflected the signals as generated, albeit minimally delayed by 5 seconds (see experimental setup in §3.3.3). A few times, the monitors showed incorrect information. During asystole, it is not possible for HealthPatch to measure heart rate or respiratory rate. During the simulation of asystole, the iPad showed '--' (two dashes) and the web interface showed the last observed value. Only after the user navigated to another page of the web interface and then returned to the live monitoring view, the web interface would also show '--' instead of the last

observed value. Other moments during the simulation protocol showed no unexpected behaviour of the monitoring interfaces.

Both the application and the web interface showed changes in colour of the measured heart rate when a threshold was crossed. This did not occur during asystole, however.

§3.5 Discussion

This chapter described the simulation experiment performed with HealthPatch MD. Simulated heart rate and rhythm could be accurately measured, but not always automatically detected due to lack of classification algorithms. The measured frequencies diverged minimally from the simulated frequencies, assessed visually. Heart rate was measured as a moving average over a number of heart beats: a simulated instant decrease from 200 bpm to 30 bpm took 45 seconds to reach a steady state in the HealthPatch measurement. This means abrupt changes in heart rate cannot be instantly detected, but within one minute this is possible. It depends on clinical context whether or not this is too long. A reaction time of one minute may be too long for continuous monitoring in an acute care setting, where immediate detection of apnoea or asystole is of vital importance. However, monitoring at home may not require such continuity and detection speed, but may focus more on slow changes (trends) or patterns corresponding to recovery/decline. As such, a reaction time of one minute may be quick enough.

Various heart rhythms were simulated: sinus rhythm (SR), asystole, premature ventricular complexes (PVCs), ventricular tachycardia (VT), bradycardia, coarse and fine atrial and ventricular fibrillation (AFib and VFib respectively), and atrioventricular (AV) blocks. The ECG simulation amplitude was also varied. All rhythms were visible in the ECG measured by HealthPatch MD. In the heart rate measurement, it was more difficult to recognise the presence of arrhythmias. PVCs had almost no influence on measured heart rate. AFib was visible as minor variations in heart rate, which may not be distinguishable from physiological variability of heart rate. VFib was visible as a great increase in heart rate variance. The third-degree AV block was visible in the heart rate signal as a bradycardia. Upon decrease of generated ECG amplitude to 0.25 mV or lower, the measurements ceased. This can be explained because the QRS complexes in the ECG signal probably become indistinguishable from the noise, or the patch algorithms become too uncertain about R-R interval. In practical measurements performed with HealthPatch (see next chapter), it was found the amplitude of the ECG generally exceeded 0.5 mV. Based on this simulation experiment 0.5 mV is enough for measurement of ECG with HealthPatch without drop-outs.

For clinical practice, it may be required some sort of rhythm detection is available. In its current state, the HealthPatch system does not provide such functionality. However, based on the 1-lead ECG some simulated arrhythmias were clearly seen, thus automatic detection may be possible. Yet, it remains unclear whether or not detection of arrhythmias is possible in patient measurements.

The HealthWatch iOS application and web interface were found inappropriate for use in clinical practice. The application showed no measurements as soon as the HealthPatch fails to calculate them, e.g. during asystole. In this acute scenario, the web interface performed even worse, because it showed the last known correct value. This phenomenon is called Last Observation Carried Forward (LOCF). This is potentially dangerous, because it may lead to an erroneous decision to not act upon an acute situation that is falsely reported as acceptable, whereas the true situation requires timely action. LOCF should be marked with a clear time stamp indicating the measurement is historical, how much time has passed since and possibly what the cause of LOCF might be (e.g. bad skin contact, loss of internet connection, etc.). Thus, asystole is not recognised by HealthPatch MD, instead LOCF or '--' are displayed. This, while in theory it is possible to recognise asystole using an integration of multiple parameters: if the skin contact is good (low

impedance) and the ECG is flat, asystole can be assumed. The same holds for respiratory depression.

Other arrhythmias, such as fibrillation or AV blocks, were clearly visible in the measured ECG. However, neither the relay software nor the web interface recognised these or reacted on them. This is because they are demonstration software provided by Vital Connect and do not incorporate detection algorithms for such patterns. The software only provides threshold notifications and alarms. More advanced (e.g. the aforementioned integration of signals) and smart alarm strategies are not available in the HealthPatch software.

The Fluke ProSim 4 patient simulator does not support simulation of respiration through changes in the ECG, i.e. RSA and QRS complex amplitude modulation. This resulted in the HealthPatch being unable to measure respiratory rate. However, the patch did estimate respiratory rate based on the available ECG. This estimate was inaccurate, as it seemed to correlate only with the heart rate at that particular moment in time. It did not agree with the simulated respiratory rates. The reason for this correlation is unknown, but may be due to small variations in simulated QRS-complex timings or assumptions/guesses made by the HealthPatch algorithms.

Because HealthPatch measures electrode impedance, it could have measured respiration based on chest impedance variability in this experimental setup. However, in practice the electrodes are probably too close to each other (about 10 cm apart) to measure variations in chest impedance due to respiration. Besides, the impedance measurement is sampled once per 4 seconds, i.e. at 0.25 Hz, as a consequence the maximum respiratory rate that can be determined from the impedance signal is $60 \text{ s} \times \frac{1}{4} \text{ Hz} \times \frac{1}{2} \div 1 \text{ min} = 7\frac{1}{2} \text{ min}^{-1}$ (using the Nyquist-Shannon sampling theorem [40]). This is not high enough to measure normal respiratory rate, not to mention tachypnoea.

Based on these aspects and the results, HealthPatch MD seems unreliable for measurement of simulated respiratory rate based on chest impedance. To test respiratory rate measurements by means of a simulation study, the simulator should be able to generate the physiological signals needed for the HealthPatch MD. With such a simulator, it would be possible to push the patch to its limits to obtain a better understanding of the three respiratory rate algorithms it uses (see also section §2.2.4).

Continuation of the development of the HealthWatch application and web interface is necessary, but they have the potential to become practical monitoring interfaces for both patients and caregivers, both locally (physically near the patient) and remotely, and continuously or at (ir)regular intervals.

In conclusion, the HealthPatch MD appears to adequately measure heart rate. How well it measures respiratory rate remains unknown, because simulation was not possible. Detection of arrhythmias currently requires active monitoring by a caregiver, which defeats the purpose of automatic monitoring with HealthPatch MD.

Chapter 4. Measurements with HealthPatch MD on a clinical ward

*If a man neglects education
he walks lame to the end of his life*
Plato

Measurements were performed with HealthPatch MD to obtain patient vital signs for the research described in the following chapters. The current chapter describes these measurements, the clinical setting, data storage and the signal analysis methods used in this study. Besides vital signs data from HealthPatch MD, measurements were performed with a continuous vital signs monitor for agreement analysis (Bland-Altman analysis) of the two methods. In the UMCU, the Spacelabs XPREZZON™ (Spacelabs Healthcare, Snoqualmie, WA USA) monitoring system is used (hereafter referred to as Spacelabs). The measurement methods are described in §4.1. Because the HealthPatch platform has never been used before in the UMCU a database is needed specifically for measurements with this type and quantity of data. The database design is described in §4.2. §4.3 describes the pre-processing methods the database allows to perform on the data before further analyses can be carried out. Lastly, §4.4 describes the analyses with which agreement between HealthPatch MD and Spacelabs measurements was assessed.

This chapter's main focus is the methodology to obtain and analyse the data. The results of these methods are presented mainly in the following chapter.

§4.1 Measurements on the surgical medium care ward

§4.1.1 Patients and setting

The main research question of this thesis focuses on the postoperative population. This population was chosen, because they are at elevated risk of adverse events, e.g. complications such as infection. Patients after major surgical procedures are particularly at risk of critical physiological decline. The goal was to acquire physiological signals that show both normal and abnormal physiology, while the population sample is representative of the surgical population. Intensive care patients, of whom a large fraction are surgery patients, are not an ideal population choice in this case. This is because they are often mechanically ventilated, i.e., they do not breathe spontaneously, yet measurements of spontaneous respiration are of interest in this study. This is why the surgical medium care ward was chosen for measurements with HealthPatch MD. On the surgical medium care patients breathe spontaneously, although some respiratory support is available, e.g. continuous positive airway pressure (CPAP). Many of these patients have undergone major surgery after traumatic injury, vascular disease or oncology. They require some form of monitoring because of their (risk of) vital instability or fragility, or for diagnostic or therapeutic reasons.

Patients were selected based on the following criteria. Inclusion criteria were age ≥ 18 years, an indication for postoperative admission to the surgical medium care ward, continuous monitoring of heart rate and respiratory rate with Spacelabs and expected length of stay (LOS) ≥ 24 hours on this ward. Exclusion criteria were allergy for sticking plasters (medical adhesives or latex), any wound or irritation on or near the application site, an electric cardiac implant, isolation (for infection control) and pregnancy or lactation of a newborn.

Medical ethical board approval was obtained for measurements on this ward. This meant a participation ended upon transfer from the surgical medium care ward to a regular or intensive care ward, or another hospital.

§4.1.2 Measurement protocol

After patients had been reviewed for inclusion or exclusion based on their electronic medical records (EMRs, also called electronic health records (EHRs)), the nurse responsible for the patient judged whether or not the patient was able to cooperate. Patients with neurological or psychiatric disorders, e.g. delirium or psychosis, may not be able to give informed consent or might unintentionally remove the HealthPatch from their chest. In these cases, the patient's partner or direct family were informed and asked to give consent if a measurement was feasible.

Informed consent was acquired from all participants by providing both verbal and written information about the study goals, the measurements and the implications for their personal data (anonymous treatment) and hospital care (none). Patients were granted time to review their decision to participate or not.

Before the measurements

Upon participation, a HealthPatch MD was applied to the patient's skin. In the first number participants of this study measurements were performed with the patch oriented in position 3 (see also §2.6), i.e., over the left midclavicular line near the 6th and 7th intercostal spaces. This was based on earlier student research that indicated that position as preferred. Later patients received the patch on position 1, i.e., at a 45° angle on the left m. pectoralis major. This was based on recommendations given by Vital Connect (the HealthPatch manufacturer) after we reported some data discontinuities (see also §4.5.2). Also, at first gentle grade HealthPatches were used, which was changed to the active grade variant for better skin adhesion. All HealthPatches were medical grade (with the MD name suffix).

At the start of a measurement a Bluetooth connection with the relay device was established. Apple iPad minis were used as relay devices. They ran the latest available HealthPatch application available for their iOS version. The relay devices were configured to connect to the Vital Connect web interface (VitalCloud) through a secured UMCU Wi-Fi connection. The iPads were positioned behind the patient bed, within a distance of less than 2 metres and connected to the power socket to ensure they were powered at all times.

During and after the measurements

After the measurements started, a number of patient characteristics were documented in a database, which is described in §4.2.2. That section contains the details of all the collected data for every measurement. The database structure was as follows: patients–participations–measurements–annotations.

Any events, such as temporary or unexpected removal (e.g. unexpected end of participation) of the patch were documented during the measurements. After every measurement, the patch was removed from the skin. The reusable sensor module was removed from the patch and cleaned for the next use and the used patch was disposed. The measurements' raw data files were collected and stored in the database.

§4.2 Data acquisition and storage

A database was set up to facilitate acquisition, storage and pre-processing of all data. This section covers the first two of these aspects, §4.3 covers the third.

§4.2.1 Data acquisition

Measurements were done with two vital signs monitors: HealthPatch MD and Spacelabs XPRES-ZON. These are completely different systems with different underlying data storage systems; they are described separately below.

HealthPatch MD data acquisition

The way the HealthPatch system stores its measurements is described in Chapter 2. In short, HealthPatch data is stored on the relay device in JSON format; on the web interface the data can be obtained in CSV format. The CSV files contain a time stamp (time of measurement), heart rate (in bpm), respiratory rate (in brpm) and a few of the other measured signals. The JSON files also contain this data and much more technical details about the measurements, but require more physical steps to collect. They are extracted from the iPads through a computer running iTunes. This was tested and it was found the JSON files could not reliably be found on the iPads: many of them were missing, even though the corresponding data was indeed correctly transmitted to the remote data server and accessible through the web interface. Vital Connect was contacted about this issue, but at the time of writing their technical support has not been able to provide a solution.

Because of these reasons and because in this study only heart rate and respiratory rate were of interest, the CSV files were sufficient for the further analyses in the next chapters. They were acquired reliably from the web interface. The web interface provides one HealthPatch CSV file per day. For example, if a measurement starts on day one at noon and ends on day two at noon (24 hours), the data is contained in two CSV files, i.e., one for each (partial) day starting and ending at midnight. The raw data files were organised in a folder and the paths to the files were stored in the database under every corresponding HealthPatch measurement entry.

Spacelabs XPREZZON data acquisition

The Spacelabs system used in the UMCU has a server backend in the hospital through which all data can be accessed, both live and retrospectively. The surgical medium care ward's Spacelabs data is stored on these servers for 24 hours, after which the data are lost. It was possible to obtain daily reports from the server. Reports were set up to be generated automatically just after midnight, in which all vital signs data of the previous day was stored. The reports were generated in a custom spreadsheet format, which is easy to read manually, but may be challenging when extracting data automatically (see also §4.3.1). By automatically e-mailing the reports, filtering the e-mails and moving the attached spreadsheets to the raw data files folder, the Spacelabs data were acquired quite easily. Unfortunately, the report generating server proved unreliable at times, resulting in Spacelabs data loss where HealthPatch data could be available.

In principle one Spacelabs vital signs file was generated per day, starting and ending at midnight. Every Spacelabs file contained vital signs sampled once per 60 seconds. The values stored in the report were the numbers that happened to be on the Spacelabs monitor screen on the instance a sample was taken. The raw data files were organised in a folder and the paths to the files were stored in the database under every corresponding Spacelabs measurement entry.

§4.2.2 Data storage

The data storage structure was chosen to closely resemble the study design: patients were included in participations, measurements were performed in each participation and annotations were made of the measurements. This database design is elucidated in the following sections.

Patients could participate more than once

Every patient inclusion equals a participation. If—after having participated before—the patient underwent another operation, was readmitted to the surgical medium care ward and they met the inclusion criteria, they could be included another time. Thus, every patient could participate more than once. This resulted in a 1: n association between patients and participations. This is reflected by the database: one patient contains records to one or more participations.

Participations could contain two or more measurements

Measurements could take multiple days, because an admission on the surgical medium care ward could last for days. As a consequence, a HealthPatch MD could need to be replaced, e.g. when the battery ran out of power. Replacement of a HealthPatch MD was considered as the start of a new measurement within the same participation. In this way, it was possible to analyse every HealthPatch measurement individually, or analyse all measurements (a participation) together as one long measurement.

Spacelabs XPREZZON vital signs were also collected throughout every participation, which is a different measurement device. Therefore, every participation comprised of two (at least one HealthPatch MD and the Spacelabs) or more measurements. Similar to the patient-participation association, participations and measurements were linked 1: n .

Measurements could contain any number of annotations

Because the main goal of this thesis is to investigate how postoperative physiological changes and decline can be measured with HealthPatch MD, annotations were made about any relevant findings during the measurements. For example, when a complication had occurred it was relevant to annotate this on the corresponding date and time within the corresponding measurement. This allowed to compare clinical findings from the EMR with measured vital signs. Annotations could be made in time, i.e., they had a start date and time and (optionally) an end date and time. For example, if a patient had suffered from tachycardia during the course of multiple days, this could be annotated with one annotation from start to end.

Naturally, any measurement could contain any number of annotations about clinical findings. Similar to the patient-participation and participation-measurement associations, the measurements and annotation were linked 1: n .

Database records

The aforementioned sections illustrate the structure of the gathered data. Figure 11 shows the database structure in a class diagram based on the Unified Modeling Language (UML). This is a diagramming language that can be used to represent a database model such as the one in this study. Every block represents a class, an object in the database system. The objects' properties are listed as well.

Note how the database class (top left in the figure) is associated with the patient, participation, measurement and annotation classes, i.e., those classes are stored in properties of the database class. These classes also have associations between them. Annotations are always associated with one measurement; a measurement is always associated with one participation; a participation is always associated with one patient.

Some associations may be shared, denoted by the \diamond association symbols. For example, the database contains patients, participations, measurements and annotations of this associations type. This means a database may contain a particular patient, but another instance of a database may contain the same patient. This is useful to perform various analyses on different subsets of the entire database. For example, to study both HealthPatch and Spacelabs measurements, the entire database could be used; a subset could contain the measurements by the HealthPatch only, for example. The patient, participation, measurement and annotation classes are associated with the \diamond symbols. This symbol indicates a composition, meaning that for example a patient is a component of a participation, i.e., every patient 'belongs to' a participation and vice versa. Moreover, this can only be one patient, hence the '1' under the 'patient ID no.' property in the association. The other way around, i.e., from patient to participation, there must be one or more participation denoted by '1.*'. Without any participation, a patient cannot exist in this data model and a participation cannot exist without a patient either. The same is true for measure-

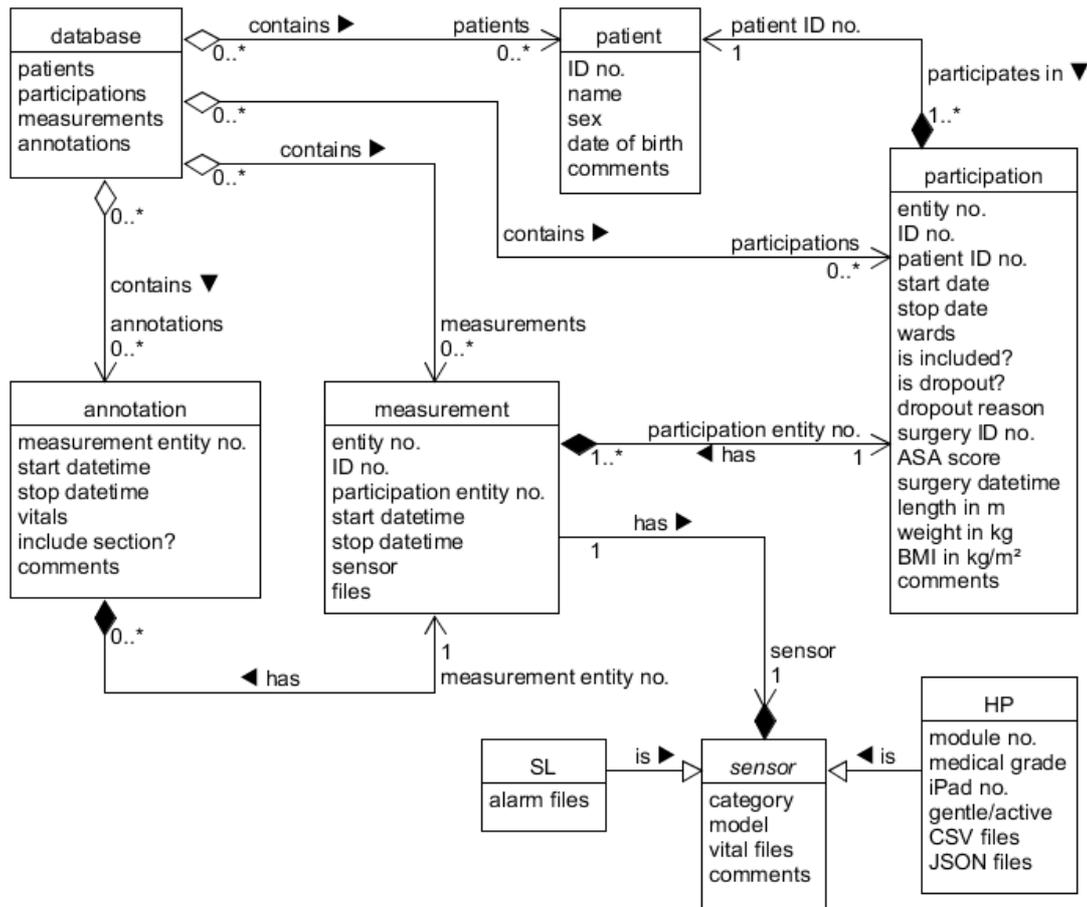


Figure 11 – The database is represented in a class diagram based on the Unified Modeling Language (UML). This is a simplified version of the true database representation. Every block represents a class. Class properties are shown in every class block. Associations are shown between the classes. Note the patient–participation–measurement–annotation associations in particular. TM: telemonitoring, ID: identification, SL: Spacelabs, HP: HealthPatch.

ments and participations. Annotations, however, may not exist: measurements can have zero annotations, but when an annotation exists, it is associated with one measurement only.

This database model allows for thorough manipulation and extraction of its contents. To illustrate this, consider the following example. The entire database contains measurements by HealthPatch and Spacelabs sensors. If a subset of only the HealthPatch measurements is required, removing all the Spacelabs measurement implies removal of all the associated annotations as well, because they must have an associated measurement (as modelled). If the participations associated with these deleted measurements contain no other measurements, they are removed too. In turn, if their removal results in patients without participations, these patients are removed from the database.

The diagram lists the properties stored in the database. The following list provides some more details about these properties.

- Patient:
 - General:
 - Identification number (from EMR)
 - Date of birth
 - Sex:

- Female
 - Male
 - Comments (any patient-specific notes, but not specific to a participation)
- Participation:
 - General:
 - Identification number (a unique positive integer incremented for each participation)
 - Start date (without time)
 - End date (without time)
 - Inclusion/exclusion status
 - Unexpected reason for end:
 - Yes
 - No
 - If unexpected end, the reason could be noted
 - Comments (any participation-specific note, but not specific to a measurement)
 - Medical:
 - Surgery identification number (from EMR)
 - Surgery date and time (of incision)
 - American Society of Anaesthesiologists (ASA) physical status classification (1–5, 6 is N/A as such patients do not survive surgery)
 - Emergency surgery priority:
 - A (within 2 hours)
 - B (within 6 hours)
 - C (within 1 day)
 - D (within 7 days)
 - N/A (not an emergency surgical procedure, i.e., elective)
 - Patient characteristics during participation:
 - length (metres)
 - weight (kilograms)
 - body-mass index (BMI, $\text{kg}\cdot\text{m}^{-2}$)
- Measurement:
 - General:
 - Identification number (a positive integer unique for each measurement within a participation)
 - Start date and time, note that this may differ from participation start date
 - End date and time, note that this may differ from the participation end date
 - Ward, i.e., surgical medium care⁸
 - Date and time of admission to the ward
 - Sensor:

⁸ Measurements took place on one ward in this study. However, plans were made to perform measurements on more than one ward, but not realised during the research for this thesis. For this reason and for future research on multiple wards, the database took record of the ward of participation. In the study design, a measurement was meant to end when a patient was transferred to another ward while still participating in the study. In that case one participation would have contained multiple measurements on multiple wards, although the monitoring devices (especially the wearables) could be the same.

- Device model:
 - Spacelabs
 - HealthPatch MD
 - Device category:
 - Reference (Spacelabs)
 - Index (HealthPatch MD)
 - For HealthPatch MD only:
 - Relay device number (iPads were numbered 1–4)
 - Sensor module number (modules were numbered 1–8)
 - Vital data files (paths to data files containing raw vital signs)
 - Comments (any measurement-specific note, but not as specific as an annotation)
- Annotation:
 - Vital signs:
 - Start date and time
 - End date and time
 - Vital signs to annotate:
 - heart rate
 - respiratory rate
 - all (for general annotations)
 - What to do with annotated section:
 - Include (specifically include this section)
 - Exclude (specifically exclude this section)
 - Nothing, i.e., just annotate
 - Comment: any text to annotate the measurement with

§4.3 Data pre-processing

Before further analyses could be performed, it was necessary to extract the data from the database. This was not a trivial task, because the database was custom-made and the data at this point were still contained in the raw data files. Further analyses were done in MATLAB (The MathWorks Inc., Natick, MA USA), which is software that provides tools and a programming language for scientific computing, statistics, data analysis, graphics, etc. For this thesis, MATLAB R2014b was used, which was the version available in the UMCU. A database interface was written in MATLAB that could import all data from the database. In essence it was written to be a mirror image of the database: it contains patients, participations, measurements and annotations in a structure similar to the UML class diagram in Figure 11. Hereafter, this mirror image is referred to as mirror, to avoid confusion with the database by which we mean the actual database in which all data were manually entered.

Besides providing an interface to the database, all analyses in the following chapters were performed in MATLAB. To facilitate this, all pre-processing steps were written embedded in the mirror. Pre-processing steps written for the mirror included, but were not limited to:

- Reading raw data (§4.3.1)
- Resampling and filtering (§4.3.2)
- Synchronising HealthPatch and Spacelabs (§4.3.3)
- Graphing participations for exploratory analysis (§4.3.4)

These steps contain a number of important methods that need to be elucidated. This is done in the following subsections.

The mirror was programmed based on the object-oriented programming (OOP) paradigm, in which classes (programs) define objects (which can represent anything) with certain properties,

methods (actions) and events (reactions). For example, a patient can be represented as an object having properties such as a name, date of birth, EMR identification number and their sex, similar to what was shown before in the UML diagram. The mirror itself was also programmed as an object, with the patients, participations, measurements and annotations as its properties. The various methods were programmed to perform pre-processing of the data.

The OOP paradigm allowed for flexible extension and modification of all the required calculations in MATLAB. It was written with future research in mind, to allow for easy enhancement and extension of the mirror when needed.

§4.3.1 Reading raw data

Raw (unprocessed) data files were organised in the database in a text field within each measurement. For example, a measurement with a HealthPatch of three days would contain a text field with paths to the three corresponding HealthPatch CSV files. To gather the vital signs data from these files, the following processing pipeline was programmed:

1. Loading raw data files
2. Removing invalid data and retaining gaps
3. Uniform sampling

Loading raw data files

First, every raw data file was read into a table. Because there were different data file sources (HealthPatch or Spacelabs), these were different methods depending on the source. For Spacelabs data, an additional step was necessary: because the report generating server was unreliable (see §4.2.1): there could be zero or more than one report of the same day. To handle this, the raw time series from all files was read, but only the unique time stamps were kept. In other words: the same sample could have been read multiple times from multiple files, but it was only kept once.

Removing invalid data and retaining gaps

At this point the data had been read in a raw format that needed to be processed into a usable format. For example, many empty or invalid (not-a-number: NaN in MATLAB) values may have been read; such values may exist in the raw data files for various reasons. All these undesired values were removed to obtain continuous vectors of vital signs samples with the corresponding time stamps.

It is known HealthPatch records heart rate and respiratory rate at one sample per four seconds. Due to the removal of missing or invalid samples, the measurements could contain gaps that were of interest to study. For example, if the skin contact impedance was high, the HealthPatch may have had trouble measuring vital signs, resulting in missing samples. These gaps were detected by looking at the time vectors of the signals: a gap was detected where the difference between consecutive sample times was greater than the expected difference of four seconds. Closer inspection of the raw data revealed the sample rate was 0.25 Hz on average, but never exactly that value. The true sample rate varied from sample to sample; it was off by a fraction of a second, but on average the sampling time was 4 seconds. In other words: the raw HealthPatch signals were sampled non-uniformly, which can be challenging for many analyses. A time difference threshold was used to detect gaps: where the time between samples was greater than 1.5 times the median sample time, a gap was defined. With this algorithm, time differences of a little longer than 4 seconds due to the variable sample rate were allowed. However, gaps were detected where a sample was truly missing, i.e., where the time between consecutive samples was about 8 seconds or longer.

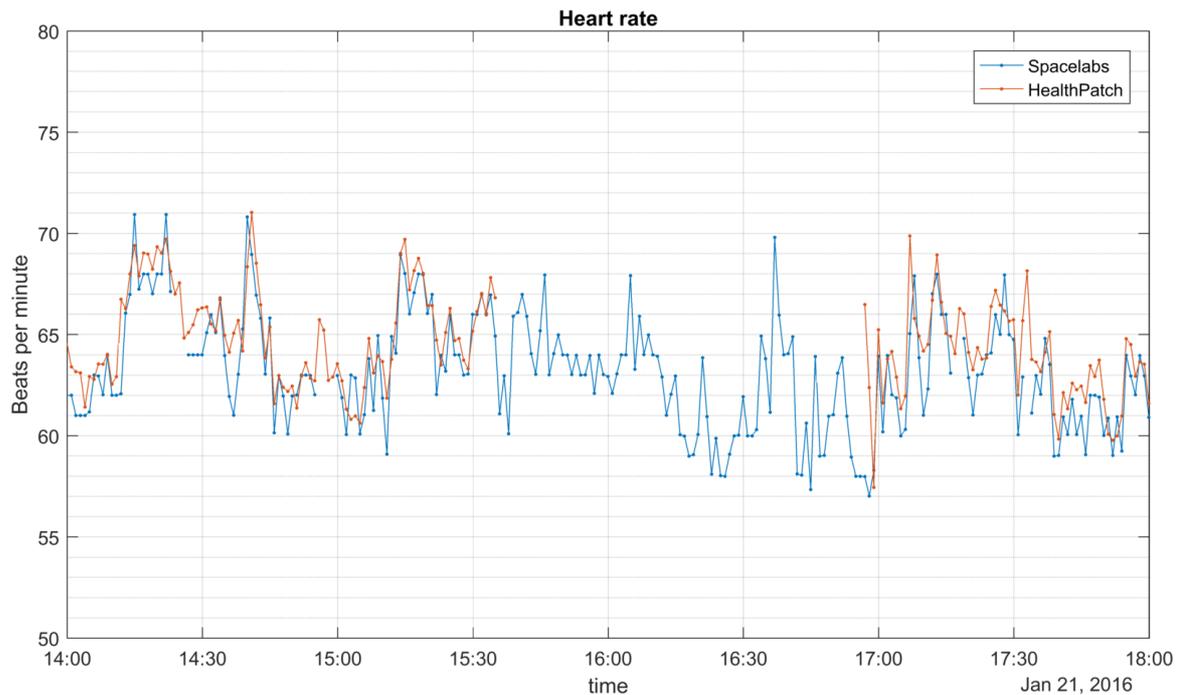


Figure 12 – Discontinuities in the data are graphed as gaps instead of continuous connections. The plotted signals are a section of an actual participation. Note that both HealthPatch MD and Spacelabs show gaps: the Spacelabs measurement is discontinuous just before 14:30, the HealthPatch MD measurement from about 15:30 to 17:00. A gap may be as short as one sample and gaps can occur in both the HealthPatch and Spacelabs measurement simultaneously.

Gaps are useful to retain in the signals, because they allow the gaps to be graphed. Graphing software, such as MATLAB, needs to explicitly know where to draw a line between samples (no gap) and where to draw no line (gap). If no gaps would be drawn, a line would be graphed at all times, which insinuates samples were measured at all times, while actually no measurement took place at some points in time. Gaps are illustrated in Figure 12, where a short section of a measurement by both HealthPatch and Spacelabs is shown.

Non-uniform sampling was found for Spacelabs vital signs as well: the data were sampled at an average sample rate of 1/60 Hz, i.e., one sample per minute, but time stamps could be a few seconds off. The same gap detection and retention methods as used with HealthPatch data were performed on the Spacelabs vital signals.

Uniform sampling

The non-uniform sample rates of the devices are impractical for many statistical and signal analysis steps, so a uniformly sampled data set was desired. Uniform signals were obtained by interpolating the non-uniform signals on a uniform grid sampled at the desired rate. Linear interpolation was used, which introduced minor errors in the sampled data. However, since following analyses will not use as frequently sampled data, but instead use low-pass filtered or averaged aggregates of samples, these errors are filtered out, therefore negligible.

§4.3.2 Resampling and filtering

With the raw data read, gaps preserved and signals uniformly sampled at the devices' particular sampling frequencies, a comparison between HealthPatch and Spacelabs (see §4.4) cannot be readily made, yet. After all, HealthPatch MD signals were sampled once per 4 seconds, i.e., at

0.25 Hz, while Spacelabs data was sampled less often at about 0.0167 Hz, i.e., once per minute. This poses a challenge which is solved by resampling the signals to the same sampling frequency.

Resampling can be done with various methods. In the current study, the resampling goal was to decimate, which means that less samples are retained from a more frequently sampled signal while applying appropriate filtering techniques. For example, one resampling method would be to decimate the HealthPatch signals to 1/60 Hz to obtain the same sampling frequency as the Spacelabs signals. The simplest—however erroneous—method is to take one sample, leave out a few and repeat: this process is called downsampling. It may cause aliasing, which is the artificial introduction of frequency components in a signal when it is sampled with a sample rate lower than twice the Nyquist frequency [40]. The Nyquist frequency is largest frequency component of a signal. Because the signals to be resampled are already sampled and the underlying continuous physiological processes can be quite variable over time, it is assumed the raw HealthPatch and Spacelabs signals contain spectral power up to the Nyquist frequencies. This was verified with spectral analysis (Fourier analysis, spectrograms and periodograms). This means that before decimation can be performed a filter must be applied to remove the spectral power from the signals beyond the new Nyquist frequency, i.e., half the new sampling frequency. Such a filter is called an anti-aliasing filter, which is a low-pass filter that keeps only low frequencies up to the Nyquist rate.

Various anti-aliasing filters can be designed, depending on the desired filter properties. MATLAB provides the `resample` function, which incorporates a finite impulse response (FIR) filter based on a Kaiser window. This window is best thought of as a moving average filter that gives more weight to original samples close to the desired samples. The resulting signal is not the resampled original signal with lots of information lost and aliasing introduced; instead, it contains much information about many of the old samples around the new sample while aliasing is prevented.

The downside of the window used in the anti-aliasing filter is that it is not robust for data with gaps. If, for example, a HealthPatch signal sampled once per 4 seconds needs to be resampled to once per minute (i.e., keep 1 in every 15 samples), a Kaiser window of 15 times the FIR window length would be used as the anti-aliasing filter. If only one sample were missing (a gap of 8 seconds) in the original HealthPatch data, the filter cannot be applied without introducing larger gaps. The result then becomes a gap as well, but since we're downsampling by a factor 15 times the window length in this example, the gap becomes a lot larger. This induces relatively large gaps in the downsampled signals, which is a form of artificial data loss. This is the price of using an appropriate resampling method. This adverse effect is not as bad when the gap in the original data consists of many samples, but is still present. The implications of this loss for the analyses are shown in §5.2, where the results of resampling are presented.

This resampling method was performed in the forthcoming analyses on both HealthPatch and the Spacelabs signals, depending on the desired sampling frequencies.

§4.3.3 Synchronising HealthPatch and Spacelabs

With the raw data of both HealthPatch MD and Spacelabs resampled to the same sampling frequency, a comparison can still not be readily made. Resampling the signals to the same uniform grids does not mean the signals are aligned, in other words: the signal time grids may be off by a (non-integer) number of samples and seconds. The raw data come from two different sources, being the iPads for the HealthPatch measurements and the UMCU servers for the Spacelabs data. These are separate computer systems whose internal clocks are not necessarily synchronised. It may well be possible that while either device's clock e.g. is at midnight, the other device's clock is a few minutes from midnight. As a consequence, the pre-processed time series are not synchronised. Comparing two samples with the same time stamp could be a comparison of samples recorded a few minutes apart, simply because the recording devices' internal clocks do

not tell the same time. Because it is unknown what time points are the same in a pair of Spacelabs and HealthPatch measurements, they must be synchronised based on the vital signs values. Assuming both devices measure the same signals with a decent accuracy and precision, it can be assumed both devices to some degree reflect the same underlying true physiological values of the vital sign being measured. Based on this assumption, both devices' measurements contain similar signals, albeit shifted in time. It is this time shift we are interested in. If the time shift is known, either device can be delayed or shifted forward in time to align the time series, i.e., synchronise the measurements. Because Spacelabs measurements were used as reference, Spacelabs time was taken as reference time to shift to, thus HealthPatch measurements were aligned to Spacelabs measurements.

Synchronisation algorithm

The time series notation

The HealthPatch and Spacelabs time and vital signs data are time series, denoted by (t_i, v_i) and (t_r, v_r) respectively. In these notations t is the time vector, i.e., timestamps of a measurement, v denotes the vector of measured vital sign values (heart rate or respiratory rate in their corresponding units). The subscripted i stands for index device (HealthPatch) and r stands for reference device (Spacelabs). The goal of the synchronisation algorithm is to align t_i to t_r by adding or subtracting a delay. Therefore, this delay must be estimated. The delay can be expressed either in samples, or in units of time. The sample and time domain units are associated by $[n] = [t] \cdot f_s$, with n in samples, t in the time unit and f_s the sampling frequency in Hertz.

Aligning signals in the time and sample domains

The final step in synchronisation is based on maximisation of the cross-correlation function (CCF) of the HealthPatch and Spacelabs signals (CCF maximisation). Before CCF maximisation can be performed, the time grids must be aligned in the sample domain. To understand this, consider the following.

CCF maximisation estimates n_{delay} , the number of samples of delay between the signals, i.e., in the sample domain. In other words, the CCF does not take into account the time domain. If, for example, a HealthPatch measurement started after a Spacelabs measurement (which was common, because Spacelabs measurements often were started upon arrival on the surgical medium care ward, whereas HealthPatch measurements were started later during the admission), their sample numbers (which simply count from 1 to the number of elements in the vital signs vectors) represent an entirely different domain than their time vectors: the first sample in t_i , i.e., $t_i[1]$ needs not be equal to $t_r[1]$. This is illustrated with example signals in Figure 13. Note in particular the different domains in the middle and lower panels, being the time and sample domains respectively. In all three panels the same data is plotted, but in the first and second panels this is done in the time domain, whereas the third is plotted in the sample domain. Note that with these particular example signals, the synchronisation in the time domain should result in a shift of the index signal (red) to the left (backward) in the time domain, but to the right (forward) in the sample domain. This paradox is caused by the difference in domains, the difference in start time (in this example $t_i[1] > t_r[1]$) and the different lengths of the signals.

The time vectors point to a moment in time of the index or reference clock. The sample numbers indicate a point in the vital signs vector, starting at 1. Thus, sample 1 of a HealthPatch measurement could point to a moment in time hours before or after the time of the first Spacelabs sample. It is up to the researcher to make sure the samples used in CCF maximisation are aligned on their respective time grids. This is what is achieved with the following steps.

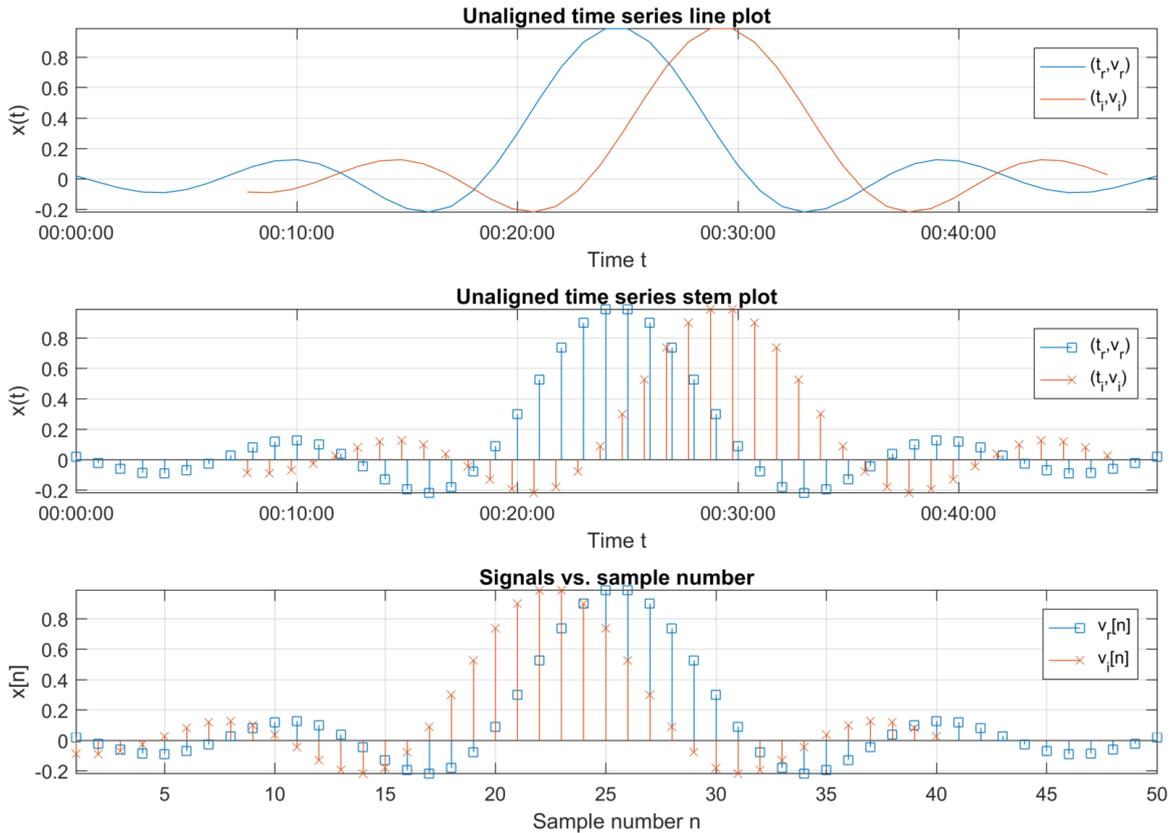


Figure 13 – Example reference and index signals containing the same information are illustrated with various methods. The upper panel shows a line plot of the signals vs. time. The middle panel shows the same data in a stem plot, which makes it easier to see individual samples in time. Note the sample time grids of reference and index do not align on whole samples. The lower panel shows the signals plotted vs. their sample number. This is a different domain than the time domain. Note how the index signal contains the same pattern as the reference, albeit shifted in both time and sample number. Also, the lengths of the signals differ. For the purpose of the example no noise was added to the signals, whereas in reality additive noise does exist. The time unit is arbitrary.

Because the sampling grid of t_i might not match the reference grid, a template time vector, denoted by t_i^* , is created that is aligned to the sampling grid of t_r . This mismatch is illustrated in the middle panel of Figure 13: note the sample grids do not align in time; they are off by a fraction of a sample. There are three possibilities for this template based on the first samples of the HealthPatch and Spacelabs time vectors:

1. $t_i[1] < t_r[1]$, i.e., the first index sample was measured before the first reference sample
2. $t_i[1] = t_r[1]$, i.e., the first index sample was measured at the same time as the first reference sample
3. $t_i[1] > t_r[1]$, i.e., the first index sample was measured after the first reference sample

As noted before, this ‘being ahead’ or ‘being behind’ only means the first *sample* of the measurement comes later in time, which not necessarily means the *actual signal* is ahead or behind the other.

Situation 2 is the simplest: we set $t_i^* = t_i$, because the template time vector grid is already equal to the reference grid. We then proceed to CCF maximisation (see the next section).

In situations 1 and 3 the time vector template t_i^* needs to be shifted so that its grid aligns to the reference grid. In these situations, we shift t_i forward by a fraction of a sample so that it aligns with the reference time grid.

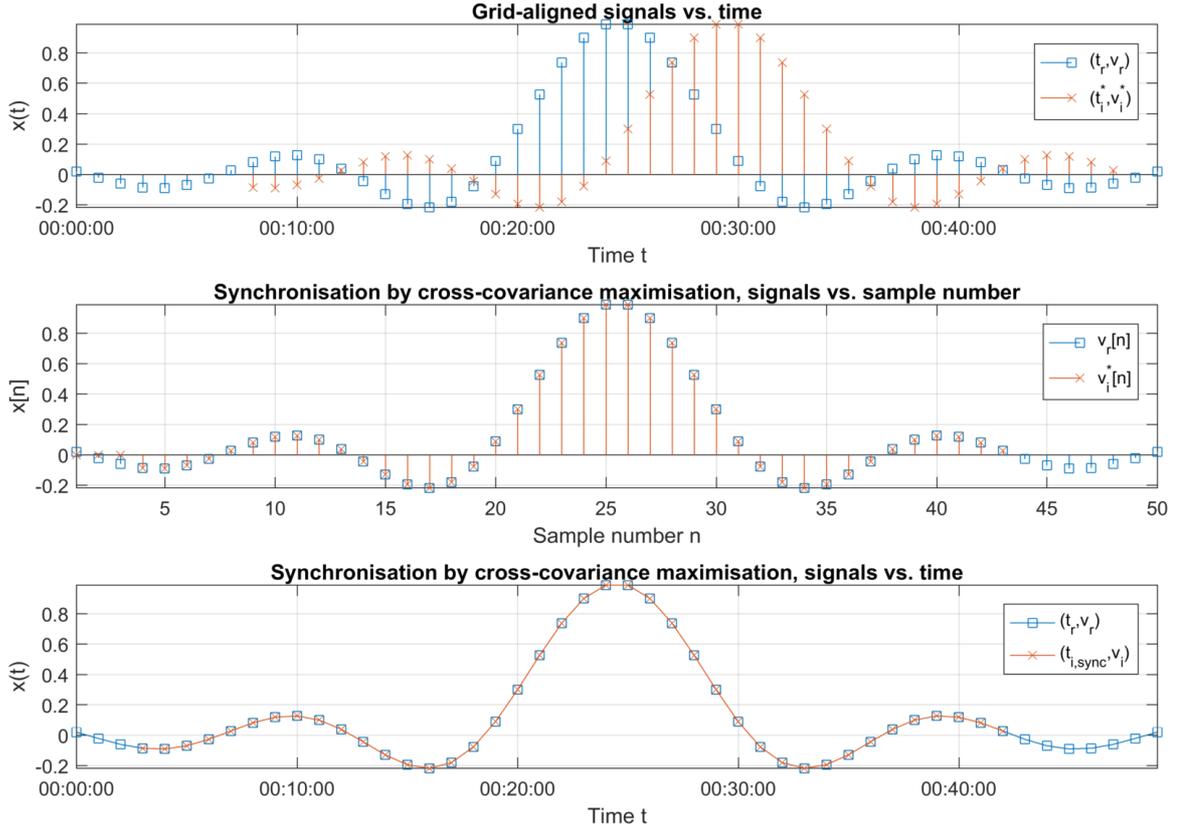


Figure 14 – The synchronisation algorithm is illustrated with the example signals of Figure 13. The upper panel shows the grid alignment step. The middle and lower panels demonstrate the cross-covariance maximisation step in the sample and time domains respectively. Note that in middle panel, the index signal has been padded with three zeroes at the beginning, hence the signal's name is v_i^* instead of v_i . The zeroes were inserted for the purpose of visibility in this figure and are not used in the true methods, in which empty samples are used instead.

In situation 1, the last sample in t_i smaller than $t_r[1]$ is used, its sample number is denoted by n^* . The time difference between these samples therefore is $d^* = t_r[1] - t_i[n^*]$. In situation 3 something similar is done, but the other way around. Here, the first sample number in t_r greater than $t_i[1]$ is n^* . The grid-alignment time shift then becomes $d^* = t_r[n^*] - t_i[1]$. The grid-aligned template then is set to $t_i^* = t_i + d^*$. In situation 3, the index time series also needs to be padded with dummy samples, so the first $n^* - 1$ samples of t_r are inserted before t_i^* and the same number of samples consisting of empty values⁹ are inserted before the index vital signs vector v_i . These empty values cannot be plotted, as they contain no numerical value, but they do take the space of one sample each, which is a property used in the next step of the synchronisation process.

Note that in the second situation no grid alignment is necessary, therefore $d^* = 0$. The grid-aligned HealthPatch vital signs time series is denoted by (t_i^*, v_i^*) , shown with the example signals in the upper panel of Figure 14.

The result of the grid alignment process is a minor shift in time (d^*) that must be accounted for. This shift is $0 \leq d^* < T_s$, where $T_s = 1/f_s$ is the sampling time in seconds. In samples: $n^* = d^*/T_s$, $n^* \in [0,1)$. In other words, a fraction up to one sample shift occurs in the grid alignment process.

⁹ Not-a-Number, NaN in MATLAB

Cross-correlation and cross-covariance maximisation in the time domain

To estimate the time delay between the time series an algorithm was used based on cross-correlation. Cross-correlation measures similarity of two signals. It is defined as

$$\text{ccf}_{fg}[n] = (f \star g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[m+n],$$

for real and uniformly sampled signals f and g with mean $\mu_f = \mu_g = 0$. Where the cross-correlation function (CCF) has a global maximum, the corresponding sample number is $-n_{\text{delay}}$, which can be formulated as (by convention, we take the negative to indicate delay)

$$n_{\text{delay},fg} = -\arg \max_n (\text{ccf}_{fg}[n]),$$

where $n_{\text{delay}} \in \mathbb{Z}$, an integer. It is the number of samples by which the HealthPatch signal must be delayed to align it to the Spacelabs signal.

The algorithm used was as follows. It can be assumed the iPad time and Spacelabs server time were not too far apart. To reduce computation time, it was assumed a maximum cross-correlation could be found within delay of ± 10 minutes, i.e., a window of ± 600 seconds. The synchronisation was done at the default sample rate of the Spacelabs measurements, to which HealthPatch measurements were resampled (1/60 Hz). At this sampling frequency, a 10-minute delay equals a maximum delay of ± 10 samples. This reduces the equation for cross-correlation to

$$\text{ccf}_{fg,10}[n] = \sum_{m=-10}^{10} f[m]g[m+n],$$

which requires much less computation time and memory for signals of many samples, such as the relatively long measurements with HealthPatch and Spacelabs. Because the grid alignment step (the previous step) is performed before the CCF maximisation step, the total time shift may become greater than the maximum of 10 samples after both steps. The total shift after both steps is $n_{\text{total}} = n^* + n_{\text{delay}}$. Because $-10 \leq n_{\text{delay}} \leq 10$ samples and $0 \leq n^* < 1$ sample, it follows that n_{total} may become greater than 10. This is not a problem, just an effect of the synchronisation processes that should not surprise the researcher. If it were undesired, one could limit the cross-correlation summation domain to be one sample smaller.

Now, the time vectors in (t_r, v_r) and (t_i^*, v_i^*) are grid-aligned and ready to be synchronised by CCF maximisation. In this step, the mean of the signals must be zero, because otherwise the summation of multiplications in $\text{ccf}_{fg}[n]$ will result in infinity or be undefined¹⁰. Subtracting the mean is only valid if a constant mean can be assumed and the signals are wide-sense stationary (WSS). If a trend or other process influences the mean, this could influence the CCF and result in an incorrect delay estimation. However, because the signals used in this study can be very long, the CCF maximisation algorithm is robust for a minor violation of this assumption. Cross-correlation of signals whose mean has been subtracted is the same as the cross-covariance function,

$$\text{cov}_{fg}[n] = \sum_{m=-\infty}^{\infty} (f[m] - \mu_f)(g[m+n] - \mu_g).$$

Alternatively, if a linear trend exists throughout a measurement, it can be subtracted instead of subtracting a mean assumed to be constant. See Chapter 6 for more about time series components, one of which is a trend.

¹⁰ A question (initiated by the author), discussion and answer about the difference between and confusion about the statistical and signal processing definitions of cross-correlation and cross-covariance can be found on <http://dsp.stackexchange.com/questions/34778>.

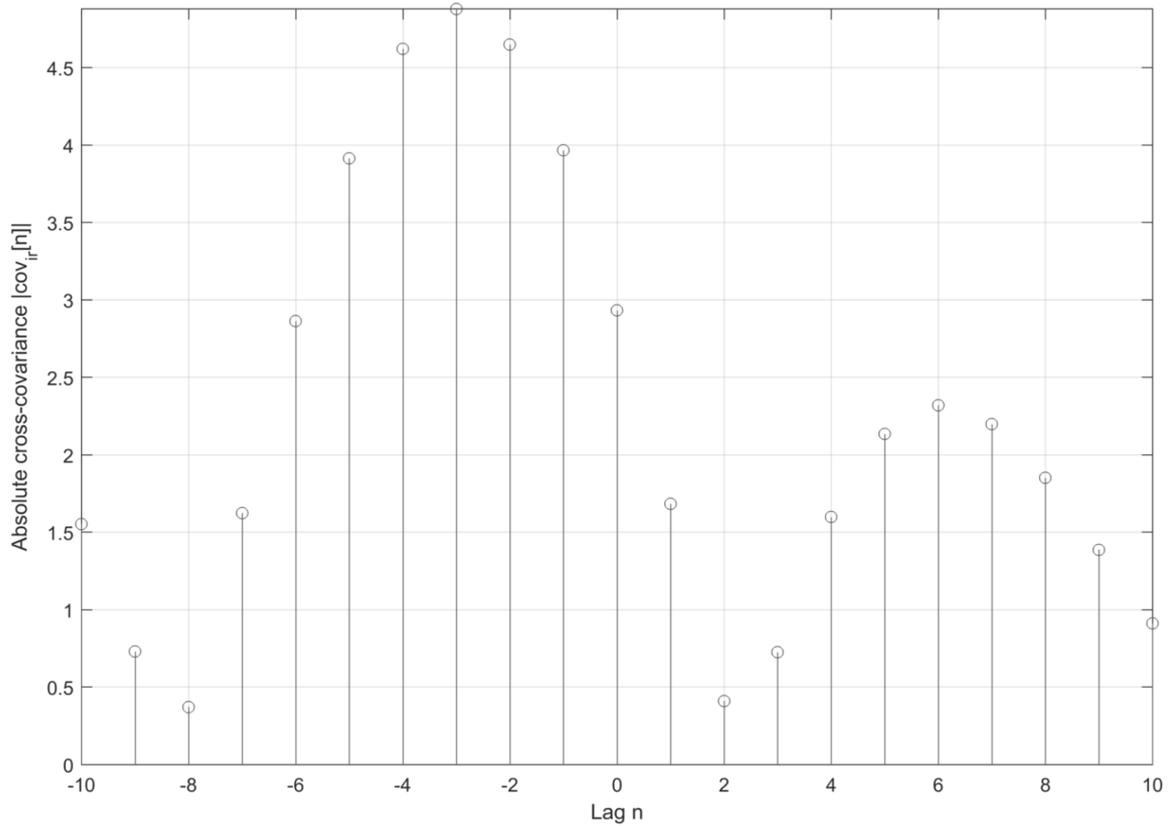


Figure 15 – The cross-covariance function for the example signals of Figure 13 is graphed for a maximum absolute lag of 10 samples. The maximum is found at lag $n = -3$, hence $n_{\text{delay}} = +3$ samples.

In the example signals of the previous figures, the mean is close to, but not exactly zero¹¹. Therefore, cross-covariance maximisation was performed with Figure 15 as a result of the process.

The synchronisation time shift

The final step is the actual calculation of the synchronised time vector of the index measurement. It is the original index time vector plus the grid-alignment delay and the synchronisation time delay.

$$t_{i,\text{sync}} = t_i + d^* + t_{\text{sync}} = t_i^* + t_{\text{sync}},$$

where

$$t_{\text{sync}} = \frac{n_{\text{delay}} + (n^* - 1)}{f_s},$$

in situation 1 and

$$t_{\text{sync}} = \frac{n_{\text{delay}} - (n^* - 1)}{f_s},$$

in situation 3.

The synchronised time series now are (t_r, v_r) (unaltered) and $(t_{i,\text{sync}}, v_i)$. For the example signals the synchronised time series are plotted in the lower panel of Figure 14. Note that the

¹¹ The mathematically inclined reader will recognise the example signals as the cardinal sine (sinc) function shifted in time and knows their mean actually is zero. However, because we consider only a short window of samples, their mean is not zero per se.

vital signs in v_i are not altered at all. The intermediate step where v_i^* is created is only used to accommodate for the changed length of t_i^* .

§4.3.4 Graphing participations for exploratory analysis

Trimming HealthPatch and Spacelabs signals

With the HealthPatch and Spacelabs signals synchronised they can be graphed to visually analyse the vital signs data. Some measurements were done while either of the measurements had not yet started or had already ended. These sections beyond the start and end are not of interest to be graphed. For better graphs of the data, these sections were trimmed from the data. Note that this was only done for the analyses that compared HealthPatch and Spacelabs data together, as it would be a waste of data to remove parts of the HealthPatch measurements when they are not compared to the reference signals.

Plotting signals of participations

In the database, every patient participates one or more times. Every participation belongs to one admission on the surgical medium care ward. The vital signs measured during one participation are of interest, so when graphing them, each participation is considered separately. One participation may encompass one or more measurements with both a HealthPatch and the Spacelabs monitor. As a consequence, plotting one participation results in two (one index plus one reference) or more measurements being graphed.

§4.4 Analysis of validity of measurements

The validity of measurements by HealthPatch MD in clinical practice is of interest in this study. By validity is meant: a description of accuracy and precision of the measured vital signs compared to a reference standard, often termed agreement [50]. That is, it is of interest how good heart rate and respiratory rate are measured by the patch, compared to another vital signs monitor of which the performance is known. If the patch's performance is similar to that of the other monitor, the agreement is said to be high. How high an agreement must be to be considered acceptable depends on clinical judgement and the nature of the measured quantities [51].

Clinical validation studies of diagnostic methods require a comparison with the 'gold standard'. This is defined as the best (available) diagnostic for the measurand of interest. Note that this does not always refer to the best available method, because that method could require a harmful intervention on the patient (e.g. arterial blood pressure via an arterial catheter, which is invasive and may not be desirable or feasible) or is only possible during autopsy (e.g. confirmation of diagnosis). For the validation in the current study, the Spacelabs XPREZZON™ system is the reference monitor. It is the current bedside reference standard for continuous measurement of heart rate and respiratory rate on the surgical medium care ward in the UMCU and used for clinical decision support. Hence, it was chosen as the reference for assessment of agreement.

The goals of this section are 1) to evaluate what methods are available to perform agreement analysis, 2) to evaluate how suitable they are for the type of data obtained with wearable continuous vital signs monitoring devices, and 3) to describe the agreement analysis that was performed on data from HealthPatch MD and Spacelabs measured in the postoperative population. The results of this analysis are presented in §5.2 in the next chapter.

§4.4.1 Agreement analysis

We need to define agreement before it can be assessed. In the arguably classic and famous papers by Bland and Altman [50,52] they present a methods comparison method. In modern medi-

cine, it is often eponymously called Bland-Altman Analysis (BAA), although others have used similar methods in other fields of science before them (e.g. John Tukey, another important statistician). Other names for BAA are limits of agreement (LOA) analysis or simply agreement analysis.

Bland and Altman defined measures of agreement for individual measurements on individual subjects by two methods to measure a certain physical quantity, the measurand. Examples of such quantities are blood pressure, heart rate, respiratory rate, etc. It is not always an easy task to measure the true value of such quantities. Instead, methods have been developed to derive the quantity of interest from other physical measurements. For example, heart rate—expressed in beats per minute—is best derived indirectly from the ECG, which is a measurement of electric potential generated by the heart, *but not the heart rate itself*.

An example of an experiment in which BAA could be performed is as follows. If one would want to assess the validity of heart rate derived from pulse measured manually on the a. radialis, one could perform BAA to compare pulse heart rate with ECG heart rate. Measurements could be performed on n subjects. Each subject's heart rate would be measured as the number of heart beats in one minute. Ideally, their ECG would be measured simultaneously by the two methods to ensure the same true heart rate was measured. BAA would then give the agreement between the two methods. How this agreement is defined and calculated is described in the following section.

Bland-Altman Analysis of bias and limits of agreement

This section is a summary of the LOA analysis by Bland and Altman [50–52]. It is summarised here, because in the following sections further specifications of BAA are described, for which a good understanding of their methods is essential.

In BAA the agreement between two methods to measure a physical quantity is of interest. Measurements are performed on multiple subjects by the two methods and used in this analysis. The result is an observation pair. Each observation pair is an estimate of the true value of the measurand, since both methods may contain an (un)known error. The error consists of a number of components. They need to be well-understood, because later we will perform BAA on data with repeated measurements, which affects how the error is measured. After all, it is the error we're interested in, for if we know how great the error is, we know how precise a measurement method is. The error components arise from the quantity to measure and the two methods themselves. The quantity to measure, i.e. the true value, varies by nature from subject to subject (e.g. heart rate) and as such has some variance. Furthermore, the two methods each have their measurement variance as well, representing their uncertainty of the measured quantity. If we denote the two methods' measurements by X and Y and assume the methods' variances are independent from the measurand, the variances of X and Y are

$$\text{var}(X) = \sigma_X^2 = \sigma_t^2 + \sigma_x^2,$$

and

$$\text{var}(Y) = \sigma_Y^2 = \sigma_t^2 + \sigma_y^2,$$

where σ_t^2 is the variance of the true values and σ_x^2 and σ_y^2 are the methods' variances.

Besides error—which tells us about precision—the relative bias of the two methods is of interest. Bias informs us about accuracy. It is possible the two methods do not measure the same values on average, thus a bias exists between them. If we know both bias and error, we have a measure for agreement of the two methods, agreement being both accuracy and precision.

The bias is estimated as the mean difference between the values as measured by the two methods. If we denote the difference as $D = X - Y$, then the bias is $\bar{D} = \bar{X} - \bar{Y}$, where the overline denotes the mean. Note that $\text{var}(D) = \sigma_D^2 = \sigma_X^2 + \sigma_Y^2 = \sigma_x^2 + \sigma_y^2$ and thus σ_D^2 is independent of σ_t^2 . This is due to the subtraction in D , in which we remove the variance of the true value.

This means that if we analyse the differences, we can estimate to which degree the two methods (dis)agree, independent of the true value of the measurand. This is a nice property of BAA, because the true value often is unfortunately unknown. If it were known BAA would not be necessary and we could resort to simpler agreement analysis methods.

The differences in D are likely to be normally distributed, independent of the distribution of the true values themselves [51]. Thus, 95% of the differences lie in the range $\bar{D} \pm 1.96s_D$, where s_D is the sample standard deviation of D and value of 1.96 is the 97.5 percentile point of the normal distribution, which represents 95% of the area around the mean (excluding 2.5% in each tail). These 95% limits around \bar{D} are called the limits of agreement (LOA). They represent a reference interval for the observed differences, not to be confused with a confidence interval (CI) despite the seemingly similar result, but in fact, their calculations are very different.

An important tool in BAA is the mean-difference plot. This is a graph in which the differences of observations are plotted against their mean, i.e. $D = X - Y$ against $\mu = (X + Y)/2$. With this graph, it is possible to assess the range of differences between the methods, where possible outliers exist and to see the bias. The bias and both upper and lower LOA can be plotted in the same graph [52].

Ideally, the bias is not significantly different from zero and the LOA encompass a small range, which should be defined beforehand by clinical judgement. However, a large bias is not a large problem, for it can be easily corrected by simply subtracting it from X or adding it to Y .

An extension to BAA is the calculation of the CI of the bias and the LOA. Since the bias and LOA are results of a sample (patients) from a larger population, they are estimates of the true bias and LOA, hence they have a CI. For the methods of the CI calculation the reader is referred to the papers by Bland and Altman [50,51]. An implementation to perform BAA was written in MATLAB and published online, the details of which can be found in Appendix A.

Bland-Altman Analysis for repeated measurements of a changing quantity

Repeated measurements on the same subject can be taken of a constant quantity or a changing quantity. An example of a constant quantity is body fat percentage measured on the same day. An example of a changing quantity is blood pressure, which can vary greatly throughout a day. This section focuses on BAA for repeated measurements of changing quantities only, because heart rate and respiratory rate are the subject of this thesis, which are changing quantities during the time period of the measurements.

BAA for repeated measurements of changing quantities is described first in the 1999 paper by Bland and Altman [51] and corrected in their 2007 publication [53]. The analysis is based on one-way analysis of variance (ANOVA), in which a linear model is used to estimate the influences of various sources of variance in the measurements by the two methods (components of variance method). As described before, there are multiple sources of variance in the measurements. One source is the variance of the true value, σ_t^2 , which was differenced out of the equation when taking the difference between measurements X and Y . When X and Y contain multiple observations per subject, their variances, σ_X^2 and σ_Y^2 , are influenced by both the particular variances of these subjects and the variances between them. The total variance thus is a sum of these sources of variance, which is written as follows.

$$\begin{aligned}\sigma_X^2 &= \sigma_t^2 + \sigma_{x,\text{between}}^2 + \sigma_{x,\text{within}}^2, \\ \sigma_Y^2 &= \sigma_t^2 + \sigma_{y,\text{between}}^2 + \sigma_{y,\text{within}}^2\end{aligned}$$

and therefore

$$\sigma_D^2 = \sigma_{x,\text{between}}^2 + \sigma_{x,\text{within}}^2 + \sigma_{y,\text{between}}^2 + \sigma_{y,\text{within}}^2.$$

This poses a problem, because we are not interested in the variance within or between each subject, but in how good methods X and Y perform with respect to each other independent of the subjects.

To illustrate this problem, consider the following extreme and trivial example. To assess methods X and Y , we could measure one observation pair in 1000 subjects and perform regular BAA. This results in some LOA and bias. Now, in the extreme repeated measurements case, we could have taken 1000 observation pairs from only one subject. This data can be used in regular BAA as well, but the results probably are not as interpretable as in the case with 1000 subjects. This is because we are determining the LOA and bias between X and Y on measurements that show variance and bias that could be heavily influenced by this particular subject, i.e., the within-subject variance is much greater than the between-subject variance (which is zero when only using data from one subject).

Theoretically, we could be measuring the subject's heart rate and it could happen to be 60 bpm on average with all 1000 measurement in the range 50–70 bpm. As a consequence, our results do not inform us well about how good the agreement is in general, but only for subjects with this particular heart rate range. Furthermore, if either method would be very bad (or good) at measuring heart rate in this range (or in this particular subject, for whatever reason), regular BAA would lead to the conclusion that the method is bad (or good), while it could be very good (or bad) in a more representative population sample and at different heart rates, i.e., a wider range of values.

To address this problem, Bland and Altman revised their original methods for replicated measurements and describe an ANOVA method, which is summarised in the following section [51,53].

The one-way ANOVA method by Bland and Altman

Bland and Altman base their analysis of agreement for repeated measurements of a changing quantity on one-way ANOVA [51]. They model the observed differences between methods X and Y as

$$D_{ij} = B + I_i + E_{ij},$$

where D_{ij} is the difference between observations X_{ij} and Y_{ij} on subject i , B is the bias (overall mean difference), I_i is the subject-method interaction and E_{ij} a normally distributed error of observation j of subject i . In this model, the variance of D is

$$\sigma_D^2 = \sigma_{D,\text{between}}^2 + \sigma_{D,\text{within}}^2.$$

It can be estimated using a components of variance technique, which estimates the between- and within-subject variances separately. In the one-way ANOVA calculations, a table of sources of variance can be constructed. In this table, the mean square error (MSE) is an estimator of $\sigma_{D,\text{within}}^2$. The sum of squared errors (SSE) is used to calculate MSE.

$$SSE = \sum_i^n \sum_j^{m_i} (D_{ij} - \bar{D}_i)^2,$$

$$\hat{\sigma}_{D,\text{within}}^2 = MSE = \frac{SSE}{N - n},$$

where n is the number of subjects, N the total number of observations such that $N = \sum m_i$ with m_i the number of observations of subject i , and \bar{D}_i the within-subject mean difference between the methods. Note that the subjects need not have equal numbers of observations (m_i) and as such the ANOVA design can be balanced (all m_i equal) or unbalanced (different values of m_i).

The between-subject variance of the difference $\sigma_{D,\text{between}}^2$ can be estimated with similar methods as for the within-subject variance. It is estimated by the difference between the mean squared subject effect (MSS) and MSE divided by a factor depending on the number of subjects and the number of observations per subject.

$$SSS = \sum_i^n m_i (\bar{D}_i - \bar{D}_{..})^2,$$

$$MSS = \frac{SSS}{n-1},$$

$$\hat{\sigma}_{D,\text{between}}^2 = \frac{MSS - MSE}{n_0},$$

where SSS denotes the sum of squared subject effects, $\bar{D}_..$ is the global mean, i.e., the bias, and

$$n_0 = \frac{N^2 - \sum m_i^2}{(n-1)N},$$

a factor depending on the number of observation and subjects.

Now, we have estimators for both components of variance, hence the total variance of the difference is estimated by $\hat{\sigma}_D^2 = \hat{\sigma}_{D,\text{between}}^2 + \hat{\sigma}_{D,\text{within}}^2$. This is the variance of differences between individual observations by both the measurement methods, independent of subject or number of observations within the subjects. That is, assuming the sample of the subjects is representative of the population being studied and the performance (variance and bias) of the methods does not change over time. These are general assumptions in agreement analysis. The rest of the calculations can now be done as in the regular LOA calculations.

Non-constant bias and limits of agreement

Because there may be an association between mean and difference of the measurements of the physical quantities of interest, e.g., measurements by two devices of large values are further off, the assumption of a constant bias (overall mean) may not be adequate. This also means the LOA may be non-constant. Furthermore, the variance of the difference may be associated with the mean, for example, the differences may show larger variability for larger means (heteroscedasticity). Using constant LOA then also is not truthful to the data, because the limits should be closer to the bias where variance of the differences is lower and further away where variance is greater.

To address this issue, Bland and Altman state that simple linear regression lines can be used instead of constants [51]. The bias then has an intercept and a slope. The bias line has a certain goodness-of-fit and all the data not fit by the line are the residuals. It may be interesting to study the residuals, because they can show some varying behaviour too. For example, in the heteroscedastic case the variance of the residuals may increase or decrease with increasing mean. This means the LOA should diverge or converge to truly fit the data. If the residuals are significantly heteroscedastic, another simple linear regression line can be fit to them to obtain the non-constant LOA.

Another method proposed by Bland and Altman is the use of logarithmic transformation [51]. This is suitable for when the differences may show an increasing variance of the differences with increasing mean. Log-transformation is similar to calculation of the ratio of the observation pair instead of their difference.

§4.4.2 Agreement analyses of HealthPatch and Spacelabs

Agreement between HealthPatch MD and the clinical reference monitoring standard at the bedside (Spacelabs) was investigated using the BAA methods for repeated measurements as described in the previous section. Because the measurements are of heart rate and respiratory rate sampled over multiple hours up to days, the methods were adjusted to assume a varying true value, i.e., it was not assumed the true value was constant (which otherwise is an assumption in BAA). Because the lengths of the measurements varied, different numbers of observations were available per subject for agreement analysis. Therefore, the BAA methods for repeated measurements incorporated an unbalanced ANOVA design, whereas equal numbers of observations per subject would have been analysed with a balanced design. To examine the effect of using the incorrect assumptions of regular BAA, i.e., not accounting for repeated measurements, the

data was analysed with regular BAA as well to compare the results and the effect of the incorrect assumptions.

The outcomes of interest were the mean-difference plot and statistics. The bias and LOA with—where available—their 95% CIs were graphed. Any trends in the difference statistics were investigated using the methods proposed by Bland and Altman [51]. Acceptable agreement was defined as the respiratory rate LOA being between -2 and 2 brpm around the bias [54] and heart rate LOA in the range of -10% to 10% of the mean around the bias. The latter range allows for easy verification using the ratio between HealthPatch and Spacelabs values instead of the difference. Ideally, the ratio bias is equal to 1, so the 95% LOA of the ratio should be no greater than the bias ± 0.1 , i.e., $\pm 10\%$.

For the analyses, the data were resampled to $f_s = 1/60$ Hz, i.e., one sample per minute, using the methods described in §4.3.2. This is the native Spacelabs sampling frequency. Individual data points from all subjects were included only if both data points from HealthPatch and Spacelabs (an observation pair) were available.

Because in this study it is of interest to measure changes in vital signs, the dynamic validity of HealthPatch MD is to be evaluated. However, BAA is a static analysis: it does not quantify how well changes in vital signs are measured by either of the compared methods. What can be done with BAA is an analysis of the extreme values in the vital signs data. This still is no performance measure for dynamic validity of HealthPatch MD, but it provides some information about its performance when vital signs are out of their normal range. After all, it is of particular interest to measure extreme vital signs correctly, because they can potentially be used to recognise the deteriorating patient. For this reason, BAA was performed on the lowest vital signs up to the 10th percentile and separately on the highest vital signs from the 90th percentile upward. If agreement is high in these subsets, then this provides evidence about the performance of HealthPatch compared to Spacelabs near these extreme values. The 10th and 90th percentiles were calculated from the Spacelabs data, because in this study it is the reference.

The results of the agreement analyses described in this section are presented in §5.2 of the next chapter.

§4.5 Discussion

This chapter described the measurements performed with HealthPatch MD and Spacelabs XPREZZON on the surgical medium care ward. Furthermore, the design of the database and the interface with MATLAB were outlined. Next, this chapter described some of the most important methods used to pre-process the raw data into a format that facilitates further analyses. Lastly, these analyses were described by summarising Bland-Altman analysis methods and which of these were used in this study.

§4.5.1 Experiences with HealthPatch MD in practical measurements

During the measurements with HealthPatch MD it was found the placement of the thermistor has a disadvantage. The thermistor, shown in Figure 5 of the previous chapter, is on the patient's skin side of the electronics inside the adhesive patch. This creates a protrusion that is visible and tangible through the patch's outer material. This design choice may have been made because in this position the thermistor probably has better skin contact, because it protrudes up to a few millimetres from the adhesive surface into the skin. The theoretical disadvantage of this placement is that the protrusion could damage the skin and cause irritation. Because thermometry was not of interest in this study, it was not further investigated how the thermistor placement affected patient comfort and temperature measurement quality. No patients reported irritation by the patch when asked, which leads to the conclusion that in this study's measurements the thermistor placement was irrelevant.

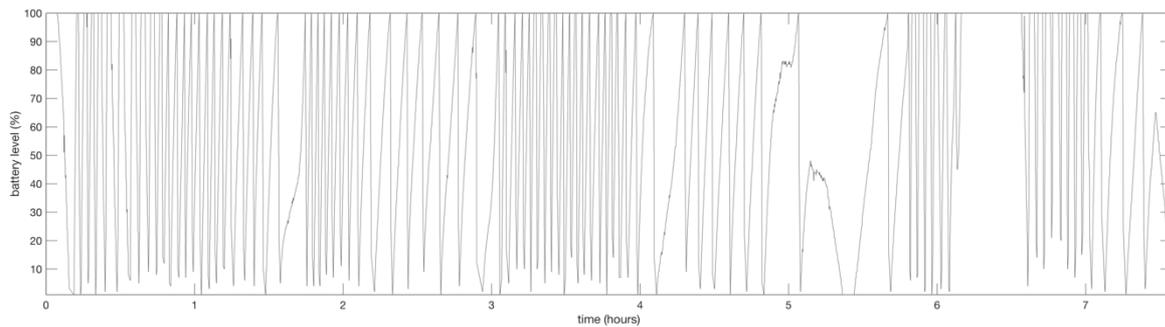


Figure 16 – The battery phenomenon is shown of a measurement containing almost 8 consecutive hours of the problem. The characteristic sawtooth shape of the battery level can be seen. The battery level can change either slowly or almost instantly from 100% to 0% and the other way around.

§4.5.2 The battery phenomenon

In early stages of this study a number of HealthPatch measurements showed problems recording data. Many short and long gaps, ranging from seconds to hours, were present in the data. Explorative analysis of the raw data files of these measurements revealed the battery level showed strange behaviour. The level could drop rapidly (and even instantly) from 100%, which is a normal level throughout the first days, to 0%. After some time at 0%, which could range from seconds to hours, the battery level instantly rose to 100% again and the measurement would continue normally. However, the battery dropouts would often occur shortly after one another. This resulted in a measurement with many short and long gaps, because the HealthPatch turned on and off repeatedly for arbitrary periods of time. It was unknown what the quality was of the sections of data that were available, especially the short sections of a few samples between battery dropouts.

The first time this phenomenon was witnessed, we thought it could have been caused by a faulty HealthPatch or by some electrical influence of the Spacelabs ECG electrodes that by chance happened to be located near the patch location¹². Unfortunately enough, we found the same strange battery behaviour in later measurements as well. Plots of the battery level versus time showed a sawtooth shape, which we named the ‘battery phenomenon’. Figure 16 shows an example of what the phenomenon looks like.

Various hypotheses for a cause were tested, but the battery phenomenon seemed to occur randomly and all attempts to prevent the phenomenon from occurring appeared futile. We contacted Vital Connect’s technical support, who after inspection of the data and some unsuccessful advice suggested that the phenomenon could be caused by the battery design. As described in Chapter 2, the patch incorporates a zinc-air cell battery as its power source. It requires oxygen for the redox with zinc that generates the electric potential for the patch to function. Vital Connect hypothesised the oxygen supply was deficient in the patch positioned over the left midclavicular line near the 6th and 7th intercostal spaces. They suggested we’d perform future measurements with the patch at a 45° angle on the left m. pectoralis major. Since their advice, all measurements were performed with the patch on the suggested application site. The battery phenomenon has never occurred since. Our current hypothesis is that the previous patch location on the skin was (more) prone to obstructing airflow towards the battery due to clothing, blankets, or the patients themselves. Fact is that not only our patient require oxygen.

¹² The patient in question had wounds near the regular Spacelabs electrode locations, hence the different placement.

Because shortly after discovery of the battery phenomenon (but not yet the solution) a lot of HealthPatches showed problems. The resulting large amount of data loss was unacceptable. Therefore, it was decided to study the reliability of HealthPatch measurements. Later, after the solution was found and the amount of data loss went down, this reliability analysis was redone. The results of these analyses of HealthPatch reliability can be found in Appendix B.

From all the aforesaid, it may be concluded that the position under the nipple is more likely to suffer from oxygen depletion of the battery than the more cranial location on the chest. The oxygen depletion could be caused by the patient's blankets and clothing covering the patch more often and more thoroughly. Also, a patient laying on their side or in pronation, which may happen e.g. while washing, could cause suffocation of the battery. Furthermore, it happened a few times that the active grade HealthPatch, which is not meant to be reapplied after removal, was reapplied and did not adhere too well any more. Consequently, the nurse put a transparent sticker (Tegaderm) over the patch, which impeded the oxygen supply enough to cause the phenomenon. Although Tegaderm is permeable for oxygen, it apparently impedes the supply enough to cause the phenomenon.

These practical implications of unforeseen circumstances and human behaviour indicate how challenging the design of a wearable vital signs monitor can be. It also shows that it is important to research and develop in close cooperation with manufacturers, researchers, technicians, nurses, clinicians and patients. If this technology is to be applied out of the hospital, e.g. in a patient's home, it is probable that more unforeseen practical issues will emerge.

§4.5.3 Annotating plots of participations or measurements

In the current implementation of the methods to plot the vital signs, no support was built in yet to add annotations to the graphs. For further studies, it is recommended information in annotations is added to the plots. This can be very informative when assessing the data visually. For example, when a patient with stable heart rate and respiratory rate is also stable according to their medical record, it may be worth annotating that moment or period in time. It is even more interesting to annotate medical findings that may influence the vital signs. For example, when patients develop a cardiac arrhythmia, that will be seen by nurses on the continuous Spacelabs monitor, resulting in a potential annotation, and confirmed by doctors, resulting in another annotation. Any treatments or interventions could be recorded as annotations as well. Complications, such as suspected infections or anastomotic leakage, would be very interesting to record along the HealthPatch measurements. If all that clinical information also shows up in the plots, it may help indicate where interesting events happen and recognise what trends or patterns can be seen in the time around such events. This may aid future developments towards the recognition of the deteriorating surgical patient.

The task of annotating plots with manually extracted information from the medical records can be tedious. It may be aided by future developments in natural language processing to extract information from the EMR automatically [55].

§4.5.4 Bland-Altman analysis

Autocorrelation in repeated measurements violates the assumption of independence

In an experiment where multiple measurements are performed on the same individual by two methods—instead of just one measurement per method per individual as with regular BAA—the assumptions of BAA are possibly violated. As with many analysis methods for sampled data, in BAA it is assumed that samples are taken independently. Repeated measurements of vital signs from the same individual likely violate this assumption when the sample rate is high relative to the natural changes of the measurand. For example, your current heart rate is probably about

the same as your heart rate of one minute ago; heart rate is autocorrelated, hence heart rate samples taken shortly after one another are not independent: knowledge of a previous sample provides information about a following sample. Repeated measurements are likely to be correlated unless the measurand is truly random, which may not be possible for physiological quantities. Moreover, zero correlation does not imply independence, since two variables may be dependent in a way correlation, which merely is a measure of linear association, cannot detect, e.g. non-linear associations.

To correct for autocorrelation or autoregression, as it is more generally called (more on this subject in Chapter 6, the data should be decorrelated to a point where it is no longer considered to be autocorrelated. The question is if this leads to analyses that can practically answer the questions being investigated.

To illustrate this, consider the following example. Heart rate is measured every minute throughout a day, just like in this study. Heart rate values are autocorrelated as exemplified before. The autocorrelation function (ACF) can be estimated using the methods from §4.3.3, specifically using the autocovariance function. The resulting graph will show high (~80–100%) correlation at and near zero lag and low correlation at lags further away from zero. In this graph, a certain threshold of, say, 10% could be used to find the lag, i.e., the time at which we consider the data to be uncorrelated to itself. For the purpose of the example, let's say the interval found is 3 hours. Then, samples taken independently at 3-hour intervals can be considered to have low autocorrelation. Sampling at this lower frequency will however diminish the time resolution of the measurement. Originally, we were interested in measuring every minute for good reasons, e.g. to be able to timely detect acute and critical changes in heart rate. This temporal resolution is lost if we sample at a lower rate merely to obtain uncorrelated samples.

In the example, it is not possible to resample the minute-sampled data down to one sample per three hours. This would either incur the penalty of aliasing, or when using an anti-aliasing filter (see §4.3.2) the data would become correlated again due to every filter being some function of current and past samples, somewhat comparable to a weighted moving average filter.

The current study therefore did not consider techniques to remove autocorrelation from the measured time series, but treated them as if they had zero autocorrelation. How this affects the analyses is unknown, but it is probable the estimations of CI and LOA in BAA are affected.

Alternative agreement analysis methods

Besides the well-known Bland-Altman analysis and the variations explained and used in this chapter, other agreement analysis methods exist. For longitudinal data, such as used in this study, others have suggested methods to accommodate the classical BAA methods. Myles and Cui propose a method based on methods commonly used in social studies: random effects models [56]. In their methods, the studied subjects are treated as different groups of observations, similar to the ANOVA method in BAA for repeated measurements. The individual observations within one subject are treated as variables that change with time, where time is a predictor in the model too. It is a different method to estimate the variation within each subject, which is indeed the factor that makes the regular BAA methods inappropriate. The random effects modelling method also allows to adjust for other covariates that the researcher deems relevant in the assessment of agreement between the two methods. As such, the random effect model is an extension of the ANOVA method. This study did not consider the random effects model, but it is recommended to study its use and methods in future agreement analyses. A bootstrapping study by Van Loon et al. showed that the random effects model by Myles and Cui produces comparable results to Bland and Altman's repeated measurements method, although the LOA range can be a bit smaller [57].

Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPREZZON

*People love chopping wood
in this activity one immediately sees results*
Albert Einstein

The previous chapter described the measurement methods, raw data storage and pre-processing, signal analysis and Bland-Altman analysis (BAA) methods used in this study. This chapter presents and analyses the obtained results.

§5.1 Measurement results

§5.1.1 Results of the patient sample

In total, 33 patients were included. Their vital signs were measured in 35 participations, as two patients were admitted more than once and could be included a second time. Therefore, the following patient characteristics are calculated for the 35 participations, not for the 33 patients. In other words, the patients are treated as a group of 35, where two of them are included twice.

Patient age at the start of the participation ranged from 23–80 years and was 60.8 years on average. Female-to-male ratio was 2:5. Patient length, weight and BMI were not available in the EMR of all patients. When two of the three parameters were available, the third could be calculated. As a result, length was known for 31 patients and was 1.77 m on average. Weight was available for 32 patients with an average of 81.40 kg. BMI was known in 31 cases and was 26.0 kg/m² on average. ASA (American Society of Anesthesiologists) score was unavailable for many participations, because some patients underwent emergency surgery where determination of the score was usually not possible. Additional descriptive statistics are listed in Table 3. Table 4 lists these statistics for individual patients per participation. Note that because two patients participated twice, their statistics are listed twice as well, although some statistics may change from one participation to another.

Of the patient sample with 35 participations, a total of 36 HealthPatch measurements and 35 Spacelabs measurements were obtained for a total of 71 measurements. There was one extra HealthPatch measurement, because one patch failed to transmit any useful data (see also the discussion about the ‘battery phenomenon’ in §4.5.2), hence it was replaced with another as soon as the problem was discovered. The patient was still expected to stay on the ward for a long enough time to start a second HealthPatch measurement. Because of the database design

Table 3 – Included patient characteristics are listed as descriptive statistics. BMI: body-mass index. IQR: inter-quartile range. ASA: American Society of Anesthesiologists.

| Characteristic | Number | Mean | Standard deviation | Median | IQR | Range (min–max) |
|--------------------------|--------|------|--------------------|--------|------|-----------------|
| Patients | 35 | – | – | – | – | – |
| Male | 25 | – | – | – | – | – |
| Female | 10 | – | – | – | – | – |
| Age (years) | 35 | 60.8 | 13.6 | 63 | 13.8 | 23–80 |
| Length (m) | 31 | 1.77 | 0.09 | 1.75 | 0.12 | 1.57–1.96 |
| Weight (kg) | 32 | 81.4 | 16.6 | 82.5 | 18.3 | 47.7–145 |
| BMI (kg/m ²) | 31 | 26.0 | 4.5 | 25.7 | 4.2 | 17.6–40.2 |
| ASA score | 20 | 2.7 | 0.8 | 3 | 1 | 1–4 |

as described before, it is allowed to have any number of measurements of any number of wearables within one participation.

The total duration of all Spacelabs measurements was over 65 days. The HealthPatch yielded 70 days, 22 hours and 45 minutes of measurements. On average, this was 1 day and 20 hours

Table 4 – Individual patient characteristics are listed per participation. BMI: body-mass-index. ASA: American Society of Anesthesiologists.

| Participation No | Sex | Age (years) | Length (m) | Weight (kg) | BMI (kg/m ²) | ASA Score |
|------------------|--------|-------------|------------|-------------|--------------------------|-----------|
| 8 | male | 48 | 1.75 | 88.6 | 28.93 | 1 |
| 9 | male | 76 | 1.8 | 76.3 | 23.55 | unknown |
| 10 | male | 59 | 1.72 | 62.7 | 21.19 | unknown |
| 11 | male | 58 | 1.78 | 66 | 20.83 | unknown |
| 12 | female | 58 | 1.75 | 68 | 22.20 | unknown |
| 13 | male | 28 | 1.89 | 92 | 25.76 | unknown |
| 14 | male | 65 | 1.78 | 85.1 | 26.86 | 3 |
| 15 | male | 53 | 1.67 | 67 | 24.02 | 2 |
| 16 | female | 64 | 1.69 | 70 | 24.51 | unknown |
| 17 | male | 75 | – | – | – | unknown |
| 18 | male | 75 | – | 81 | – | unknown |
| 19 | female | 73 | 1.68 | 75 | 26.57 | unknown |
| 20 | male | 71 | 1.7 | 85 | 29.41 | unknown |
| 21 | male | 71 | 1.8 | 74 | 22.84 | 2 |
| 22 | male | 63 | 1.9 | 145 | 40.17 | 4 |
| 23 | female | 57 | 1.67 | 100.5 | 36.04 | 3 |
| 24 | male | 23 | – | – | – | 3 |
| 25 | female | 61 | 1.75 | 84 | 27.43 | 3 |
| 26 | male | 71 | 1.8 | 57 | 17.59 | 2 |
| 27 | male | 66 | – | – | – | unknown |
| 28 | female | 63 | 1.65 | 81 | 29.75 | 4 |
| 29 | female | 37 | 1.90 | 87.7 | 24.29 | 2 |
| 30 | female | 37 | 1.90 | 87.7 | 24.29 | 2 |
| 31 | male | 75 | 1.75 | 78.8 | 25.73 | 2 |
| 32 | male | 63 | 1.72 | 90 | 30.42 | 3 |
| 33 | male | 67 | 1.69 | 70 | 24.51 | 3 |
| 34 | male | 77 | 1.78 | 87 | 27.46 | 3 |
| 35 | male | 52 | 1.96 | 92 | 23.95 | unknown |
| 36 | female | 58 | 1.57 | 47.65 | 19.33 | 2 |
| 37 | male | 70 | 1.82 | 85.2 | 25.72 | unknown |
| 38 | male | 49 | 1.84 | 88 | 25.99 | unknown |
| 39 | male | 80 | 1.88 | 97.7 | 27.64 | unknown |
| 40 | male | 63 | 1.76 | 69.5 | 22.44 | 3 |
| 41 | male | 60 | 1.74 | 90.5 | 29.89 | 4 |
| 42 | female | 62 | 1.64 | 75 | 27.89 | 3 |

per patient for Spacelabs and a little over 2 days for HealthPatch. Three tables in Appendix D list the numbers of samples per participation of the raw, uniformly sampled and ensemble (both HealthPatch and Spacelabs resampled, synchronised and trimmed to each other) data sets; the second and third table also list total measurement duration.

Because measurements could contain gaps, it was of interest to quantify the amount of data loss. This loss is expressed in time of measurement without data and can be expressed as a percentage of the total measurement duration as well. The data loss of the HealthPatch measurements is summarised in Table 5. This table shows the data loss duration (time without data) and the percentage of the total measurement time that was lost. Because most measurements had no data loss, the median data loss also was zero. For this reason, a subset of measurements was taken to calculate the same statistics: only HealthPatch measurements with nonzero data loss were considered in the subset.

For reference, plots of all vital signs as measured by HealthPatch MD and Spacelabs XPRESZON in this study are included in Appendix C. As an example, one participation is shown in Figure 17. The figure shows the vital signs as measured by both Spacelabs and HealthPatch during one participation. Heart rate and respiratory rate are graphed in different panels. The vertical axes show heart rate in bpm (top panel) and respiratory rate in brpm. The horizontal axes are identical and show date and time: the participation took place over the course of almost 4½ days. Of this participation, the case is described below.

Case study: one participation explored

The participating patient is a 60-year-old male that underwent open nephrectomy for suspected carcinoma in situ one day before the start of the participation. The patient has an extensive cardiac and vascular medical history, including restrictive cardiomyopathy, pulmonary hypertension, obesity, diabetes mellitus type 2, peripheral vascular disease with most notably lower leg amputation and renal insufficiency.

On the day of surgery, atrial flutter (AF) was seen. The patient was admitted to the surgical medium care ward for hemodynamic monitoring on the day after surgery. For heart rate and respiratory rate, not many interesting developments transpired on this day. However, on the next morning (September 8th) the patient required more norepinephrine to maintain tension and elevations of heart rate were seen. Later that morning AF was seen again. However, the ECG did not seem worse than what was seen before. In the evening the heart rhythm became very irregular with tachycardia up to 110 bpm. A cardiologist concluded the same AF as seen before was the main cause. On September 9th, some bradycardias as low as 44 bpm were seen during haemodialysis. The next day, even slower heart rhythms of 35–40 bpm were witnessed.

Table 5 – HealthPatch measurement data loss statistics are summarised. The lost time is given in hours:minutes:seconds format. The percentage indicates how much of the total measurement duration is lost.

| Data set | Total loss | Mean loss | Median loss | Min loss | Max loss |
|---|----------------------|---------------------|--------------------|--------------------|---------------------|
| All HealthPatch measurements (<i>n</i> = 36) | 101:15:24 (5.9%) | 02:48:45 (5.9%) | 00:00:00 (0%) | 00:00:00 (0%) | 59:59:08 (52.4%) |
| Subset: HealthPatch measurements with any data loss (<i>n</i> = 17) | 101:15:24 (15.5%) | 05:57:22 (15.5%) | 00:53:36 (2.3%) | 00:00:08 (0.0%) | 59:59:08 (52.4%) |

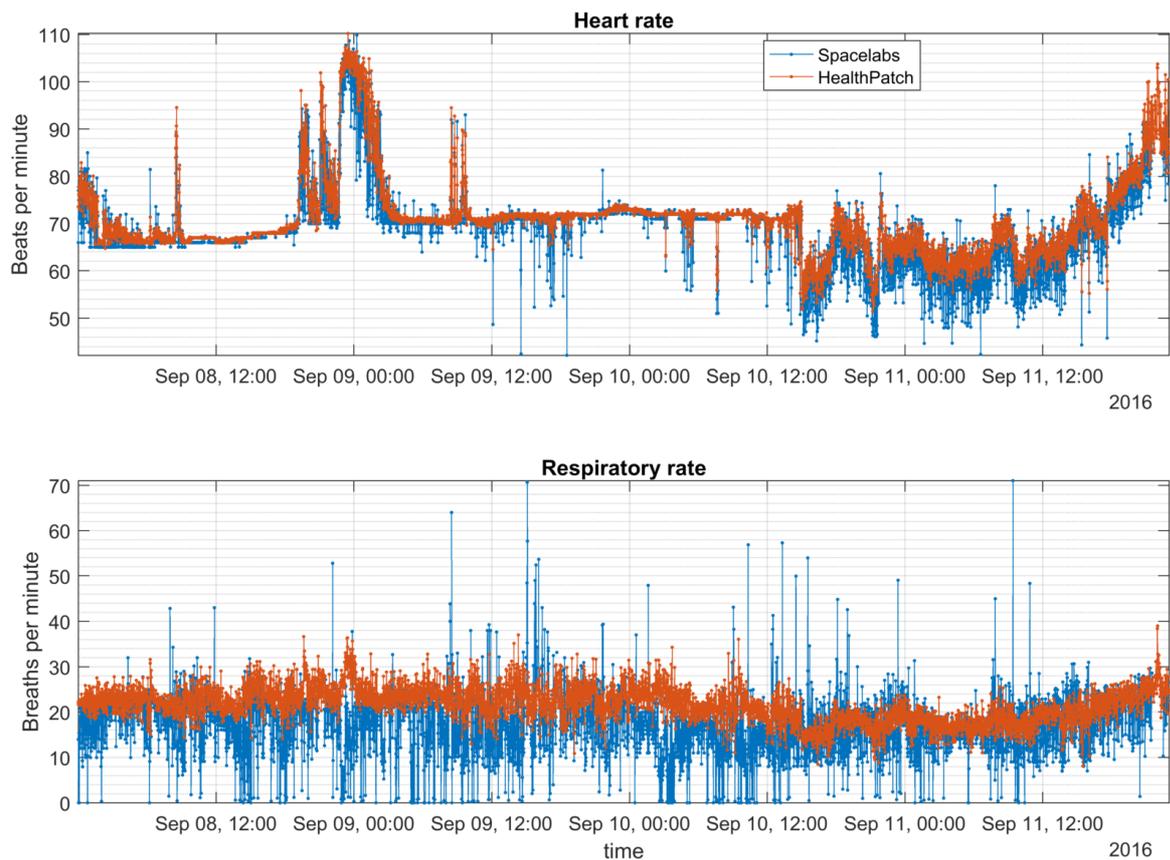


Figure 17 – A participation of multiple days is shown. The upper panel shows heart rate in beats per minute, the lower respiratory rate in breaths per minute. Both the measurements by HealthPatch (red) and Spacelabs (blue) are shown.

Throughout September 10th and 11th, AF and bradycardias were regularly seen. Norepinephrine was still required and was gradually increased over the course of the day. During the day, the patient declined slowly until the nurse became worried and activated a urologist to see what was going on. The doctor suspected haemorrhage and around midnight of September 12th an angiographic procedure confirmed the diagnosis. The right testicular artery was coiled subsequently. After this procedure, the patient was admitted to the IC and the participation ended.

To illustrate the course of this patient’s physiological decline, consider the Early Warning Score (EWS) values that were recorded during this patient’s admission on the surgical medium care: Figure 18 shows these values. In Chapter 6, §6.1 goes into more detail about the EWS.

§5.2 Agreement analysis results

The results of the agreement analysis as described in §4.4.2 are presented in this section. The resampling and observation pair matching required for these analyses resulted in partial exclusion of data (see also Table 11 in Appendix D). After exclusion of data points without observation pairs in either HealthPatch or Spacelabs measurements, the data of 26 of the 35 participations remained. This may seem as a large exclusion, but this was not caused by the observation pair matching. The actual reason for exclusion of these participations was that, after resampling, zero data points were available in both the HealthPatch and Spacelabs measurements. Reasons for this were that either the Spacelabs or the HealthPatch data files were unavailable (mostly Spacelabs, see also Appendix D), so no common measurement times were possible; or that the measurements were either too short or contained relatively many gaps, which made the resampling process impossible without removing all data (see also §4.3.2).

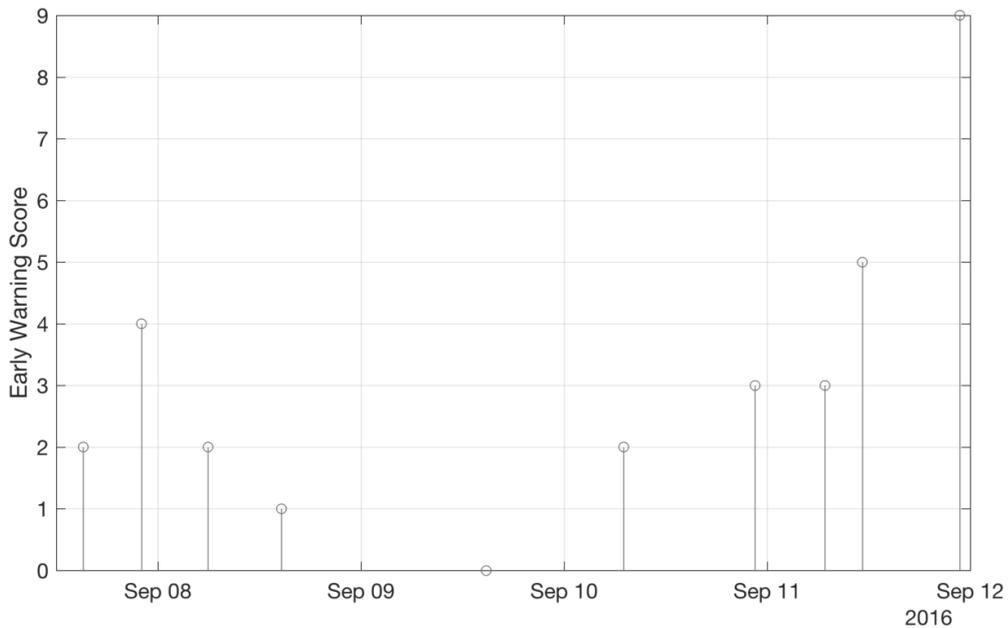


Figure 18 – The Early Warning Score throughout the course of the case of Figure 17 is plotted. Note the increase to a value of 9 in the last 48 hours of the participation.

In total, $N = 55565$ observation pairs were available of heart rate and $N = 56674$ observation pairs were available for respiratory rate. These total numbers were distributed between the 26 patients as follows. The numbers of samples per patient ranged from 249 to 5501 for heart rate and for respiratory rate the numbers ranged from 243 to 5507. On average, there were 2137.1 and 2179.8 observation pairs per patient respectively. The numbers of observation pairs are listed in Table 6.

§5.2.1 Agreement analysis of heart rate

Regular Bland-Altman analysis of HealthPatch MD and Spacelabs heart rate

Regular BAA on all data as if each observation pair came from an individual subject resulted in a bias of $\bar{D} = 1.13$ bpm and 95% LOA ranging from -6.38 bpm to 8.64 bpm on the domain of about 40–140 bpm. The range of the LOA is $8.64 - -6.38 = 15.02$ bpm. Using the agreement criterion of 10% on either side of the bias, the agreement of HealthPatch MD with Spacelabs is acceptable for heart rates over $10 \cdot 15.02/2 = 75.10$ bpm. Figure 19 shows the mean-difference plot created with regular BAA performed on all data.

In a regular mean-difference plot used in BAA a scatter plot of the observation pairs would be made. However, due to the very large number of samples, this cannot be visualised. Therefore, a honeycomb plot was used instead. A honeycomb plot is a two-dimensional binning method using hexagonal bins instead of regular square bins. Hexagonal bins have the advantage that they are rounder and fit into one another to form a compact grid of bins. The bin colour indicates the number of samples in the particular bin. Altogether, the colour cloud in the honeycomb plot reveals the underlying distribution of the data. Because no (correct) MATLAB implementation existed to create honeycomb plots, one was written and published. The details of this can be found in Appendix E.

Note that the graph's vertical axis, i.e., the difference, is limited from about -9 bpm to about 11 bpm for better visibility of the bias and LOA lines and the point cloud. Although some observations lie out of the graph's vertical range, most data points are plotted; in fact, 95% of the observations lie between the LOA, so more than 95% of the data is shown in the figure. Note that the 95% CI of the bias and LOA are graphed as red error bars on the very right edge of the plot. The CI are tiny, because their calculation is based on division by $\sqrt{N} \approx 236$, which is a large number relative to $\sqrt{n} \approx 5$. Furthermore, the line of equality (zero difference) is drawn for reference. Lastly, the graph's legend shows the Spearman rank correlation of the differences with the means and its p -value. The correlation is significant and is slightly negative (-0.16).

A large amount of data points is seen near the bias line. Because of the large amount of points, the centre of the point cloud is in itself clouded by the points, which makes it difficult to see individual points. Near the right end of the graph a group of points appears to be concentrated, while looking a bit separated from the main point cloud. This lesser cloud could be the result of a number of observations from one patient near an average heart rate about 130–140 bpm.

Table 6 – The numbers of observations per patient available for Bland-Altman analysis are listed. Every sample represented a minute's worth of data.

| Patient No | Number of observation pairs | |
|------------|-----------------------------|------------------|
| | Heart rate | Respiratory rate |
| 1 | 5230 | 4584 |
| 2 | 1114 | 1304 |
| 3 | 741 | 758 |
| 4 | 5501 | 5507 |
| 5 | 910 | 1512 |
| 6 | 3947 | 4891 |
| 7 | 1989 | 2019 |
| 8 | 2307 | 2555 |
| 9 | 2356 | 2392 |
| 10 | 1121 | 1124 |
| 11 | 4790 | 4685 |
| 12 | 2113 | 2181 |
| 13 | 1038 | 1049 |
| 14 | 1312 | 1352 |
| 15 | 2306 | 1930 |
| 16 | 4263 | 4276 |
| 17 | 1342 | 1343 |
| 18 | 249 | 243 |
| 19 | 1900 | 1727 |
| 20 | 1103 | 1208 |
| 21 | 2206 | 2233 |
| 22 | 547 | 561 |
| 23 | 1903 | 1920 |
| 24 | 1429 | 1438 |
| 25 | 1414 | 1432 |
| 26 | 2434 | 2450 |

Mean-difference plot of observations ($n = 55565$ subjects and observation pairs)

M_1 : HP HR (bpm), M_2 : SL HR (bpm)

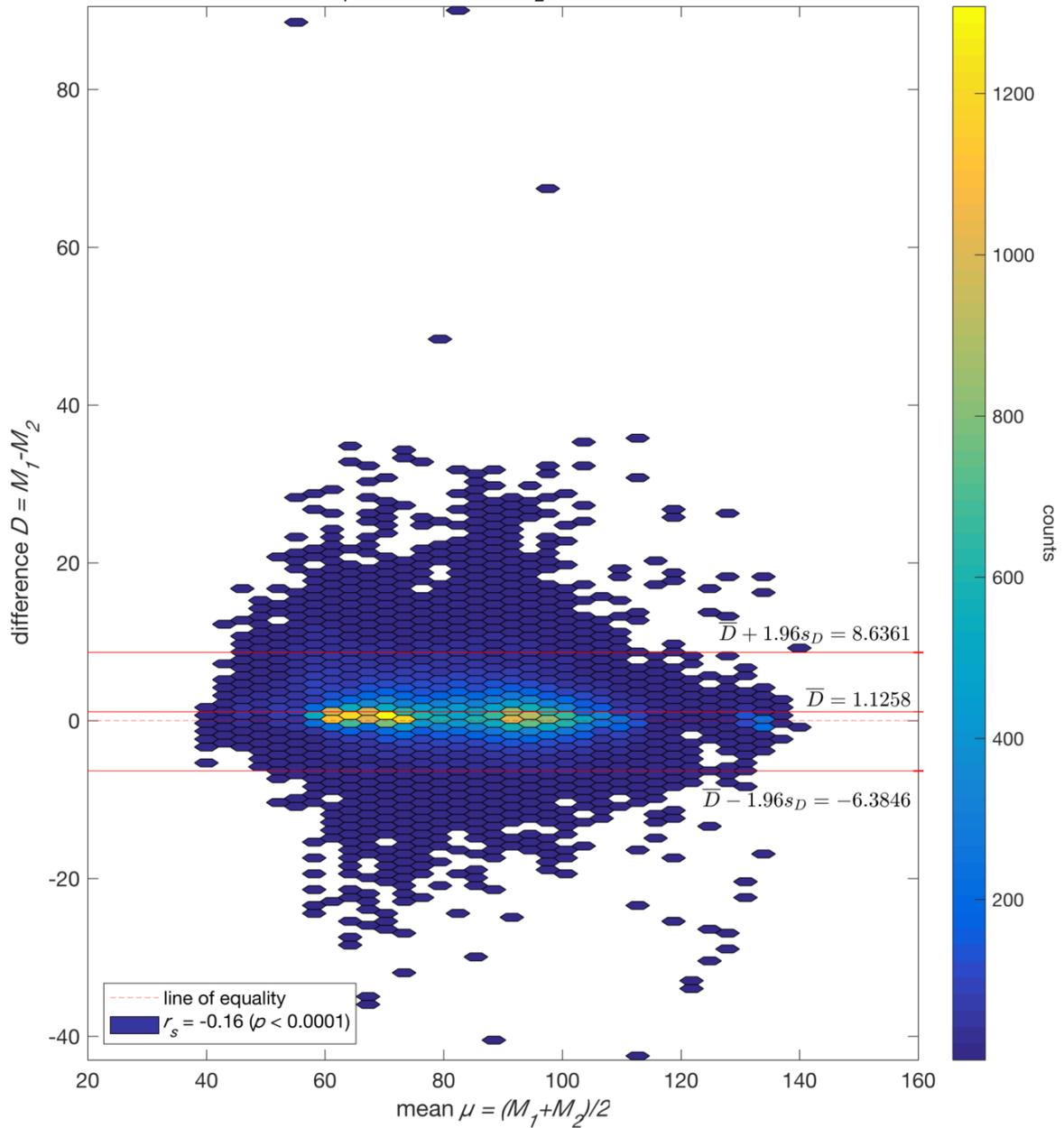


Figure 19 – The mean-difference plot of regular Bland-Altman analysis of heart rate is shown. The analysis was performed with all heart rate data, as if each data point was an individual observation from an individual subject. Because the number of samples is large, a honeycomb plot is made instead of a scatter plot. The agreement between the two measurement methods (M_1 and M_2) is given by the bias and the limits of agreement (LOA), drawn as horizontal lines. Note the tiny 95% confidence intervals to the right end of the lines for the 95% limits of agreement and bias: they are so small due to the large number of observations; they appear as flat lines (compare with the next figure). HP: HealthPatch. HR: heart rate. SL: Spacelabs. r_s : Spearman rank correlation.

Mean-difference plot of observations ($n = 26$ subjects, $\Sigma m = 55565$ observation pairs)

M_1 : HP HR (bpm), M_2 : SL HR (bpm)

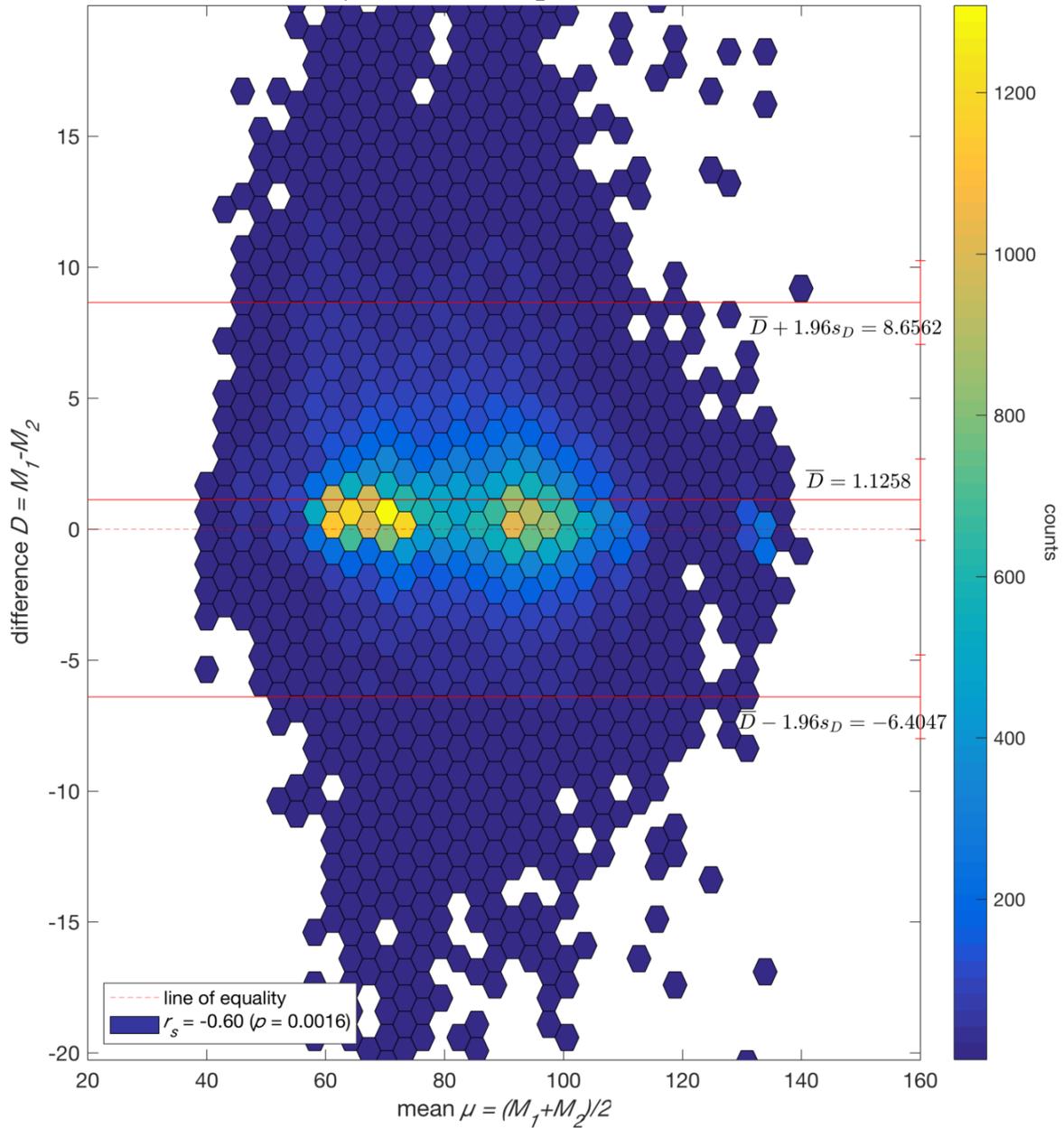


Figure 20 – The mean-difference plot of Bland-Altman analysis for repeated measurements of heart rate in beats per minute (bpm) is shown. The bias and limits of agreement (LOA) are shown as horizontal lines. The 95% confidence intervals of the bias and LOA are shown as error bars on the right ends of the lines. Note the vertical axis has been scaled compared to Figure 19. HP: HealthPatch. HR: heart rate. SL: Spacelabs. r_s : Spearman rank correlation.

Mean-difference plot of subject means ($n = 26$ subjects, $\Sigma m = 55565$ observation pairs)
 M_1 : HP HR (bpm), M_2 : SL HR (bpm)

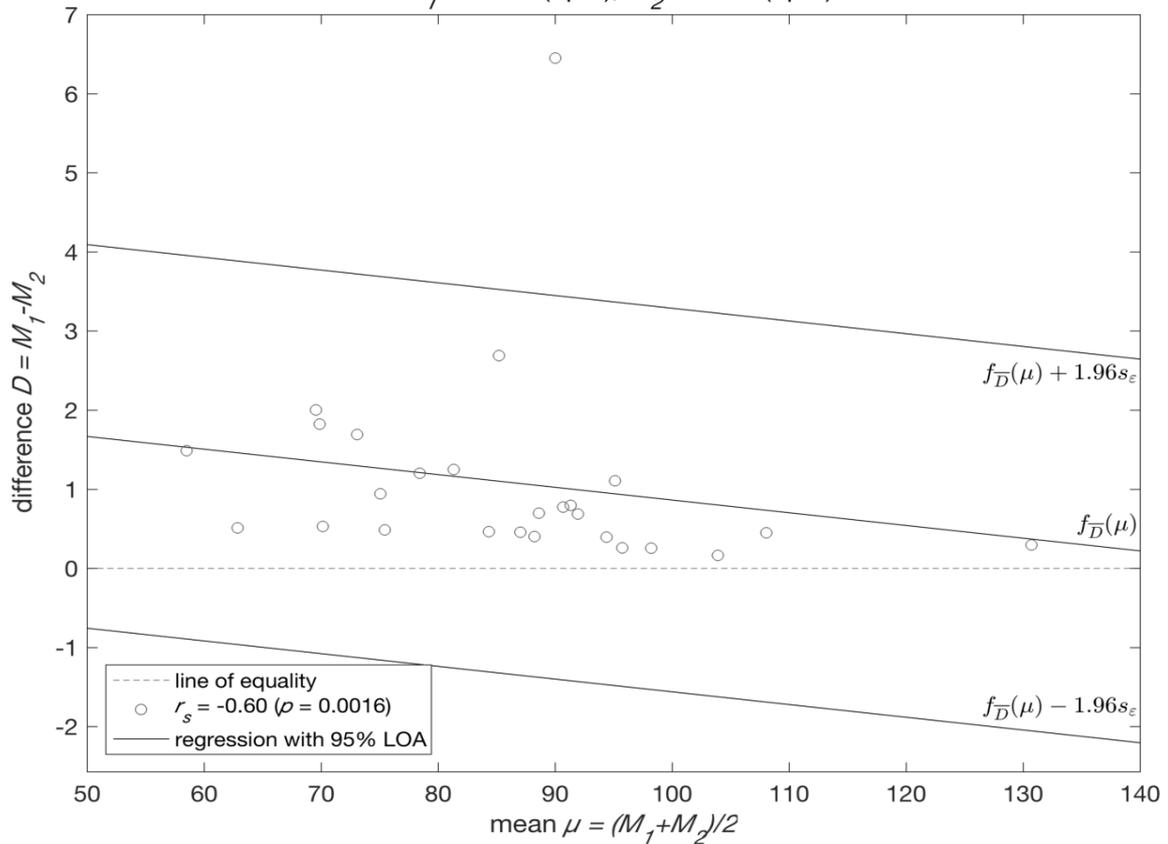


Figure 21 – Bland-Altman analysis using simple linear regression lines for bias and limits of agreement is shown. Instead of plotting all observations, only patient means are plotted. HP: HealthPatch. HR: heart rate. Bpm: beats per minute. SL: Spacelabs. r_s : Spearman rank correlation.

Bland-Altman analysis for repeated measurements of heart rate by HealthPatch MD and Spacelabs

Figure 20 shows the mean-difference plot of BAA for repeated measurements of heart rate while accounting for changing true values. The difference with the previous figure is clear: there are much less data points to be plotted, because only the subjects' mean differences are plotted versus their mean. The same elements from Figure 19 are plotted in Figure 20. The bias was $\bar{D} = 1.13$ bpm, which equals the bias of the regular BAA method due to the identical calculation. The LOA are slightly different, but similar nonetheless, at -6.40 and 8.66 . The biggest difference is seen in the 95% CI of the bias and LOA, which have become much larger. Using the CI of the bias, it cannot be concluded that the bias is different from the line of equality at a 5% significance level, because the line of equality lies within the interval.

The domain in the graph ranges from about 60 bpm to about 130 bpm. The narrower domain compared to the previous figure is a result of the averaging of values within each subject. One subject had an average heart rate of over 130 bpm, which is concordant with the group of data points around 130 bpm in the previous figure.

Like in the regular BAA, the Spearman rank correlation is found to be negative in the repeated measurements analysis; it is -0.60 with a p -value of 0.0016 . Because both analyses point out a negative linear trend in the differences with increasing mean, it may be interesting to adjust the bias and LOA lines to these trends. Two analyses were done in which both bias and LOA were calculated as simple linear regression lines of difference on mean. In one analysis, the residual variance was assumed to be constant, in the other it was assumed to be variable. Because the

residual variance did not appear to be variable, only the results of the constant residual variance analysis are given here. These results are summarised in Figure 21, in which the mean values per patient are plotted instead of using the honeycomb plot of the previous figures. This has no influence on the calculation of the bias or LOA, but is a different way of visualising the data set. Bland and Altman also plot the subject means in their articles [51,53].

The bias in this figure a simple linear regression line; the intercept is 2.47 bpm and the slope is -0.016 (dimensionless). The line ranges from 1.67 bpm to 0.22 bpm over the 50–140 bpm domain. The upper and lower LOA regression lines are the bias line plus or minus a constant based on the standard deviation of the residuals. This constant only influences the intercept. The upper 95% LOA has an intercept of 4.90 bpm and the lower 95% LOA's intercept is 0.05 bpm. Note the LOA are closer to the bias than in the constant bias and LOA method of the previous figure. The LOA are 2.42 bpm above and below the bias regression line.

Using ratio R instead of difference as the agreement statistic resulted in a bias of $\bar{R} = 1.017$ (dimensionless) and 95% LOA of 0.893 and 1.140.

Instead of analysing the entire heart rate range, the lowest and highest heart rates were of interest to investigate as well. Applying BAA for repeated measurements on heart rates up to the 10th percentile (61.1 bpm) resulted in a non-significant bias of -3.1 bpm with 95% LOA of -24.4 bpm and 18.1 bpm ($n = 13$ subjects, $\Sigma m = 5483$ observation pairs). The same analysis on heart rates greater than the 90th percentile (101.9 bpm) resulted in a significant bias of 8.4 bpm with 95% LOA of -15.8 bpm and 32.6 bpm ($n = 21$ subjects, $\Sigma m = 4742$ observation pairs).

§5.2.2 Agreement analysis of respiratory rate

Regular Bland-Altman analysis of HealthPatch MD and Spacelabs respiratory rate

The results of the regular Bland-Altman method are shown in the mean-difference plot in Figure 22. The bias and 95% LOA are 2.28 brpm and -10.95 brpm to 15.52 brpm respectively. These limits of agreement are too wide to be considered as acceptable by the predefined limits of 2 brpm around the bias. The Spearman rank correlation between difference and mean is significantly negative at -0.16 .

Bland-Altman analysis for repeated measurements of respiratory rate by HealthPatch MD and Spacelabs

As with heart rate, a more appropriate method was BAA for repeated measurements assuming a varying true value. The results of this analysis are summarised in Figure 23. The bias is identically calculated as with the regular method bias; it is 2.28 brpm and it is not significantly different from zero based on the 95% CI. The lower and upper 95% LOA are -11.00 brpm and 15.56 brpm respectively, which are similar to the regular method LOA. The Spearman rank correlation between difference and mean is not significant anymore with a p -value of 0.31.

It can be seen that a clear heteroscedasticity is present in the central area of the honeycomb plot: low mean respiratory rate values show low differences on average than high mean respiratory rates, i.e., most low mean values lie close to the bias whereas high mean values do less so. Because of this, the repeated measurements analysis was extended by allowing non-constant bias and LOA to better fit the data. Using the simple linear regression approach, the mean-difference graph in Figure 24 was constructed. The bias regression line has an intercept of -0.49 brpm and a slope of 0.16 (dimensionless). The lower 95% LOA has an intercept of 3.11 brpm and a slope of -0.26 . The upper 95% LOA has an intercept of -4.06 brpm and a slope of 0.58.

Instead of analysing the entire respiratory rate range, the lowest and highest respiratory rates were of interest to investigate as well. Applying BAA for repeated measurements on respiratory rates up to the 10th percentile (10.9 brpm) resulted in a non-significant bias of 1.3 brpm

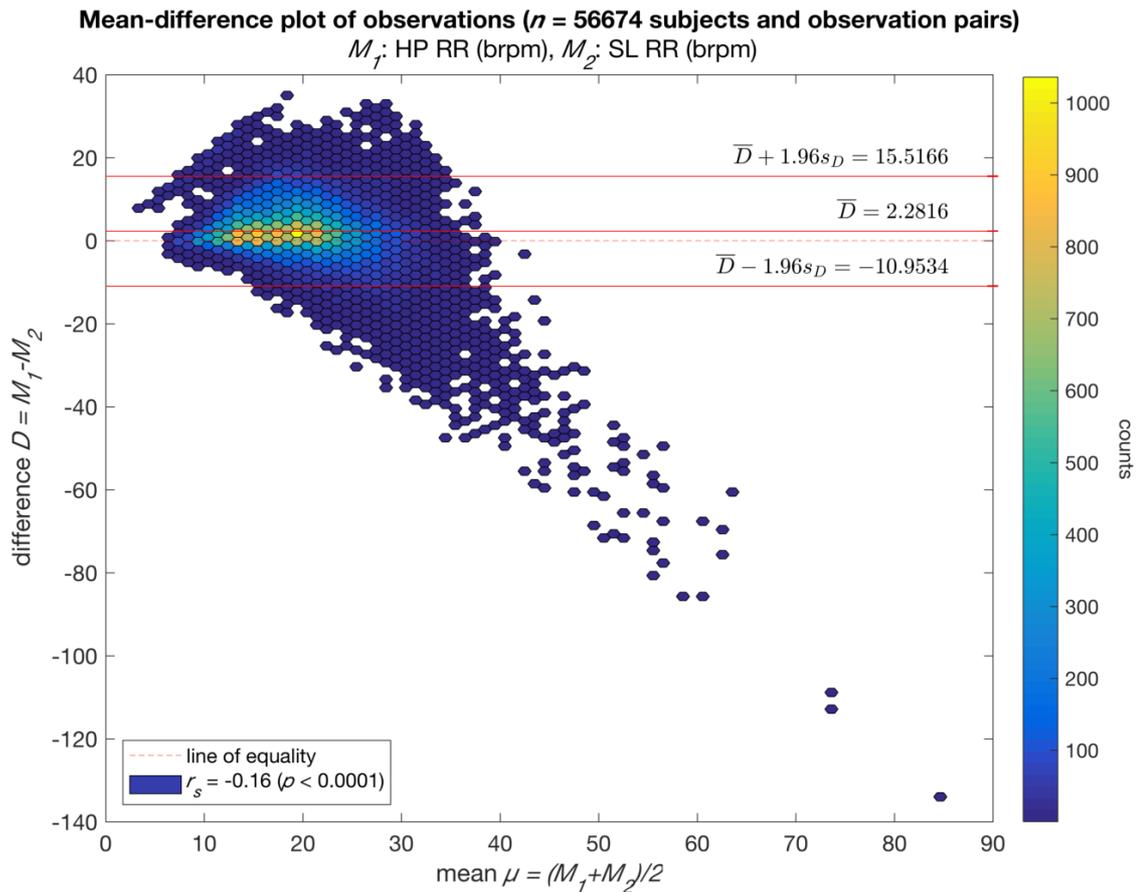


Figure 22 – Respiratory rate (RR) measured by HealthPatch (HP) and Spacelabs (SL) was analysed for agreement with regular Bland-Altman Analysis. The results are summarised in the plot, where the bias and 95% limits of agreement are shown as horizontal lines. Note the tiny 95% confidence intervals to the right end of the lines for the 95% limits of agreement and bias: they are so small due to the large number of observations; they appear as flat lines. Brpm: breaths per minute. r_s : Spearman rank correlation.

with 95% LOA of -15.3 brpm and 17.8 brpm ($n = 26$ subjects, $\Sigma m = 5554$ observation pairs). The same analysis on respiratory rates greater than the 90th percentile (27.0 brpm) resulted in a non-significant bias of 2.7 brpm with 95% LOA of -13.6 brpm and 19.0 brpm ($n = 26$ subjects, $\Sigma m = 4533$ observation pairs).

§5.3 Discussion

§5.3.1 Case study: exploration of vital signs of a deteriorating patient

The case study in §5.1.1 is a great example of how vital signs can aid in the recognition of a deteriorating patient. This particular patient underwent major surgery and had a relatively bad condition and comorbidity before surgery. Due to the complicated care and heavy procedures this patient needs, it is this kind of patient that has very high risk of not recovering well or experiencing an adverse event, as was the current case. Especially the last 48 hours of the participation show that the vital signs do contain some interesting features that may be sensitive for a deteriorating patient.

Heart rate was generally very stable throughout September 9th of the measurement, but on the 10th it became less stable and, at times, very low. Furthermore, from September 10th until confirmation of the complication, an upward trend emerged in heart rate. While these absolute

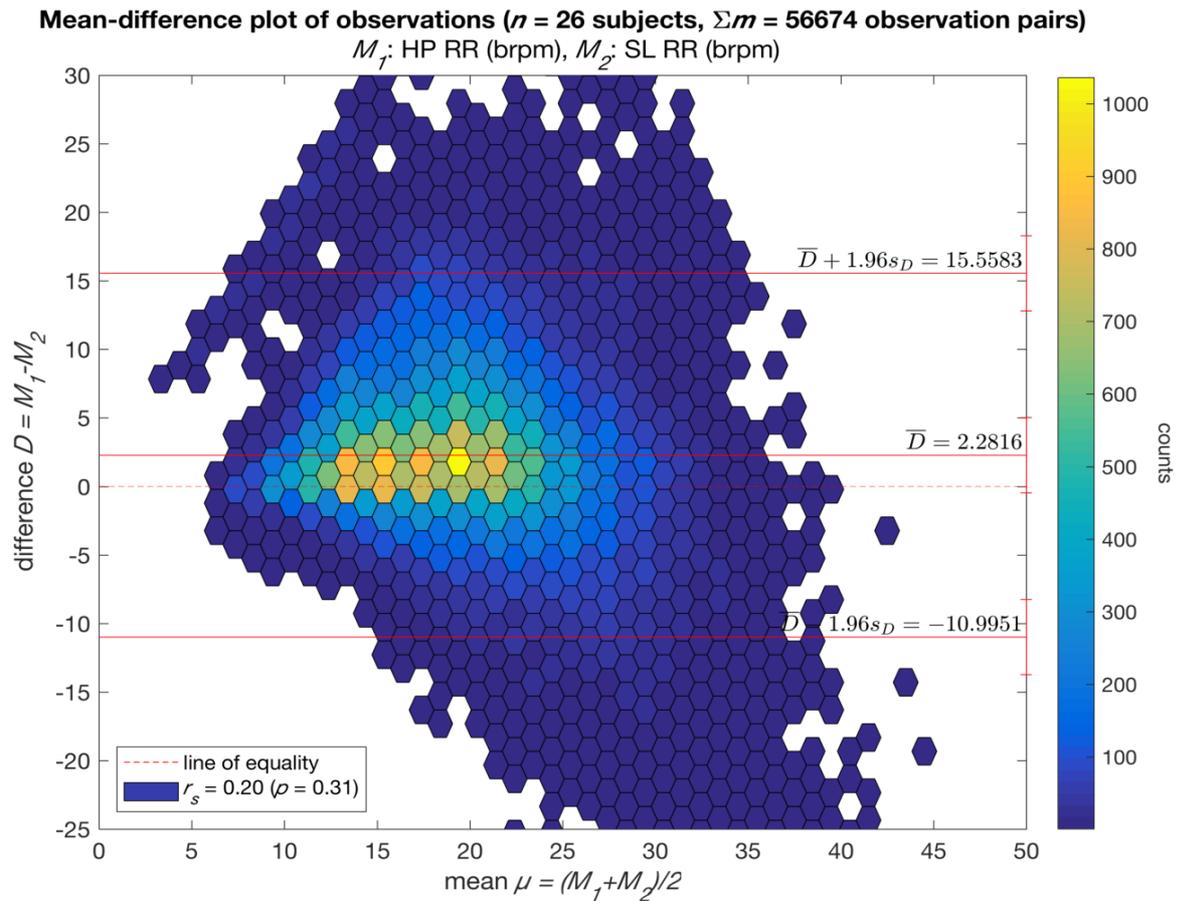


Figure 23 – The results of Bland-Altman analysis for repeated measurements of respiratory rate (RR) in breaths per minute (brpm) by HealthPatch (HP) and Spacelabs (SL) are shown. The bias, 95% limits of agreement and their 95% confidence intervals (error bars on the right edge) are drawn in the mean-difference plot. r_s : Spearman rank correlation.

heart rate values are not an indication of a (developing) complication, the dynamical changes may be informative.

Respiratory rate showed less information in general, because of the greater variance of the apparent high frequency noise compared to the slower signal components. Nevertheless, the last 48 hours of the respiratory rate measurements show the same upward trend.

Although it seems such trends are informative for developing adverse events, it should be noted that this particular patient specifically was a great example. Not only did the vital signs show interesting dynamical behaviour, but the participation was long as well and the measurements were very stable while the patient seemed to be too. This is not so much the case in the other measurements, in which almost all patients simply recovered from their surgery and were dismissed to a regular ward. Nonetheless, this specific case study shows that monitoring of vital signs with a wearable vital signs monitor such as HealthPatch MD may provide information that otherwise may not be available, e.g. at home or on a regular hospital ward.

§5.3.2 Agreement analysis

In this chapter, the results of agreement analysis applied to heart rate and respiratory rate measurements by HealthPatch MD and Spacelabs were presented. Agreement was found to be acceptable for heart rate, but not for respiratory rate. The following sections discuss the analyses performed on heart rate and respiratory rate, as well as some methodological considerations.

Mean-difference plot of subject means ($n = 26$ subjects, $\Sigma m = 56674$ observation pairs)
 M_1 : HP RR (brpm), M_2 : SL RR (brpm)

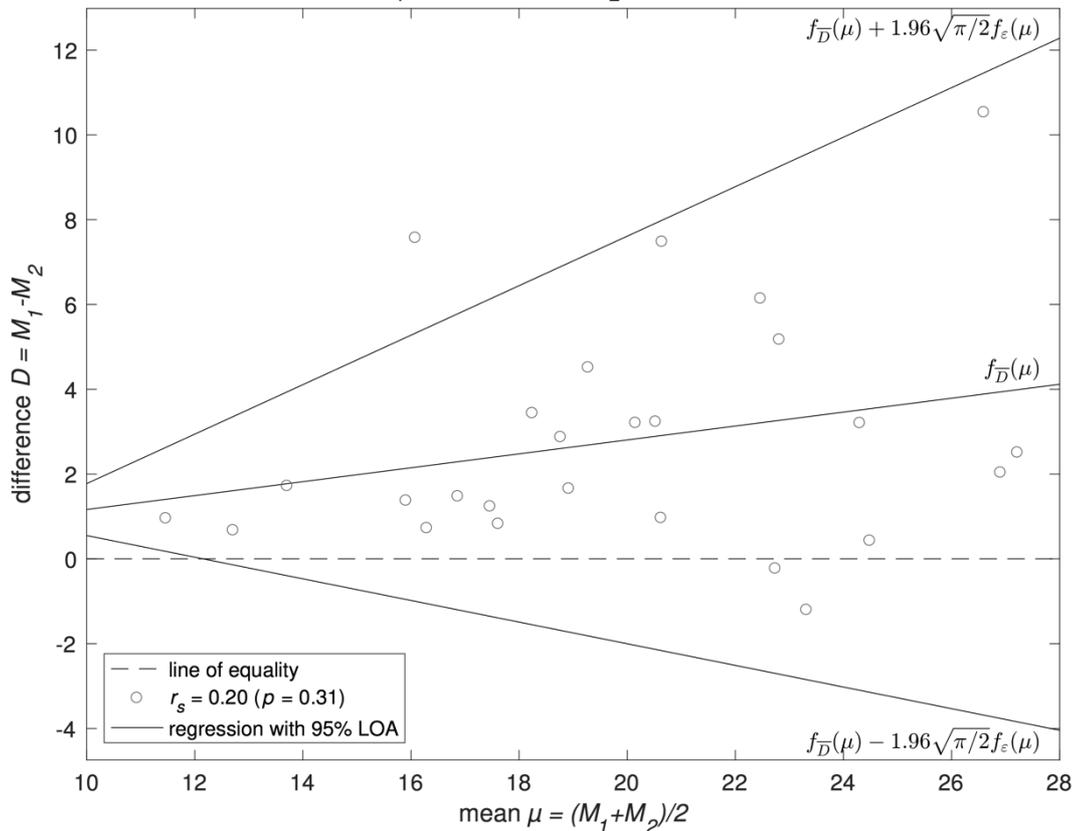


Figure 24 – Simple linear regression lines were used to fit the 95% limits of agreement and the bias to the observed respiratory rate (RR). The mean values per patient are plotted instead of a honeycomb plot for all observation pairs. HP: HealthPatch. Brpm: breaths per minute. SL: Space-labs. r_s : Spearman rank correlation.

Heart rate agreement was acceptable, but not for extreme values

The Bland-Altman analysis methods performed on the heart rate data showed acceptable agreement for most of the range of heart rate values. Five analyses were performed: 1) regular BAA, 2) BAA for repeated measurements of a changing true value, 3) simple linear regression modelling of bias and the LOA, 4) using ratio instead of difference as the agreement statistic, and 5) BAA on the lowest and highest lowest heart rates.

The first analysis was methodologically incorrect, but yielded insight in what the agreement could be and how the incorrect method performs, which then could be compared to a more correct method. The bias and LOA found were very close to those found in the second analysis. The bias and LOA do not always have to be similar when using these two methods. They were similar in this study, because the within-subject variance was not much greater than the between-subject variance (repeated measurements on one subject did not vary much more than the measurements from subject to subject). If this were not the case, the LOA estimated by the first analysis could become wider than those found in the second analysis. The latter analysis accounts for this effect [53].

The LOA appear to be too wide in the second analysis at first sight; the data points are not too far from the bias, but the 95% LOA encompass a much wider range. However, the LOA are correct. Consider an incorrect analysis to correct for repeated measurements: one could simply

take each subject's average difference and mean and perform regular BAA. This would result in too narrow LOA which may falsely lead to the conclusion of acceptable agreement.

In the third analysis, it was tried to more truthfully fit the data by allowing simple linear regression lines instead of constants to be fitted on the data. If the data truly shows a linear trend, this method allows the bias and LOA to better fit the data than in the analyses with constant bias and LOA. This is seen as a reduction of the LOA range. This was the case in the third analysis. Compared to the second analysis the 95% LOA range (difference between upper and lower LOA) went down from 15.02 bpm to 4.85 bpm. The latter result leads to the conclusion that HealthPatch agrees with Spacelabs heart rate over the entire range of observed heart rates (50–140 bpm). The downside of the regression lines is that they are more difficult to interpret compared to the constant bias and LOA. The constant values allow for easy understanding of the agreement over the entire observed range, whereas the lines require some calculations.

Because the acceptable range was defined as 10% above or below the bias, this range varies with the mean. A mean heart rate of 50 bpm accepts smaller LOA (± 5 bpm) than those at 130 bpm (± 13 bpm). For this reason, the fourth analysis was performed with ratio as the agreement statistic instead of difference. The ratio inherently shows the factor by which the methods agree at a certain mean heart rate, so LOA between ± 0.1 around the bias were considered acceptable agreement. The bias of $\bar{R} = 1.017$ and the 95% LOA of 0.893 and 1.140 (range: 24.70%) are marginally unacceptable at the predefined range of $\pm 10\%$.

The last analysis was done on the lower and upper end of the measured heart rate range. This was done because BAA does not quantify dynamic validity of the methods being compared. Quantifying agreement in the extreme ends of the data set is not a dynamic performance measure either, but it does provide information about the performance when it is most critical. For heart rate, BAA applied separately to the lowest and highest values, did not result in acceptable agreement between HealthPatch and Spacelabs: the 95% LOA were too wide in both analyses. This leads to the conclusion that either HealthPatch MD does not measure extreme heart rates as well as normal heart rates. This is a concern if this wearable is to be used in clinical practice and out of the hospital in the future. Primarily, extreme heart rates are of interest to diagnose the deteriorating patient. It is due to the large amount measurements in the normal heart rate range that heart rate agreement between HealthPatch and Spacelabs seemed acceptable.

Respiratory rate agreement was not acceptable

The respiratory rate agreement analysis results lead to the conclusion HealthPatch MD and Spacelabs do not agree within the limits of ± 2 brpm. There is still room for improvement in the pre-processing of the data, because some artefacts are present that could be removed for this analysis (but also read the discussion in §7.1.2).

Note the following peculiarities in Figure 22. In the top left part, a linear cluster of data points is visible which appears detached from the main point cloud. This linear area is an artefact due to observations where Spacelabs measured 0 brpm or a value very close to zero. When Spacelabs measures 0 brpm, the difference with the value measured by HealthPatch simply is that value. The mean of these sample pairs then always is half of the HealthPatch value. As such, these points draw a line with slope 2 in the top left corner of the graph. This artefact may influence the bias and LOA found. Removal of the corresponding data points would however be manipulation of the observed data. Because BAA aims to analyse agreement between two measurement methods, flaws in either method influence the results. Removal of these flawed observations is not recommended, because in practice they do exist and are not removed, thus their removal would embellish the results artificially.

If one would want to remove the artificially low measurements by Spacelabs, the mean-difference graph can be used to determine the threshold below which to exclude samples. Note that an area exists between the main point cloud and the flawed linear cloud where very few

observations are seen. This is a good candidate area to draw a line (with slope 2), which represents the threshold below which Spacelabs samples are considered to be artificially low, hence to be excluded.

Another part of the graph that draws attention is the long tail of points towards the bottom right corner of the graph. Here, a similar mechanism gives rise to this artefact. Instead of the Spacelabs values being close to zero, they are close to unnaturally large values (greater than 50 brpm). It is likely the respiratory rate signal mistakenly took variation in chest impedance due to heart contractions for respirations, which results in values of the same order as heart rate. These are measurement errors that result in a strongly negative difference with HealthPatch, hence the long negative linear tail towards the bottom right corner of the graph. The same reasoning to not exclude these observations holds for these artificial observations. However, if these observations would have to be excluded, they can be by using some filtering method that removes outliers. A simple threshold near 42 brpm may suffice. This number is based on the following clinical and technical considerations. Clinically, a respiratory rate of greater than 40 brpm is not absolutely of interest anymore: the severe tachypnoea is proven and of interest, but the number is not very informative of the severity. More interesting would be detection of the breathing pattern, which is not possible with HealthPatch MD. Technically, the HealthPatch is known to measure up to 42 brpm (see §2.2.4), which makes measurements by the Spacelabs monitor greater than a value near 42 brpm unnecessary.

Because relatively few observations make up these measurement error tails, their influence is not too large in the repeated measurements analysis seen in Figure 23. The 95% LOA are similar in both the regular and repeated measurements analyses, which supports the conclusion of too little agreement between HealthPatch and Spacelabs respiratory rate.

Another observation in the mean-difference graph of Figure 22 is the leftmost part of the main point cloud near the bias line: heteroskedasticity can be clearly seen. The variance around the bias is very low at low respiratory rates compared to higher rates. This suggests the agreement is better than the constant LOA suggest at lower respiratory frequencies, and may be worse at higher rates. This situation lends itself for logarithmic transformation of the data before performing the analyses, or for calculation of the ratio instead of differences of observations. However, due to a number of Spacelabs samples being so close to zero, some of the ratios would be artificially high (or low if Spacelabs values were the numerator) and result in unreliable calculation of the bias and LOA. Logarithmic transformation also isn't possible when the data contains zeroes, because the logarithm is undefined at zero.

For this reason, a third BAA was performed while accounting for repeated measurements to construct the mean-difference graph shown in Figure 24. The bias and 95% LOA are simple linear regression lines. Constant residual variance (after bias regression) was not assumed, so that the LOA were allowed to have different slopes to accommodate for the heteroskedasticity.

The resulting LOA diverge with increasing mean, which indicates respiratory rate agreement was better for lower means. Throughout the domain of 11–27 brpm, the 95% LOA range (upper minus lower) is 2.06 brpm to 15.49 brpm. In other words, near the lower end of the domain the agreement is acceptable, because the LOA lie about 1 brpm above and below the bias. It is at these lower frequencies where the absolute value is critical. The difference between 3 and 5 breaths per minute is relatively larger than the difference between 23 and 25. The clinical consequences of the agreement at the low end are more important than at the higher end.

The agreement analyses applied to the lowest and highest respiratory rates as well. This was done for the same reasons as with heart rate: BAA does not assess dynamic agreement. The resulting agreement was poor in both the low and the high data set. A possible cause is that artefacts are certainly present in both these data sets. In the low respiratory rate data set, Spacelabs regularly measured a value of (almost) zero when HealthPatch did not. It remains unknown which of the two devices measured a value closer to the true value, although it seems

unlikely that Spacelabs is correct because of the seemingly high frequency of the low values. In the high value subset, the low agreement may have been caused by artefacts certainly caused by Spacelabs: the respiratory rates over 42 brpm.

Overall, the results of BAA applied on respiratory rate measurements by HealthPatch MD and Spacelabs lead to the conclusion that agreement was not acceptable. However, the regression approach resulted in good agreement between the devices for low respiratory rates. For example, this can be done by performing a simulation study followed by analysis of measurements performed on patients with bradypnoea (e.g. respiratory depression due to morphine overdose).

Chapter 6. Predicting risk of adverse events

*All models are wrong
but some are useful*

George Box

As illustrated in Chapter 1, timely recognition of deterioration is of utmost importance to adequately manage the deteriorating surgical patient. To investigate whether or not the HealthPatch MD is able to aid in this recognition, this chapter elaborates on methods that can be used on the particular type of data produced by the HealthPatch. However, this chapter does not result in a model for the prediction of risk of adverse events.

Up front it is important to realise the amount and quality of objective information available to a clinician decline with every step down the path from surgery to home. Normally, monitoring of vital signs is unavailable at home. Other than a patient or their direct peers (e.g. spouse) actively calling a caregiver when it may be already too late, very little options that provide continuous and objective information on a patient's status are available. Hospitalised patients are monitored more often than patients after discharge, but on regular hospital wards this monitoring is usually done manually and infrequently. Only on high care wards monitoring is done more often or continuously and automatically. On wards where monitoring is infrequent, a system such as the Vital Connect platform may increase the quality, amount and timeliness of critical information while supporting nursing staff in their monitoring tasks. These situations are reasons why systems like HealthPatch are of interest to be used as continuous monitoring solutions. As of yet it is unknown how much added information and thus potential value such a continuous monitoring device has. It is hypothesised the use of such monitoring devices has the potential to provide early warnings before adverse events, such as deterioration, occur. If true, this theory can enhance patient safety in and out of the hospital.

The goal of the current study is to assess methods to accurately predict risk of adverse events in the surgical population using wearable, wireless, nearly continuous and non-invasive monitoring devices. Note that predicting risk of adverse events differs from prediction of the events themselves. For example, a postoperative patient who develops sepsis in the first week after surgery will probably show abnormal vital signs compared to their normal physiology, such as tachycardia and tachypnoea. However, many other diseases and complications may cause similar symptoms. These physiological deviations are not specific to sepsis, neither do they have 100% sensitivity for the condition. This makes prediction of a specific adverse outcome difficult. Yet, physiological abnormalities in apparently normal physiological conditions—e.g. a patient resting in bed with elevated heart rate, temperature and respiratory rate—indicate some dysfunction, which can be of interest. Deviations from normal values are not always sensitive for a specific cause, but they are specific for dysfunction in general. This may be enough information to promptly inform and activate a caregiver, who can then further assess the *status praesens* and may decide to more specifically investigate the causes of the abnormalities. Especially when comparing the current situation to previous moments in time, the relative changes indicate the patient's physiology is changing. This does not only show the patient's current physiology, but tells what changes have happened as well. Some of these changes precede further decline, indicating elevated risk of adverse events. This risk is not specific for a certain disease or any other negative outcome, e.g. (re)admission to an ICU, but nonetheless tells a caregiver whether or not a patient deserves additional attention relative to other patients. It is this risk of deterioration we are interested in to predict.

A good risk measure should provide information over the course of time about the severity of disease and chances of adverse events taking place in the (near) future. In literature, adverse events often are conveniently defined as dichotomous outcomes [6,10,58–66], e.g. patient survival: the event either occurs or not. Because many adverse events can be defined and changes in vital signs are not specific for any in particular (unless extreme) and the number of patients measured in this study is limited, it is futile to fit a prediction model for one outcome. Fitting a model to one particular outcome, e.g. readmission to the ICU, would lead to overfitting to the particular predictors for those few observations in which the event occurred. Chances are the model would become hardly generalizable for other observations (external validity would be low). Therefore, searching for methods to predict risk of deterioration and risk of adverse events in general is more feasible.

One standard used in clinical practice for the early recognition of adverse events is the Early Warning Score (EWS) [67]. This is a score that encompasses a number of physiological and clinically relevant parameters that add points to the score if deviant from a normal range of values. An EWS is also used on the surgical medium care ward of the UMCU as an additional tool for nurses and doctors to assess a patient's risk of adverse events. In theory, if HealthPatch MD were able predict an EWS, it may be used to indicate patients at risk of postoperative decline. Note that the EWS neither predicts adverse events—originally it was just intended as a tool to bring expert attention to patients who most need it [68]—nor does it have perfect sensitivity and specificity for all conceivable adverse outcomes [13,60,61]. To raise medical awareness for patients with elevated risk is, of course, the goal in both an in-hospital situation and after discharge, where patients can otherwise be monitored insufficiently. Therefore, if it were possible to predict the EWS accurately using HealthPatch data, we would know whether or not a patient is of interest to bring the attention to, but not really if they are truly deteriorating. Further medical examination and diagnostics due to this directed attention may increase chances of early diagnosis and as a consequence prevent adverse events. It is hypothesised that in this way, prediction EWS will enhance patient safety in and out of the hospital.

The following sections describe what the EWS is, how it works and what methods may be used to predict it using measurements by the HealthPatch.

§6.1 Early Warning Scores

The introduction to this chapter mentions the EWS as a standard. However, various EWS systems exist, so the EWS can hardly be called standardised. Furthermore, the type of EWS used varies from country to country and even differs between hospitals within countries, e.g. in the Netherlands. The original EWS was first published on a poster at an intensive care conference [67]. Later it was modified and named the Modified Early Warning Score (MEWS) accordingly [58,69]. Further implementations and adaptations have been made to enhance the score's ability to indicate patients at risk of various adverse events, e.g. the standardised EWS (SEWS) [59,60]. Early warning scores have been adjusted to different populations, e.g. for paediatrics [70] and obstetrics [71]. Currently, the National Early Warning Score (NEWS) is used in the United Kingdom [72]. The system used on the surgical medium care ward in the UMC Utrecht is based on MEWS and NEWS, but differs from both on some points. It is an adaptation of a guideline for Dutch intensive cares¹³. The EWS chart used in the UMCU is summarised in Table 7.

¹³ The UMCU EWS protocol is available on the hospital's intranet and as a form for nurses to use in HiX, the EMR used by the UMCU. However, a difference between the two protocols was found by the author. An IC nurse responsible for patient safety organisation in the UMCU was notified about this potential danger, who in turn notified the appropriate people to correct the HiX EWS protocol.

Basically, all EWS systems are point accumulation algorithms where points are added for deviating vital parameters. In other words, no points are added for vital signs that are considered normal. For example, heart rate is considered normal in the range of 51–90 bpm in NEWS. A point is added if the heart rate is low, e.g. in the range of 41–50 bpm. Even lower values (<40 bpm) result in addition of 3 points to the total score. A similar system of ranges applies to increased heart rates. Similar amounts of points are accumulated for deviating respiratory rates, blood oxygen saturation (SpO₂), need for any supplemental oxygen, core temperature, systolic blood pressure and the level of consciousness (AVPU, Alert, Verbal, Pain, Unconscious) [72].

In the UMCU protocols an EWS < 3 indicates little risk of deterioration in the next few hours to a day. Elevated EWS, i.e., ≥ 3, indicates increased risk of deterioration of the patient requiring activation of a doctor. More frequent monitoring is advised if the patient has a high EWS. Patients with EWS ≥ 7 have a strong indication for continuous monitoring on a medium or intensive care ward, because some major systemic dysfunction is causing physiological distress that may lead to worse outcomes, e.g. irreversible organ damage or death if not adequately treated.

In short, EWS systems are algorithms that aid caregivers in the early recognition of patients who require more attention than otherwise regarded adequate. Because the HealthPatch system measures two important elements of the EWS algorithms, it may be possible to predict EWS with adequate accuracy based on heart rate and respiratory rate alone. If the accuracy is high enough, the HealthPatch EWS (primary EWS) could aid in the direction of medical attention to patients who most need it. For this to be possible, the accuracy of a HealthPatch EWS must be studied. Various methods can be used for this. The following sections elaborate on methods from various sciences that aim to predict future time series data based on current and past values: forecasting.

Table 7 – A summary of the Early Warning Score chart as used in the UMCU is shown. Deviations in vital signs increase the value of the score. A score of zero indicates a low risk of near-future deterioration. By protocol, a score of three or greater requires an inquiry of a doctor. * AVPU: Alert, Verbal, Pain, Unconscious. AVPU is a consciousness scoring system, where alert indicates an EMV (eye, movement, verbal) score of E4M6V5 and unconscious corresponds to E1M1V1. Verbal is a reaction to spoken language and pain is a reaction to a painful stimulus.

| Score | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
|---|--------|-------|---------|-----------|-----------|---------|-------|
| Respiratory rate (brpm) | ≤ 8 | – | 9–11 | 12–20 | – | 21–24 | ≥ 25 |
| Blood oxygen saturation (%) | ≤ 90 | 91–93 | 94–95 | ≥ 96 | – | – | – |
| Heart rate (bpm) | – | ≤ 39 | 40–50 | 51–100 | 101–110 | 111–130 | ≥ 131 |
| Systolic blood pressure | ≤ 69 | 70–80 | 81–100 | 101–200 | – | ≥ 201 | – |
| Core temperature (°C) | ≤ 35.0 | – | 35.1–36 | 36.1–38.0 | 38.1–39.0 | ≥ 39.1 | – |
| Consciousness | – | – | – | A* | V* | P* | U* |
| Nurse concern | – | – | Yes | No | – | – | – |
| Urine production in the past four hours | – | – | < 75 mL | ≥ 75 mL | – | – | – |

§6.2 Methods to predict risk of deterioration using repeated measurements of vital signs

§6.2.1 Matching the number of inputs and outputs

The first step towards predictive modelling is data acquisition from HealthPatch MD, which is described in a previous chapter in this thesis. With the available data, it may be possible to predict a desired outcome: the EWS as measured by surgical medium care nurses. By protocol, EWS is measured at least once every work shift, but more often if indicated. That means up to about every 8 hours an EWS value is determined for a patient. This value is usually based on whatever values are currently displayed by the bedside vital signs monitor (Spacelabs).

Because measurements with HealthPatch MD contain many samples in time, the number of inputs for a predictive model is very large compared to the number of EWS observations; the latter are a few per day. Either the number of inputs (also known as independent variables, regressors, predictors) must decrease, or the number of outputs (also known as dependent variables, regressands, outcomes) must increase. In previous research time windows have been used to equalise the number of inputs to the output to predict [10,73]. In the study by Cuthbertson et al. [10], in each window the last measured vital signs are used to predict the next outcome, which was sampled hourly. Taking only the last known value in an hour-long time window is to some degree a waste of the vital signs measured earlier in the hour. However, the last value before the outcome could be the most representative for the physiological state of the patient at that moment in time. This approach allows for easy modelling and interpretability is high. On the other hand, the largest part of the data (all but the last known value in the time window) must contain some information about the current or future state of the patient as well. Still, the number of input samples must match the number of output samples. Therefore, features can be extracted based on the entire time window (see next section). Tarassenko et al. [73] used the median of the vital signs (sampled per minute) over time windows of 4, 8 and 12 hours. This uses all data available in a time window, but does not consider other information such as trends.

Instead of 'fixed' time windows, other window functions can be used. It is probable that the last value in a time window is most informative of the next EWS to predict and a sample from the past less so. This could be modelled using a window function with decaying memory, i.e., it weighs samples from recent history more than samples from longer ago. An exponential decay is an example of a historically weighed window.

Besides reducing the number of inputs using windows, the number of outputs can be increased to match the number of input observations. This can be achieved through imputation where the predictors have values, but the outcome does not. Imputed values can be interpolated, extrapolated, carried forward, etc. Whatever method used for imputation depends on the implicit assumptions, especially what they mean clinically. True EWS physiology could be very different from the imputed values. Where predictors are missing, samples can be imputed using similar methods, although great care must be taken when imputing missing data in general [74].

§6.2.2 Feature extraction and exploration

The inputs for the prediction model are heart rate and respiratory rate. Instantaneous heart rate and respiratory rate are just two of the many pieces of information in the data that can be used to predict risk of deterioration. This instantaneous information merely reflects the current state of a patient, but tells little about where they come from and where their physiological parameters might be going. We need to extract features in the data that contain information over time. This is where predictive modelling gets both complicated and interesting.

Many physiologically reasonable and unreasonable features can be extracted from the heart rate and respiratory rate signals measured by HealthPatch MD. For example, based on EWS cri-

teria in the UMCU, a heart rate in the range 101–110 bpm adds one point to the score. Similarly, a predictive model should increase the predicted EWS if heart rate is too high (or too low). However, not all patients with a high EWS show an elevated heart rate, nor can a high heart rate purely be used to predict a high EWS. To complicate things further, the vital signs used as inputs to a predictive model are time series, i.e., they consist of many points in time. This means they can potentially be used to continuously¹⁴ predict EWS. However, an important question arises: what time window(s) should be considered in the vital signs? A short time window is easy to interpret: e.g. the mean or median heart rate over the last 15 minutes could be informative about the current EWS. However, the 15 minutes before this time window might be informative for the current value as well, and could contain information about where the EWS is going. Who knows, maybe an hour, two hours, a work shift (eight hours) or a day are interesting because they contain useful information? Clinical investigations, experience and judgement say that somewhere between a few minutes up to a few days are potential windows of interest. This is why feature extraction and exploration are required; it can indicate the time windows in which information is found; it can reveal where the signal content is interesting.

Another complicating factor is the number of features that can be extracted. The possibilities are limitless. One could use time windows in which to take the mean, mode, median, variance, standard deviation, minimum value, maximum value, signal energy, signal power, power spectral density (PSD), frequency contents, harmonics, estimated signal-to-noise-ratio (SNR), and so on. Many of these features are synthetic and few of them are directly related to clinically relevant information. Furthermore, if we were to use such a large number of features taken from many time windows in data measured in many patients over multiple days, the amount of data to analyse would be beyond a current normal computer's power. Recent advances in big data analysis and machine learning techniques can provide methods to cope with such relatively large amounts of data. The downside of these methods is the difficult clinical interpretability of the models, since they may become very complex, nonlinear, etc.

It is likely an approach using windows to match the number of inputs to the number of outputs results in a model that can be used to predict EWS, or be used to better understand how to achieve a model to predict the EWS.

§6.2.3 Predictive modelling for time series

Now that it has been described how features can be extracted for the prediction of repeated measurements of the outcome of interest, we elaborate on methods to actually create the prediction models. The methods available for use depend on the type, distributions and assumptions about the input and output data [75]. For example, the output can be modelled either as continuous, i.e., any real number, or categorical, i.e., any of a number of classes. The EWS is an increasing degree of risk of deterioration with fixed levels, being nonnegative integers 0, 1, 2 and so on up to 20. Technically this is categorical data, but it is ordinal categorical data that can quite easily be modelled as continuous. For example, if in the prediction model a (continuous) EWS of 3.37 were the outcome, it is still interpretable: the EWS is greater than three, but not by a lot. Even though in reality such an EWS is never seen, the score is still usable, even if non-integer. This allows some flexibility in the types of prediction models used.

Classically, in medical statistics continuous variables are often predicted using simple or multivariate regression, whereas categorical outcomes are predicted using logistic regression tech-

¹⁴ Continuous is actually termed incorrectly, because the vital signs time series are sampled, hence discrete and not continuous. However, the sampling rate (once every few minutes or greater) is much greater than the usual EWS sampling rate (once every few hours or less). In this perspective, the sampling rate is nearly continuous.

niques (multinomial regression for categorical output with more than two classes, such as EWS). For time series data, other techniques are available from various sciences.

Binomial and multinomial logistic regression for time series

To predict categorical variables, logistic regression is an option. Because EWS can be modelled as categorical, e.g. low vs. high or in more than two categories, various logistic regression methods can be used. For the case with two classes (binary), binomial logistic regression is used. If more than two EWS classes are modelled, multinomial logistic regression is available. The latter is an extension of the first, so the first needs to be explained before the second. In literature, the word binomial is often omitted from binomial logistic regression to indicate a binary model, although binomial and multinomial are two different types of logistic regression.

Binomial logistic regression models a dichotomous outcome by modelling the probability of an observation being of one class, $P(Y = 1|X)$, where Y is the class and X the observation of model parameters $[1, x_1, x_2, \dots, x_{p-1}]$ (the 1 is the intercept term). Because there are only two possible outcomes, this means the probability of being of the other class is $P(Y = 2|X) = 1 - P(Y = 1|X)$. The probability is modelled through a function f that links the simple linear regression $f = \beta X$ to the probability of being of a class, where β contains the model coefficients. The most common link functions in logistic regression are logit and probit. Logit is

$$\text{logit}(\pi_1) = \ln\left(\frac{\pi_1}{1 - \pi_1}\right),$$

which is then set to equal βX ,

$$\ln\left(\frac{\pi_1}{\pi_2}\right) = \beta X,$$

where $\pi_1 = P(Y = 1|X)$ and β is a vector of model coefficients such that $\beta X = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}$. The logit function is the log ratio of probabilities, i.e., the log odds for outcome $Y = 1$ versus $Y = 2$. Probit is $\Phi^{-1}(\beta X)$, with Φ^{-1} the inverse of the standard normal cumulative distribution function.

Using logit, the log odds of either outcome versus the other are modelled. In other words, the other outcome is the outcome 'not of interest', i.e., the reference outcome. This is an important category for multinomial logistic regression, which models the probability of more than one outcome class. Also, note that the probabilities of all classes must sum to one, so if there are k classes, then $\sum_1^k \pi_k = 1$. Logically, it follows that $k - 1$ models must be fitted, because if the probability of all but one class is modelled, the probability of the remaining class is known.

In multinomial regression, a reference class is used like with binomial logistic regression, often the first or the k^{th} class in the model. The logit model from above then becomes

$$\text{logit}(\pi_j) = \ln\left(\frac{\pi_j}{\pi_k}\right) = \beta_j X, \quad j \in [1, 2, \dots, k - 1].$$

Note π_j , the probability of $Y = j$, the j^{th} outcome and π_k , the probability of the reference outcome. Also, there can be different model parameters β_j for the j models.

The outcome $Y = [y_1, y_2, \dots, y_k]$ not necessarily indicates an observation belongs to a particular class. Other possibilities exist in multinomial logistic regression. For example, consider EWS. Every EWS value can be modelled individually (nominally), but because we know there is an order of the categories, we can model the outcome to be ordered as well (ordinal). For example, we know that an EWS of 4 is greater than 3, etc. This means a prediction model could predict $P(\text{EWS} > 3)$ instead of $P(\text{EWS} = 3)$. Possible predicted outcomes are

1. the probability of $Y = y_j$;
2. the probability of $Y \leq y_j$ or similarly $Y \geq y_j$;
3. a combination of the first two options: the probability of $Y = y_j$ given $Y \leq y_j$ or given that $Y \geq y_j$.

To predict if an observation belongs to one class (option 1), a model could be fit that models that class versus all other classes grouped as one reference class. This is a one-versus-all approach, which essentially creates the same situation as with binomial logistic regression. This is a valid approach for nominal outcomes. However, EWS is ordinal, so options 2 or 3 can be more interesting.

The second option models being of at least (or at most) a given class. In the logit link function then the range of classes of interest is put in the numerator of the probability fraction, the range of classes not of interest are put in the denominator. Then, the logit link function is written as follows.

$$\ln\left(\frac{P(Y \leq y_j|X)}{P(Y > y_j|X)}\right) = \beta_j X,$$

or

$$\ln\left(\frac{P(Y \geq y_j|X)}{P(Y < y_j|X)}\right) = \beta_j X.$$

Note that the differences between the two equations are the flipped greater/less than symbols, which depends on the model choice of the researcher. This model may be a good representation of the EWS in the UMCU, because in that hospital an EWS of 3 or greater is by protocol a reason to activate a supervising doctor for further attention, while an EWS of 2 or less is not. This is modelled exactly by the second equation, which models the log odds of having an EWS equal to or greater than j versus an EWS less than j . The model for $j = 3$ then would closely resemble the UMCU EWS protocol.

The third and last option models the outcome being $EWS = j$, given that it is at least or at most j , i.e., it is not less or greater than $j - 1$. This is a hierarchical ordering of the outcome classes. The logit link function combines options 1 and 2.

$$\ln\left(\frac{P(Y = y_j|X)}{P(Y > y_j|X)}\right) = \beta_j X,$$

or

$$\ln\left(\frac{P(Y = y_j|X)}{P(Y < y_j|X)}\right) = \beta_j X.$$

Again, note that the flipped greater and less than symbols are the difference between the two equations depending on the researcher's choice. This hierarchical model would, for example, predict the probability of the EWS being 3 or—using the second equation—otherwise lower.

One of the ever-existing dangers of fitting models to data is overfitting, which may occur if a model can adapt too much to the data in ways that aren't realistic. If a model contains too many parameters, for example, it could fit to particular peculiarities in the data that in truth may not be interesting to model, but to consider as error. Such models in general suffer from low generalisability. A multinomial logistic regression model consists of $k - 1$ models for every class to be predicted. Every model has its own parameter set β_j , each with p (for the p variables in X). So, a multinomial logistic regression model has $p(k - 1)$ parameters. To reduce this number, the proportional odds assumption can be made. This assumes every input in X has the same influence, i.e., the same β , on the outcome, but the intercept in the models is allowed to vary. The proportional odds model is

$$\ln\left(\frac{\pi_j}{\pi_k}\right) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

with α_j the intercept for the j^{th} model and β the p model coefficients in X (without an intercept term). All β are the same in the $k - 1$ models. The proportional odds model probably fits the data worse than a model with more parameters, but it is less prone to overfitting with its $p + k - 1$ parameters.

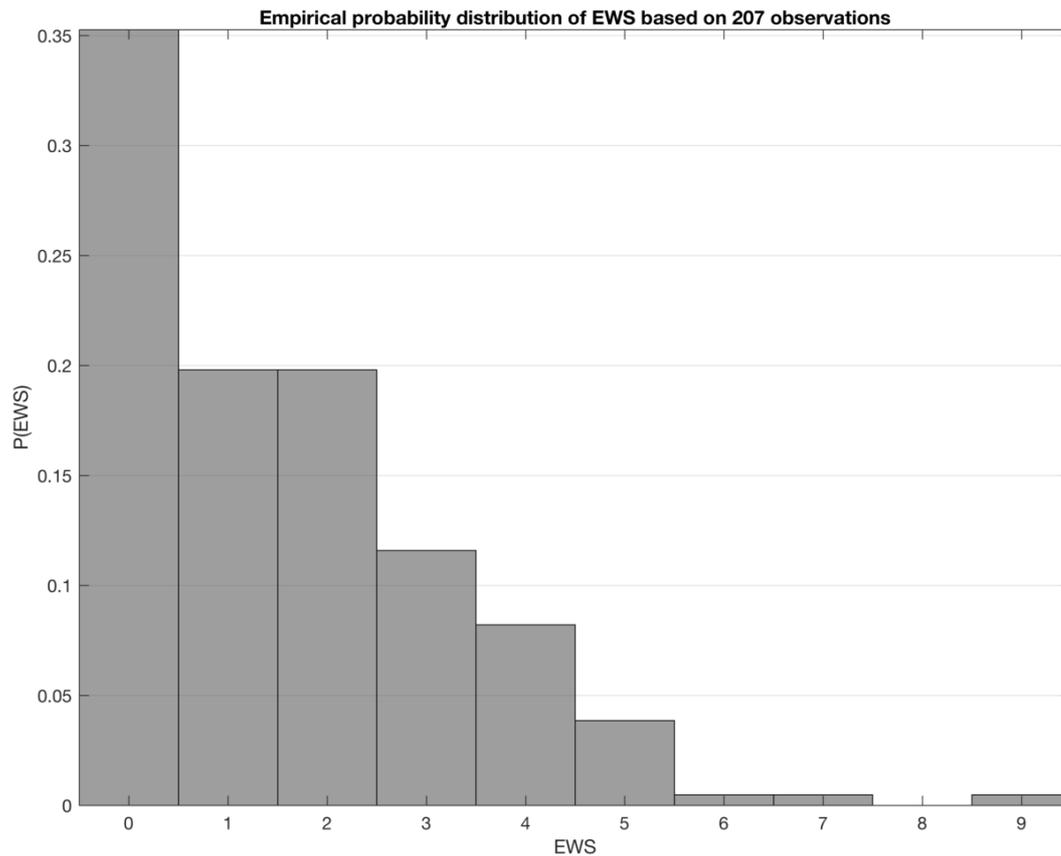


Figure 25 – The Early Warning Scores as recorded by nurses during participations of patients in this study are graphed in a histogram. The distribution of Early Warning Scores can be estimated from the graph. The 207 values are the result of all Early Warning Score observations during all participations, i.e., some participations have contributed more scores than others.

The aforementioned methods can be applied on single observations of X and Y . If multiple observations have been made in time of both X and Y , they are time series. For example, Cuthbertson et al. [10] predicted a binary outcome (transfer to IC versus transfer to general ward from a high dependency unit) based on multiple observations of vital signs in the past 24 hours. Their methods included a separate logistic regression for every hour in the past day. The data from all these hours was aggregated and used as input to a single logistic regression model to predict the desired outcome. Such a model can be used for EWS prediction as well: features extracted from the past day (or any time window of interest) can be used to predict EWS. Furthermore, the model can be adjusted for time-invariant features, such as patient characteristics (e.g. sex).

Mixed-effects models

Mixed effects models are a class of models available for regression analysis with time series. More generally, Generalised Linear Mixed-Effects Models (GLMEs or GLMMs) are a class of available mixed-effects models. A main difference between mixed-effects models and GLMEs is that the first assumes normality of the outcome, the latter can assume other distributions as well [75,76]. The latter case is true for EWS, as can be seen in Figure 25. The figure shows that most EWS values are close to zero, higher EWS values have a decreasing probability of being measured. The distribution could be half normal or exponential, but no tests for any distribution in particular were performed. The distribution is a result of the artificial scoring chart that cate-

gorises various vital signs, that each have their own distributions (see Tarassenko et al. [73] for empirical examples).

Because the EWS consists of nonnegative integers it is a discretely distributed variable. Depending on the patient population, the distribution will have more or less values near zero, the most common EWS value. In a truly average population there will be relatively many EWS values of 0 and 1, while relatively few values will be 2 or greater. The population described in this thesis is relatively ill and old and every patient has recently undergone surgical procedures. Therefore, EWS values greater than 2 will be relatively common, although most of the time EWS will still be low. The distribution of EWS depends on the population. This has influence on the GLME used, because (assumptions about) the distribution must be specified by the researcher.

Another way to look at GLMEs is to see them as an extension of logistic regression, explained in the previous section. In logistic regression, every model contains fixed effects (the β vector estimated for the $k - 1$ classes). However, there may be effects that add to these fixed effects depending on a different group or a moment in time, the random effects. This is an additional set of model parameters that accommodates for changes within a subject. For example, time of measurement can be taken as a random effect [56]. Taking time as a random effect allows for an overall model to be fit with the fixed effects, but changes over time are then modelled by the random effects which are added to the fixed effects. Hence the name mixed effects models. This accomplishes a way to model correlations within patients; as mentioned before, sequential vital signs of a particular patient are autocorrelated.

Autoregressive models

In econometrics, geology, meteorology and astronomy, autoregressive (AR) models have been used extensively in the past decades for time series forecasting. Examples of applications are the stock market, where it is of interest to model price changes (increasing or decreasing trends) and modelling seasonal effects (e.g. short-term temperature prediction from hour to hour) or climate change (long-term average prediction from year to year). Autoregression means current values of a certain outcome of interest depend on previous values. This is a concept we have seen before in the discussion of Chapter 4. Current vital signs depend on previous values and because EWS is constructed from vital signs measurements it possibly is autocorrelated too. Using AR modelling techniques is an option for predictive modelling of EWS.

For this type of modelling, we need to clearly define the components of time series. Changes in a time series can result from various sources in which we might be interested. Trends are a linear non-random component. Periodic behaviour, such as increases followed by decreases during active and inactive periods, are a non-random component of interest as well. Slow periodic changes, such as day-night rhythm can be distinguished too. Finally, time series show random fluctuations about their current value which cannot be explained by a trend, fast or slow periodic behaviour [77].

Trends can be visualised using moving average (MA) filters. Depending on the desired time resolution, an averaging window of more or less samples can be used. In this way, trends can be explored over every hour, work shift, day, etc. By removing (subtracting) the trend a signal is left that consists of periodic and random behaviour. This periodic behaviour can be analysed using various time or frequency domain techniques. With such techniques, a time series can be reduced to its various components until white noise is all that is left.

The AR model of order p is $X(t) = \sum_{i=1}^p \varphi_i X(t - i) + \varepsilon(t)$, sometimes written more compactly as $\sum_{i=0}^p \varphi_i X(t - i) = \varepsilon(t)$. In these equations, $X(t)$ is the AR process, depending on p past values of X itself, multiplied by parameters φ , and white noise $\varepsilon(t) \sim N(0, \sigma^2)$.

Besides AR models, an MA model is another in the same family. The MA model of order q is also dependent on previous states, but this time it depends on previous noise: $X(t) = \varepsilon(t) +$

$\sum_{j=1}^q \theta_j \varepsilon(t-j)$, or again arranged more compactly $X(t) = \sum_{j=0}^q \theta_j \varepsilon(t-j)$. The model parameters are θ . Note that an MA model is not the same as an MA filter.

AR and MA models can be combined into an ARMA model. Noise-integrated variants exist (ARIMA). All these models only model changes over time in the output variable. These models can be extended to be dependent on input time series too. For example, the AR model with exogenous input is called ARX, the ARMA model with exogenous input is called ARMAX. These models do also depend on current and previous values of other variables. The ARMAX model of order (p, q, b, n) with input $U(t)$ is

$$\sum_{i=0}^p \varphi_i X(t-i) = \sum_{j=0}^q \theta_j \varepsilon(t-j) + \sum_{k=1}^b \eta_k U(t-n-k).$$

The model order b is the number of previous inputs the model depends on (η being the weights) and n is a fixed lag. If $n = 0$, then there is no lag between input and output. To fit model parameters φ , θ and η , various methods exist in software packages such as MATLAB. AR models are the legacy of the forecasting research by Box and Jenkins [78].

Machine learning and big data

To conclude this list of methods, machine learning (ML) and big data analysis are mentioned. ML is the field of mathematics and computer science that tries to program computers in such a way that they learn without being told precisely what to do. For example, the many features that one might come up with when using time series data, such as in this study, can be used in ML: a computer algorithm may be able to select only those features that are truly predictive for a certain outcome. In this case, the researcher only tells the computer to take the data and how to look for good information, but not what the information should be, nor what features should be.

ML is also a technique to handle big data. Big data is a vague term for every set of data that is large in volume, variable, changing in (real-)time or uninterpretable due to various complicating factors, such a very diverse data of variable quality. ML techniques are available that can find patterns in (big) data that are otherwise concealed, even from expert judgement [79].

Trivially, ML is an extension of many modelling and statistical techniques. For example, finding the association between length and weight is a trivial ML problem: the researcher asks the computer to find one or a few parameters in a model, but does not tell the computer what these parameters should be. This becomes less trivial when the model itself is unknown and the ML algorithm is instructed to look for models that may be appropriate.

The science of ML also studies how to find good models. If an ML algorithm is allowed to fit data as good as it can achieve, it will probably overfit the data: it would model variability that is not of interest. This would not be a good model, because good models perform well in both observed and unobserved data.

For the prediction of time series, such as in this study, some ML techniques can be used. One of them has been trending in the last decade, although its conception dates from halfway back in the past century. Artificial neural networks (ANNs), or neural networks (NNs) for short, model the connection between input and output data with individual nodes (neurons), each having an individual set of inputs. The basic NN consists of a number of fully interconnected layers; Figure 26 shows a schematic of a basic NN.

The first is the input layer, in which every neuron takes one model variable as its input. All neurons have an activation function that can be high or low, depending on the value at its input. The second layer consist of an arbitrary number of neurons. All these neurons take as inputs all the outputs (activations) from the input layer. All these connections are weighted, so the second layer neurons all activate differently based on the activations of the input layer. Any next layers consist of neurons as well, each layer connecting to all previous layer's neurons in a similar

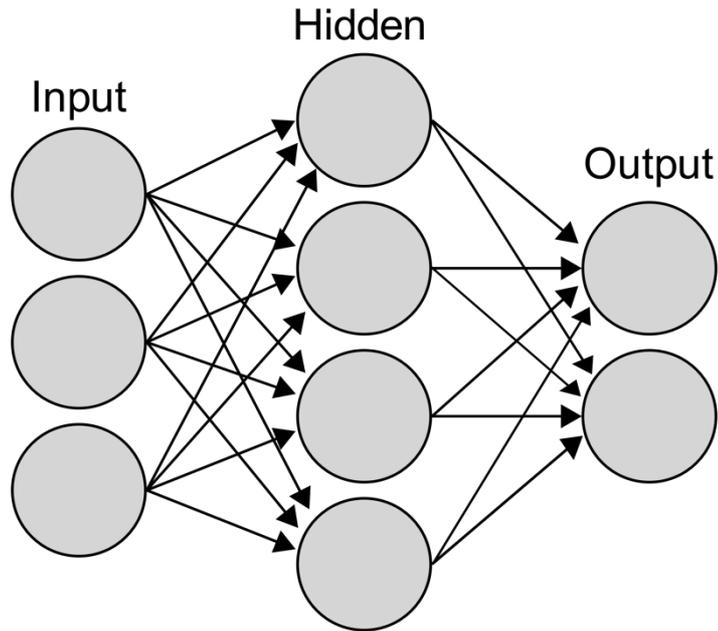


Figure 26 – A schematic artificial neural network with one hidden layer and all interconnections is shown. Image reused under license from Wikipedia Commons [96].

weighted fashion and all neurons activating to these weighted inputs. The last layer, the output layer, activates only those neurons that are associated with the predicted outcome.

When fitting a NN (using a method called backpropagation), the goal is to find the weights in the connections between all layers. If the resulting model is good, every individual neuron is a specialised model in itself for a particular feature in the data available to it. Because of the many interconnections, layers and neurons per layer, NNs can model very complex a high-dimensional data with many subtleties.

NNs have been used successfully in diverse applications, such as computer vision (face recognition, self-driving cars, etc.), music classification and generation, system control (robotics), natural language processing (written or spoken language) and medical diagnosis [80]. For some of these applications, advanced network topologies and computation strategies are used to harness the power of NNs.

§6.3 Discussion

This chapter elaborated on various available methods for the prediction of future values of time series based on current and past values (of other time series). Such methods are currently neither used in clinical practice for the goals of the current research, nor in out-of-hospital settings. Thus, they can be considered experimental. However, the methods described in this chapter have proven to be useful in other scientific areas such as digital signal processing, control system theory, machine learning and econometrics. These sciences already embrace the benefits of the use of these methods and greatly aided in their development. This indicates the potential they may have if applied to the information in the time series used in this research. Because use of information technology in medicine classically lags behind other sciences [55] and medical technologies/science are evidence-based, it remains important the promising methods from other sciences presented in this chapter are investigated promptly, but carefully considering their clinical implications.

Take, for example, the neural networks described in the previous section. They are so versatile, because they can take any form a researcher allows their network to be. This is a challenging

paradigm shift for modern medicine. In general, clinicians want to have some understanding of predictive models before they would apply them on their patients. How would a doctor, whose responsibility is to find the best treatment for every individual patient, use a model that was fitted to data that probably does not contain his particular patient population? How would this doctor know whether or not the model is applicable to his patients? This situation can unfortunately be the case in randomised controlled trials in which a specific population is studied. How well this population generalises to others is not always an easy question to answer. This is why big data and machine learning are the future of modelling in clinical practice. Why consider a small population when you can consider every individual?

The concepts used in ML can be much more difficult to understand than classical medical statistics, but the power lies in the generalisability. Because ML methods can learn from very large data sets, the models can be applicable to very large patient populations too. For example, if a very complex but correct ML system would be set up in such manner that it learned from every new patient in a hospital, the system would not only encompass the entire hospital population, but keep learning and therefore adapting to slow epidemiological changes as well.

This step towards ML may be the first in a paradigm shift away from classical medical statistics. This shift will probably take many years, if not decades. Classical medical statistics will probably never cease to exist, because their relative simplicity and interpretability will always have applications and preference where understanding of a model is required.

Before advanced predictive modelling methods can be applied to vital signs data from wearable monitors such as HealthPatch MD, it must be ensured that this type of data contains relevant information. The hypothesis that HealthPatch MD data can predict the EWS or risk of deterioration in the surgical population remains to be investigated. Only then this information can become knowledge about early recognition of adverse events in recovering postoperative patients in and out of the hospital.

Chapter 7. Discussion and conclusion

This chapter concludes this thesis. Table 8 summarises the main conclusions of every chapter's research questions.

Table 8 – The research questions studied in this thesis and their conclusions are summarised.

| Chapter | Research questions | Conclusion |
|---|---|--|
| Chapter 2. HealthPatch MD by Vital Connect | What is HealthPatch MD, what does it do and how does it work? | The validation status of HealthPatch MD for use in clinical practice is unknown. |
| Chapter 3. Simulation experiment with HealthPatch MD | How well does HealthPatch MD perform in simulated patients? | HealthPatch MD can accurately measure simulated heart rate, but not respiratory rate due to different simulation/measurement principles. |
| Chapter 4. Measurements with HealthPatch MD | How are measurements with HealthPatch MD performed in clinical practice? | Measurements are done easily and patients reported no to little perceived burden. |
| Chapter 4. Measurements with HealthPatch MD | How can data from HealthPatch MD be acquired? | Reliability of raw data acquisition was good. |
| Chapter 4. Measurements with HealthPatch MD on a clinical ward | How can data from HealthPatch MD be stored and pre-processed for further analysis? | A versatile database was set up to store, extract and pre-process both HealthPatch MD and Spacelabs XPRESSON data. |
| Chapter 4. Measurements with HealthPatch MD on a clinical ward | What methods are available to measure agreement between measurement methods of time series? | The agreement analyses by Bland and Altman are applicable to the data in this study, Bland-Altman analysis for repeated measurements is the basis for the data analysis in this thesis. |
| Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPRESSON | What do heart rate and respiratory rate measurements by HealthPatch MD look like? | A case study of one participation showed that a postoperative deteriorating patient shows changes in vital signs measured by HealthPatch MD. |
| Chapter 5. Data analysis: agreement of HealthPatch MD with Spacelabs XPRESSON | How well do HealthPatch MD signals agree with the bedside reference standard? | Heart rate is measured acceptably accurate, but respiratory rate is not. However, extreme values are measured less accurately. |
| Chapter 6. Predicting risk of adverse events | How is risk of deterioration in the surgical patient measured? | Early Warning Scores may be used in prediction modelling of risk of adverse events. About one in four measured Early Warning Scores was 3 or higher. |
| Chapter 6. Predicting risk of adverse events | How can risk time series be predicted based on other time series? | Various contemporary methods for prediction modelling are described; machine learning is a potential candidate to predict risk of deterioration from the complex patterns seen in vital signs. |

§7.1 Discussion

For the purpose of evaluating HealthPatch MD as a near-continuous, automatic, non-invasive and wearable vital signs monitor in postoperative patients, the device was studied on a number of subjects in this Master thesis. These subjects were the technical specifications, simulated performance, measurements in clinical practice on a surgical medium care ward, data processing (acquisition, storage, pre-processing, graphing), agreement analysis with the practical reference and prediction modelling of risk of postoperative decline.

This chapter concludes this thesis and discusses all topics from the previous chapters. These chapters each have a discussion section too, where many topics have already been addressed.

§7.1.1 Wearable vital signs monitoring technology in literature

In short, there is little literature about the use of wearables for the purposes of this thesis. Many studies are focused on the engineering of wearable vital signs monitoring technology [46,81,82]. Other studies promise clinical validity while performing measurements in very limited or healthy populations, or their methods are in controlled environments, or their data analyses have limitations or violate assumptions being made [36–38,41,42,83–86]. This leads to the belief that there is little evidence about the validity of current developments in this field. This study is the first to perform a validation of a wearable vital signs monitor by performing long-term, spontaneous measurements in a true clinical setting with a realistic patient population and using a passive measurement protocol, i.e., not performing controlled measurements. Although one monitoring system, EarlySense (EarlySense Ltd, Ramat-Gan, Israel), does have quite extensive validation studies in clinical practice, it is not a wearable system and consequently unfit for truly continuous monitoring of vital signs [87–90].

§7.1.2 Strengths and limitations

Benchmark for wearable vital signs monitors

The simulation experiment clearly showed where knowledge and practical possibilities were lacking to test a wearable vital signs monitor. For example, respiratory rate could not be simulated for the methods the HealthPatch uses measure it. For future research with wearable vital signs monitors, some technical experimental setup must be available to assess such devices. With such a setup, it would be possible to put these devices to the test before they are applied in patient measurements. Extreme physiology, accuracy, wear and lifespan can be tested in such experiments. Because no benchmark exists for wearable vital signs monitors, it would be a unique benefit to have one available. With this benchmark it would be possible to analyse the performance of many manufacturers' wearable devices for clinical and non-clinical applications, for both research purposes and as a service to third parties.

ECG electrode ageing

In this study, it was not investigated if an ageing effect is present in the HealthPatch MD ECG electrodes and whether or not it influences its measurements. It is known typical ECG electrodes used in clinical practice show an ageing effect when applied to the skin. This is usually not a problem, because these electrodes are regularly replaced (based on experiences from nurses). By protocol, regular ECG electrodes should be replaced every 48 hours. However, HealthPatch MD and other similar devices are designed to be attached to the skin continuously for prolonged times up to days. In theory, what could happen is that both skin contact electrodes age differently. As a consequence, the electrical specifications of the patch change over the course of days.

If and how electrode ageing affects HealthPatch MD ECG recordings and further calculations of physiological parameters needs further investigation. In this research, it was assumed the ageing effect had negligible influence, which was a limitation.

Tailored database

A strength in this study was the database design (§4.2.2) and implementation in MATLAB (§4.3). The database was set up to reflect reality. This allowed for easy interpretation of the digital environment in this study due to its similarity to reality. Patients could participate in multiple participations, participations could contain any number of measurement by any number of devices, etc. The difficulty in this is that programming such a system requires some experience with writing software, which we did not have very much at the beginning of this research, but gained a little through writing the database and many trials and errors¹⁵. It will therefore not be easy to keep using the database in continuation of this research with wearables, unless future researchers have experience with writing software or are willing to learn it. However, with such a database it is possible to acquire, store, pre-process and explore data in future studies. Minor and major adjustments and enhancements are possible with a tailored database system.

Artefact rejection and reduction

No explicit artefact rejection or reduction techniques were applied to the data in this study. Some artefacts were removed in the pre-processing steps, such as outlier removal in the downsampling process with an anti-aliasing filter, but no attempts to remove particular artefacts were made. This is a limitation, because it was clearly visible many measurement errors were present in the data.

Mainly Spacelabs data showed large variability in both heart rate and especially respiratory rate measurements. These samples were obviously artefacts, because their values were not plausible. For example, sometimes the respiratory rate algorithm of the Spacelabs monitor would pick up heart rate. Of course, a respiratory rate of greater than 50 brpm is quite rare, especially when it is just one sample and all others in the vicinity have more logical values. Also, very low values (below 2 brpm) were seen often too. These values emerged when patients breathed very superficially, e.g. when sleeping, and Spacelabs missed a breath.

A reason for not performing artefact rejection was the resampling method. The current resampling method was not robust for gaps in the data (see also §4.3.2). Rejecting an artefact would mean inducing a gap, which is undesirable for the resampling method. Instead of removing a sample as an artefact, it can be replaced with an estimated value. However, such practices are always risky, because one might be manufacturing their own data and results with or without knowing. Moreover, assuming outliers are artefacts can be dangerous in situations where physiology is actually extreme. In §4.4, another reason to keep measurements with their artefacts is given: when comparing methods, measurements including artificial flaws are to be used. It was therefore decided no artefact removal techniques would be applied besides the artefact reduction incorporated in some of the pre-processing methods of Chapter 4.

Spacelabs XPRESSON is not a gold standard

In validation studies a comparison of the method being validated with a gold standard is typical. In Chapter 5 a comparison of HealthPatch MD vital signs was made with the same vital signs as measured by the Spacelabs XPRESSON bedside monitoring system. For both heart rate and respiratory rate the Spacelabs data was not a gold standard.

¹⁵ Especially errors, certainly thousands

The ECG-based heart rate monitor of Spacelabs may be near gold standard validity, but the data consisted of one sample per minute. This makes Spacelabs to some degree an impractical reference standard for near-continuous heart rate measurement.

For respiratory rate, capnography is the gold standard. However, capnography is impractical to perform in a continuous postoperative setting. An alternative would be respiratory inductance plethysmography, in which stretchable conducting coils are strapped around the patient's chest and abdomen. Because the Spacelabs monitoring system was the most practical bedside reference standard in clinical practice in the UMCU, its respiratory rate signals were chosen as the reference for the comparison analyses.

Because of these reasons, the comparison study was limited by the limited validity of the reference monitor. This is always the case in Bland-Altman analysis, but the agreement is lowered by the Spacelabs monitor probably having a larger variability of its respiratory rate measurements.

Generalisability of this study's results is limited

Because not all research goals in this thesis have been fully addressed, the potential for HealthPatch MD to be used as a diagnostic tool for early recognition of deteriorating patients is not yet known. The studied population was chosen, because it was expected many abnormal vital signs would be measured and that some adverse outcomes would be observed. However, there are different settings and populations, e.g. at home, where the potential of a system such as HealthPatch may be different. Furthermore, because the included population is relatively ill, the effects of specific comorbidities, medication use and medical interventions during participation have been measured, but have not specifically been studied. How a wearable system such as HealthPatch MD will behave in other settings and populations is a matter of performing many more measurements in various situations to gain experience and find more evidence.

Failure to transmit in HealthPatch MD measurements was variable

The survival analyses presented in Appendix B show that failure to transmit was common during measurements with HealthPatch MD. However, most patches did not lose any data, while others lost relatively large amounts. It seems the HealthPatch system is prone to complete failure to transmit once the preconditions for failure are met. The battery phenomenon as described in §4.5.2 was a main cause of the complete failure of some measurements. However, other influences on data transmission are yet to be investigated. For example, in an ambulatory setting at home, the connection with a relay device (e.g. iPad) cannot be guaranteed.

§7.1.3 Early Warning Scores may be used to estimate sample sizes for future studies

In Chapter 6, measurements of the Early Warning Score (EWS) were presented in a histogram. From this histogram, the distribution of EWS values as seen on the surgical medium care in the UMCU can be estimated. When EWS is greater than or equal to 3, the UMCU protocol is to activate a doctor for further assessment of the patient. If a similar activation system could be established using a wearable vital signs monitor such as HealthPatch MD, then patient safety may be enhanced in and out of the hospital. However, more observations of complications and other adverse events are needed before patient outcomes can be estimated or predicted using measurements by such a device.

The EWS distribution allows to do a recommendation for future research. In the histogram it can be seen that about 25% of EWS values are 3 or greater. This means that about one in four EWS measurements requires activation of a doctor. These 25% of observations can be of interest to investigate into more detail with HealthPatch measurement. If the vital signs measured by

HealthPatch contain information that can be used to estimate that EWS is high, this information can also be used to activate a doctor. If, for example, a future study would want to witness 50 doctor activations, then about 200 EWS measurements are expected to have to be witnessed for this number of values to be 3 or higher. Compared to the current study, about 30–40 participations will be required to observe these EWS values.

Note that the surgical medium care ward measures EWS values once per nurse shift by protocol. However, on regular wards, it is probable that true EWS values are lower on average, but EWS measurements are mostly done when a high value is suspected. How many samples would be needed on such a ward to observe enough outcomes of interest is more difficult to say.

§7.2 Conclusion

To conclude this Master thesis, we return to the original research question.

Can HealthPatch MD be used in clinical practice for early recognition of the deteriorating surgical patient?

The answer has not been definitively given nor found in this study. However, many steps were taken towards answering parts of the question. The HealthPatch MD system was analysed thoroughly based on technical specifications, simulation of physiology, measurements in the postoperative population and early validation by a comparison with the clinical bedside reference standard. From these studies, it can be concluded the HealthPatch MD system is able to measure heart rate with a marginally acceptable agreement and respiratory rate with acceptable agreement in a more advanced comparison analysis. These agreements may improve if properly validated heart rate and respiratory rate monitors are used instead, which do not have the limitations of the reference monitor used in the current study.

The amount of data loss in the current study is debatable. Some measurements by the HealthPatch MD system proved reliable, but others suffered from complete failure. The current system is still too prone to transmission failure to be used reliably in clinical practice.

An overview of methods for the predictions of risk of adverse events based on the Early Warning Score chart was presented. Various methods from different sciences may prove useful to model adverse events based on wearable vital signs monitors. It will be interesting to see what value techniques such as machine learning add to the future of this research.

Bibliography

1. Buist M, Bernard S, Nguyen T V., Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation* [Internet]. 2004 Aug;62(2):137–41. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0300957204001236>
2. Jacques T, Harrison G a., McLaws M-L, Kilborn G. Signs of critical conditions and emergency responses (SOCCER): A model for predicting adverse events in the inpatient setting. *Resuscitation* [Internet]. 2006 May;69(2):175–83. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0300957205003813>
3. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. [Internet]. Vol. 36, *Critical care medicine*. 2008 [cited 2013 Nov 7]. 296-327 p. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18158437>
4. Schmidt PE, Meredith P, Prytherch DR, Watson D, Watson V, Killen RM, et al. Impact of introducing an electronic physiological surveillance system on hospital mortality. *BMJ Qual Saf* [Internet]. 2015 Jan 1;24(1):10–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25249636>
5. Churpek MM, Yuen TC, Huber MT, Park SY, Hall JB, Edelson DP. Predicting Cardiac Arrest on the Wards. *Chest* [Internet]. 2012 May;141(5):1170–6. Available from: <http://dx.doi.org/10.1378/chest.11-1301>
6. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med* [Internet]. 2016 Feb;44(2):368–74. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003246-201602000-00016>
7. Barwise A, Thongprayoon C, Gajic O, Jensen J, Herasevich V, Pickering BW. Delayed Rapid Response Team Activation Is Associated With Increased Hospital Mortality, Morbidity, and Length of Stay in a Tertiary Care Institution. *Crit Care Med* [Internet]. 2016 Jan;44(1):54–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26457753>
8. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* [Internet]. 2016 Feb 23;315(8):801. Available from: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.0287>
9. Fuhrmann L, Lippert A, Perner A, Østergaard D. Incidence, staff awareness and mortality of patients at risk on general wards. *Resuscitation* [Internet]. 2008 Jun;77(3):325–30. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S030095720800052X>
10. Cuthbertson BH, Boroujerdi M, McKie L, Aucott L, Prescott G. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? *Crit Care Med* [Internet]. 2007 Feb;35(2):402–9. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003246-200702000-00010>
11. National Guideline Clearinghouse (NGC), National Institute for Health and Clinical Excellence (NICE). Surgical site infection: prevention and treatment of surgical site infection. [Internet]. Welsh A, editor. RCOG Press; 2008. Available from: <https://www.guideline.gov/content.aspx?id=13416>
12. National Confidential Enquiry into Patient Outcome and Death. Just Say Sepsis! [Internet]. London, UK: NCEPOD; 2015. Available from: www.ncepod.org.uk
13. Hollis RH, Graham L a., Lazenby JP, Brown DM, Taylor BB, Heslin MJ, et al. A Role for the Early Warning Score in Early Identification of Critical Postoperative Complications. *Ann Surg* [Internet]. 2016 May;263(5):918–23. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-900000000-97169>
14. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* [Internet]. 2009 Oct;361(14):1368–75. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00132586-201010000-00033>
15. World Health Organization. Safe Surgery Saves Lives [Internet]. WHO Guidelines for Safe Surgery 2009. 125 p. Available from: http://whqlibdoc.who.int/publications/2009/9789241598552_eng.pdf
16. Ireland CJ, Chapman TM, Herbison GP, Zacharias M. Continuous positive airway pressure (CPAP) in the postoperative period for the prevention of postoperative morbidity and mortality following major abdominal surgery. In: Ireland CJ, editor.

- Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2011. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008930.pub2/full>
17. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Griffiths P, Busse R, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet* [Internet]. 2014 May;383(9931):1824–30. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0140673613626318>
 18. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat A-HS, Dellinger EP, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *N Engl J Med* [Internet]. 2009 Jan 29;360(5):491–9. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJMs0810119>
 19. Lawson EH, Hall BL, Louie R, Ettner SL, Zingmond DS, Han L, et al. Association Between Occurrence of a Postoperative Complication and Readmission. *Ann Surg* [Internet]. 2013 Jul;258(1):10–8. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-201307000-00003>
 20. Hawari FI, Nazer LH, Addassi A, Rimawi D, Jamal K. Predictors of ICU Admission in Patients With Cancer and the Related Characteristics and Outcomes. *Crit Care Med* [Internet]. 2016 Mar;44(3):548–53. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003246-900000000-97073>
 21. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Sermeus W. Nurses' reports of working conditions and hospital quality of care in 12 countries in Europe. *Int J Nurs Stud* [Internet]. 2013 Feb;50(2):143–53. Available from: <http://dx.doi.org/10.1016/j.ijnurstu.2012.11.009>
 22. Griffiths P, Jones S, Bottle A. Is “failure to rescue” derived from administrative data in England a nurse sensitive patient safety indicator for surgical care? Observational study. *Int J Nurs Stud* [Internet]. 2013 Feb;50(2):292–300. Available from: <http://dx.doi.org/10.1016/j.ijnurstu.2012.10.016>
 23. Neuraz A, Guérin C, Payet C, Polazzi S, Aubrun F, Dailier F, et al. Patient Mortality Is Associated With Staff Resources and Workload in the ICU. *Crit Care Med* [Internet]. 2015 Aug;43(8):1587–94. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S113023991630013X>
 24. Falk A-C, Wallin E-M. Quality of patient care in the critical care unit in relation to nurse patient ratio: A descriptive study. *Intensive Crit Care Nurs* [Internet]. 2016 Aug;35:74–9. Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=27117560>
 25. Subbe CP, Gao H, Harrison D a. Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med* [Internet]. 2007 Mar 23;33(4):619–24. Available from: <http://link.springer.com/10.1007/s00134-006-0516-8>
 26. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Higgins B. A review, and performance evaluation, of single-parameter “track and trigger” systems. *Resuscitation*. 2008;79(1):11–21.
 27. Graham KC, Cvach M. Monitor Alarm Fatigue: Standardizing Use of Physiological Monitors and Decreasing Nuisance Alarms. *Am J Crit Care* [Internet]. 2010 Jan 1;19(1):28–34. Available from: <http://ajcc.aacnjournals.org/cgi/doi/10.4037/ajcc2010651>
 28. Cvach M. Monitor Alarm Fatigue : An Integrative Review. *Biomed Instrum Technol* [Internet]. 2012 Jul;46(4):268–77. Available from: <http://www.aami-bit.org/doi/abs/10.2345/0899-8205-46.4.268>
 29. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ* [Internet]. 2016 Mar 17;352352:i1174. Available from: <http://dx.doi.org/10.1136/bmj.i1174>
 30. Yilmaz T, Foster R, Hao Y. Detecting Vital Signs with Wearable Wireless Sensors. *Sensors* [Internet]. 2010 Dec 2;10(12):10837–62. Available from: <http://www.mdpi.com/1424-8220/10/12/10837/>
 31. Khan Y, Ostfeld AE, Lochner CM, Pierre A, Arias AC. Monitoring of Vital Signs with Flexible and Wearable Medical Devices. *Adv Mater* [Internet]. 2016 Jun;28(22):4373–95. Available from: <http://doi.wiley.com/10.1002/adma.201504366>
 32. Dorsey ER, Topol EJ. State of Telehealth. *Campion EW*, editor. *N Engl J Med* [Internet]. 2016 Jul 14;375(2):154–61. Available from: <http://www.nejm.org/doi/10.1056/NEJMra1601705>
 33. HealthPatch® MD [Internet]. 2015 [cited 2016 Sep 27]. Available from: <http://www.vitalconnect.com/healthpatch-md>
 34. Choi A. Developers (developer access required) [Internet]. 2014 [cited 2016 Sep 27]. Available from: <https://vitalconnect.atlassian.net/wiki>
 35. Clinical Resource Guide. Revision B. Vital Connect;

36. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2013. p. 4058–61. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6610436>
37. Chan AM, Selvaraj N, Ferdosi N, Narasimhan R. Wireless patch sensor for remote monitoring of heart rate, respiration, activity, and falls. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS. 2013*;6115–8.
38. Selvaraj N. Long-term remote monitoring of vital signs using a wireless patch sensor. In: 2014 IEEE Healthcare Innovation Conference (HIC) [Internet]. IEEE; 2014. p. 83–6. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7038880>
39. Yasuma F, Hayano J. Respiratory Sinus Arrhythmia. *Chest* [Internet]. 2004 Feb;125(2):683–90. Available from: <http://dx.doi.org/10.1378/chest.125.2.683>
40. Wikipedia contributors. Nyquist–Shannon sampling theorem [Internet]. Wikipedia, The Free Encyclopedia. 2016 [cited 2016 Sep 27]. Available from: https://en.wikipedia.org/w/index.php?title=Nyquist–Shannon_sampling_theorem&oldid=740641472
41. Selvaraj N, Narasimhan R. Automated prediction of the apnea-hypopnea index using a wireless patch sensor. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society [Internet]. IEEE; 2014. p. 1897–900. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6943981>
42. Selvaraj N. Psychological acute stress measurement using a wireless adhesive biosensor. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2015. p. 3137–40. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7319057>
43. FDA. HealthPatch MD FDA summary & clearance 510(k) K132447 [Internet]. 2014. Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K132447>
44. FDA. HealthPatch MD FDA summary & clearance 510(k) K141167 [Internet]. 2014. Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K141167>
45. FDA. HealthPatch MD FDA summary & clearance 510(k) K152139 [Internet]. 2015. Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K152139>
46. Vital Connect Revolutionizes Wearable Medical Biosensors with New Product Launch at the 2016 HIMSS Conference & Exhibition [Internet]. 2016 [cited 2016 Sep 27]. Available from: <http://www.vitalconnect.com/news/vital-connect-revolutionizes-wearable-medical-biosensors-with-new-product-launch-at-the-2016-himss-conference-exhibition>
47. VitalPatch® [Internet]. 2016 [cited 2016 Sep 27]. Available from: <http://www.vitalconnect.com/vitalpatch>
48. Puurtinen M, Viik J, Hyttinen J. Best Electrode Locations for a Small Bipolar ECG Device: Signal Strength Analysis of Clinical Data. *Ann Biomed Eng* [Internet]. 2009 Feb 20;37(2):331–6. Available from: <http://link.springer.com/10.1007/s10439-008-9604-y>
49. Fluke Biomedical. ProSim 4 Vital Signs Patient Simulator [Internet]. [cited 2016 Sep 29]. Available from: <http://www.flukebiomedical.com/biomedical/usen/patient-simulators/prosim-4-vital-signs-patient-simulator.htm?PID=72621>
50. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986 Feb;327(8476):307–10.
51. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* [Internet]. 1999 Jun 1;8(2):135–60. Available from: <http://www.ingentaselect.com/psv/cgi-bin/cgi?ini=xref&body=linker&reqdoi=10.1191/096228099673819272>
52. Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. *Stat* [Internet]. 1983 Sep;32(3):307. Available from: <http://www.jstor.org/stable/2987937?origin=crossref>
53. Bland JM, Altman DG. Agreement Between Methods of Measurement with Multiple Observations Per Individual. *J Biopharm Stat*. 2007 Jul 2;17(4):571–82.
54. van Loon K, Breteler MJM, van Wolfwinkel L, Rheineck Leyssius a. T, Kossen S, Kalkman CJ, et al. Wireless non-invasive continuous respiratory monitoring with FMCW radar: a clinical validation study. *J Clin Monit Comput* [Internet]. 2015 Sep 30; Available from: <http://link.springer.com/10.1007/s10877-015-9777-5>
55. Kalkman CJ. Medical Devices, Electronic Health Records and Assuring Patient Safety: Future Challenges? In: Koornneef F, van Gulijk C, editors. *Computer Safety, Reliability, and Security*:

- 34th International Conference, SAFECOMP 2015, Delft, The Netherlands, September 23-25, 2015, Proceedings [Internet]. Cham: Springer International Publishing; 2015. p. 3–6. Available from: http://dx.doi.org/10.1007/978-3-319-24255-2_1
56. Myles PS, Cui J. I. Using the Bland Altman method to measure agreement with repeated measures. *Br J Anaesth*. 2007 Jun 7;99(3):309–11.
57. Van Loon K, Van Zaane B, Kalkman CJ, Peelen LM. Comparison of statistics for methods comparison studies with repeated measurements: A simulation study. In: Annual Meeting of the International Anesthesia Research Society (IARS), San Francisco, May 2016. 2016.
58. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM* [Internet]. 2001 Oct 1;94(10):521–6. Available from: <http://www.qjmed.oupjournals.org/cgi/doi/10.1093/qjmed/94.1.0.521>
59. Barlow GD, Nathwani D, Davey PG. Standardised early warning scoring system. *Clin Med (Northfield II)*. 2006;6(4):422–3.
60. Paterson R, MacLeod D, Thetford D, Beattie A, Graham C, Lam S, et al. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin Med (Northfield II)* [Internet]. 2006 May 1;6(3):281–4. Available from: [internal-pdf://13.6.254.173/Paterson-2006-Prediction of in-hospital mortal.pdf%5Cnhttp://www.clinmed.rcpjournals.org/content/6/3/281.full.pdf](http://13.6.254.173/Paterson-2006-Prediction%20of%20in-hospital%20mortality.pdf)
61. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS—Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* [Internet]. 2010 Aug;81(8):932–7. Available from: <http://dx.doi.org/10.1016/j.resuscitation.2010.04.014>
62. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform* [Internet]. 2013 Oct;46(5):837–48. Available from: <http://dx.doi.org/10.1016/j.jbi.2013.06.011>
63. Finlay GD, Rothman MJ, Smith R a. Measuring the modified early warning score and the Rothman Index: Advantages of utilizing the electronic medical record in an early warning system. *J Hosp Med* [Internet]. 2014 Feb;9(2):116–9. Available from: <http://doi.wiley.com/10.1002/jhm.2132>
64. Houthoofd R, Ruysinck J, van der Hertten J, Stijven S, Couckuyt I, Gadeyne B, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intell Med* [Internet]. 2015 Mar;63(3):191–207. Available from: <http://dx.doi.org/10.1016/j.artmed.2014.12.009>
65. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A Severe Sepsis Mortality Prediction Model and Score for Use With Administrative Data. *Crit Care Med* [Internet]. 2016 Feb;44(2):319–27. Available from: http://journals.lww.com/ccmjournal/Abstract/publishahead/A_Severe_Sepsis_Mortality_Prediction_Model_and.97110.aspx
66. Le Manach Y, Collins G, Rodseth R, Le Bihan-Benjamin C, Biccari B, Riou B, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM). *Anesthesiology* [Internet]. 2016 Mar;124(3):570–9. Available from: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000542-201603000-00016>
67. Morgan R, Williams F, Wright M. An Early Warning Scoring System for detecting developing critical illness. *Clin Intensive Care* [Internet]. 1997;8(2):100. Available from: <http://www.tandfonline.com/doi/abs/10.3109/tcic.8.2.93.110>
68. Morgan RJM, Wright MM. In defence of early warning scores. *Br J Anaesth* [Internet]. 2007 Sep 27;99(5):747–8. Available from: <http://bj.oxfordjournals.org/lookup/doi/10.1093/bja/aem285>
69. Stenhouse C, Coates S, Tivey M, Allsop P, Parker T. Prospective evaluation of a modified Early Warning Score to aid earlier detection of patients developing critical illness on a general surgical ward. *Br J Anaesth* [Internet]. 2000;84(5):663. Available from: <http://connection.ebscohost.com/c/articles/44420088/prospective-evaluation-modified-early-warning-score-aid-earlier-detection-patients-developing-critical-illness-general-surgical-ward>
70. NHS. Paediatric Early Warning Scores (PEWS) [Internet]. 2015 [cited 2016 Nov 13]. Available from: http://www.institute.nhs.uk/safer_care/paediatric_safer_care/pews.html
71. Nirmal D, Goodsell R. Trust Guideline for the use of the Modified Early Obstetric Warning Score (MEOWS) in detecting the seriously ill and deteriorating woman. [Internet]. 2016. Available from: <http://www.nnuh.nhs.uk/publication/modified-early-obstetric-warning-score-meows-mid33-ao13-v4-1/>
72. National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. London: RCP; 2012.
73. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of

- vital signs. Resuscitation. 2011;82(8):1013–8.
74. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* [Internet]. 2009 Sep 1;338(jun 29 1):b2393–b2393. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.b2393>
 75. Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis [Internet]. Second Edi. Springer International Publishing; 2015. (Springer Series in Statistics). Available from: <https://books.google.nl/books?id=94RgCgAAQBAJ>
 76. The MathWorks. Generalized Linear Mixed-Effects Models [Internet]. 2016 [cited 2016 Dec 1]. Available from: <https://mathworks.com/help/stats/generalized-linear-mixed-effects-models.html>
 77. Falk M, Marohn F, Michel R, Hofmann D, Macke M, Spachmann C, et al. A First Course on Time Series Analysis with SAS [Internet]. 2012 [cited 2016 Dec 11]. Available from: http://www.statistik-mathematik.uni-wuerzburg.de/wissenschaftsforschung/time_series/
 78. Box GEP, Jenkins GM. Time series analysis: forecasting and control. Holden-Day; 1976.
 79. Van Der Heijden F, Duin RPW, De Ridder D, Tax DMJ. Classification, Parameter Estimation and State Estimation, An Engineering Approach using MATLAB. John Wiley & Sons, Ltd; 2005.
 80. Ganesan DN, Venkatesh DK, Rama DMA, Palani AM. Application of Neural Networks in Diagnosing Cancer Disease using Demographic Data. *Int J Comput Appl* [Internet]. 2010 Feb 25;1(26):81–97. Available from: <http://www.ijcaonline.org/journal/number26/pxc387783.pdf>
 81. Dionisi A, Marioli D, Sardini E, Serpelloni M. Autonomous Wearable System for Vital Signs Measurement With Energy-Harvesting Module. *IEEE Trans Instrum Meas* [Internet]. 2016;1–12. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7398038>
 82. Chetelat O, Ferrario D, Proenca M, Porchet J-A, Falhi A, Grossenbacher O, et al. Clinical validation of LTMS-S: A wearable system for vital signs monitoring. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2015. p. 3125–8. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7319054>
 83. Granholm A, Pedersen NE, Lippert A, Petersen LF, Rasmussen LS. Respiratory rates measured by a standardised clinical approach, ward staff, and a wireless device. *Acta Anaesthesiol Scand* [Internet]. 2016 Nov;60(10):1444–52. Available from: <http://doi.wiley.com/10.1111/aa.12784>
 84. Hernandez-Silveira M, Wiczorkowski-Rettinger K, Ang S, Burdett A. Preliminary assessment of the SensiumVitals®: A low-cost wireless solution for patient surveillance in the general wards. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2015. p. 4931–7. Available from: <http://ieeexplore.ieee.org/document/7319498/>
 85. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. IEEE; 2015. p. 8099–102. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7320273>
 86. Steinhubl SR, Marriott MP, Wegerich SW. Remote Sensing of Vital Signs: A Wearable, Wireless “Band-Aid” Sensor With Personalized Analytics for Improved Ebola Patient Care and Worker Safety. *Glob Heal Sci Pract* [Internet]. 2015 Sep 10;3(3):516–9. Available from: <http://www.ghspjournal.org/cgi/doi/10.9745/GHSP-D-15-00189>
 87. Zimlichman E, Shinar Z, Rozenblum R, Levkovich S, Amital H, Shoenfeld Y. Using Continuous Motion Monitoring Technology to Determine Patient’s Risk for Development of Pressure Ulcers. *J patient saf.* 2011;7(0):181–4.
 88. Klap T, Shinar Z. Using Piezoelectric Sensor for Continuous-Contact-Free Monitoring of Heart and Respiration Rates in Real-Life Hospital Settings. *Comput Cardiol* (2010). 2013;(40):671–4.
 89. Zimlichman E, Szyper-Kravitz M, Shinar Z, Klap T, Levkovich S, Unterman A, et al. Early recognition of acutely deteriorating patients in non-intensive care units: Assessment of an innovative monitoring technology. *J Hosp Med* [Internet]. 2012 Oct;7(8):628–33. Available from: <http://doi.wiley.com/10.1002/jhm.1963>
 90. Brown H V, Angeles L, Zimlichman E. Improved Outcomes and Reduced Costs with Contact-free Continuous Patient Monitoring on a Medical-Surgical Hospital Unit. *White Pap.* 2010;
 91. Huizinga E. Bland-Altman Analysis [Internet]. 2016. Available from:

- <https://github.com/erikhuizinga/Bland-Altman-Analysis>
92. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* [Internet]. 1958 Jun;53(282):457. Available from: <http://www.jstor.org/stable/2281868?origin=crossref>
93. Cox DR. Regression Models and Life-Tables. *J R Stat Soc* 1972;34(2):187–220.
94. Huizinga E. honeycomb [Internet]. 2017. Available from: <https://github.com/erikhuizinga/honeycomb/>
95. Wikipedia contributors. Hexaxial reference system [Internet]. Wikipedia, The Free Encyclopedia. 2015 [cited 2016 Sep 30]. Available from: https://en.wikipedia.org/w/index.php?title=Hexaxial_reference_system&oldid=692334075
96. Wikipedia contributors. Artificial Neural Network [Internet]. Wikipedia, The Free Encyclopedia. 2011 [cited 2016 Dec 19]. Available from: https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

Appendix A. Bland-Altman Analysis in a MATLAB implementation

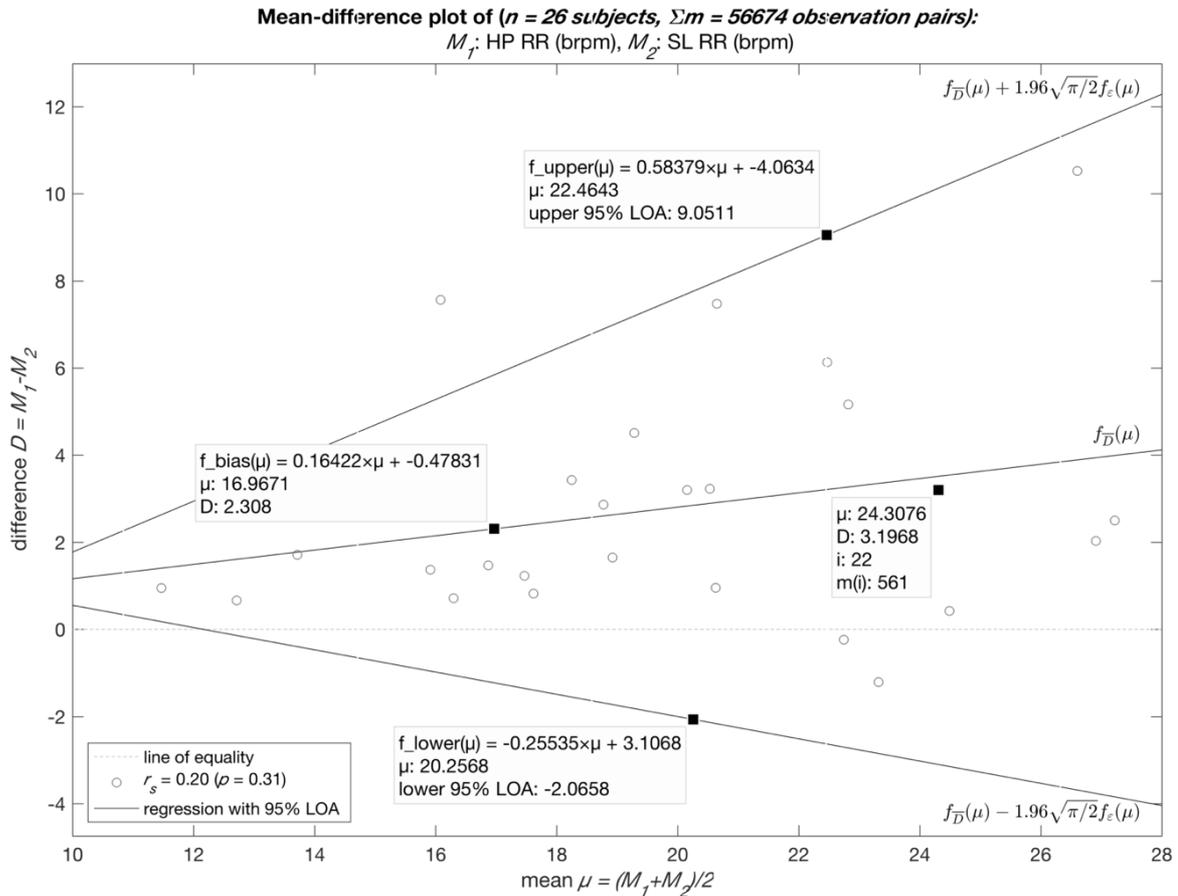


Figure 27 – Some of the graphical capabilities of the Bland-Altman analysis tool programmed in MATLAB are shown. Note the data tips added to the graph show additional information about the lines or individual data points. For example, the regression line coefficients are given, the mean and difference are shown, the subject number and number of observations are given.

The methods used for agreement analysis of HealthPatch MD and Spacelabs XPRESSON vital signs signals are described in §4.4.2. Because all analyses were performed in MATLAB, an implementation of Bland-Altman Analysis (BAA) was required for the specific type of data in this study. Basic BAA implementations are available online¹⁶, but none implement methods for the various cases of repeated measurements (equal/unequal number of replicates, assuming constant or variable true values), nor any of the other more advanced methods found in literature from Bland, Altman and others on the subject. Therefore, a MATLAB implementation was written and published on GitHub [91].

The latest main¹⁷ and development¹⁸ releases are available for download. This software is still under development and is in an early release stage where some desired functionality has not yet been implemented. Nonetheless, it is already a versatile, customisable and very fast piece of software (the calculations in §4.4.2 take about four seconds).

¹⁶ <https://mathworks.com/matlabcentral/fileexchange/?term=bland> (accessed 24 October 2016)

¹⁷ <https://github.com/erikhuizinga/Bland-Altman-Analysis/releases/latest>

¹⁸ <https://github.com/erikhuizinga/Bland-Altman-Analysis/archive/develop.zip>

The version used for the analyses of this thesis has been labelled as the TM (telemonitoring) release¹⁹. The features supported by the TM release are listed below. A demonstration of the graphical output is shown in Figure 27.

- Regular BAA is supported
 - Various agreement statistics between the two methods are supported
 - Difference (default)
 - Ratio
 - Standard deviation (of either method, or their difference)
 - Pearson correlation (with p -value)
 - Limits of agreement (LOA) are calculated
 - LOA confidence intervals (CI) at a configurable significance level are calculated
 - Bias (mean difference) is calculated
 - Bias CI at a configurable significance level is calculated
 - Various statistics of the input data are calculated
 - Standard deviation of the agreement statistic
 - Spearman rank correlation of the agreement statistic with mean (with p -value)
 - Regression of the statistic on the mean (with mean squared error (MSE))
- Creation of various graphs is supported
 - Mean-agreement statistic plot (e.g. mean-difference plot)
 - Pearson correlation plot between the two methods
 - Various statistics can be added to the graphs, such as LOA, bias, correlation, regression, etc.
 - Various plot customisation options are supported, such as setting the names of the methods
 - The graph can be configured to be very concise, or to be very information-dense
- BAA for repeated measurements is supported
 - Equal number of replicates are supported
 - Unequal number of replicates are supported
 - Assuming either a constant or varying true value is supported
 - Additional statistics for the repeated measurements situation are calculated
 - Least-squares simple linear regression line of bias and LOA
 - With and without assumption of constant residual variance
 - Within-subject variances are calculated
- Transformation of the input data is supported, e.g. logarithmic
- Exclusion of individual samples is support, e.g. to study the influence of (removal of) outliers
- A demonstration of the methods and data of Bland and Altman's 1999 and 2007 articles is implemented [51,53]. The articles contain many tables of example data from true experiments. These data and the published results were used to validate the implementation of this software

¹⁹ <https://github.com/erikhuizinga/Bland-Altman-Analysis/tree/TM>

Appendix B. Survival analysis of HealthPatch measurements

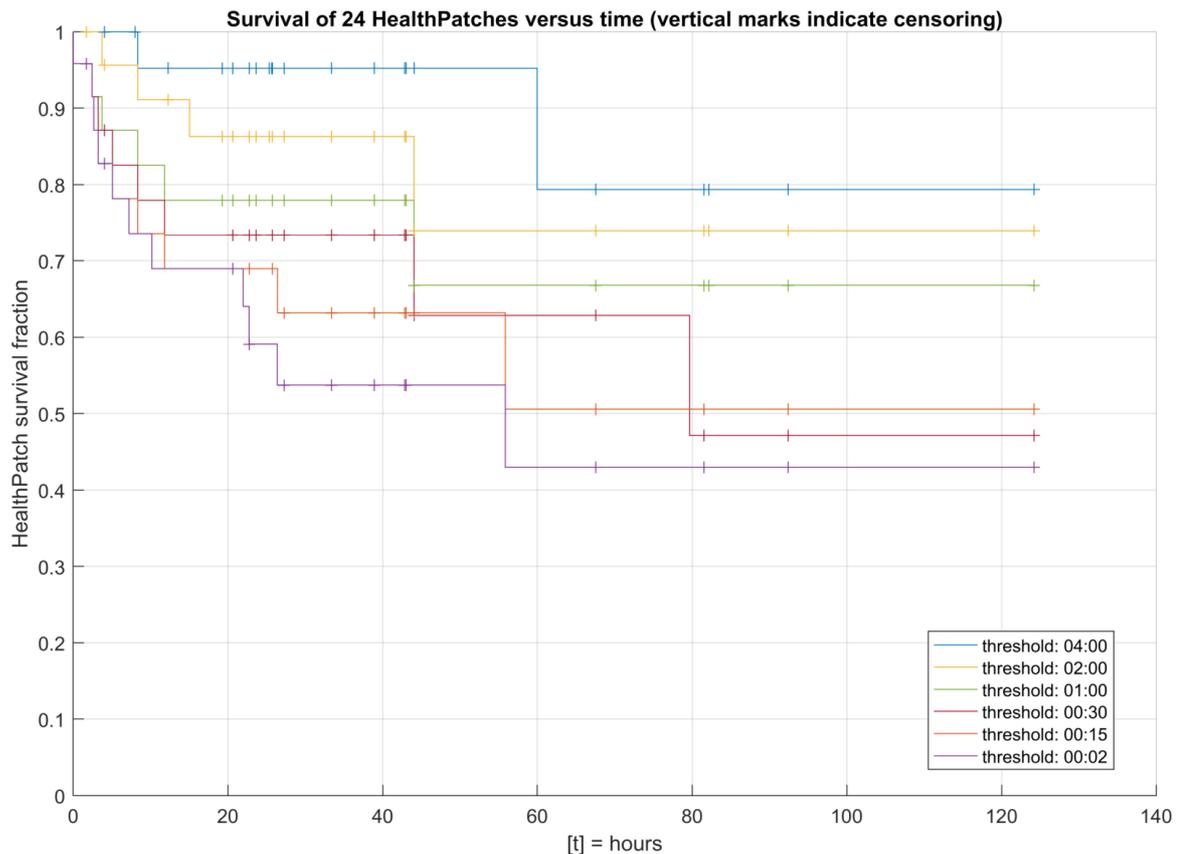


Figure 28 – Survival analysis of the first 24 HealthPatch measurements. When signal loss lasts longer than one of the various threshold times (see legend in hours:minutes format), it counts as a failure.

In an early stage of this thesis research, it was planned to perform reliability analysis of the HealthPatch MD measurements. Reliability was to be expressed as amount of data loss and time to failure. The first is already listed in Chapter 4, but the latter is not yet defined or shown. Time to failure was defined as the duration of a measurement until it contained a large enough gap to be considered as unacceptable. After all, to enhance patient safety a measurement must be somewhat continuous and gaps may not be large.

To assess the survival of measurements, it can be analysed with Kaplan-Meier analysis [92] and Cox regression [93]. In an early stage of the reliability analysis a survival plot (Kaplan-Meier) was made, shown in Figure 28. Censoring was applied when a measurement ended due to another reason than failure to transmit data, i.e., when the measurement ended before the patch could not transmit any more it was censored.

Multiple thresholds were used in the analysis, ranging from two minutes to four hours. The threshold determined how long the duration of a gap was allowed to be before the measurement was considered a failure. As can be seen, with a threshold of two minutes less than 50% of the HealthPatch measurement 'survive'. After 24 hours, about 40% have already failed. Even with a four-hour threshold about 20% fail.

A two-minute threshold is strict, but only allows the HealthPatch to have short periods of data loss. This may be practical in clinical practice, where a threshold of hours is unacceptable.

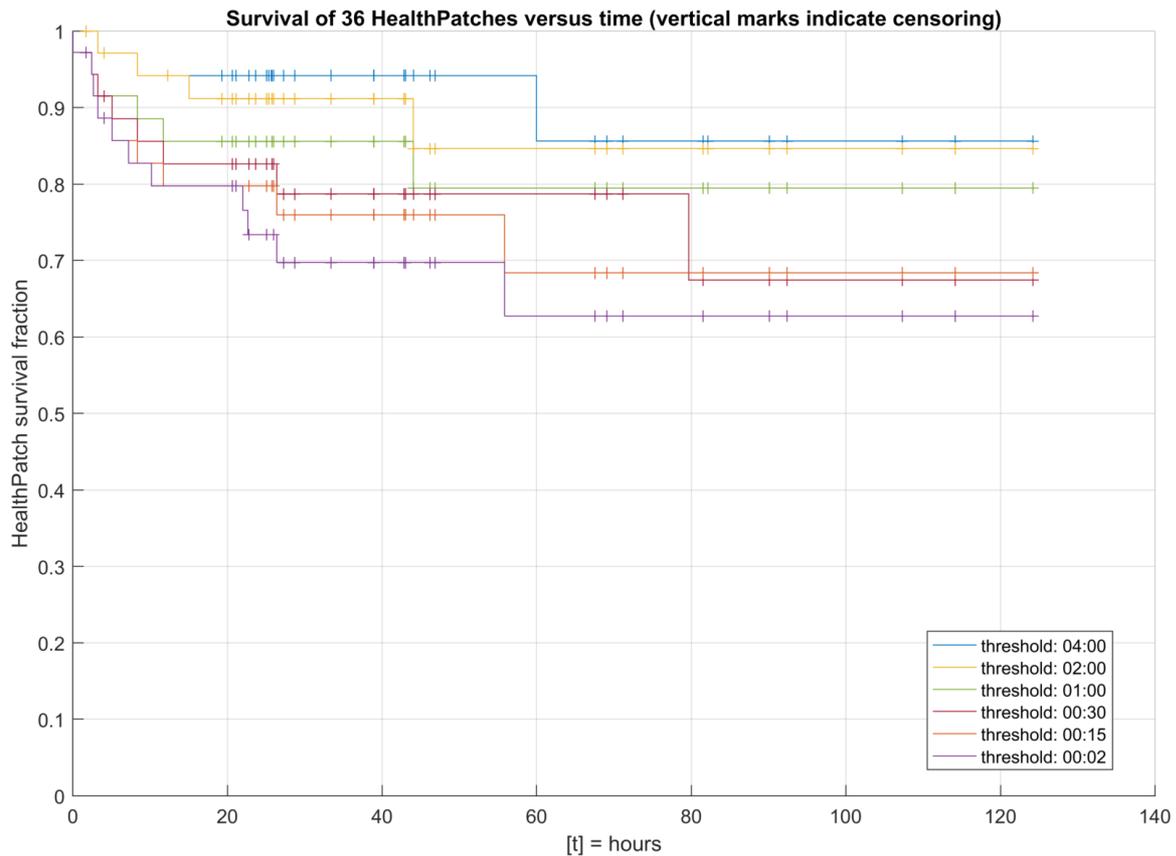


Figure 29 – Survival analysis of all 36 HealthPatch measurements. When the signal loss lasts longer than one of the various threshold times (see legend in hours:minutes format), it counts as a failure.

However, in a situation at home or outside, it may not be realistic to require gaps in the data to be shorter than a few minutes. There, a less strict threshold would be preferable.

The same analysis was redone with all available HealthPatch measurements at the end of this study, the results of which can be seen in Figure 29. It is clear, although no significance has been calculated, that HealthPatch measurements survived longer in the last 12 measurements than in the first 24. The main difference between the graphs is the amount of censoring is greater in the second figure, which shows there were relatively fewer failures among the last 12 measurements. This can also be seen in the general shape of the survival lines, which is generally the same; this indicates most failures occurred in the first 24 measurements. In this second analysis, the results are much better than in the first. About 85% survived the entire measurement with a threshold of two or four hours, which is a 10% and 5% increase respectively. The two-minute threshold 24-hour survival increased from 60% to 70%.

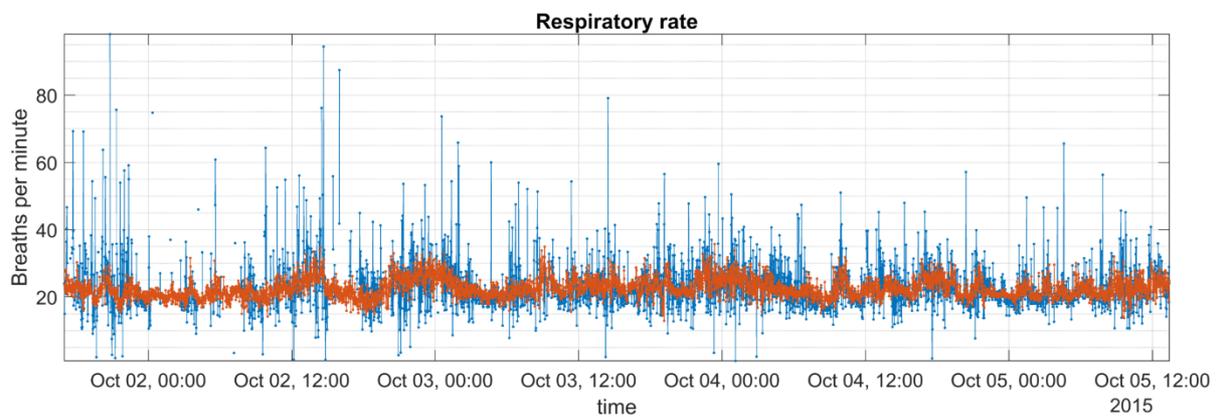
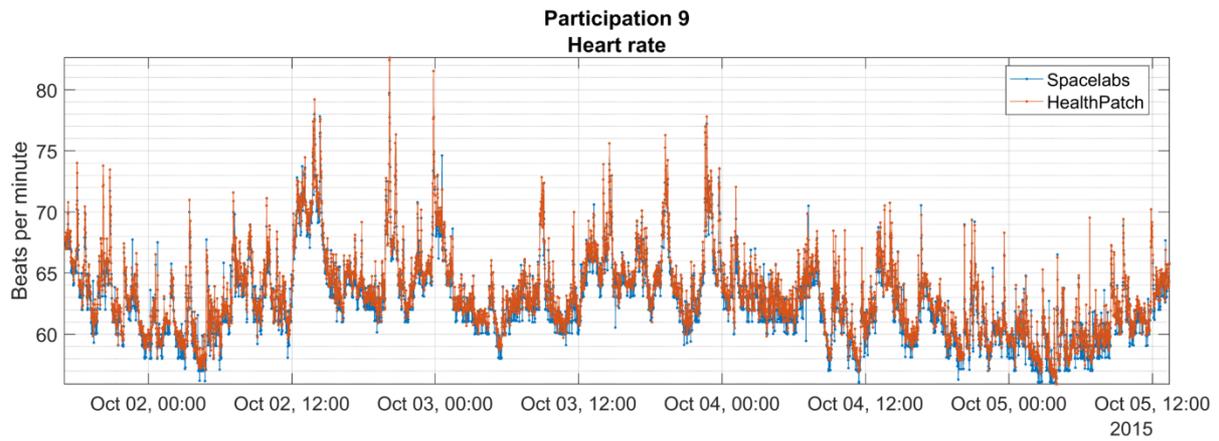
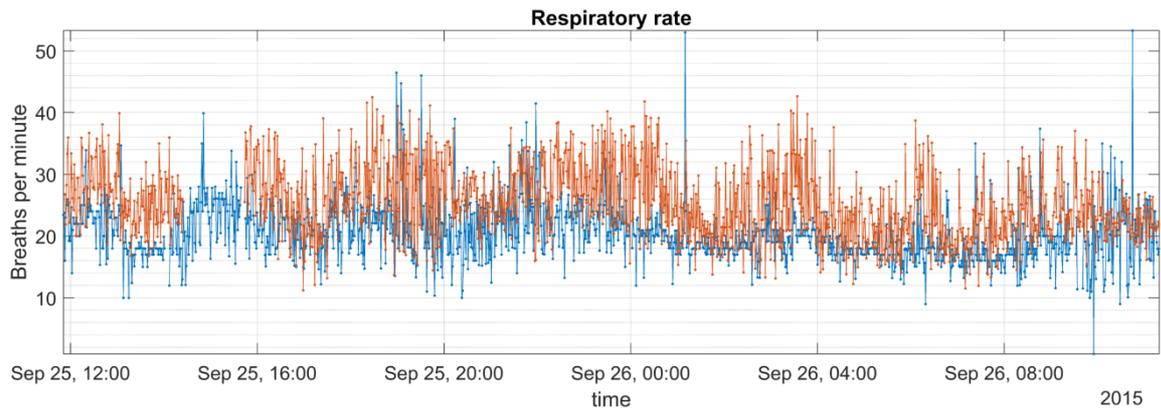
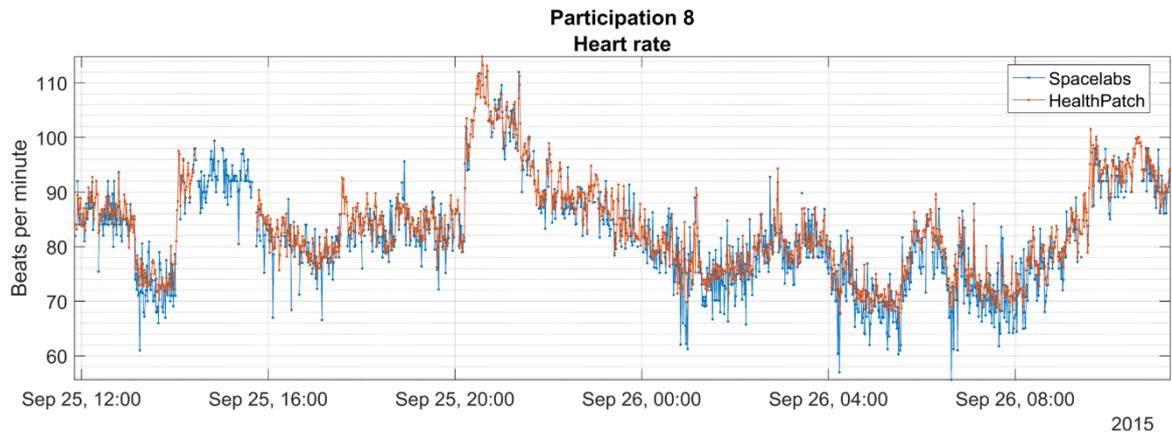
In both analyses, the large fraction of censored measurements induces a large bias when trying to assess how long the HealthPatch can measure. In theory, the patch should be able to measure 96 hours. Most measurements ended well before this mark, so we cannot tell if these patches would have made it to the 96 hour mark, or that they would have failed. A practical reason for this was that measurements took place only on the medium care ward and could not be continued on the wards where patients were dismissed to.

Appendix C. Plots of all participations' vital signs as measured by HealthPatch MD and Spacelabs XPREZZON

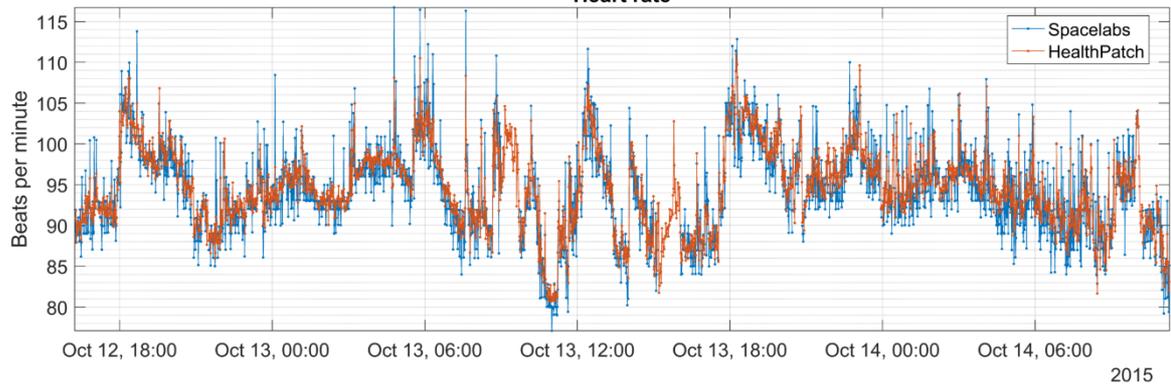
This appendix contains plots of all participations. The HealthPatch MD and Spacelabs measurements have been pre-processed as described in §4.3, i.e., the data has been uniformly sampled, resampled (to one sample per minute), synchronised (HealthPatch to Spacelabs) and trimmed (samples before start and after end of the other measurement device are removed).

Because for some measurements there was no data due to missing data files, these participations have been excluded from this section.

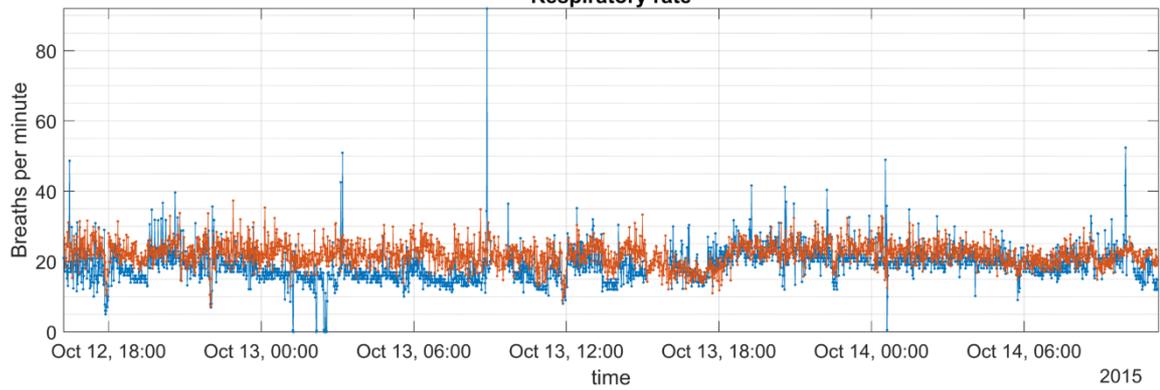
Note that all figures in this appendix show participation numbers from 8 onwards. These numbers correspond to the numbers in the database and are used here for reference. The first seven participations were excluded from this study, because they were measured in previous research using different measurement protocols.



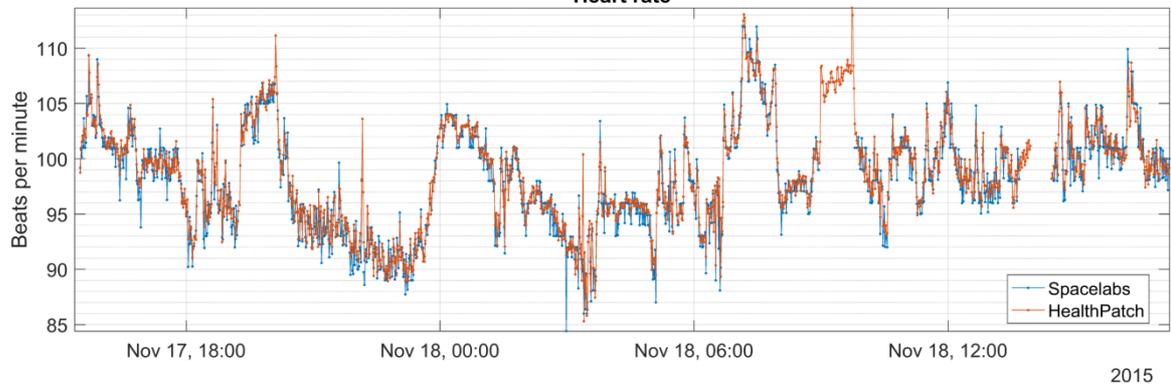
Participation 10
Heart rate



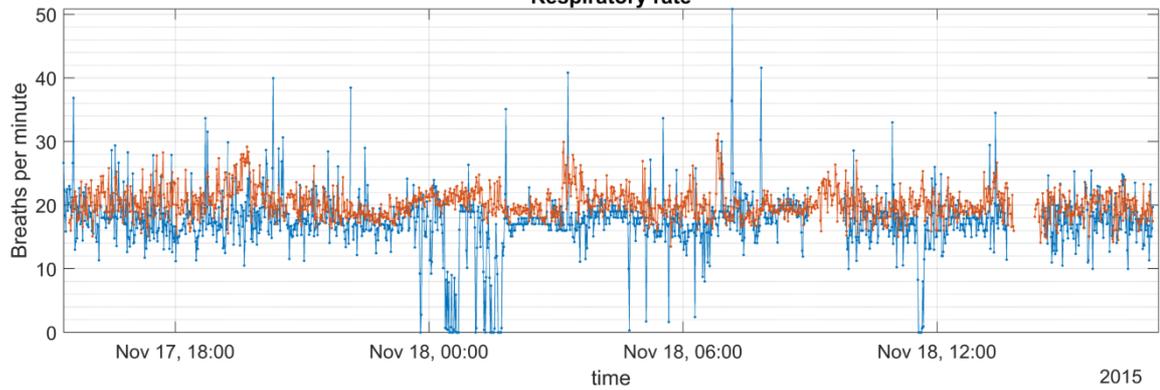
Respiratory rate



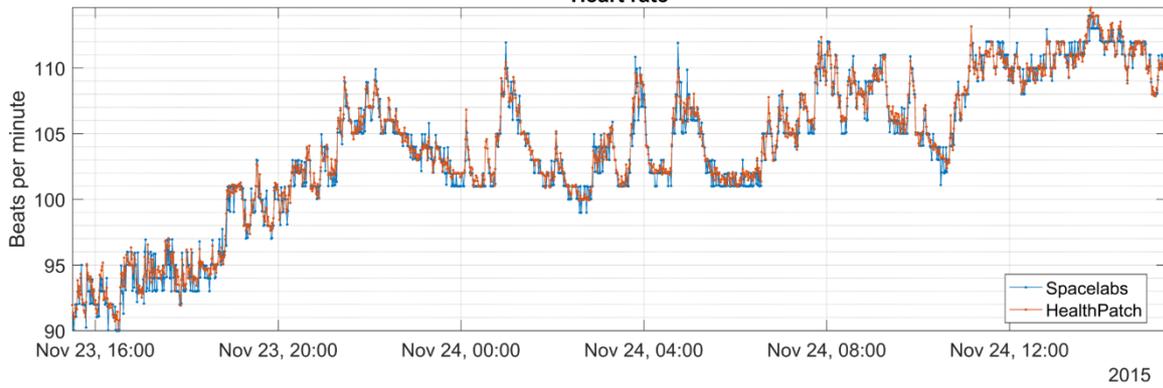
Participation 11
Heart rate



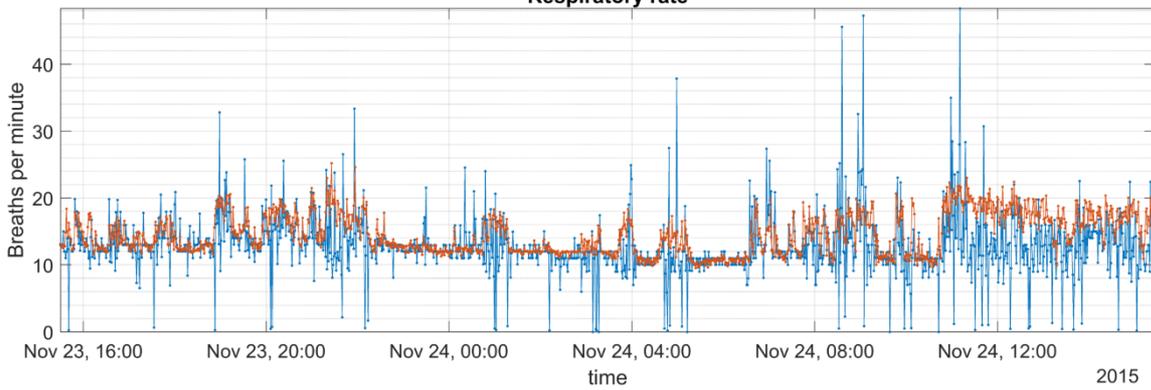
Respiratory rate



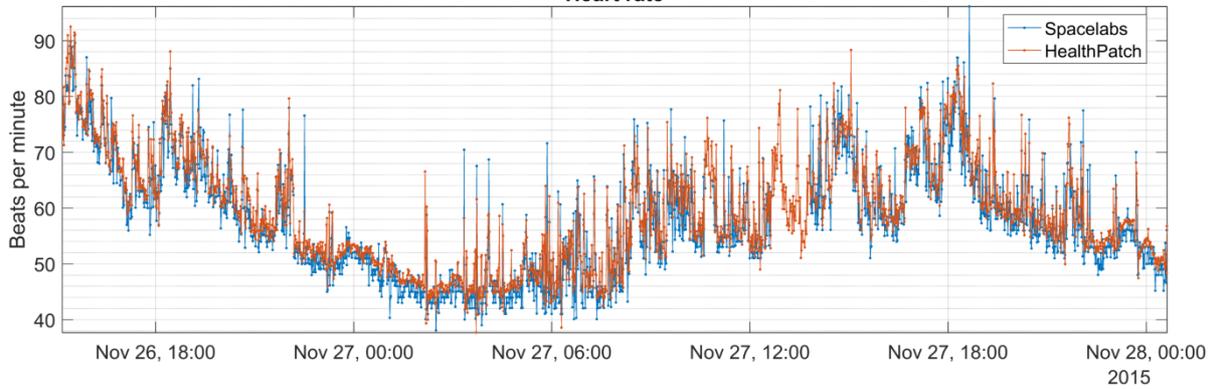
Participation 12
Heart rate



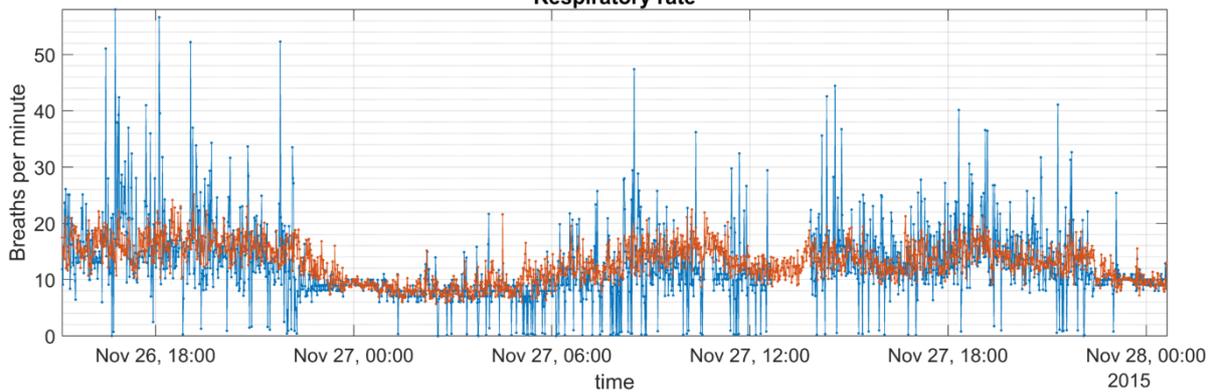
Respiratory rate



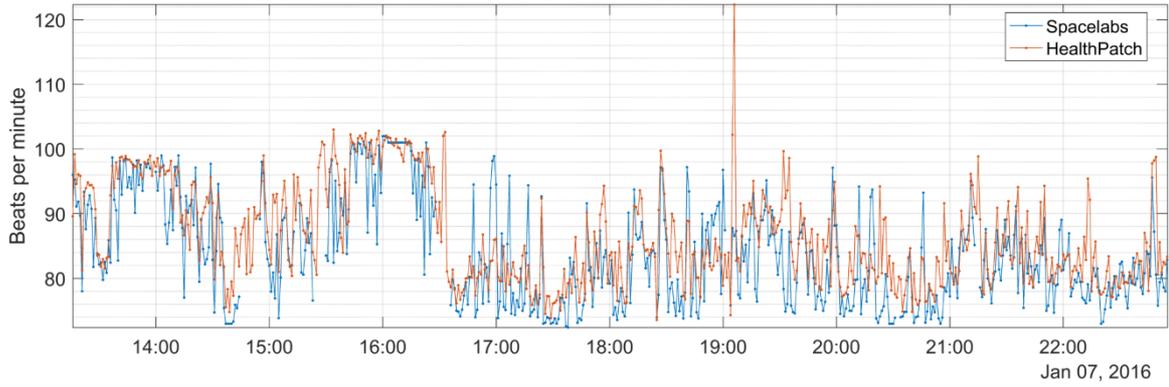
Participation 13
Heart rate



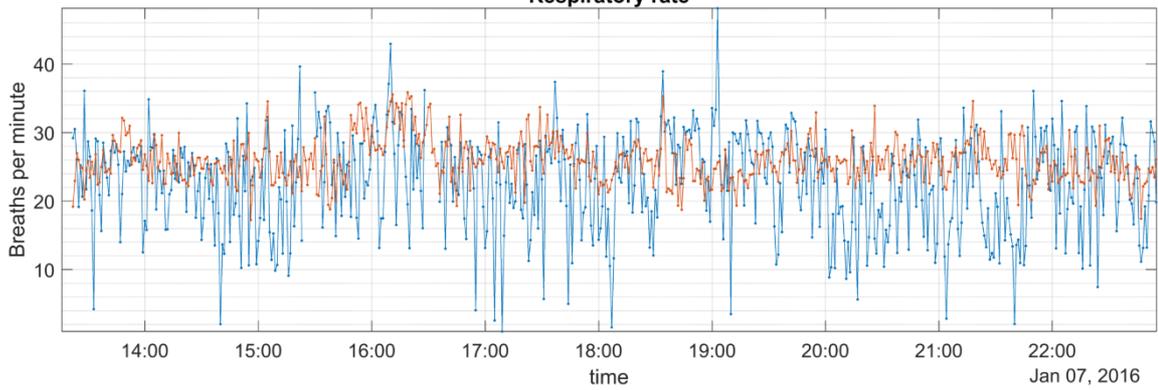
Respiratory rate



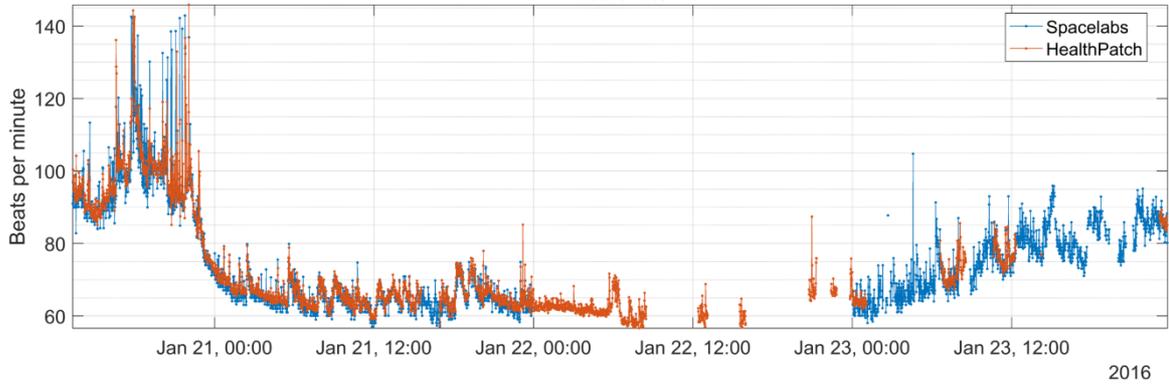
Participation 17
Heart rate



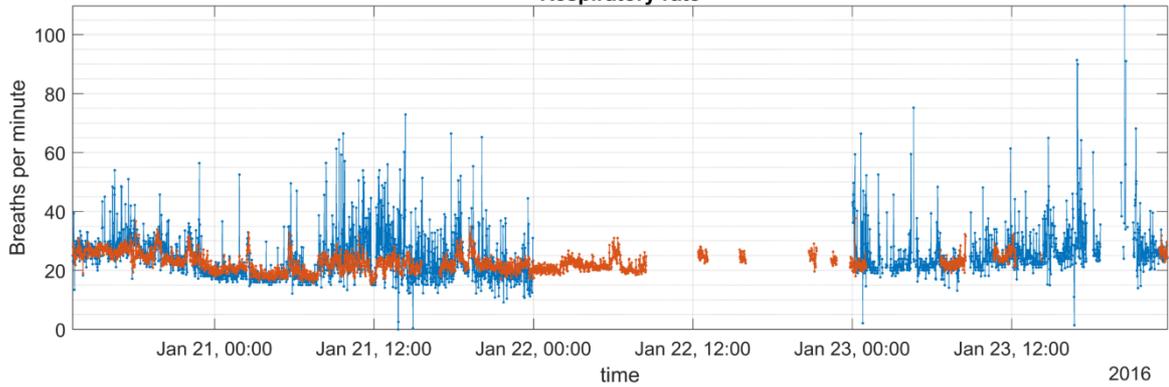
Respiratory rate



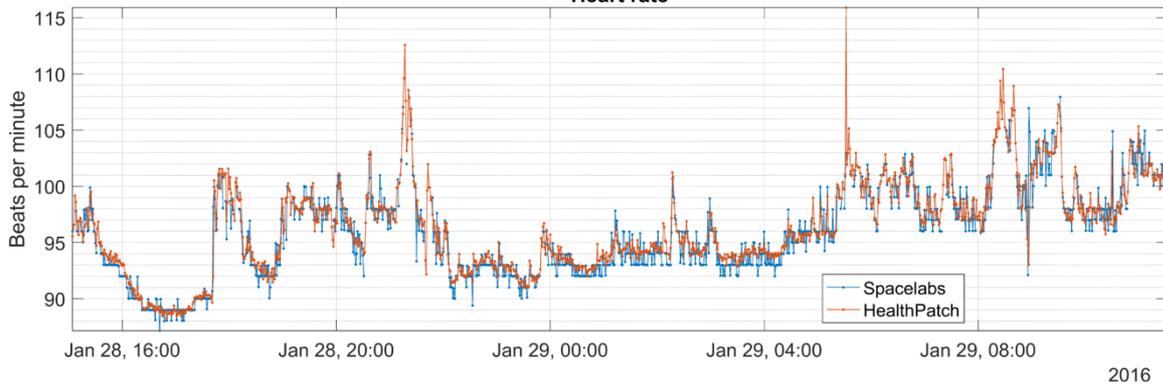
Participation 18
Heart rate



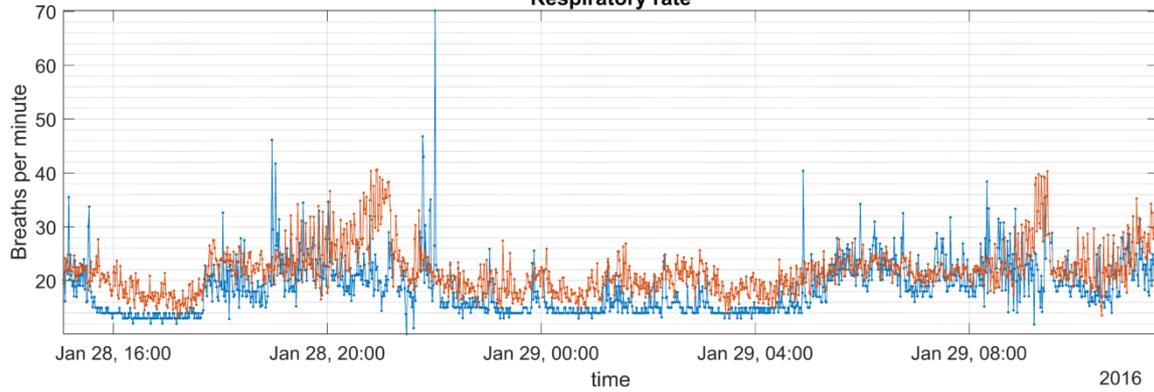
Respiratory rate



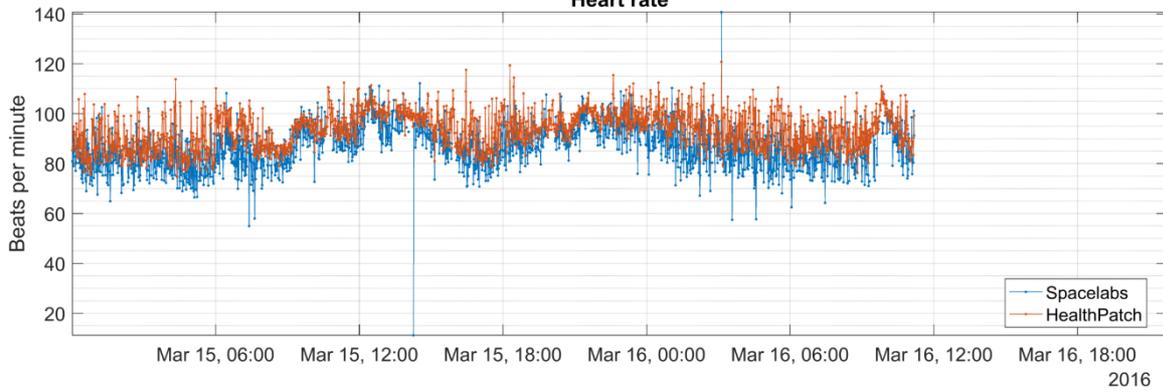
Participation 19
Heart rate



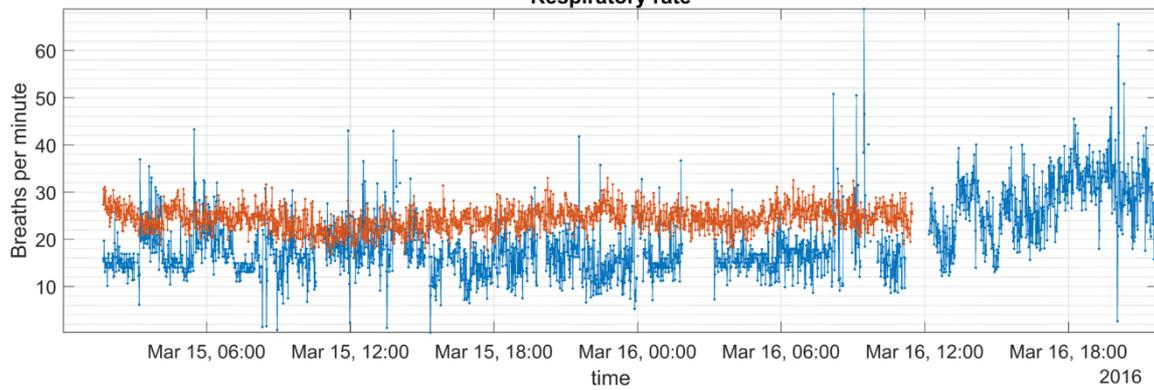
Respiratory rate



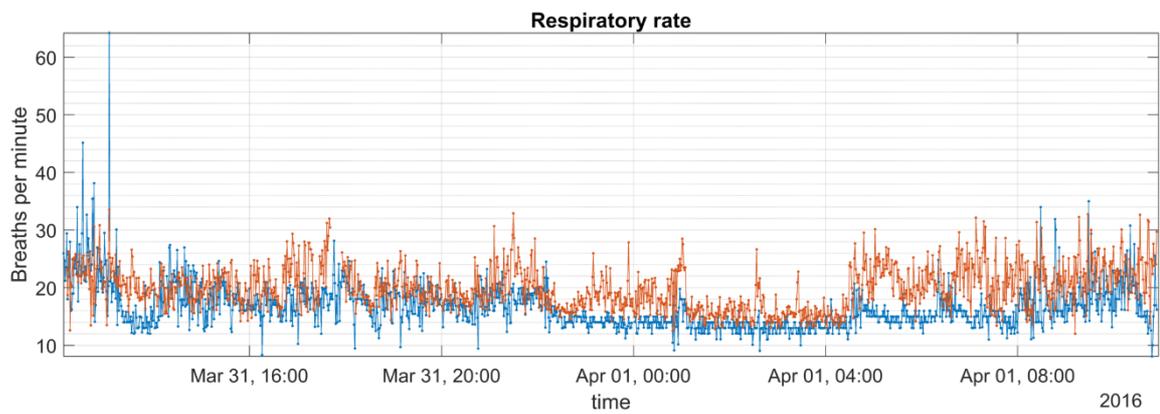
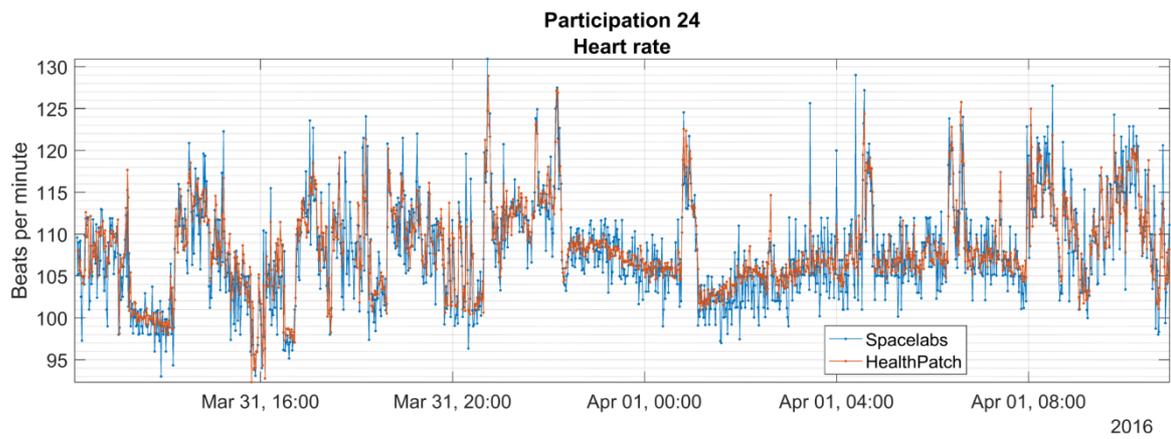
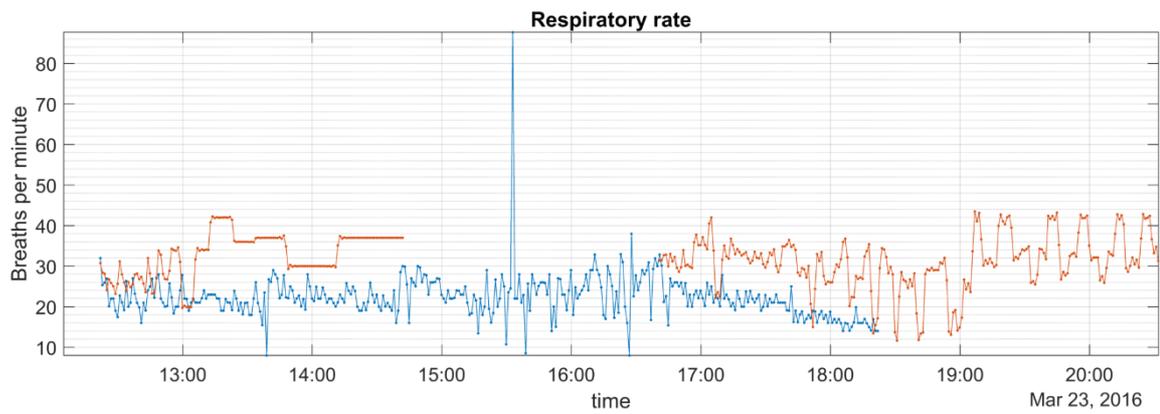
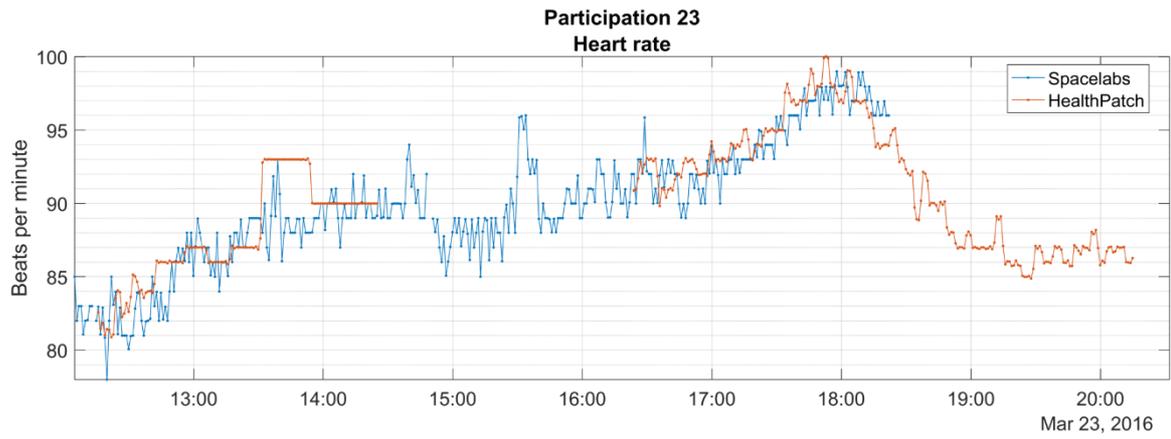
Participation 22
Heart rate



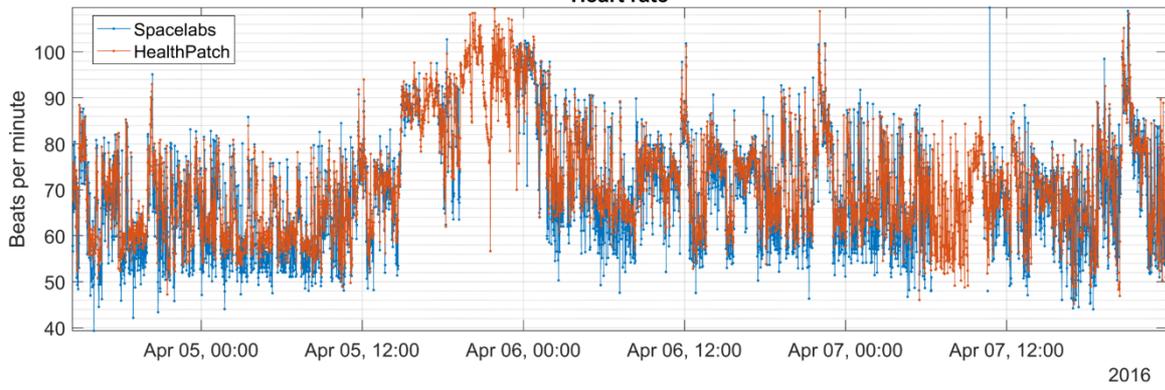
Respiratory rate



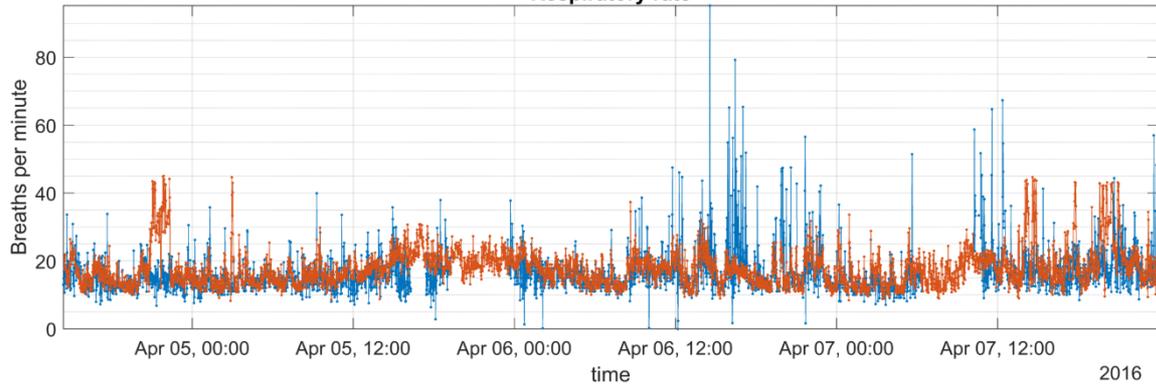
X



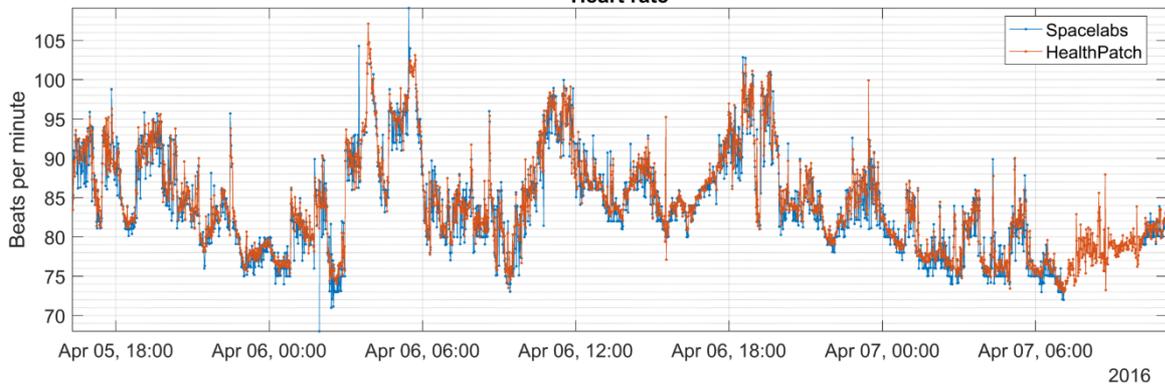
Participation 25
Heart rate



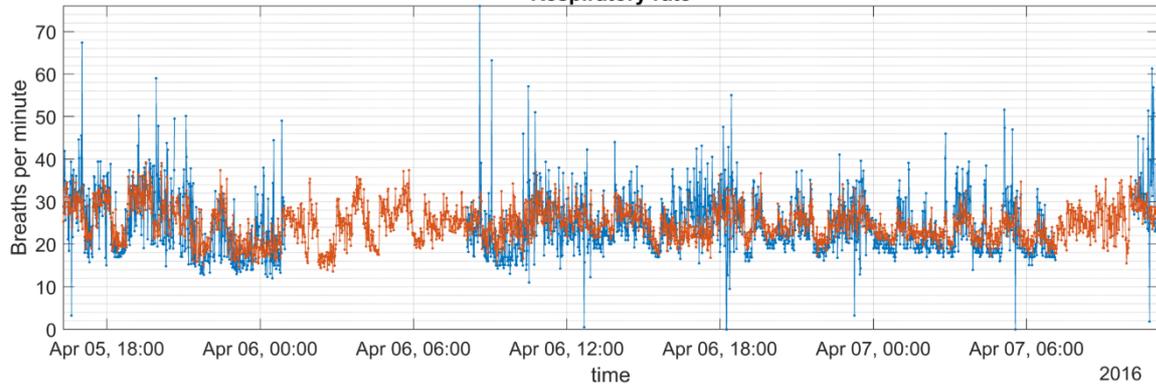
Respiratory rate



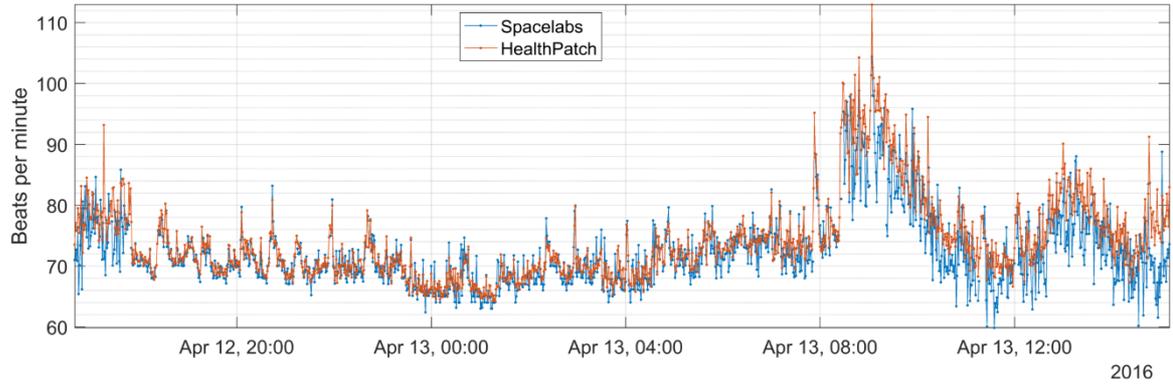
Participation 26
Heart rate



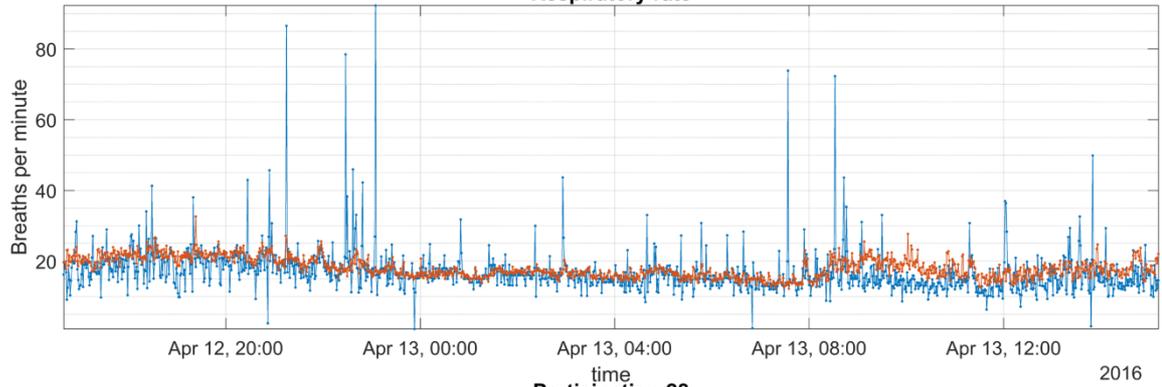
Respiratory rate



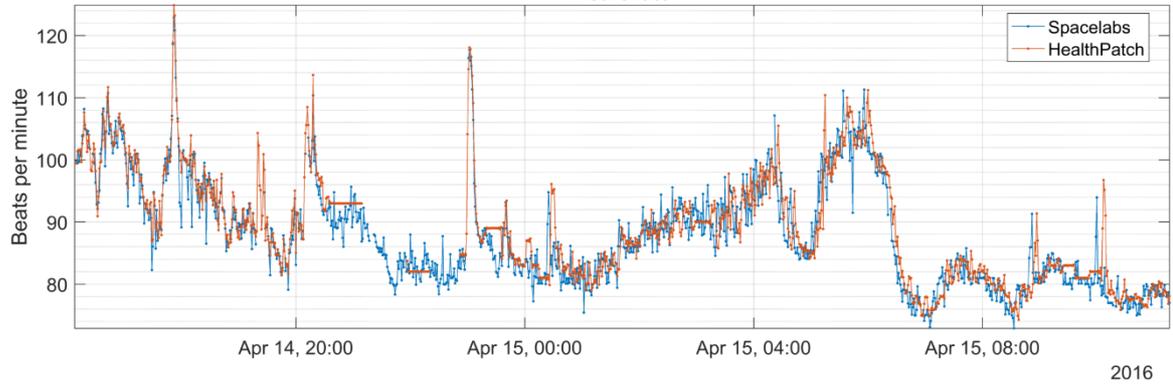
**Participation 27
Heart rate**



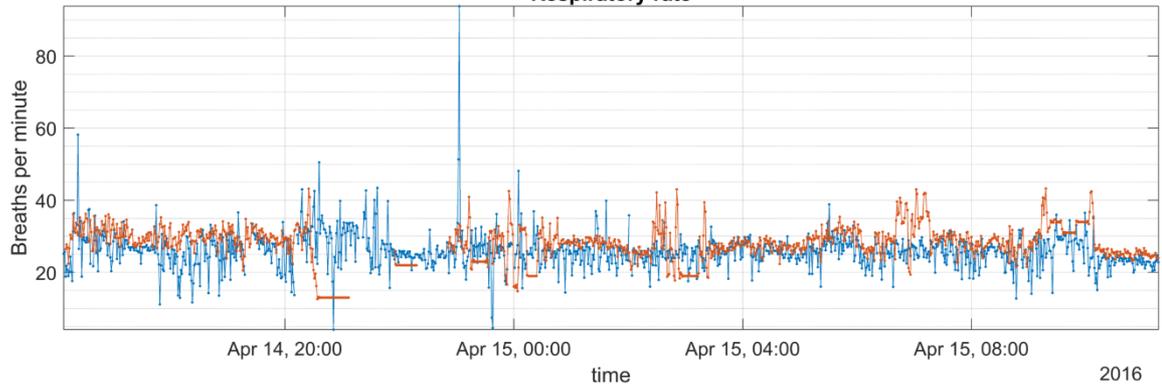
Respiratory rate



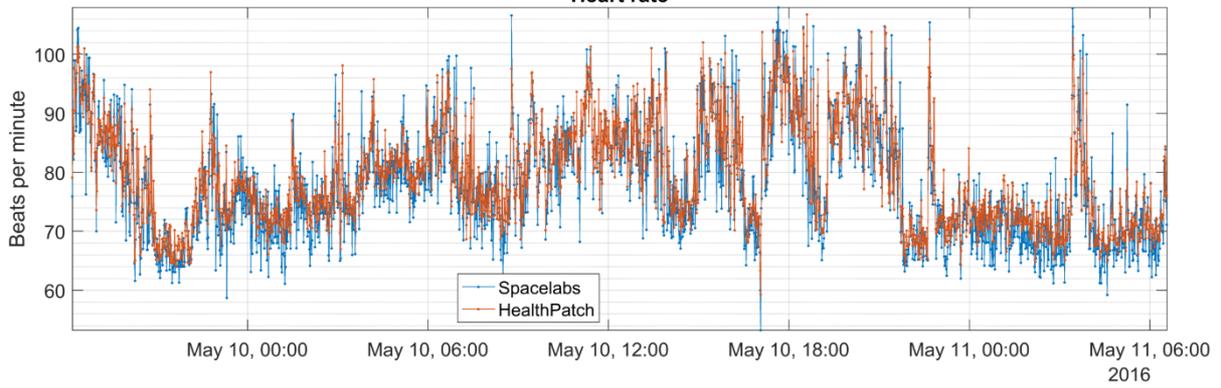
**Participation 28
Heart rate**



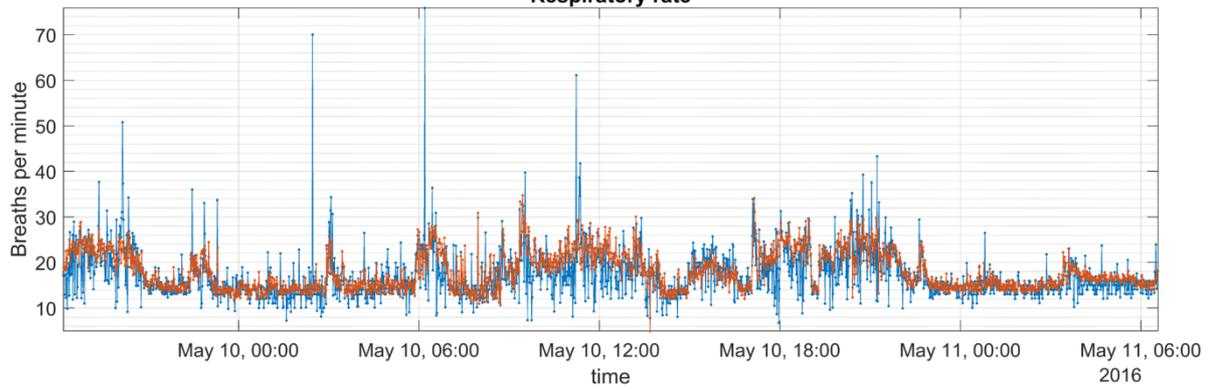
Respiratory rate



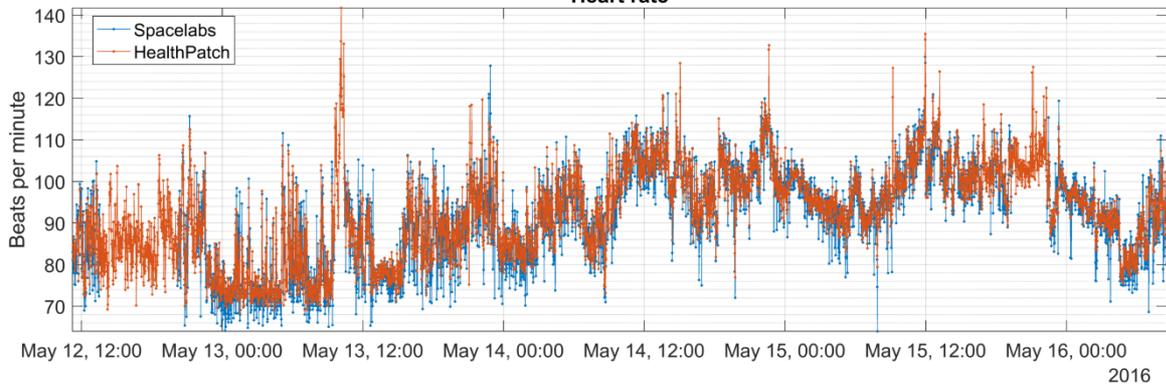
Participation 29
Heart rate



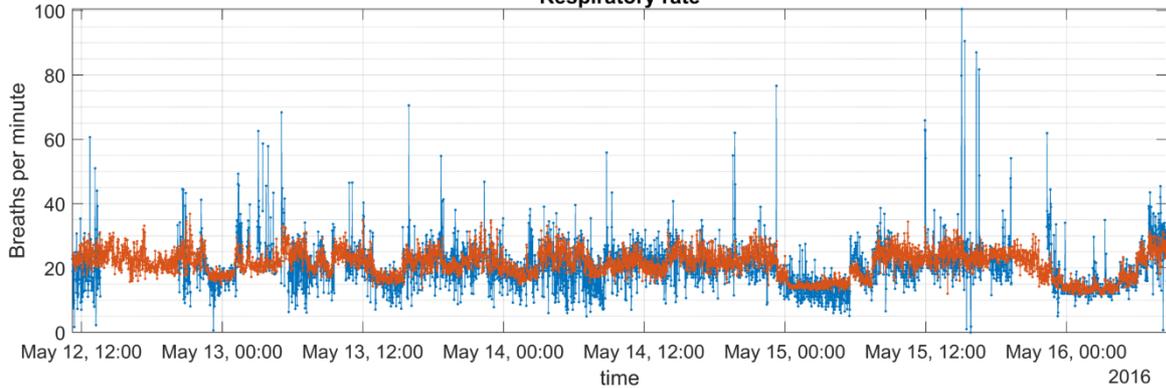
Respiratory rate



Participation 30
Heart rate

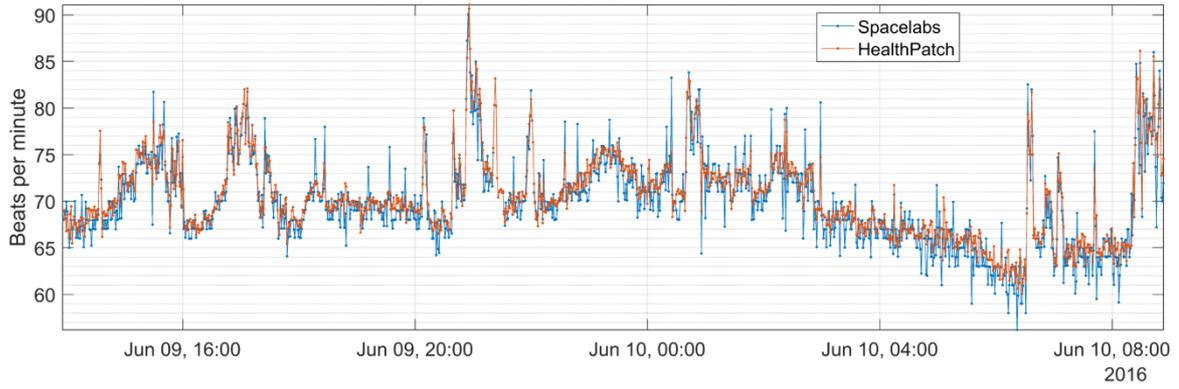


Respiratory rate

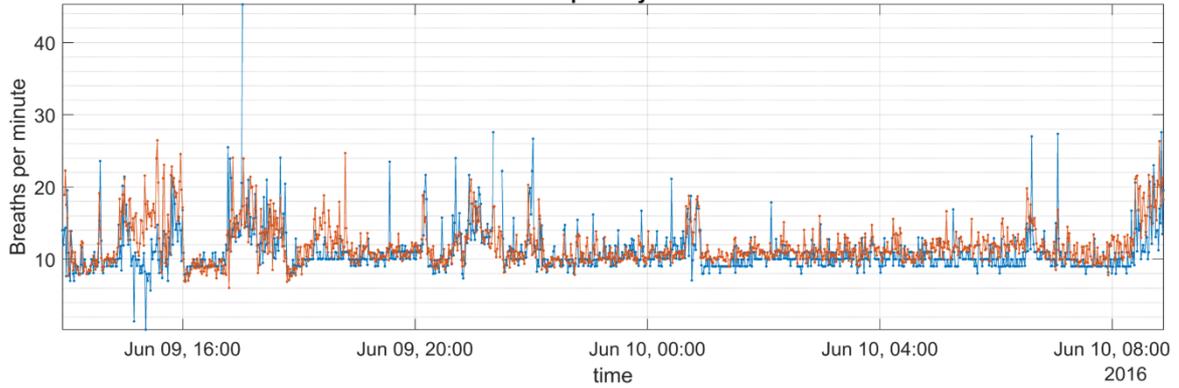


Participation 31

Heart rate

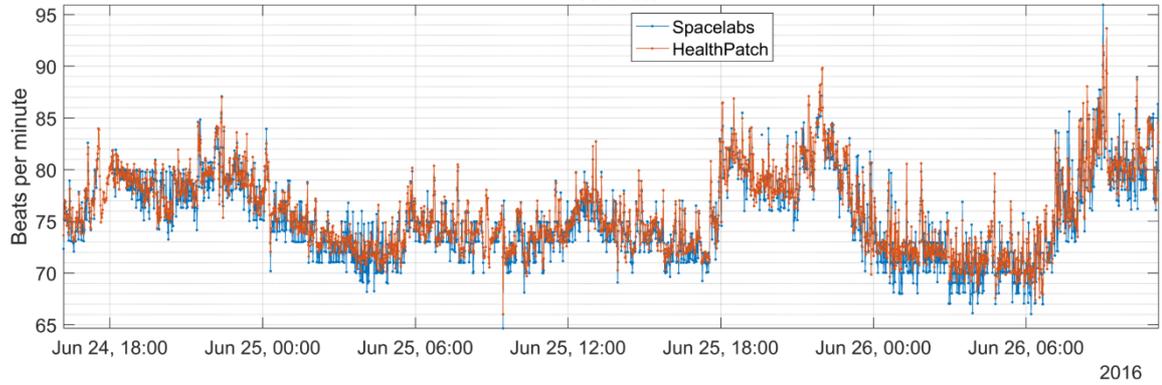


Respiratory rate

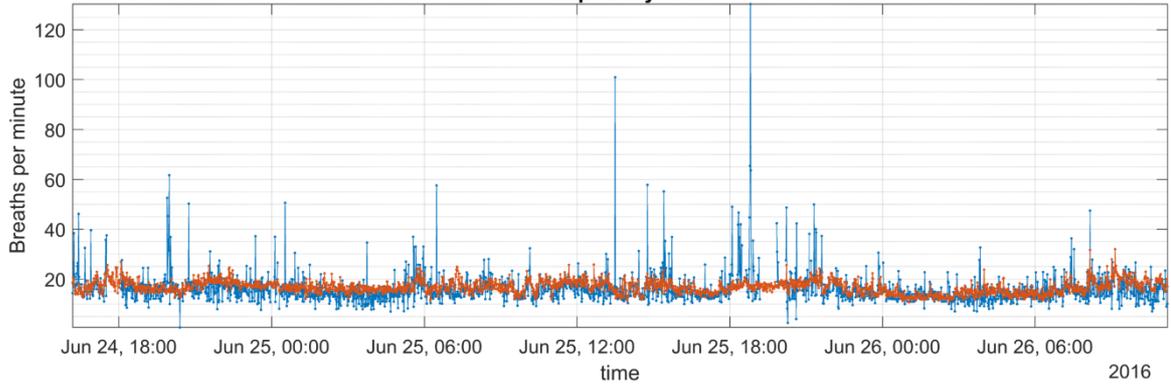


Participation 32

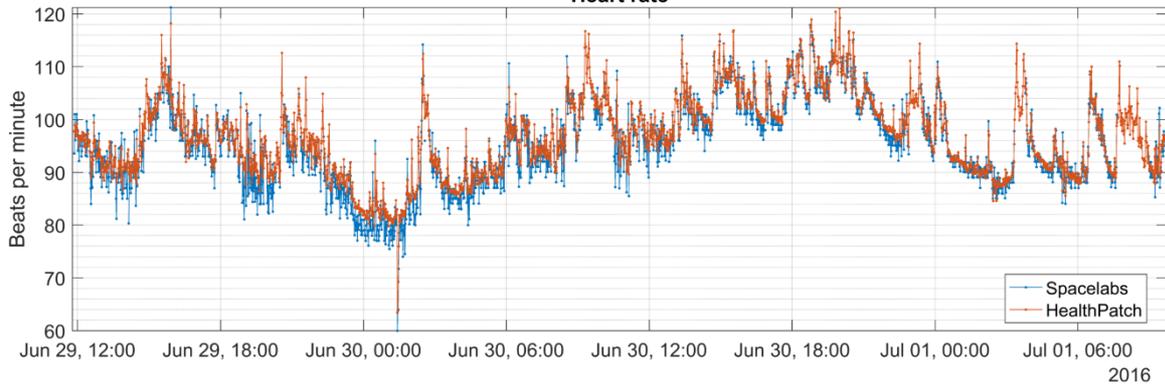
Heart rate



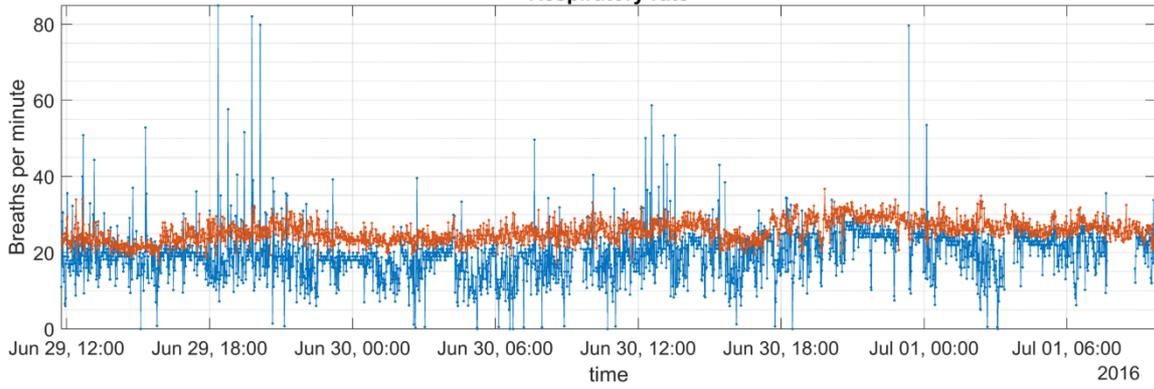
Respiratory rate



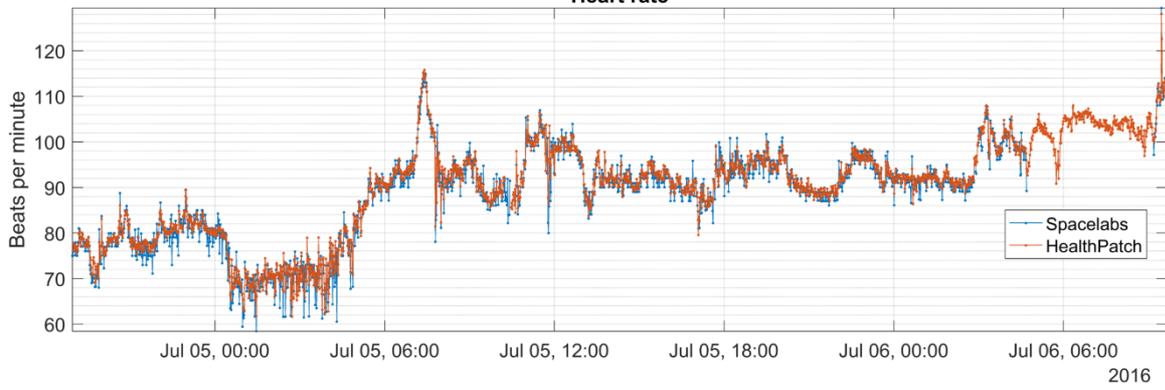
Participation 33
Heart rate



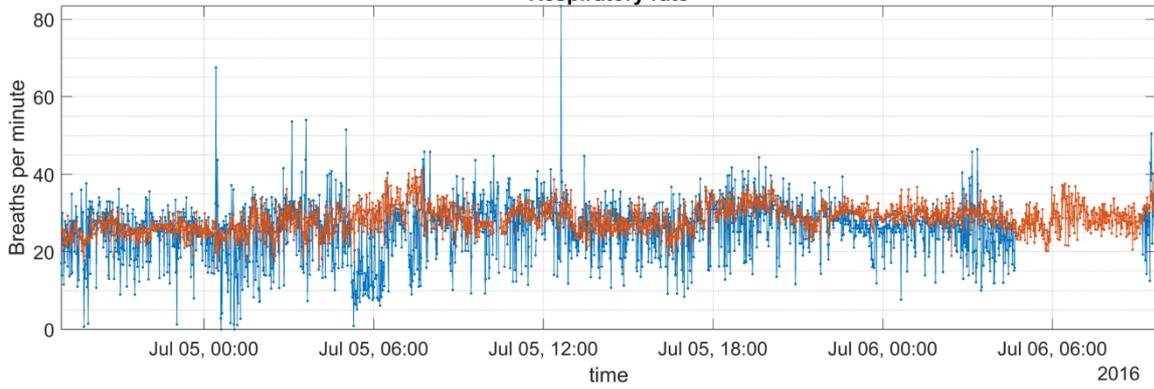
Respiratory rate



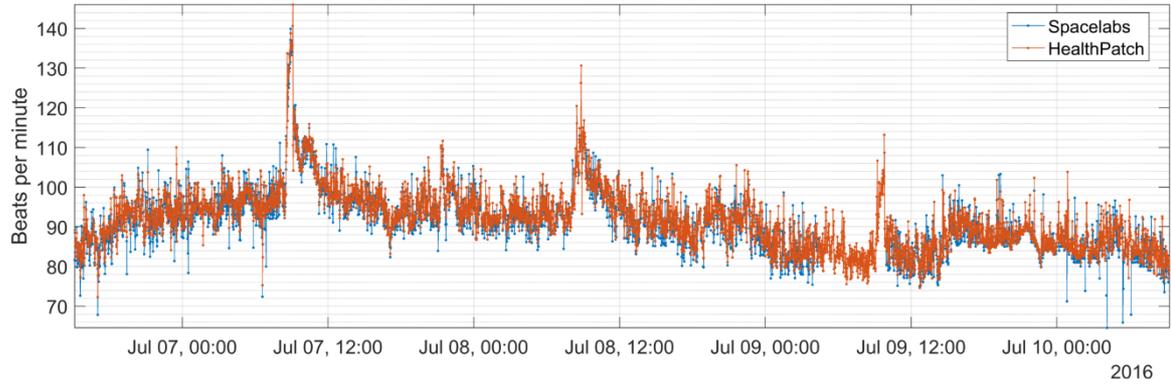
Participation 34
Heart rate



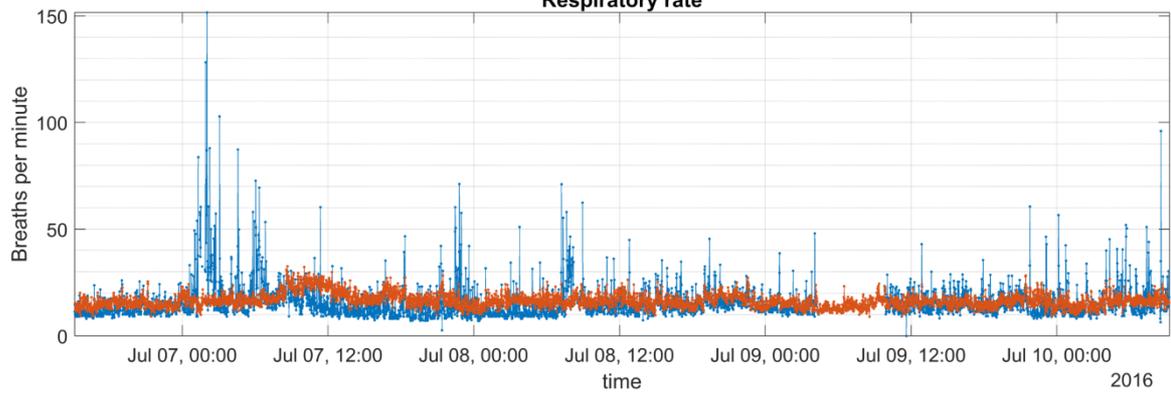
Respiratory rate



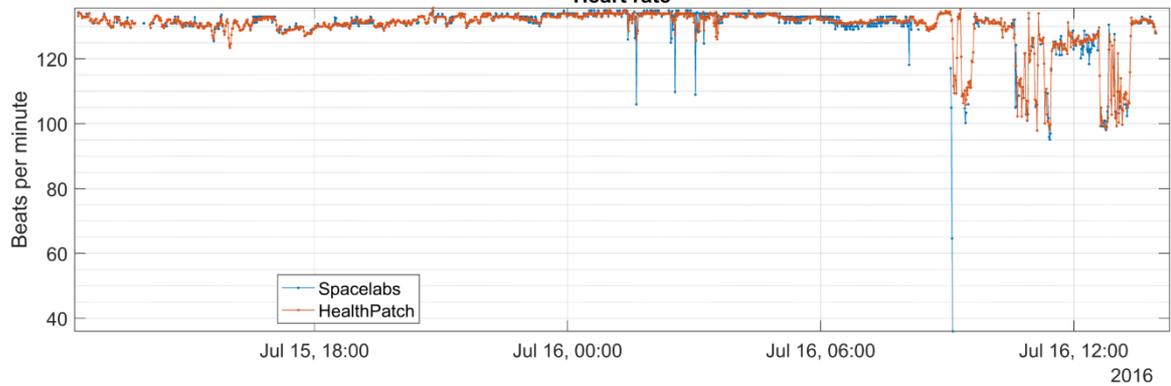
Participation 35
Heart rate



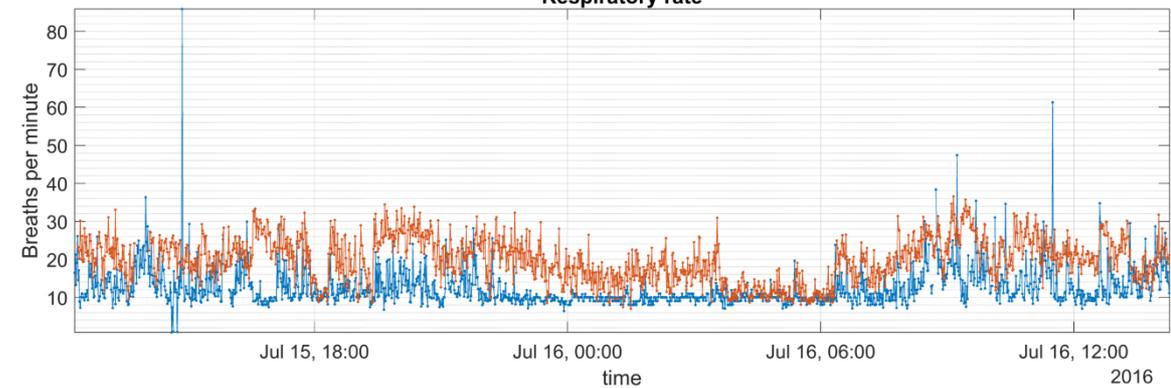
Respiratory rate

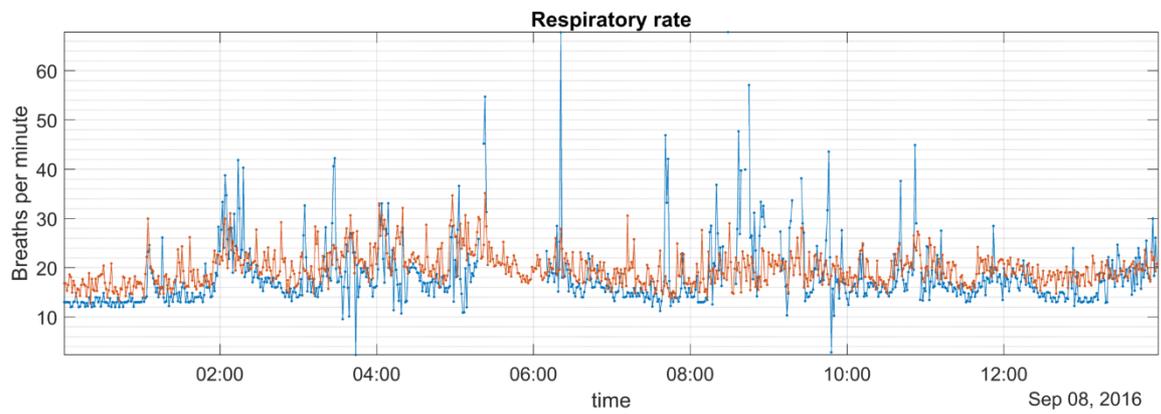
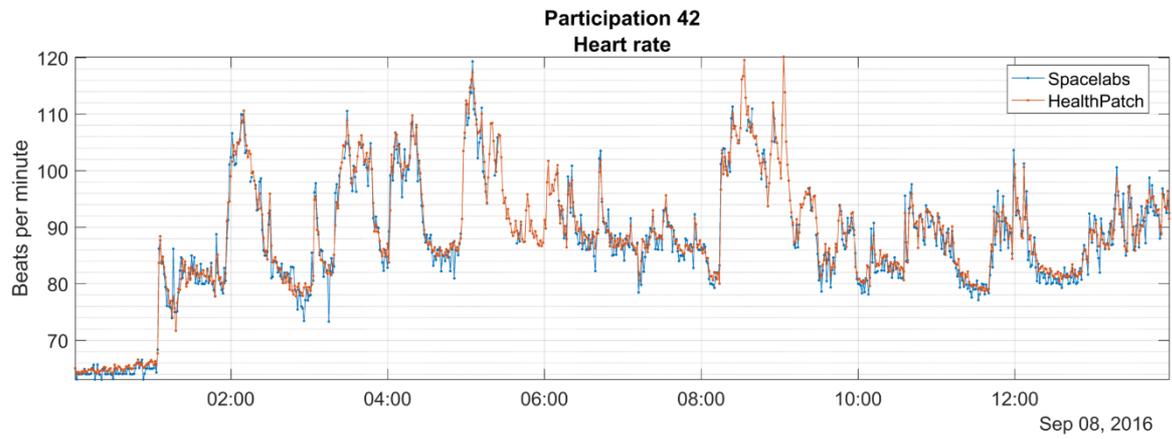
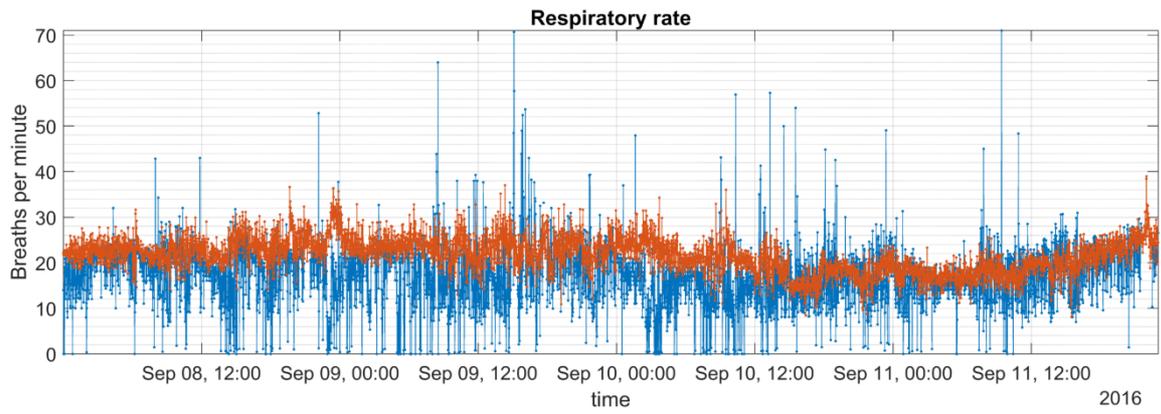
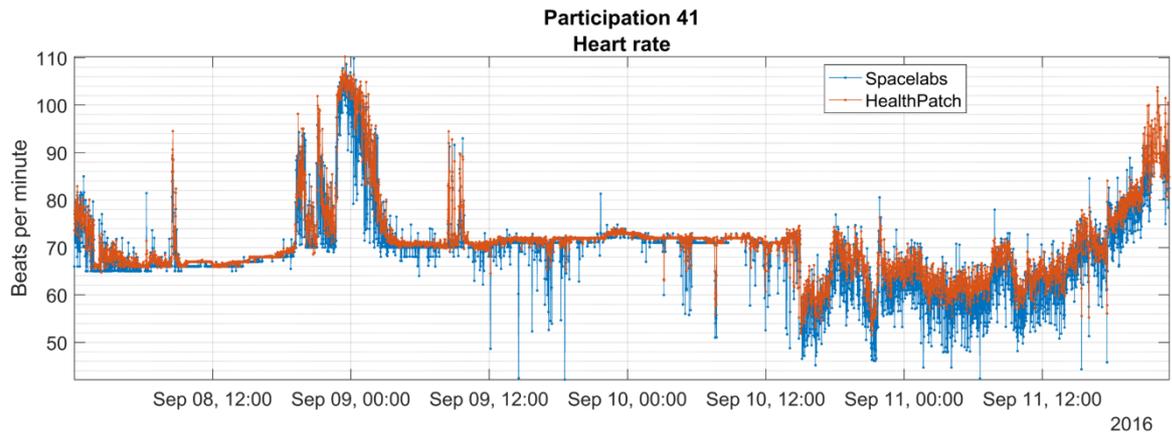


Participation 37
Heart rate



Respiratory rate





Appendix D. Tables

Note that all tables in this appendix show participation numbers from 8 onwards. These numbers correspond to the numbers in the database and are used here for reference. The first seven participations were excluded from this study, because they were measured in previous research using different measurement protocols.

Table 9 – The number of samples of the raw data obtained from participations are listed.

| Participation № | Number of samples in raw data | | | |
|-----------------|-------------------------------|------------------|----------------|------------------|
| | Spacelabs XPREZZON | | HealthPatch MD | |
| | Heart rate | Respiratory rate | Heart rate | Respiratory rate |
| 8 | 2608 | 2800 | 19710 | 19710 |
| 9 | 6215 | 5581 | 80564 | 80564 |
| 10 | 4114 | 4124 | 37683 | 37683 |
| 11 | 2308 | 2323 | 21966 | 21966 |
| 12 | 1686 | 1691 | 20707 | 20707 |
| 13 | 3170 | 3179 | 28642 | 28642 |
| 14 | 0 | 0 | 69470 | 69470 |
| 15 | 0 | 0 | 17232 | 17232 |
| 16 | 0 | 0 | 2980 | 2980 |
| 17 | 1327 | 1344 | 16081 | 16081 |
| 18 | 4079 | 3992 | 46868 | 46868 |
| 19 | 2566 | 2683 | 18067 | 18067 |
| 20 | 0 | 0 | 3784 | 3784 |
| 21 | 0 | 0 | 61375 | 61375 |
| 22 | 2676 | 2592 | 41045 | 41045 |
| 23 | 1166 | 1138 | 5844 | 5844 |
| 24 | 1755 | 1762 | 20522 | 20522 |
| 25 | 5161 | 5169 | 73604 | 73604 |
| 26 | 3619 | 3235 | 38892 | 38892 |
| 27 | 2141 | 2143 | 24788 | 24788 |
| 28 | 2239 | 2499 | 16646 | 16646 |
| 29 | 3047 | 3105 | 35282 | 35282 |
| 30 | 5500 | 5408 | 112716 | 112716 |
| 31 | 2224 | 2232 | 19085 | 19085 |
| 32 | 2951 | 2956 | 42058 | 42058 |
| 33 | 3119 | 3278 | 41653 | 41653 |
| 34 | 2452 | 2469 | 35304 | 35304 |
| 35 | 4711 | 5281 | 81767 | 81767 |
| 36 | 0 | 0 | 62759 | 62759 |
| 37 | 1826 | 2364 | 23314 | 23314 |
| 38 | 0 | 0 | 26062 | 26062 |
| 39 | 0 | 0 | 102725 | 102725 |
| 40 | 0 | 0 | 64562 | 64562 |
| 41 | 5541 | 5543 | 97091 | 97091 |
| 42 | 1111 | 1117 | 22535 | 22535 |
| Total | 79312 | 80008 | 1433383 | 1433383 |

Table 10 – The number of samples and total duration of the uniformly sampled data obtained from participations are listed. *n*: number of samples. hh:mm:ss: hours:minutes:seconds.

| Participation | Number of samples and duration of uniformly sampled data | | | | | | | | |
|---------------|--|--------------------|----------|------------------|----------|----------------|----------|------------------|--|
| | № | Spacelabs XPRESSON | | | | HealthPatch MD | | | |
| | | Heart rate | | Respiratory rate | | Heart rate | | Respiratory rate | |
| | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | |
| 8 | 2479 | 47:57:00 | 2479 | 47:57:00 | 20307 | 23:37:56 | 20307 | 23:37:56 | |
| 9 | 6078 | 109:24:00 | 6078 | 109:24:00 | 83204 | 92:26:52 | 83204 | 92:26:52 | |
| 10 | 4091 | 71:57:00 | 4091 | 71:57:00 | 38744 | 43:02:52 | 38744 | 43:02:52 | |
| 11 | 2289 | 41:14:00 | 2289 | 41:14:00 | 23012 | 25:44:56 | 23012 | 25:44:56 | |
| 12 | 1682 | 28:09:00 | 1682 | 28:09:00 | 21702 | 25:19:52 | 21702 | 25:19:52 | |
| 13 | 3154 | 56:55:00 | 3154 | 56:55:00 | 30091 | 33:26:00 | 30091 | 33:26:00 | |
| 14 | 0 | 00:00:00 | 0 | 00:00:00 | 71211 | 82:08:52 | 71211 | 82:08:52 | |
| 15 | 0 | 00:00:00 | 0 | 00:00:00 | 17698 | 35:36:28 | 17698 | 35:36:28 | |
| 16 | 0 | 00:00:00 | 0 | 00:00:00 | 2928 | 07:59:28 | 2928 | 07:59:28 | |
| 17 | 1309 | 22:54:00 | 1309 | 22:54:00 | 15517 | 25:39:12 | 15517 | 25:39:12 | |
| 18 | 4016 | 95:57:00 | 4016 | 95:57:00 | 48997 | 114:25:32 | 48997 | 114:25:32 | |
| 19 | 2489 | 47:58:00 | 2489 | 47:58:00 | 18558 | 20:37:16 | 18558 | 20:37:16 | |
| 20 | 0 | 00:00:00 | 0 | 00:00:00 | 3669 | 04:04:56 | 3669 | 04:04:56 | |
| 21 | 0 | 00:00:00 | 0 | 00:00:00 | 60794 | 67:32:52 | 60794 | 67:32:52 | |
| 22 | 2621 | 47:57:00 | 2621 | 47:57:00 | 41214 | 47:14:41 | 41214 | 47:14:41 | |
| 23 | 1148 | 23:01:00 | 1148 | 22:47:00 | 5912 | 12:15:28 | 5912 | 12:15:28 | |
| 24 | 1745 | 36:56:00 | 1745 | 36:56:00 | 20502 | 22:46:56 | 20502 | 22:46:56 | |
| 25 | 5123 | 95:57:00 | 5123 | 95:57:00 | 73379 | 81:31:52 | 73379 | 81:31:52 | |
| 26 | 3537 | 71:57:00 | 3537 | 71:57:00 | 38564 | 42:50:52 | 38564 | 42:50:52 | |
| 27 | 2088 | 39:11:00 | 2088 | 39:11:00 | 24525 | 27:14:56 | 24525 | 27:14:56 | |
| 28 | 2122 | 47:58:00 | 2122 | 47:58:00 | 16542 | 19:16:20 | 16542 | 19:16:20 | |
| 29 | 2997 | 71:57:00 | 2997 | 71:57:00 | 35025 | 38:54:56 | 35025 | 38:54:56 | |
| 30 | 5472 | 142:06:00 | 5472 | 142:06:00 | 111825 | 124:14:56 | 111825 | 124:14:56 | |
| 31 | 2209 | 47:57:00 | 2209 | 47:57:00 | 18989 | 21:05:52 | 18989 | 21:05:52 | |
| 32 | 2884 | 59:11:00 | 2884 | 59:11:00 | 42179 | 46:51:52 | 42179 | 46:51:52 | |
| 33 | 3007 | 57:44:00 | 3007 | 57:44:00 | 41580 | 46:11:56 | 41580 | 46:11:56 | |
| 34 | 2427 | 46:03:00 | 2427 | 46:04:00 | 35051 | 38:56:52 | 35051 | 38:56:52 | |
| 35 | 4289 | 96:01:00 | 4289 | 96:01:00 | 81119 | 90:07:52 | 81119 | 90:07:52 | |
| 36 | 0 | 00:00:00 | 0 | 00:00:00 | 62190 | 69:05:56 | 62190 | 69:05:56 | |
| 37 | 1621 | 43:29:00 | 1621 | 43:30:00 | 23355 | 25:57:56 | 23355 | 25:57:56 | |
| 38 | 0 | 00:00:00 | 0 | 00:00:00 | 25844 | 28:42:52 | 25844 | 28:42:52 | |
| 39 | 0 | 00:00:00 | 0 | 00:00:00 | 102725 | 114:09:56 | 102725 | 114:09:56 | |
| 40 | 0 | 00:00:00 | 0 | 00:00:00 | 64035 | 71:08:56 | 64035 | 71:08:56 | |
| 41 | 5504 | 94:59:00 | 5504 | 94:59:00 | 96569 | 107:17:52 | 96569 | 107:17:52 | |
| 42 | 1076 | 23:57:00 | 1076 | 23:57:00 | 22544 | 25:02:52 | 22544 | 25:02:52 | |
| Total | 77457 | 1568:46:00 | 79270 | 1568:34:00 | 1440100 | 1702:44:45 | 1440100 | 1702:44:45 | |

Table 11 – The number of samples and total duration of the ensemble (HealthPatch and Spacelabs resampled, synchronised and trimmed to each other) participations are listed. *n*: number of samples. hh:mm:ss: hours:minutes:seconds.

| Participation | Number of samples and duration of ensemble data | | | | | | | | |
|---------------|---|--------------------|----------|------------------|----------|----------------|----------|------------------|--|
| | № | Spacelabs XPREZZON | | | | HealthPatch MD | | | |
| | | Heart rate | | Respiratory rate | | Heart rate | | Respiratory rate | |
| | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | <i>n</i> | hh:mm:ss | |
| 8 | 1183 | 23:27:00 | 1381 | 23:28:00 | 1328 | 23:26:00 | 1328 | 23:26:00 | |
| 9 | 5230 | 92:23:00 | 4585 | 92:26:00 | 5544 | 92:23:00 | 5546 | 92:25:00 | |
| 10 | 2435 | 43:02:00 | 2452 | 43:02:00 | 2582 | 43:01:00 | 2581 | 43:00:00 | |
| 11 | 1414 | 25:44:00 | 1440 | 25:44:00 | 1516 | 25:44:00 | 1508 | 25:36:00 | |
| 12 | 1429 | 23:57:00 | 1438 | 23:57:00 | 1438 | 23:57:00 | 1438 | 23:57:00 | |
| 13 | 1903 | 33:26:00 | 1920 | 33:26:00 | 2007 | 33:26:00 | 2007 | 33:26:00 | |
| 14 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 15 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 16 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 17 | 547 | 09:39:00 | 561 | 09:33:00 | 580 | 09:39:00 | 574 | 09:33:00 | |
| 18 | 3198 | 82:23:00 | 3141 | 82:23:00 | 2975 | 82:22:00 | 2975 | 82:22:00 | |
| 19 | 1103 | 20:27:00 | 1208 | 20:28:00 | 1228 | 20:27:00 | 1229 | 20:28:00 | |
| 20 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 21 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 22 | 1913 | 35:10:00 | 2317 | 44:04:00 | 2096 | 35:10:00 | 2014 | 33:48:00 | |
| 23 | 375 | 06:17:00 | 361 | 06:00:00 | 362 | 07:59:00 | 373 | 08:10:00 | |
| 24 | 1357 | 22:45:00 | 1361 | 22:46:00 | 1351 | 22:44:00 | 1349 | 22:42:00 | |
| 25 | 4266 | 81:31:00 | 4280 | 81:31:00 | 4889 | 81:28:00 | 4888 | 81:27:00 | |
| 26 | 2306 | 42:50:00 | 1930 | 42:50:00 | 2571 | 42:50:00 | 2571 | 42:50:00 | |
| 27 | 1312 | 22:31:00 | 1352 | 22:31:00 | 1352 | 22:31:00 | 1352 | 22:31:00 | |
| 28 | 1113 | 19:06:00 | 1126 | 19:08:00 | 1068 | 19:06:00 | 1068 | 19:06:00 | |
| 29 | 2113 | 36:23:00 | 2181 | 36:23:00 | 2184 | 36:23:00 | 2184 | 36:23:00 | |
| 30 | 4791 | 93:21:00 | 4687 | 93:22:00 | 5601 | 93:20:00 | 5601 | 93:20:00 | |
| 31 | 1122 | 18:56:00 | 1126 | 18:57:00 | 1136 | 18:55:00 | 1136 | 18:55:00 | |
| 32 | 2356 | 43:01:00 | 2392 | 43:01:00 | 2582 | 43:01:00 | 2582 | 43:01:00 | |
| 33 | 2307 | 45:53:00 | 2557 | 45:58:00 | 2754 | 45:53:00 | 2757 | 45:56:00 | |
| 34 | 2001 | 38:42:00 | 2031 | 38:42:00 | 2309 | 38:41:00 | 2309 | 38:41:00 | |
| 35 | 3948 | 90:07:00 | 4892 | 90:07:00 | 5407 | 90:06:00 | 5407 | 90:06:00 | |
| 36 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 37 | 912 | 25:33:00 | 1537 | 25:57:00 | 1514 | 25:33:00 | 1533 | 25:52:00 | |
| 38 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 39 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 40 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | 0 | 00:00:00 | |
| 41 | 5501 | 94:59:00 | 5507 | 94:59:00 | 5700 | 94:59:00 | 5700 | 94:59:00 | |
| 42 | 741 | 13:57:00 | 758 | 13:56:00 | 838 | 13:57:00 | 837 | 13:56:00 | |
| Total | 56876 | 1085:30:00 | 58521 | 1094:39:00 | 62912 | 1087:01:00 | 62847 | 1085:56:00 | |

Appendix E. Honeycomb plots in a MATLAB implementation

Because honeycomb plots were beneficial for the visualisation of data in Bland-Altman analysis in this study (see §5.2), it was practical to use a MATLAB implementation. A thorough search yielded in two implementations (31 March 2017):

1. 'hexscatter.m' by Gordon Bean²⁰
2. 'Hexagonal Scatter Plot' by Salman Mashayekh²¹

Unfortunately, these releases were of no use, because one of them was implemented incorrectly and the other had different use cases. Hence, a honeycomb plot implementation was written, tested and published. No fully functional release is available yet, but version 0.1 is released on GitHub²² and published on the MATLAB File Exchange²³ and provides all basic functionality [94]. Some customisation options are available, such as specifying the number of bins in the horizontal or vertical direction.

Release v0.1 was integrated into the Bland-Altman-Analysis implementation presented in Appendix A to enable optional plotting of honeycomb plots instead of regular scatter plots.

²⁰ <https://mathworks.com/matlabcentral/fileexchange/45639-hexscatter-m>

²¹ <https://mathworks.com/matlabcentral/fileexchange/39486-hexagonal-scatter-plot>

²² <https://github.com/erikhuizinga/honeycomb/releases/tag/v0.1>

²³ <https://mathworks.com/matlabcentral/fileexchange/62355-honeycomb>

Summary

After major surgery, patients are at increased risk of adverse outcomes, such as complications and increased hospital length of stay and they have increased morbidity and mortality. To enhance patient safety, vital signs (e.g. heart rate and respiratory rate) are monitored. However, when patients move from high care wards (e.g. intensive or medium care) to regular wards, the frequency and quality of monitoring decrease. As a consequence, patient safety may be compromised. After discharge from the hospital, monitoring is virtually unavailable and the probability of timely recognition of deterioration increases substantially. However, because it is known that certain vital signs can show early changes before adverse events occur, patient safety may be enhanced during hospital admission and after discharge if these were monitored.

To monitor in an accessible manner while not encumbering patients, caregivers and medical personnel, it could be a solution to use portable, wireless and non-invasive sensors called wearables. This study has investigated one such wearable, HealthPatch MD by Vital Connect, by performing measurements with it on the surgical medium care ward at the UMC Utrecht. The goal was to study the principles that such a wearable uses, assess its performance in clinical practice and comparing its measurements with a regular patient monitor (Spacelabs XPREZZON). Measurements were done in 35 participations. Using Bland-Altman analysis for repeated measurements, it was found that heart rate accuracy is acceptable compared to Spacelabs. Respiratory rate is not measured accurately enough to be considered acceptable. However, it remains unknown if Spacelabs is a good reference for respiratory rate, because this monitor is not the gold standard.

To use this type of data, obtained by a wearable monitor, for early recognition of the deteriorating patient, a model needs to be developed to predict a measure of risk of decline. The Early Warning Score (EWS) is such a measure that is currently used in clinical practice. If EWS were predictable using a wearable vital signs monitor, that monitor may be able to predict patient decline as well. Because in this study not enough data was obtained collected, such a model could not be developed. However, various prediction modelling techniques that can be applied to this type of data are described in this thesis.

Samenvatting

Patiënten na grote chirurgie hebben verhoogd risico op slechte uitkomsten, zoals complicaties en verlengde opnameduur en hebben een verhoogde morbiditeit en mortaliteit. Om de patiëntveiligheid te vergroten wordt monitoring van vitale functies, zoals hart- en ademhalingsfrequentie, ingezet. Echter, wanneer patiënten van een bewaakte afdeling (bijv. intensive of medium care) naar een reguliere afdeling gaan, neemt de frequentie en kwaliteit van monitoring af. Patiëntveiligheid kan hierdoor in het geding komen. Na ontslag is monitoring nagenoeg afwezig en stijgt de kans op het niet of te laat herkennen van verslechtering nog verder. Juist omdat het is aangetoond dat bepaalde vitale functies afwijkingen kunnen vertonen voordat negatieve uitkomsten plaatsvinden, zou de patiëntveiligheid vergroot kunnen worden tijdens opname en na ontslag als deze wel gemonitord worden.

Om te kunnen monitoren op laagdrempelige wijze en zonder de patiënt en medisch personeel / verzorgers te belemmeren, zouden draagbare, draadloze en niet-invasieve sensoren ingezet kunnen worden, zogeheten wearables. Dit onderzoek heeft één zo'n wearable, HealthPatch MD van Vital Connect, onderzocht door er metingen mee te verrichten op de medium care van chirurgie in het UMC Utrecht. Het doel was het verkrijgen van inzicht in de principes waarmee zo'n wearable werkt, hoe een dergelijke wearable in de medische praktijk presteert en het vergelijken van de metingen met de reguliere monitor (Spacelabs XPREZZON). Er zijn metingen gedaan in 35 deelnames. Met behulp van Bland-Altman analyse voor herhaalde metingen werd bevonden dat HealthPatch MD in vergelijking met de Spacelabs monitor hartfrequentie met acceptabele precisie meet. Ademhalingsfrequentie wordt niet acceptabel genoeg gemeten. Echter, het is onduidelijk of Spacelabs een betrouwbare referentiewaarde meet, aangezien deze monitor geen gouden standaard is.

Om met dit soort data, gemeten door een wearable monitor, vroegtijdig achteruitgang van de patiënt te kunnen herkennen, dient een model te worden ontwikkeld dat een maat voor risico op achteruitgang kan voorspellen. De Early Warning Score (EWS) is zo'n maat die reeds in de kliniek gebruikt wordt. Als EWS voorspeld zou kunnen worden a.d.h.v. metingen met een wearable, dan zou met de wearable mogelijk ook het risico op negatieve uitkomsten voorspeld kunnen worden. Omdat in deze studie nog onvoldoende data verzameld is om risico op negatieve uitkomsten te voorspellen, wordt geen voorspelmodel ontwikkeld. Echter, wel worden verschillende beschikbare methoden uiteengezet waarmee dit bereikt zou kunnen worden.

Acknowledgements

*If I have seen further than others
it is by standing upon
the shoulders of giants*
Isaac Newton

I would like to give my heartfelt thanks to all my supervisors. Each of them have been valuable beyond words for my research and personal development. I can only hope that one day I will have a few of the qualities you have, both as a professional and on a personal level. Together, you gave me the possibility to do pioneering research while educating me. My gratitude goes to Hermie Hermens for the criticism he provided. For me as a Master student it is particularly valuable and inspiring to be given feedback by a professor with such an amount of experience and knowledge. To Taco Blokhuis and Cor Kalkman I give my thanks especially for their continued education in both clinical and research aspects. To Nicole Cramer Bornemann I am grateful for the (in the meantime) years she has supported me and helped me to find a way to become a better version of myself. I also thank Martine Breteler for the frequent supervision, aid and criticism. It is inspiring to see a young and talented researcher and Technical Physician take the lead in such an ambitious research project. Last but not least, I thank Erik Groot Jebbink (with an honourable mention of Hil Meijer, who was in his place before) for taking the time to review my thesis.

There are many others I would like to thank, because many colleagues at the University Medical Center in Utrecht have made many things possible for my internship. In no particular order, my special thanks go to Janet van Gisbergen with her admirable medium care staff, Jan Dirk Tiggelman, Wietze Pasma, Hans van der Brugge, Pieter Dik, Jacqueline Vromen, Kristel Schaap, Eline KleinJan, Leo van Wolfswinkel, Linda Peelen, the researchers and PhD candidates on the IC staff hallway, the OR complex doctors plus assistants and the flex room workers for all the pleasant times.