Opioid-Induced Respiratory Depression, a Comprehensive Data Analysis

MSc Thesis

F.S.Leenen

TU Delft

LUMC Leids Universitair Medisch Centrum

Erasmus MC Universitair Medisch Centrum Rotterdam

This page intentionally left blank

Opioid-Induced Respiratory Depression, a Comprehensive Data Analysis

Unravelling Opioid-Induced Respiratory Depression through Unsupervised Machine Learning on Respiratory Flow Data

Femke Leenen

Student number: 4664485

11 - 04 - 2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Delft University of Technology | Leiden University | Erasmus University Rotterdam

Supervision

Dr. M. Boon | Anaesthesiologist, LUMC
Dr.ir. J.H.G. Dauwels | Associate professor, TU Delft
Bsc. M.A. van Lemmen | PhD candidate, LUMC

Thesis committee

Prof.dr. A. Dahan | Anaesthesiologist, LUMC
Dr. M. Boon | Anaesthesiologist, LUMC
Dr.ir. J.H.G. Dauwels | Associate professor, TU Delft
Bsc. M.A. van Lemmen | PhD candidate, LUMC

Master Thesis Project
TM30004 - 35 ECTS

Dept. of Anaesthesiology
Leiden University Medical Center

14-08-2023 till 26-04-2024

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

LUMC Leids Universitair Medisch Centrum

Erasmus MC Universitair Medisch Centrum Rotterdam

This page intentionally left blank

Preface

As I write this preface, marking the final step in completing this thesis, I am concluding my 6.5 years of study. I am grateful that during this thesis, I had the opportunity to delve further into an emerging technology, within the medical field, which I had not yet had much experience with at the start. Embracing the challenge of learning the concepts of machine learning during my thesis brought its difficulties, yet, looking back, it only made me prouder of the end results.

I am thankful for the guidance provided by my supervisors, without whom this achievement would not have been possible. Specifically, I would like to thank my daily supervisor, Maarten van Lemmen, whose inspiration played a major role in the realisation of this project. Thank you for your enthusiasm, knowledge, and approachability. I also want to thank my supervisors, Martijn Boon and Justin Dauwels, who deserve special recognition for carving out time in their demanding schedules to contribute to my growth and learning. Martijn, for welcoming me back to the department and sharing your enthusiasm and expertise in guiding me once more. And Justin, for sharing your knowledge in the field of machine learning with me and your enthusiasm in guiding me during this final part of my studies. Additionally, I also wish to express my gratitude in advance to Albert Dahan for agreeing to participate in my thesis committee.

As this thesis was obtained within the research group of the anaesthesiology department at the Leiden University Medical Center, this preface would not be complete without expressing gratitude to all the PhD students who made the past months more enjoyable. Thanks for including me in the activities, the entertaining stories during lunchtime, and the much-needed coffee breaks. Additionally, I want to thank the rest of the anaesthesiology department for offering me insightful days in the clinic, enriching my practical knowledge and understanding of the medical field.

To conclude, I would like to sincerely thank my family and friends for their support, not only during this thesis, but throughout the entirety of my studies.

Femke Leenen
Delft, April 2024

This page intentionally left blank

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
CO_2	Carbon dioxide
DTW	Dynamic Time Warping
$EtCO_2$	End-tidal carbon dioxide
FS	Feature space
H^+	Hydrogen ions
HCO_3^-	Bicarbonate
IQR	Interquartile Range
LUMC	Leiden University Medical Center
ML	Machine learning
MV	Minute volume
O_2	Oxygen
OIRD	Opioid-induced respiratory depression
OUD	Opioid use disorder
PCA	Principal Component Analysis
pCO_2	Partial pressure of carbon dioxide
PCs	Principal components
SHAP	SHapley Additive exPlanations
slm	Standard liter per minute
TS	Time series
U.S.	United States
WHO	World Health Organisation

PAPER

Opioid-Induced Respiratory Depression, a Comprehensive Data Analysis

Abstract

Introduction Opioids are vital for pain management but are highly addictive and may lead to opioid-induced respiratory depression (OIRD), which is the primary cause of death related to both prescription and illicit opioid use. This study employed unsupervised machine learning (ML) to examine potential changes in cluster patterns post-opioid administration and their relationship with respiratory depression levels. Additionally, comprehensive data analysis was conducted based on questions that resulted from observations of cluster behaviour.

Methods Following preprocessing, a preliminary study evaluated three different models, each trained on the baseline and post-opioid epochs, aiming to identify the most efficient distance metrics and feature space combinations. These models included principal component analysis with Euclidean distance, computed feature space with Euclidean distance, and time series with dynamic time warping. Next, the preferred approach, determined by testing different hypotheses regarding the desired cluster behaviour, underwent further refinement, and predictions were generated for post-antagonist epochs. Finally, feature influence was determined and questions were identified for the data-analysis.

Results In total, 34 subjects were included, resulting in a total of 6630 epochs for model development. Based on the preliminary study, it was decided to opt for a fuzzy clustering approach using calculated features as input, resulting in membership values indicating the probability of an epoch belonging to a certain cluster. A change in membership values was observed post-opioid as well as a recovery to baseline values post-antagonist in the majority of subjects when clustering was obtained for each subject individually. However, it was expressed as a sudden switch followed by a prolonged plateau phase rather than a gradual transition which was expected. Conversely, when trained collectively over all subjects, the majority of subjects showed no difference, probably due to the presence of inter-subject variability. Nevertheless, SHAP value analysis identified the same feature behaviour among subjects, despite variations in orientation and positioning, hinting at the potential for adjustment using static features. Yet, no significant correlations were found between static features and feature behaviour within this study.

Conclusion In this study, a fuzzy clustering model was implemented, incorporating SHAP value analysis to enhance the interpretability of the clustering results. Although the model successfully identified changes in respiratory flow patterns associated with OIRD and subsequent recovery after naloxone administration, it requires further refinement.

This page intentionally left blank

1 | Introduction

In the field of anaesthesia, analgesics (i.e. painkillers) play a crucial role in managing postoperative and chronic pain. Opioids, a specific kind of analgesic, have long been recognised for their efficacy in pain relief. However, as they are highly addictive, they may lead to opioid use disorder (OUD), which involves the misuse or abuse of prescribed opioids, diversion of medications, and illicit opioid use [1]. The resulting persistent occurrence of overdose cases and opioid toxicity remains a global concern and is recognised by the World Health Organisation (WHO) as an international health issue, also referred to as the opioid epidemic.

The epidemic is characterised by three waves, with the latest and current wave beginning in 2013, marked by the rise of synthetic opioids and an alarming increase in overdose deaths (Figure 1). As of 2021, more than 75% of all overdoses in the United States (U.S.) were related to opioid use, nearly 88% of which were synthetic opioids. Daily, opioid overdoses claimed the lives of 220 individuals in the U.S., with prescription opioids responsible for 45 of these deaths [2].

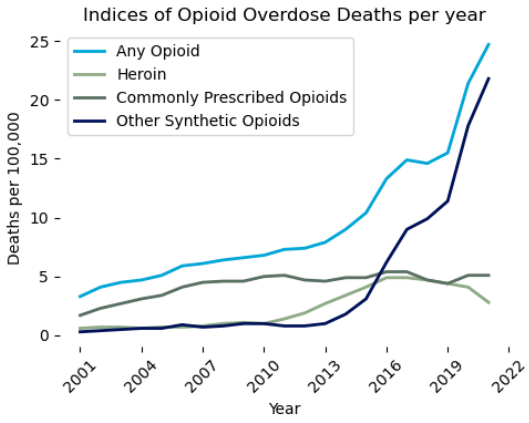


Figure 1: Raise in opioid related deaths over time in the United States [2].

Besides analgesia and sedation, the use of opioids has various effects on the respiratory system. These include decreased diaphragm activity due to decreased activity of phrenic motor neurons, reduction of hypercapnic and hypoxic ventilatory reflexes due to reduction of central and peripheral chemo receptors, inhibition of upper airway patency by inhibiting the activity of hypoglossal motor neurons and rigidity of respiratory skeletal muscles [3]. These effects results in slow, shallow, irregular breaths and in severe cases to respiratory arrest. This effect of opioid use on breathing is called opioid-induced respiratory depression (OIRD) and is the leading cause of death related to both prescription and illicit opioid use [4, 5].

Although not all cases of OIRD result in mortality, its impact can still be significant. According to a study conducted by Khanna et al. [6], patients in the U.S. who had experienced one or more episodes of OIRD had significantly longer average hospital stays of 1.4 days and incurred higher average total hospital costs of $4426. These results highlight the significant burden that OIRD places on the health care system and the potential social economic impact associated with it.

As roughly 84% of patients experience pain after surgery, of which approximately 88% are prescribed analgesic medication, predominantly opioids, the ongoing challenge in opioid research is to find a pain therapy that effectively relieves pain without causing respiratory depression [7, 8]. Until such painkillers are developed, opioids will continue to be the most commonly utilised option, and efforts to limit respiratory side effects will remain a key focus. In a study of Lee et al. [5], 97% of incidents identified in medical malpractice claims were considered probably or possibly preventable if better monitoring was used.

Since OIRD is a multi-factorial event, monitoring a single criterion may not be reliable or adequate. Readings of oxygen saturation, for example, can be confounded by administration of supplemental oxygen [9]. However, at present, multi parameters are interpreted intermittently by medical staff, leaving patients monitored insufficiently for more than 95% of the time [10]. Researchers have concluded that, to prevent possible respiratory arrest, continuous monitoring of respiration is necessary. Therefore, 'smart' algorithms, which combine multiple parameters, are being developed [11, 12, 13, 14, 15]. Although the studies report high performance metrics, none of the studies mention external testing or implementation strategies. This suggests that, as of now, none of the algorithms are currently used or tested outside the research context.

The high outcome measures of these studies suggest the potential utility of machine learning (ML) in this domain. Nevertheless, all studies used a different definition for OIRD. Although this is not a limitation on its own, it is known from previous research that the lack of standardised definitions and assessment methods makes OIRD re-

search and management difficult [9]. In addition, the majority of definitions used are binary. As the magnitude of respiratory depression correlates with the level of sedation and analgesia and thus is dose-dependent, it would be preferred if a model used a discrete or even continuous definition for OIRD [16]. However, a continuous definition for OIRD has not yet been used in literature. Also, except for two studies [13, 14], which could predict an OIRD event 10 minutes and 1 minute in advance, all models were descriptive and alerted medical personnel when OIRD was already occurring.

The study by Sunshine and Fuller [11] stands out as the sole study employing unsupervised ML. Since unsupervised ML is independent of predefined labels, it is ideal for studying OIRD, for which there is no uniform definition. It allows for the discovery of new insights and relationships within the data that could lead to a more nuanced understanding of OIRD. The study of Sunshine and Fuller showed promising results with an increase in certain clusters after opioid use. However, they did not research the possibility of clusters increasing after an increased amount of opioids was used, thus their outcome was also binary. In addition, this study was obtained using rat data, and is not yet tried with human data.

At the Leiden University Medical Center (LUMC), within the anaesthesiology research department, research is conducted on preventing and recovering from OIRD. Inspired by the study of Sunshine and Fuller [11], the data collected at LUMC has been utilised in this pilot-study to develop an unsupervised machine learning model. The primary aim was to identify any noticeable changes in cluster formations following opioid administration and to investigate how these clusters vary with the intensity of respiratory depression. Initially, a preliminary study was conducted to determine the most effective distance metric and feature space for the ML model. Subsequently, the definitive model was developed. Finally, during the final part of the study, a comprehensive data-analysis was obtained based on questions that resulted from the observed cluster behaviour.

2 | Background

2.1. Opioid-induced Respiratory Depression

2.1.1. Normal respiratory drive

The respiration drive originates from the pons and medulla located in the brain stem. These are in turn influenced by the cerebral cortex to control conscious and unconscious respiration, including activities like breath holding and speech modulation. The medulla, divided in the dorsal and ventral side, is responsible for the inhalation and airway defence (dorsal), and exhalation (ventral). Located within the medulla, the pre-Bötzinger complex is thought to be the pacemaker for respiratory rhythm and facilitates seamless transitions between different respiratory phases, while simultaneously inhibiting the activation of opposing muscle groups. The pontine grouping in the pons, controls breathing patterns and allows a smooth transition between in- and expiration by modulation of the frequency and intensity of the signals of medullary signals [17, 18]. Figure 2 provides an overview of the locations of these structures within the brain stem.

The rate and depth of respiration are influenced by the sensory input systems, consisting of the mechanoreceptors, metaboreceptors, and peripheral and central chemoreceptors. The central chemoreceptors, located within the ventral medulla and retrotrapezoid, are believed to have primary control over respiration by triggering the respiratory center in response to heightened levels of carbon dioxide (CO_2), leading to an acidic environment in the brain due to increased hydrogen ions. Information regarding the mechanical status of the lung and chest, including the breathing rate, lung space, and irritation triggers is provided by the mechanoreceptors in the airways, trachea, lungs, and pulmonary vessels. During exercise, the metaboreceptors in the skeletal muscles are activated to stimulate breathing. The partial pressure of arterial oxygen (O_2) in the blood is monitored by the peripheral chemoreceptors, located at the bifurcation of the common carotid arteries and near the arch of the aorta, called the carotid and aortic bodies, respectively. These chemoreceptors are responsive to hypercapnia (high levels of carbon dioxide) or acidosis (increased acidity), ensuring appropriate respiratory responses [17, 18].

2.1.2. Respiration after opioid administration

Opioids act mainly on the μ-opioids receptors in various regions of the central nervous system. Normally these receptors are activated by endorphins, which are hormones that are naturally released by your body when feeling pain or stress. When stimulated, the receptor activates inhibitory intracellular pathways, leading to reduced neuronal excitability and therefore inhibiting the activation of neurons that transmit pain signals to the brain. This

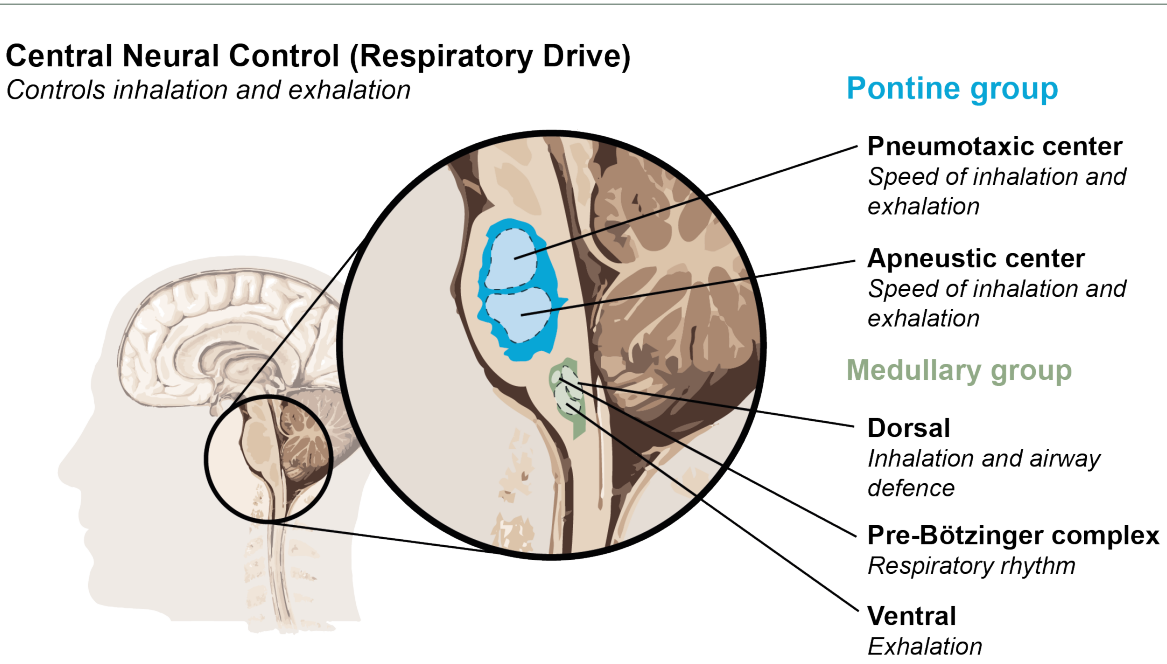


Figure 2: Overview of brainstem structures involved in central neural control of respiration [17].

pain-relieving system is part of the so-called endogenous opioid system [17, 18].

Besides pain, the endogenous opioid system also effects the respiratory control. Having different neuronal sites of action, including the medulla, pons, pre-Bötzinger complex and chemoreceptors, the activation of these receptors impacts the respiratory rhythm [17, 18]. Additionally, opioids suppress peripheral and central chemoreceptors, and neural signals to the upper airway dilator muscles, suppressing the normal response to hypoxemia (low levels of oxygen) and hypercapnia and lowering the upper airway patency, respectively. Clinically, OIRD manifests as a slow, shallow, and irregular breathing pattern. Apnea, the temporary cessation of breathing, is a characteristic of OIRD, which in its most severe form can escalate to respiratory arrest [3].

One reason why OIRD is harmful, is due to CO_2 stacking in the body, leading to respiratory acidosis [19]. This raise in CO_2 is often used as indicator for OIRD, determined using the end-tidal carbon dioxide ($EtCO_2$). When hyperventilating, more CO_2 is produced than the body is capable of eliminating, resulting in an arterial partial pressure of carbon dioxide (pCO_2) above the normal range of 38-42 mmHg. This causes an elevation in hydrogen ions (H^+) and bicarbonate (HCO_3^-), as evidenced by the equilibrium reaction of carbon dioxide shifting to the right:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+$$

The pH level of the blood is mainly determined by the concentration of H^+. During acute respiratory acidosis, the body is able to slightly compensate the sudden elevation of pCO_2. In case of an elevated H^+ concentration, HCO_3^- act as a buffer to minimise the drop in pH. In addition, to compensate for the disturbance in balance, the kidneys begin to excrete more acid in the forms of hydrogen and ammonium and reabsorb more base in the form of bicarbonate. However, the buffer system and the continues production of H^+ both have their limits. In the end, if CO_2 accumulation continues, the system cannot keep up, resulting in a net increase in H+ concentration and a decrease in pH.

2.1.3. Reversal of the opioid effect

Reversing the effects of opioids relies on the administration of antagonists, with naloxone being the first choice since it is widely available and effective. As an antagonist, naloxone binds to opioid receptors in the brain, displacing opioids and reversing their effects [20]. However, it is important to note that due to the short half-life of naloxone (30 min), the effects are relatively short compared to those of many opioids. Thus, the effect of naloxone may wear off prematurely, resulting in the possible return of respiratory depression [21,

22]. Therefore, people with opioid overdoses may need multiple doses of naloxone to fully reverse the effects of the opioids and restore normal breathing and consciousness.

2.2. Time Series

In mathematical terms, a time series (TS) can be described as a sequence of data points x_i, with each point observed at a specific time t_i. For a time series with n time steps, it is typically represented as follows:

$$TS = \{(x_1, t_1), (x_2, t_2), ..., (x_n, t_n)\} \quad (1)$$

If all time intervals are equal, then TS can be simplified as $\{x_1, x_2, ..., x_n\}$.

When working with TS data, especially in the context of resampling or filtering, understanding two fundamental concepts in signal processing is crucial: the Fourier theorem and the Nyquist frequency.

2.2.1. The Fourier transform

The Fourier theorem, a broadly used technique in signal processing, states that any periodic function can be expressed as a sum of sine functions as visualised in Figure 3. By using the (inverse) Fourier transform, one can switch between the representation in the time domain and the frequency domain. Insight into the different frequencies that make up the signal, is mainly invaluable for filtering and resampling. As noise is usually caused by certain frequencies, manipulating these frequency components results in noise reduction. In addition, when sampling a signal, one must be aware of the present frequencies to avoid aliasing, as described by the Nyquist theorem.

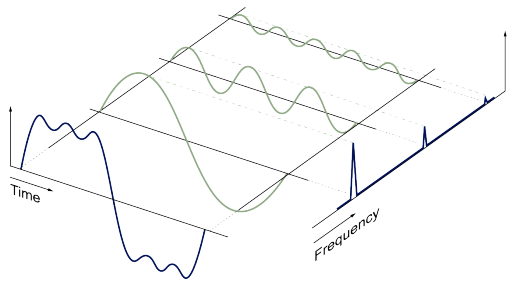


Figure 3: An illustration depicting the Fourier transform. The Fourier transform allows for the conversion between time domain and frequency domain representations, as every time-series can be expressed as a sum of sine functions.

2.2.2. The Nyquist frequency

When sampling a signal, the chosen sample rate determines the highest frequency that can be accurately captured and reconstructed from a digital signal. This limit, known as the Nyquist frequency, equals half the sampling rate. For accurate representation of a signal, it must be sampled at least at twice the rate of its highest frequency component. A lower sampling rate results in distortion during signal reconstruction, where higher frequency components are incorrect reconstructed as lower frequencies, as illustrated in Figure 4. This phenomenon is called aliasing.

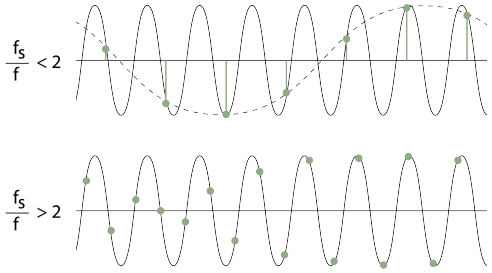


Figure 4: Example of aliasing as a result of undersampling (upper) and an example of proper sampling according to the Nyquist frequency theorem (lower).

2.3. Unsupervised Machine Learning

With supervised ML, the most common ML approach, the algorithm is provided with a desired outcome measure (e.q. labels). It seeks to identify the links between the given features and the desired outcome measure, resulting in a classification or regression model for discrete or continuous data, respectively. The model can then be used to make predictions about new, 'unlabelled' data. In contrast, unsupervised ML is not depending on predefined labels and involves discovering patterns or structures within data without guidance, seeking to uncover inherent relationships or groupings. This approach is particularly beneficial when dealing with large datasets where labelling is too costly and time-consuming or when a standardised label does not exist. Unsupervised methods excel at finding hidden correlations or clustering within the data. In medicine, unsupervised ML can, for example, identify distinct subgroups of patients based on similarities in their medical profiles, symptoms, or response to treatments.

As unsupervised ML does not have an explicit outcome measure, one challenge is to evaluate the

performance of the model. Typically, the usefulness of patterns discovered by unsupervised machine learning must be validated by human assessment. Nevertheless, certain metrics, such as the silhouette score, provide an indication of how well the clustering was performed. The metric measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation), with a higher score indicating more optimal clustering.

3 | Methods

3.1. Data acquisition

This study utilised retrospective data gathered from two different studies (ROAR & ORNAC) obtained in the LUMC, of which the characteristics and methods are listed in Appendix A. All studies used the same high frequency proximal flow sensor (Sensirion, Stäfa, Switzerland) in combination with an Oro-Nasal Mask to measure breath signals before and after opioid administration (Figure 5).

The flow sensor measures bidirectional flow in standard liters per minute (slm), with inspiration recorded as negative values and expiration as positive values. By using a controllable heating element located in the middle of a pressure-stable membrane, temperature sensors mounted symmetrically upstream and downstream measure the resulting thermal transfer of heat caused by the passing flow [23]. In combination with the Oro-Nasal Mask, which has a tight seal with almost negligible leakage, the setup results in a precise signal and has a high sensitivity to low flow rates and pressure differences.

After baseline data was collected, a bolus followed by a low-dose infusion of either fentanyl or sufentanil was administered until ventilation had decreased by a study dependent predetermined amount. After this threshold was reached, an antagonist, to reverse the effects of respiratory depression, or a placebo was administered. For the development of the model, only the data until antagonist administration was used. An example of the setup and resulting flow curve is visualised in Figure 5. In addition to flow data, $EtCO_2$, inspired oxygen fraction, minute volume (MV), tidal volume, and saturation were also recorded. All studies consisted of two study days, where the used opioid or antagonist was alternated between the two days.

3.2. Data preprocessing

3.2.1. Re-sampling

The obtained flow data had a non-uniform sample frequency ranging from 1000Hz to 2300Hz. Initially, to establish a uniform sample rate, the signal was interpolated to a higher sample frequency of 2500Hz, in accordance with the Nyquist's theorem. A 10th order low-pass filter was applied based on the observation that most interesting frequencies were below the 10Hz domain as shown in Figure 6. In addition, the lower frequency range allowed down-sampling the data to achieve faster computation times during subsequent data analysis. While the Nyquist's theorem prescribes a minimum sampling frequency of 20Hz for signals below 10Hz, a slightly higher frequency of 40Hz was chosen to minimise the chance of aliasing.

3.2.2. Artefact detection

Artefacts in the data were mainly caused by moving or talking of the participant, both resulting in fast alternating high peaks in the signal. To detect these artefacts, the moving median of the signal was used. Peaks in the absolute differences between the moving median and the original signal were flagged as artefacts when lying outside the upper fence of the Interquartile Range (IQR).

In addition, detachment of the device resulted in a flat line in the signal, as no flow passed the sensor. In contrast with apnea, which is also characterised by a period of 'no flow', the detachment leads to a completely flat line, while during apnea small flow alternations persists due to air movement caused by the beating of the heart. Therefore, using the moving maximum, values below 0.2 were indicated as the detachment of the device. Both detection methods are illustrated in Figure 5.

Both detected artefacts were initially not deleted from the signal, as this would leave gaps in the signal and interfere with the separation of breaths epochs later on. Instead, detected data points were solely marked during this stage.

3.2.3. Separate breath epochs

After preprocessing, breath epochs were segmented from the signal. Therefore, the peaks corresponding to the highest points of inspiration and expiration efforts were detected. A breath epoch was defined as the occurrence of an inspiration peak followed by an expiration peak. As the signal crossing the y-axis indicates the transition between inspiration and expiration, the crossing before the

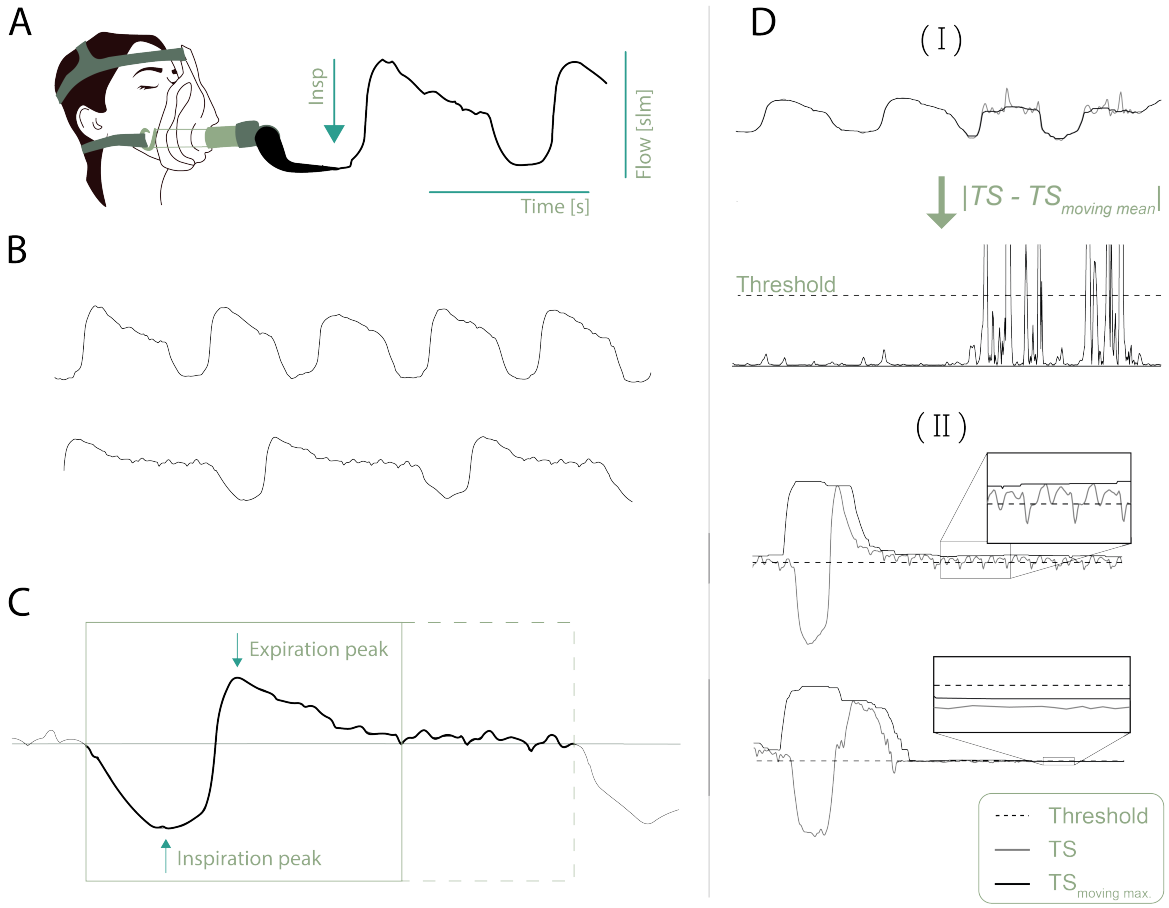


Figure 5: Method for data acquisition and pre-processing. **(A)** Measurement setup. The inspiration and expiration results in negative and positive flow, respectively. **(B)** Example of the difference in the flow of breathing during baseline (upper) and after opioid administration (lower). **(C)** Breath epochs were identified by detecting peaks in inspiration and expiration. A breath epoch was defined as the sequence of an inspiration peak followed by an expiration peak. The start and end of each breath epoch were determined by the zero-crossing before inspiration and after expiration (solid window). To encompass apneas in the epoch, all data until the subsequent breath was incorporated for data analysis (dashed window). **(D)** Example of artefact detection. (I) Noise was identified by comparing the signal variance between the moving median (black line) and the original signal (grey line). Subsequently, peaks found in the absolute variances were marked as artefacts if they were greater than the upper fence of the Interquartile Range (IQR), represented by the dotted line. (II) To differentiate between device attachment and apneas, the moving maximum was employed. If the moving maximum (black line) of the original signal (grey line) fell below the threshold of 0.2 (dotted line), it was classified as device detachment.

inspiration peak, and the crossing after the expiration peak were marked as the start and end of a breath, respectively. However, to include apneas, the entire epoch from the start of one breath until the crossing before the next inspiration peak was segmented. In Figure 5 the solid box indicate the initial detected epoch from start of inspiration till the end of expiration. The final included epoch is indicated with the extended dashed box.

In instances where a detected inspiration peak was directly succeeded by another inspiration peak, only the latter was employed to identify breaths. Nevertheless, the crossing before the first inspiration peak served as an indicator for the end of the preceding epoch. Finally, epochs containing data points marked as artefacts were excluded from subsequent analysis.

3.2.4. Selecting breath epochs

To avoid the probability that poor data quality would be the reason this pilot study would obtain unsuccessfully results, breath epochs that would be used for model development were checked by the researcher. Epochs that included multiple breaths or contained a lot of noise were manually excluded. The amount of epochs to select were later determined based on the amount of detected breaths by the algorithm. Subsequently, time series of the in-

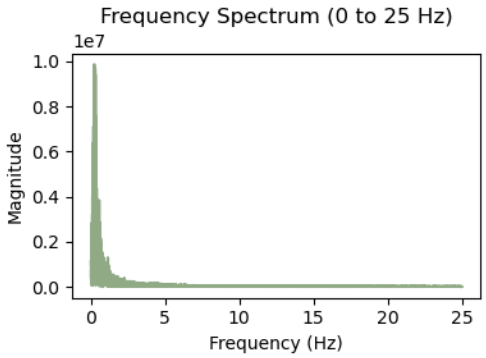


Figure 6: Frequency spectrum of the flow signal.

cluded breath epochs were stored in a matrix as followed:

$$X_{ts} = \begin{pmatrix} TS_1 = \{x_1, x_2, ..., x_n\} \\ TS_2 = \{x_1, x_2, ..., x_n\} \\ \vdots \\ TS_n = \{x_1, x_2, ..., x_n\} \end{pmatrix}. \quad (2)$$

3.2.5. Scaling breath epochs

To eliminate the influence of variable maximum flow, which may be subject specific, z-score normalisation was applied to each breath epoch. This normalisation process adjusts the data so that the resulting distribution has a mean of zero and a standard deviation of one and is performed as followed:

$$z = \frac{x - \mu}{\sigma}. \quad (3)$$

where:

μ = Mean
σ = Standard Deviation.

3.2.6. Feature / data exploration

Initially, prior to conducting the preliminary data analysis, the data underwent exploration by assessing correlations among specific features of the signal. These features were chosen based on literature and a visual interpretation of the signals and included: the minute volume based on one breath; the length, volume and maximum flow of inspiration, expiration & the ratio between both; the length of apnea and total breath & the ratio between both; and the tangent of the flow between inspiration and expiration.

Furthermore, we examined correlations between these features and the changes in $EtCO_2$ and MV relative to baseline, both individually for each subject and collectively. To assess differences in feature variation, the features were normalised using z-score normalisation (Equation 3). Subsequently, the variation in these features was visualised by plotting their values in a boxplot, distinguishing between feature values at baseline and post-opioid administration.

3.3. Preliminary data-analysis

During the preliminary data-analysis the objective was to find the best feature-space and distance metric to cluster the breath epochs. Therefore, three different approaches were applied. The first approach was based on the rat study of Sunshine and Fuller [11], were they observed a difference in the appearance of clusters in baseline and after opioid administration using unsupervised clustering with flow data of rats obtained with full-body plethysmography. The latter two approaches are extensions of this first approach.

3.3.1. Model development

While the different approaches varied in feature-space and distance metric, all used the same clustering method: hierarchical clustering. Hierarchical clustering, as implied by its name, aims to organise clusters into a hierarchical structure. This can be achieved through either a top-down (divisive) or bottom-up (agglomerative) strategy, with the latter being utilised in this study. Therefore, the process begins with each data point considered as an independent cluster, which are then progressively merged based on their similarity until all objects are encompassed within a single, comprehensive cluster. This process yields a linkage matrix, typically represented as a dendrogram where the height of each link shows the distance (or dissimilarity) at which the linked clusters were merged (Figure 7).

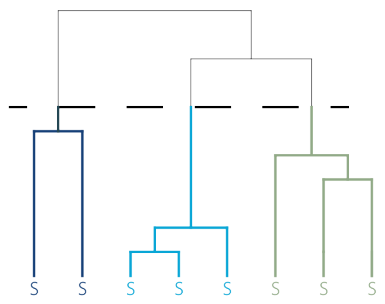


Figure 7: Example of a dendrogram used during hierarchical clustering.

This process requires determining two crucial components: the linkage method and the distance metric. The linkage method determines how clusters are merged at each step of the hierarchical process, whether it be through measuring the distance between the closest points (single linkage), farthest points (complete linkage), or minimising variance (Ward's linkage). In this study, the latter technique was adopted. Additionally, the distance between points required for the linkage method is determined by the distance metric, which varies among the different approaches within this study.

One of the advantages of hierarchical clustering is that it does not require to specify the number of clusters in advance. A variety of methods are known to determine the amount of clusters. For this study the dendrogram, as well as the silhouette score were used.

In the following paragraphs the different approaches will be explained. All approaches used the matrix as described in Equation 2. In contrast to approaches 1 and 3, which both employed the z-score normalised matrix, approach 2 used the matrix without z-score normalisation.

Approach 1: Principal component analysis + Euclidean distance

As this approach required the time-series to be of the same length, zeros were padded to the end of the time-series when needed. Initially, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the dataset. PCA transforms the original features of a dataset into a new set of uncorrelated variables called Principal Components (PCs), each explaining a portion of the overall variance within the data.

PCs were selected from highest explaining variance to lowest until a cumulative variance of at least 90% was reached, creating a new feature space:

$$X_1 = \begin{pmatrix} PC1_1, & PC2_1, & ..., & PCn_1, \\ PC1_2, & PC2_2, & ..., & PCn_1 \\ & \vdots & & \\ PC1_n, & PC2_n, & ..., & PCn_n \end{pmatrix}. \quad (4)$$

Then, the linkage matrix was obtained using the Euclidean (point-to-point) distance metric.

Approach 2: Calculated feature space + Euclidean distance

Instead of using each time stamp as an individual feature, a feature-space was created containing calculated features that described the waveform. These features were chosen based on the results of the feature / data exploration phase, excluding features that showed a strong correlation with other features to allow a smaller feature space.

In total, for all time series, all the normalised features (Equation 3) were stored in a new matrix:

$$X_2 = \begin{pmatrix} F1_1, & F2_1, & ..., & Fn_1, \\ F1_2, & F2_2, & ..., & Fn_1 \\ & \vdots & & \\ F1_n, & F2_n, & ..., & Fn_n \end{pmatrix}. \quad (5)$$

Features deemed more significant based on data exploration were assigned a weight of 2, while others were assigned a weight of 1. Finally, this approach also utilised the Euclidean distance metric to obtain the linkage matrix.

Approach 3: Time series + dynamic time warping

Instead of using the Euclidean distance metric utilised in the other two approaches, dynamic time warping (DTW) was chosen. By allowing local deformations in the alignment, DTW is less dependent on time shifts compared to the Euclidean distance, which computes the straight-line distance between corresponding points in time series.

The linkage metric was obtained using the time series matrix described in Equation 2, without any alterations. Since DTW allows the comparison of time series with varying lengths, substituting zeros at the end is not necessary.

3.3.2. Preferred approach

To identify the most effective method, various hypotheses were formulated regarding the desired clustering behaviour for it to be considered an adequate model. Then, the selection of the preferred approach was based on a series of visualisations and tests designed to either support or refute these hypotheses.

Initially, it was expected that clusters would show a significant difference in $EtCO_2$ relative to baseline, indicating that certain clusters are related to a more severe respiratory depression, resulting in an elevated $EtCO_2$. The choice of statistical tests was determined based on the distribution of the

data. First, the variance in EtCO2 values relative to baseline were examined among the four clusters. If found to be significant, the differences between adjacent clusters were assessed, which were visualised together with a boxplot.

In addition, it was also examined whether the clustering was independent of subjects. When a clustering approach is subject-dependent, clusters tend to predominantly appear in specific subjects instead of being uniformly distributed between all subjects. This suggests that the clustering is influenced more by individual subjects breathing patterns rather than the contrast between baseline and post-opioid characteristics.

Lastly, to determine the cluster stability and prevent serendipitous findings, 'bootstrapping' was employed. This involved conducting iterative clustering, with a single subject excluded during each iteration. To allow comparison between iterations, the resulting clusters were ordered based on their median $EtCO_2$ relative to baseline. Then, the variation in clustering for each epoch was measured against the median assigned cluster of that epoch over all iterations.

3.4. Main data-analysis

The preferred method that was identified from the preliminary study was used for the final data analysis, with adjustment where needed. After the model was trained, predictions were made for epochs after antagonist administration. For this, only data from the ROAR study was used, as for the ORNAC study it was not known whether an antagonist or placebo was administered.

The results of the final clustering, as well as the additional antagonist results, were interpreted by the researcher. It was investigated which features, as described by the feature exploration phase, had the greatest influence on the clustering results. Furthermore, based on the observations of the researcher, additional questions regarding the clustering results were identified for the comprehensive data-analysis.

4 | Results

4.1. Data characteristics

On January 17 2024, the day of final data extraction, 42 participants (31 ROAR; 11 ORNAC) were included in the LUMC dataset, of which 39 completed both the study days at that time.

4.1.1. In- / exclusion breath epochs

When creating the databases used for data-analysis 8 measurements, all from different subjects, were excluded based on: Sensirion not used (n=2), other measurement settings used (n=4), and additional errors (n=2). Nevertheless, since all measurements were from subjects who completed both study days, no subjects were excluded during this part.

After breath detection by the algorithm an additional of 4 subjects were excluded, since they had less than 50 breaths detected in baseline (n=1) or less than 70 after opioid administration (n=3).

After the manual assessment of breath epochs, subjects were included if they had more than 70 breaths in baseline and 125 breaths after opioid administration. When this applied on both study days of a subject, the first one was included. During this process, four additional subjects were excluded due to not meeting the inclusion criteria of either 70 breaths at baseline (n=1) or 125 breaths post-opioid (n=3). In total, 34 subjects were included for training, resulting in a total of 6630 breath epochs for model development. In Table 1 an overview of the characteristics of the included subjects is given.

4.1.2. Data exploration

Figure 8 illustrates the correlation between the features as well as the variation. As expected, a positive correlation of 1 is observed between the *inspiration volume* and *expiration volume*. The *maximum flow ratio* and *volume ratio* both showing only low correlations, indicating that these ratios are not solely dependent on the inspiration or expiration. The *length ratio* and the *total epoch length* are mainly influenced by the expiration, characterised by moderate to strong correlations.

From the variation plot, it can be concluded that the variation within the features are quite similar, with *volume ratio* showing the least variation. *Minute volume* has a high variation and shows a noticeable shift in medians between baseline and post-opioid. Regarding the *apnea length*, the variation is more pronounced when considering the ratio compared to the *total epoch length*.

The collectively analyses between the features and both $EtCO_2$ and MV change relative to baseline, indicated that only the features of *total epoch length*, and *apnea length* showed a noticeable correlation with the $EtCO_2$ level. Specifically, when analysing solely all healthy subjects, the correla-

Table 1: Characteristics of patients included in the data-analysis

	ROAR (n = 23)	ORNAC (n = 11)
Gender, male (n, %)	11 (48%)	3 (27%)
State, non-opioid users (n, %)	10 (43%)	11 (100%)
Age, years (mean ± SD)	39.7 ± 15.1	23.5 ± 1.8
Length, cm (mean ± SD)	177.1 ± 8.7	181.7 ± 6.0
Weight, kg (mean ± SD)	78.4 ± 12.9	75.0 ± 11.0
BMI, kg/m2 (mean ± SD)	25.1 ± 4.0	22.6 ± 2.4
Baseline EtCO2, mmHg (mean ± SD)	4.9 ± 0.5	4.8 ± 0.2
Baseline MV, L (mean ± SD)	8.2 ± 2.4	8.6 ± 3.4

tion of features associated with the duration of breath epoch (inspiration, expiration, total, and apnea) showed a higher correlation. In contrast, MV did not display any notable correlation among the features. It was notable that some subjects had low correlation across all features, whereas others showed mainly high correlations, which could not be explained as a difference between the healthy subjects and opioid users. Additionally, some features showed high correlation in some subjects but were reversed correlated in others, resulting in a negligible overall correlation. A heatmap illustrating the correlations can be found in Appendix C.

4.2. Preliminary data-analysis

4.2.1. Model specifications

For both the first and second approach specifications emerged from steps during model development or were determined based on results of the data exploration, respectively.

For approach 1, this included identifying the number of PCs required to achieve a cumulative variance of 90%. The PCA results indicated that five principal components were required, leading to a cumulative variance of 91%. The eigenvalues of these PCs were all greater than 1.

For approach 2, features with high correlations to other features (Figure 8 A) were excluded (n=3), leaving the following 11 features included for the model development:

1. Minute volume based on one breath
2. Length of inspiration
3. Maximum flow of inspiration
4. Volume of inspiration
5. Length of expiration
6. Length of total epoch
7. Ratio of volume between ex- and inspiration
8. Ratio of maximum flow between ex- and inspiration
9. Ratio of length between ex- and inspiration
10. Ratio of length between apnea and total epoch
11. The tangent of the flow between inspiration and expiration.

In addition, based on the variation within a feature and the difference between baseline and post-opioid (Figure 8 B), it was decided to assign the features *minute volume based on one breath* and *ratio of length between apnea and total* a weight of two.

4.2.2. Clustering result

The silhouette scores and dendrograms that resulted from the model development are visualised in Figure 9 and Appendix D, respectively. In contrast to these results, which conclude that two clusters would be ideal for each approach, a cluster amount of four was chosen. This was decided since one of the aims of this research was to determine whether different clusters would appear more after more opioid was administered. Using only two clusters, it would not be possible to test this.

The results of the clustering, are visualised in Figure 10 were each cluster is represented as the same color across all three plots. For statistical analysis, non-parametric tests were used, since the $EtCO_2$ relative to baseline was not normally distributed within the clusters. First, the difference within the approach was evaluated using the Kruskal-Wallis test, which yielded a p-value of less than 0.001 for all approaches. Furthermore, the Mann-Whitney U test was applied to assess the differences between adjacent clusters, which are visualised in the boxplot. Except for the comparison

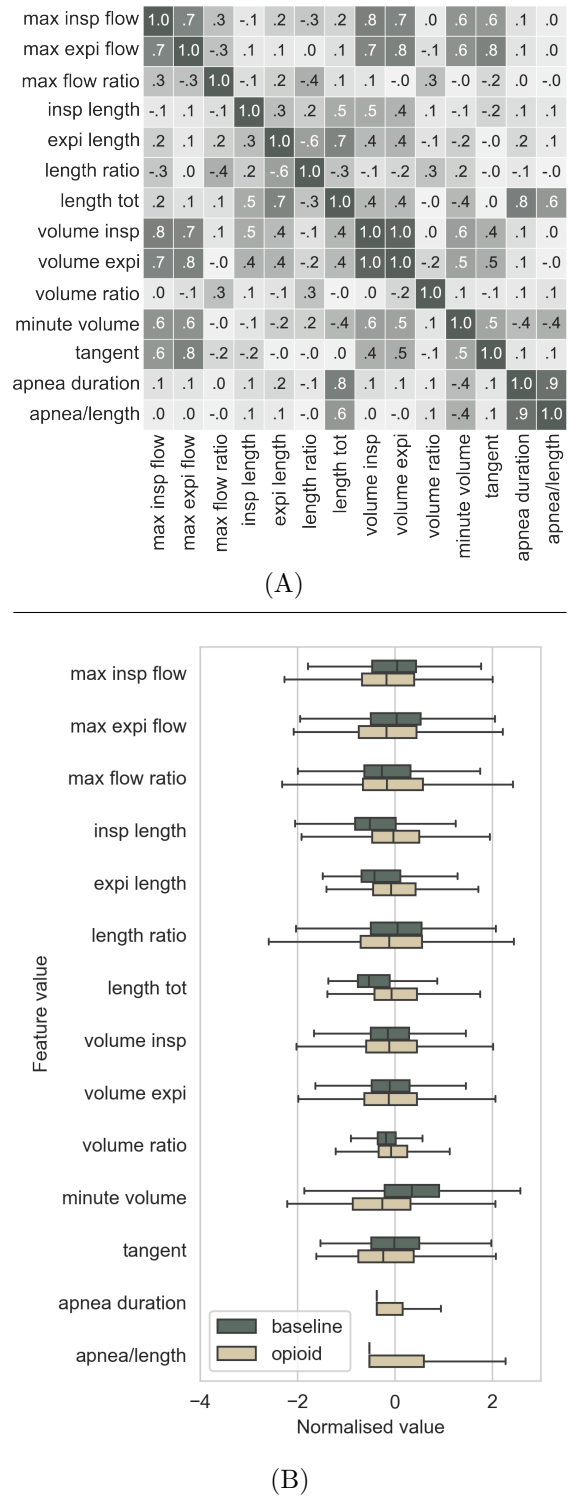


Figure 8: Visualisations of data exploration. (A) The correlation between different features. (B) The distribution of different features broken down into baseline and after opioid administration.

between 3 and 4 in the first approach, significant differences were observed for all adjacent cluster pairs ($p < 0.0001$). Despite these statistical findings, all boxplots show a large overlap, with the PCA approach showing the smallest trend.

The bar plots reveal that clusters with higher overall $EtCO_2$ levels relative to baseline tend to occur less frequently during baseline and more frequently post-opioid. Although the second and third approach show similar results in terms of the percentage occurrence of the clusters, this phenomenon is less obvious in the PCA approach. Nevertheless, it's important to note that none of the approaches had a cluster that was exclusively present post-opioid.

None of the approaches showed a subject-dependent clustering. However, the third cluster in the PCA approach and the fourth cluster in the DTW approach were absent in the majority of subjects, which can be explained by the overall lower incidence of these cluster in comparison to the other clusters (n=434 & n=124). An overview of the cluster incidence per subject can be found in Appendix D.

The bottom graph in Figure 10 shows the temporal distribution of clusters, with each row representing an individual subject and the black line marking the onset of opioid administration. The absence of color indicates periods were no breaths were included. This chart was evaluated visually, revealing a shift in cluster occurrence post-opioid administration when subjects are examined individually. However, when analysing the data collectively across all subjects, it becomes evident that a cluster predominant in the post-opioid phase for one subject might correspond to a prevalent baseline cluster for another. Consequently, while within-subject cluster patterns may indicate the post-opioid phase, these patterns are not consistent across different subjects and thus not reliable for inter-subject comparisons.

From the temporal distribution, it can also be observed that the cluster assignment differences between the approaches. Of the 6630 epochs, only 1356 were consistently assigned the same cluster across all approaches. In 2564 epochs, two approaches assigned the same cluster, while the third approach assigned the epoch to an adjacent cluster. Of these, the first approach (PCA) deviated in 846 epochs, the second (FS) in 756 epochs, and the third (DTW) in 962 epochs.

The training duration's varied across the clustering method, with the DTW method being the most time-consuming, requiring approximately 20 minutes to complete. In contract, both PCA and

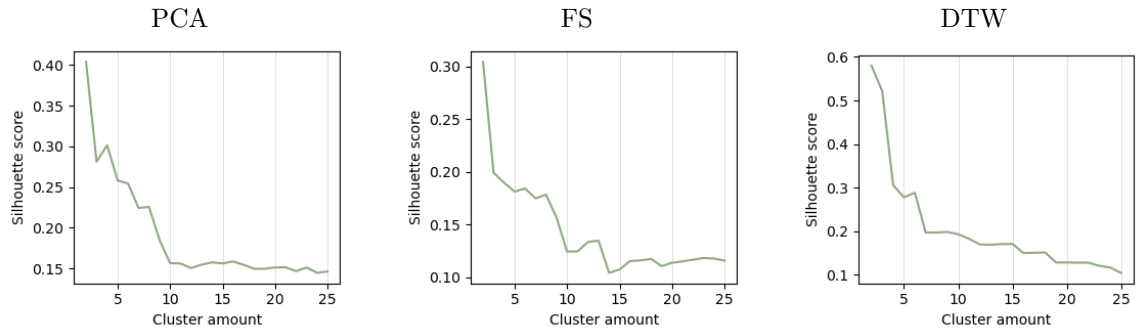


Figure 9: Silhouette score for each approach.

FS were notably faster, completing training in 40 seconds and 20 seconds, respectively.

4.2.3. Cluster stability

To evaluate the cluster stability, bootstrapping techniques were utilised. A plot showing the temporal distribution of clusters for each iteration is attached in Appendix C. The findings of the variation in clustering are illustrated in Figure 11. Generally, epochs either showed no change in assigned cluster or were assigned to an adjacent cluster. Notably, in the PCA and FS approaches, shifts to the adjacent cluster were more observed than remaining the same cluster, suggesting these approaches was the least stable. Therefore, the DTW approach being slightly more stable due to its reduced incidence of two-cluster shifts.

4.3. Main data-analysis

It is noteworthy that during bootstrapping in the preliminary study, the clustering deviated by only one cluster for the majority of epochs. Furthermore, the clusters were observed to align sequentially in a matrix kernel density estimate plot (Appendix E). These observations could argue for opting for regression over fixed clustering. Therefore, it was chosen to implement fuzzy clustering for the main data-analysis, which quantifies the probability of a breath epoch belonging to a cluster, referred to as the membership value, yielding a continuous result in terms of clustering.

In the preliminary data analysis, both the FS and DTW approaches yielded similar results, outperforming PCA. Nevertheless, as there were no Python package available to allow DTW as a distance metric for fuzzy clustering, the choice was made to utilise the FS approach.

Based on the observed inter-subject variability in the results of the preliminary study, it was decided to train the final model also individually for each subject in addition to solely collectively. This approach allows the cluster behaviour within each individual subject to be examined independently, without being influenced by other subjects.

4.3.1. Model development

For the development of the model, the *cmeans* function from the *skfuzzy* library was applied. The corresponding function *cmeans_predict* allows new data to be fitted to the trained model and will be used to analyse the data of the period after antagonist administration.

The probability that a breath epoch is part of a specific cluster is indicated by its membership value, of which the sum for an epoch across all clusters always equals 1. The value is determined using a membership function:

$$u_{ij} = \left(\sum_{k=1}^{C} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1}. \qquad (6)$$

where:

u_{ij} = the membership of the ith data point in the jth cluster.
x_i = the ith data point
v_j = the center of the jth cluster
v_k = the center of the kth cluster
C = the total number of clusters
m = the fuzziness parameter, controlling the level of cluster fuzziness

The membership function is non-linear, meaning that the decrease in distance causes a more significant increase in membership as points get closer

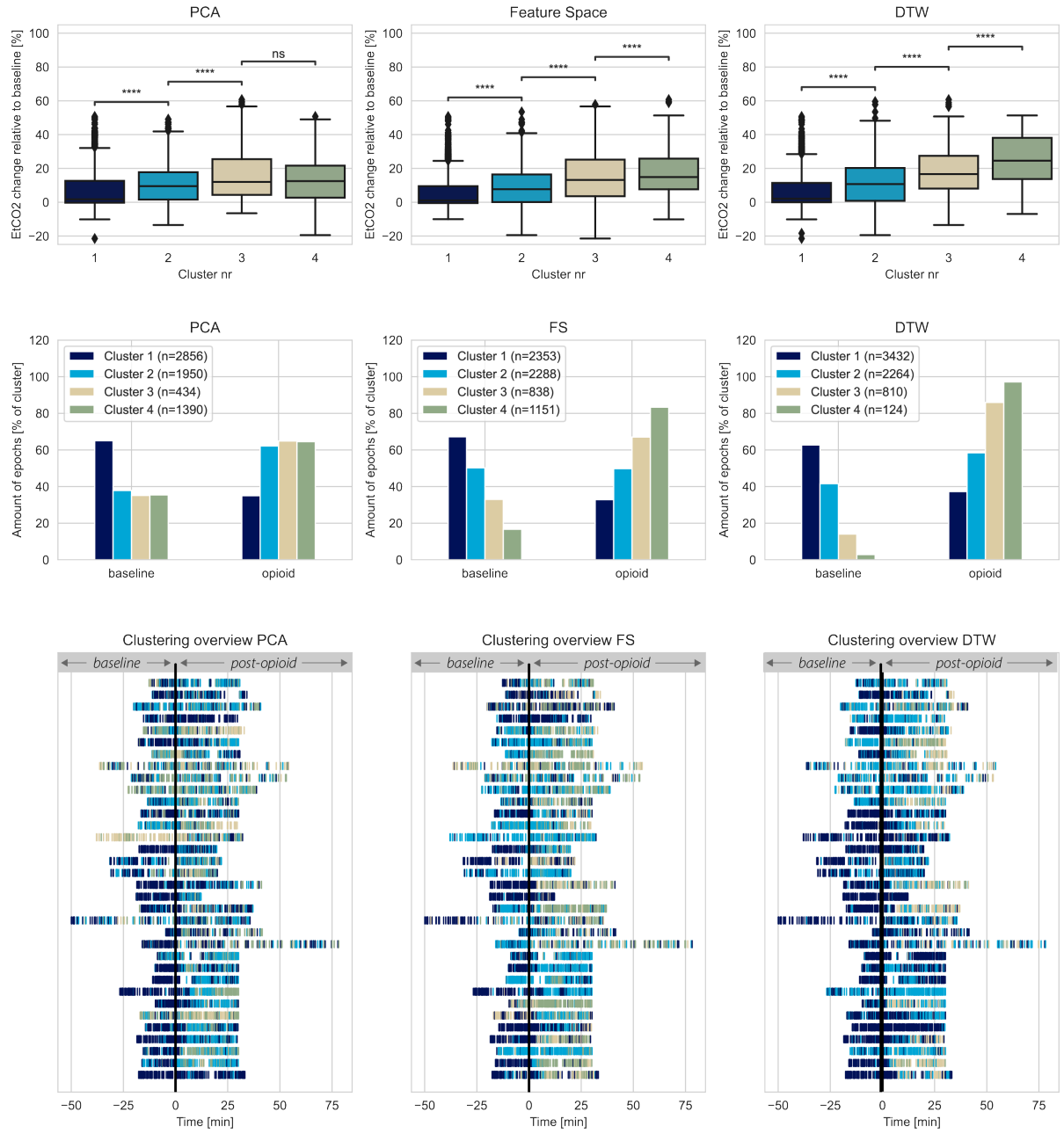


Figure 10: Clustering results per approach. Each plot represents every cluster using the same color. Upper: boxplots showing the change in $EtCO_2$ relative to the baseline (ns : $p > 0.05$; ****: $p < 0.0001$); Mid: bar plots indicating the prevalence of the clusters; Lower: temporal plots visualising the change in cluster prevalence over time. Each row represents an individual subject. The vertical black line indicates the time of opioid administration.

to the cluster center. The fuzzy parameters m, which typically has a value of 2, defining the level of 'fuzziness' in the clustering. As m approaches 1, the algorithm becomes more like hard clustering, while a higher value of m results in memberships values that are more evenly distributed across clusters. For this study, it was chosen to use the standard value of 2.

The process of recalculating membership values and corresponding cluster centers continues iterative until the change in membership between consecutive iterations falls below a specified threshold. Next, these cluster centers were used to calculate the membership values of the data points after antagonist administration using *cmeans_predict*.

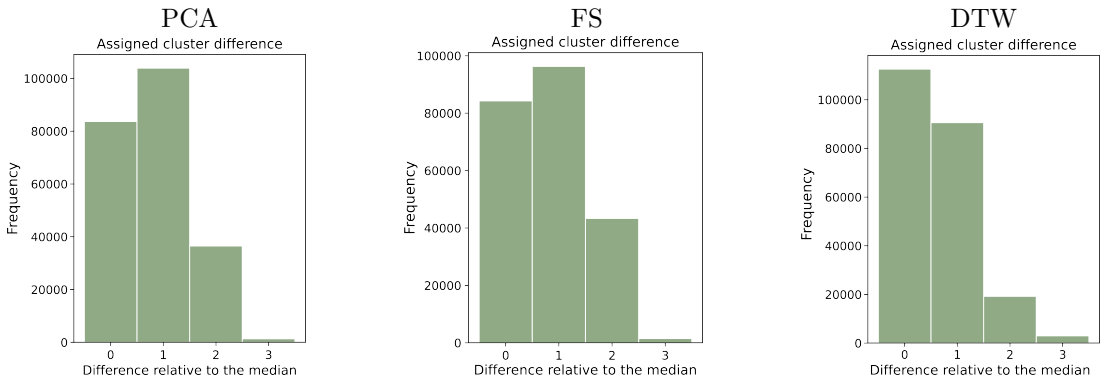


Figure 11: Clustering stability based on bootstrapping.

4.3.2. Model analysis

Another advantage of using the FS approach, is the possibility to use SHAP (SHapley Additive exPlanations) value analysis, which is a method used in machine learning to interpret model predictions [24]. The SHAP values of each feature describe the contribution of that feature in the difference between the predicted value of the model for the given epoch ($f(x)$) and the mean of all predictions ($E[F(X)]$). Thus, the sum of all SHAP values will be equal to $E[f(x)]$—$f(x)$. A visual example of how the SHAP values describe the model outcome is shown in Figure 12.

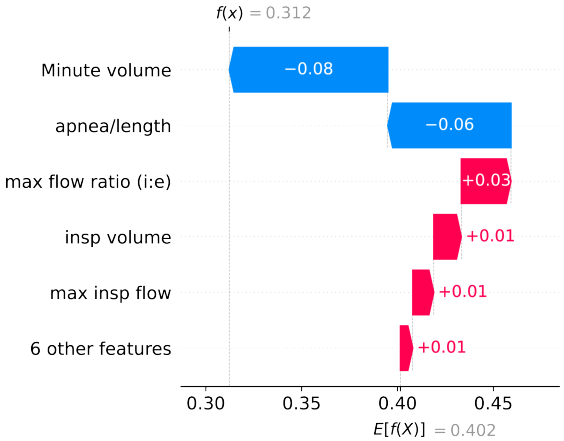


Figure 12: Illustration of SHAP Analysis: the x-axis represents the membership values. $E[f(x)]$ is the expected outcome, equal to the mean of all predictions, and $f(x)$ the predicted value of the model for the given input epoch x. The rows depict the impact of each feature on the prediction outcome, with positive contributions in red and negative in blue, transitioning from the expected output to the specific prediction output.

To allow integration between the Python packages *skfuzzy* and *SHAP*, modifications were made to the *cmeans* and *cmeans_predict* functions of *skfuzzy*, transforming them into a Python class structure with defined *fit* and *predict* methods. In addition, whereas the SHAP value analysis is designed for a single outcome value, the fuzzy cluster model produces two outcome values (e.g., membership values for each cluster). Therefore, the *predict* method was designed to return only the membership values of one cluster, making it a single outcome value. It was chosen to let $f(x)$ correspond to the membership values assigned to the cluster most predominant post-opioid, where higher $f(x)$ values indicate a more severe respiratory depression. Since the sum of the membership scores is equal to 1, analysing the other cluster separately is unnecessary as its outcome would be the exact opposite.

4.3.3. Model results

Since the model was trained collectively across all subject as well as for each subject individually, the clustering results were divided into collective- and individual-clustering results, respectively. For both, temporal plots were established per subject, showing the change in membership values over time. Based on the observed behaviour of these plots, subjects were divided into four subgroups as followed:

1. *Complete switch*: distinction between cluster memberships in baseline as well as post-opioid, showing a change in the dominating cluster.
2. *Observed change*: distinction between cluster memberships in baseline in combination with a notable difference (moving towards a change in dominating cluster or continue alternating) post-opioid.

3. *Reversed change*: alternating dominant cluster in baseline in combination with a distinction between cluster membership post-opioid.
4. *No difference*: no difference in the dominating cluster or an alternating dominating cluster in baseline as well as post-opioid.

In addition, the same was performed for the observed behaviour after naloxone administration, classifying subjects into: *complete recovery*, *incomplete/temporary recovery* and *no recovery*. Examples for both subgroup classifications are visualised in Figure 13. All temporal plots together with the corresponding $EtCO_2$ and MV can be found in Appendix E. The collective- and individual-clustering results are described separately in the following two paragraphs.

Collective-clustering

The temporal plots showed a change post-opioid in 21 (62%) subjects, of which 6 (18%) showed a *complete switch*, 12 (35%) an *observed change* and 3 (9%) a *reversed change*. The remaining 13 (38%) subject were classified as *no difference*. Regarding the baseline, the same cluster was dominant for 27 (79%) subjects. In contrast, in one healthy subject, the other cluster was dominant in baseline, however showing *no difference* post-opioid. The remaining five (15%) subjects showed an alternating dominant cluster in baseline, of which three were classified as *no difference* and two a *reversed change* post-opioid.

Apnea/length showed to be the most important feature, with *minute volume* being second, according to their mean absolute SHAP values of 0.12 and 0.08, respectively. The other features all had mean absolute SHAP values around the 0.01. At the level of individual subjects, three subjects had, in contract to the overall result, *minute volume* as their most important feature, positioning *apnea/length* as second. The order of importance of the features other than *apnea/length* and *minute volume* alternated among subjects. There was no variation in feature importance among *male/female*, *user/healthy*, *fentanyl/sufentanil* or *ORNAC/ROAR* differences.

To determine clustering stability, bootstrapping was applied by leaving one subject out every iteration. Then, the differences among the iterations were determined for the membership values as well as the SHAP values, showing a stable clustering. Regarding the membership values, the overall difference in membership values was 0.007, with the maximum observed difference between two epochs being 0.14. *Minute volume* showed the most alternation between iterations, having an overall difference of 0.011, with a maximum of 0.06 between two epochs. For the other features, the overall difference in SHAP values among epochs was <0.0025.

Individual-subject clustering

A *complete switch* and *observed change* was observed in 16 (47%) and 14 (41%) subjects, respectively. In contrast to the collective-clustering results, no subjects were classified as *reversed change*. However, there were still four (12%) subjects in whom *no difference* was observed. These four subjects all showed alternation between the prominent cluster in both baseline and post-opioid and were classified as *no difference* during collective-clustering as well.

For 20 (59%) subjects, *apnea/length* was identified to be the most import feature, consistently followed by *minute volume* as the second most significant. For the remaining 14 (41%) subjects, *minute volume* was the most important feature. However, among these, only two subjects had *apnea/length* as their succeeding feature. For the rest, the second most important feature differed among subjects with some showing an equal significance across multiple subsequent features.

As bootstrapping by *'leaving one subject out'* was not possible, as the model was developed on solely one subject, bootstrapping was applied by randomly excluding different epochs each iteration. The membership values showed a mean difference of 0.002 and a maximum difference of 0.015. Regarding the features, the *minute volume* SHAP values showed the most alternation, having a mean and maximum difference of 0.013 and 0.053, respectively, similar to the collective-clustering results. In addition, the remaining features also had an overall difference in SHAP values of <0.0025.

For subjects of the ROAR study, membership values of epochs within one hour after naloxone administration were predicted. Six subjects were excluded from prediction because they showed *no difference* between baseline epochs and epochs after opioid administration (n=3) or no data was available post-naloxone administration (n=3). Finally, six subjects were identified as having a *complete recovery*, whereas nine showed only an *incomplete/temporary recovery*. The remaining two subjects showed *no recovery* in the temporary plot. However, they did showed a recovery in $EtCO_2$.

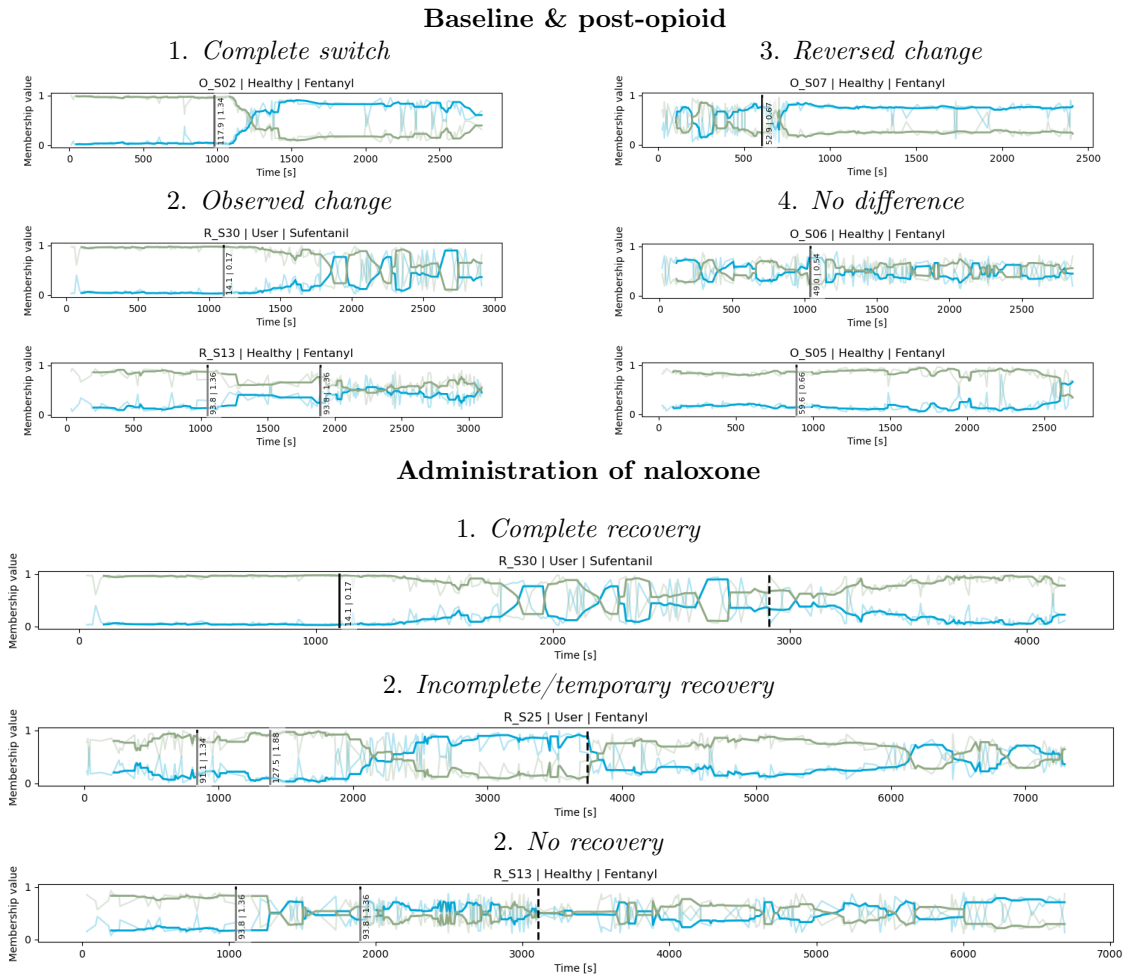


Figure 13: Examples of the temporal membership plots for each classification. The thicker lines represent the median over 7 epochs, while the background lines depicts the original membership values. The solid and dotted black vertical lines annotate the time of opioid administration, including the given dose (total dose | dose/kg), and the time of antagonist administration, respectively.

4.3.4. Data analysis

Based on the observations of the clustering results both collective and individual, questions were formulated. These questions guided the execution of various data analyses aimed at providing answers. The questions included:

- Can the discrepancy between collective- and individual-clustering be explained?
- What happens at the moment a switch occurs post-opioid in the individual-clustering results?
- What is the reason there is a difference in clustering behaviour, indicated by the different classifications, among subjects?

Discrepancy collective- and individual-clustering

To test if there was similarity between the collective- and individual-clustering results, the differences between membership values as well as the feature behaviour were investigated.

Regarding the four classification, the temporal plots of 20 (59%) subjects were classified the same for both the collective- and individual-clustering method. For each epoch, the difference between the assigned membership value resulting from the collective- and individual- clustering were computed, yielding an average difference of 0.026. When the average difference among epochs was calculated for each subject separately, it ranged from 0.012 to 0.049. Including solely the subjects that were classified the same for both collective- as individual-clustering the mean difference

decreased to 0.020 (range: 0.012 – 0.035).

The overall feature importance, as well as the range observed among subjects, are listed in Table 2. In both the collective- and individual-clustering, *minute volume* and *apnea/length* are the most important features. Nevertheless, when clustering individually, the range of the feature importance among subjects is wider. In collective-clustering, *minute volume* and *apnea/length* consistently show relatively high importance in every subject, with a minimum observed importance of 0.05 compared to a maximum of 0.05 for other features. Conversely, in individual clustering, these features show low significance in at least one subject. Moreover, in individual clustering, there is an increase in the maximum observed significance of the less important features when compared to their importance in collective clustering.

Table 2: Feature importance (*mean(|SHAP value|)*)

Feature	**Importance** *mean (range: min - max)*	
	Collective	Individual
Minute volume	0.08 (.05 - .12)	0.10 (.02 - .16)
apnea/length	0.12 (.07 - .20)	0.09 (.00 - .17)
insp length	0.01 (.01 - .05)	0.02 (.00 - .04)
expi length	0.01 (.00 - .02)	0.02 (.01 - .05)
length ratio	0.01 (.00 - .01)	0.02 (.00 - .04)
total length	0.01 (.01 - .02)	0.02 (.01 - .05)
insp volume	0.01 (.00 - .01)	0.01 (.00 - .04)
volume ratio	0.01 (.01 - .02)	0.01 (.00 - .03)
max insp flow	0.01 (.01 - .01)	0.02 (.00 - .04)
max flow ratio	0.01 (.00 - .02)	0.01 (.00 - .02)
tangent	0.01 (.01 - .02)	0.01 (.00 - .04)

In addition, partial dependence plots were obtained for each feature. These plots illustrate how the predicted outcome varies with changes in feature values by plotting the feature values against the associated SHAP values. Figure 14 shows the partial dependence plots for the two most important features: *minute volume* and *apnea/length*. Here, the grey area represents the distribution of observations from the collective-clustering results, while the mean of each individual-clustering results is depicted as a black line.

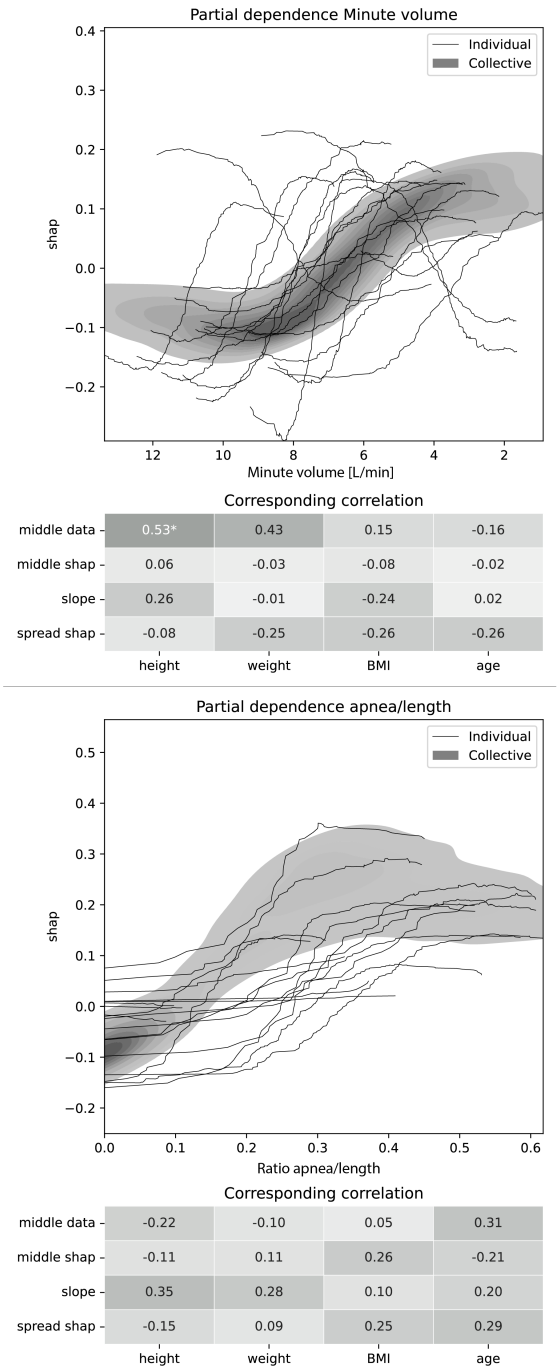


Figure 14: (A) Partial dependence plot showing the collective-clustering results (grey area) as well as the individual-clustering results for every subject (black lines indicating the mean trend). Included are the corresponding correlations between various aspects of the individual-clustering result and static features of the subjects.
*: $0.01 < p <= 0.05$.

The plots reveal that, despite variations in orientation and positioning, most subjects showed similar shapes and behaviour of the feature importance. To explore whether these variations across subjects could be attributed to static

characteristics, including height, weight, BMI, and age, correlation analyses were conducted. These analyses assessed the relationships between the static features and various aspects of the partial dependence plot, including: the middle feature value, middle SHAP value, SHAP value slope in the middle, and the range of SHAP values. Specifically, for *minute volume*, height was found to have moderate correlations of 0.53 (p-value = 0.013), respectively, indicating that taller individuals tend to have higher minute volumes. Nevertheless, after adjusting for multiple testing using Bonferroni, the correlation is no longer found to be statistically significant (0.05/16 = 0.003). For *apnea/length* no significant correlations were found. The partial dependence plots for the remaining features can be found in Appendix E. Solely *max flow ratio* showed a significant correlation of 0.65 following Bonferroni correction, indicating the relationship between the slope of the partial dependence plot and the age of the subject (p = 0.001).

Regarding *minute volume*, three subjects showed the exact opposite behaviour, having a decreasing instead of increasing SHAP value when the feature value becomes less. When excluding these aberrant subjects from the collective-clustering, the mean difference between the membership values resulting from the collective- and individual-clustering remained unchanged. Specifically, it was 0.026 (range: min 0.012 – max 0.049) before exclusion and 0.026 (range: min 0.012 – max 0.045) after exclusion.

What occurs at the time of the *complete switch*?

For this analysis, solely the individual-clustering results were utilised. The majority of these temporal plots showed a difference post-opioid, characterised by a sudden switch followed by a prolonged plateau phase. In Figure 15.A a detailed visualisation is shown, depicting the change between the epochs before and after the switch. The corresponding effect of the most significant feature of that subject, *minute volume*, was addressed by including the *minute volume* effect plot (Figure 15.B). It can be observed that the *minute volume* of the subject was 10.29 before the switch and 8.07 after, bypassing the values where a gradual change in membership value might be expected based on the effect plot. In the effect plot (Figure 15.B), the effect of the feature value on the resulting membership value appears to stagnates before 8.07 and after 10.29.

In certain subjects, a scatterplot visualisation of the epochs, with a different feature on each axis and $EtCO_2$ levels indicated through color hues, revealed a notable pattern as depicted in Figure 16. In these subjects two groups could be distinguished, of which one was associated with the lower values of $EtCO_2$ and the other with the higher values of $EtCO_2$. However, within the groups, no pattern was observed of a gradual change in $EtCO_2$ for both features. Although the clarity of this observation differed among subject, it could not be related to why some subjects showed a *complete switch* and others did not.

Differences among classification

To identify the factors influencing the variations in subject behaviour observed in the temporal plots, various test were obtained regarding the classification. Initially, it was tested whether there were class-associated factors. Therefore, the opioid involved, opioid usage status, gender, and the type of study from which the data originated were evaluated using the Chi-Squared Test, resulting in no significant differences for both collective- and individual-clustering (p-value > 0.05).

In addition, it was explored whether there was a difference in received opioid dose and the maximum reached change in $EtCO_2$ and MV relative to baseline for the collective- and individual-clustering. The results were visualised as boxplots in Figure 17. Significant difference following the Kruskal–Wallis test was found in the maximum decrease in MV among the classifications when collective clustering was obtained (p = 0.009). While for collective-clustering there was also a trend observed in the maximum raise of $EtCO_2$, it was not significant (p = 0.8). When the clustering was obtained individually, the difference observed between classification 1 and 2 during collective-clustering disappeared. All remaining tests resulted in p-values > 0.05.

For the individual clustering results, the difference in feature behaviour and importance was also considered. No difference was found in the feature important among the classifications, referring to the alternation of the most significant feature among subjects. Conversely, a variation in feature behaviour was detected. Subjects who were classified as showing *no difference* exhibited periods of typical post-opioid feature behaviour in baseline, including apneas and extended expiration duration's. On the other hand, subjects who were classified as *observed change* demonstrated post-opioid periods with feature behaviours that were

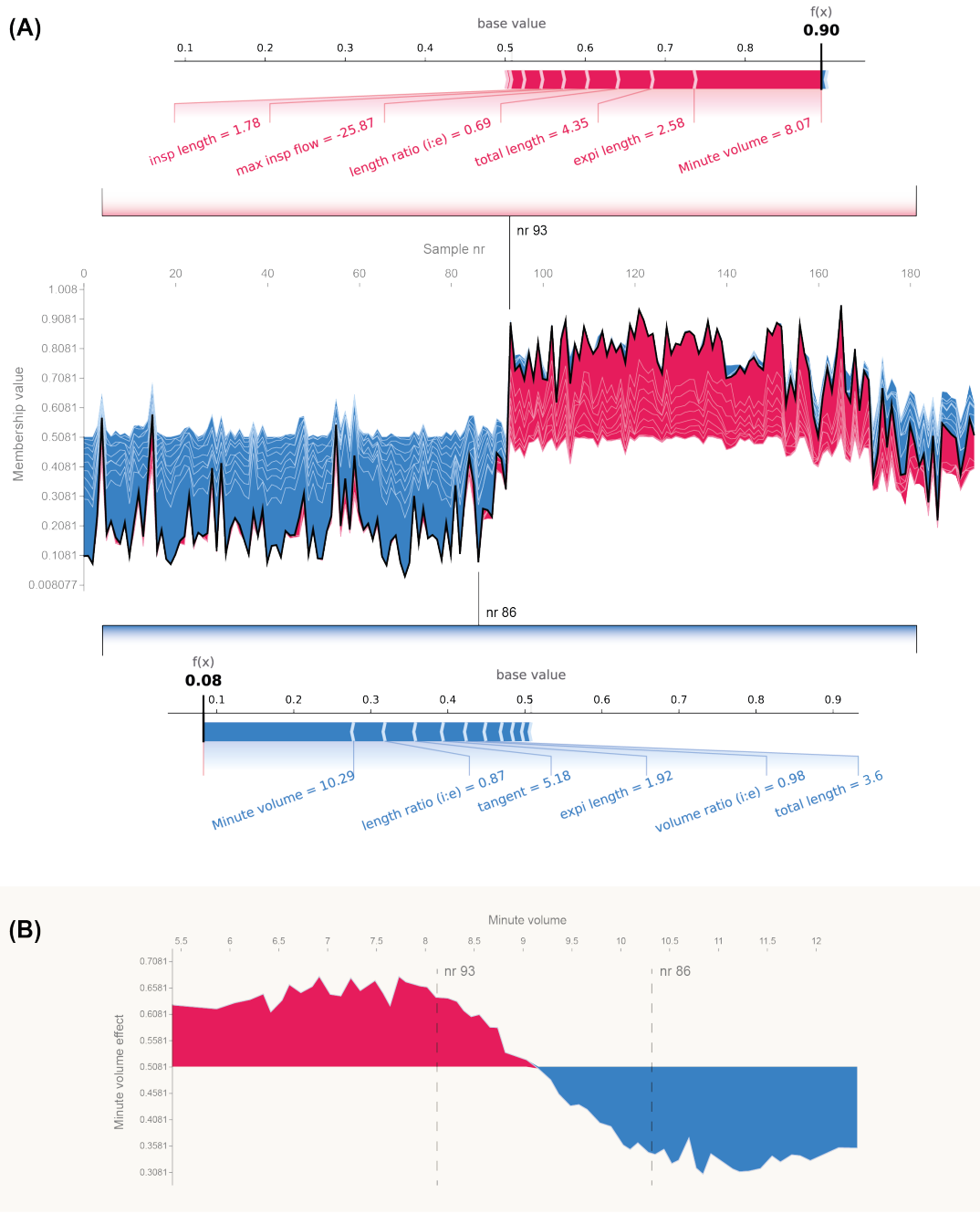


Figure 15: Example of feature behaviour of a subject identified as *complete switch*. **(A)** Detailed visualisation (SHAP value force plot) showing how feature contributions vary in the epochs before (nr 86) and after (nr 93) the switch, including specific values for key features. The arrangement of these features relative to other epochs is depicted using an additive force layout (showing the SHAP value force plot simpler for each individual epoch) organised temporally, with the black line indicating the resulting membership value. **(B)** Plot showing the positive (red) and negative (blue) effects of the *minute volume* on the resulting membership value relative to base value (0.5081). The values of the epochs before and after the switch are indicated using their corresponding numbers.

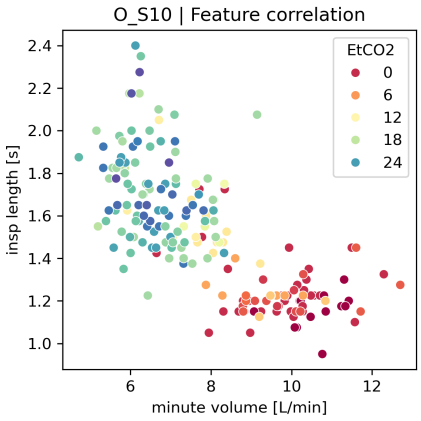


Figure 16: Scattersplot showing the relationship between two features and the corresponding $EtCO_2$ of one subject.

similar to their baseline epochs.

5 | Discussion

In this pilot study, an unsupervised fuzzy clustering model was introduced, which demonstrated consistent stability across different bootstrapping iterations. In contrast to the expected gradual transition in cluster membership, the temporal plots showed a sudden switch in membership values followed by a prolonged plateau phase. The temporal plots were divided in the following four classifications of clustering behaviour: *complete switch*, *observed change*, *reversed change* and *no difference*.

In addition, a discrepancy was observed between the collective- and individual-clustering results. Although the partial dependence plots showed the same behaviour among most subjects, there were noticeable variations in orientation and positioning. Based on the clustering results, an in-depth data-analysis was conducted by generating and answering corresponding questions. Although no direct answers to the questions were found, the following section will propose hypotheses derived from the results.

5.1. Interpretation of results

While the research was inspired by the study conducted by Sunshine and Fuller [11], their method using PCA appeared to be the least promising during the preliminary phase. Instead, FS and DTW demonstrated more promising results. While DTW showed slightly better results in the distinction regarding the change in $EtCO_2$ relative to baseline,

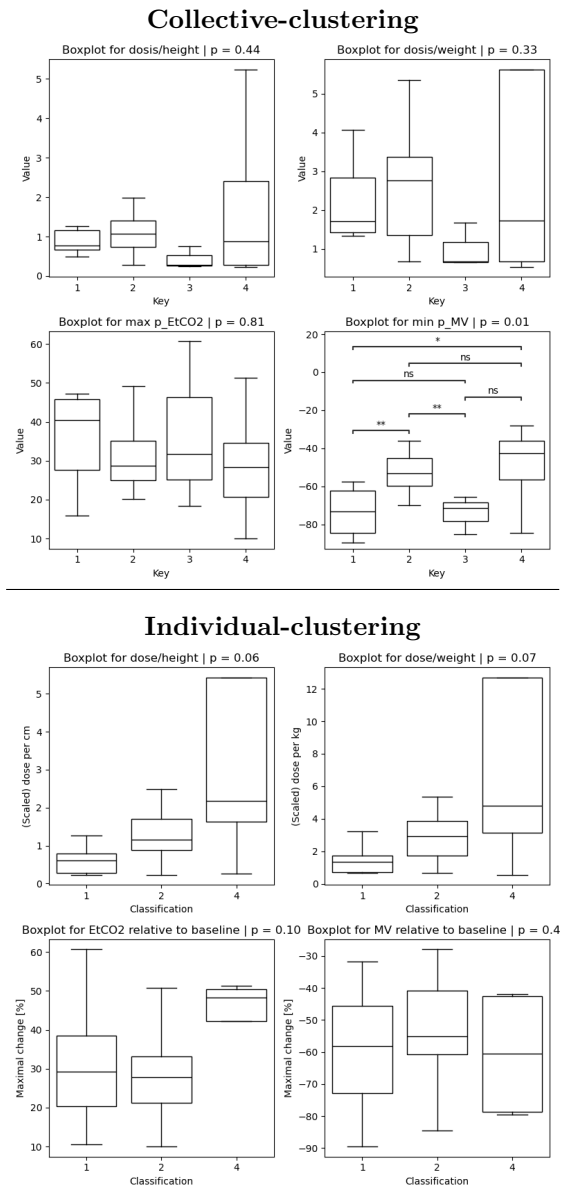


Figure 17: Boxplots visualising the difference among classifications in received opioid dose and the maximum reached change in $EtCO_2$ and MV relative to baseline for the collective- (upper) and individual-clustering (lower).
ns: $p > 0.05$.
* : $0.01 < p <= 0.05$.
** : $0.001 < p <= 0.01$.

FS had much faster computation times. The similar performance of FS and DTW indicates that high-frequency time series data used during this research may not be essential to detect OIRD, as the FS method does not rely on the high-frequency data attributes.

Fuzzy clustering was utilised as the preliminary results suggested a possible regression. The expectation was to observe a gradual transition in clus-

ter membership, indicating a progressive decline in respiratory function. However, instead an abrupt switch in membership was observed, followed by a prolonged plateau phase which remained until an antagonist was administered (*complete switch*) or alternated with abrupt switches between different prolonged plateau phases (*observed change*).

The findings of this study suggest two possible interpretations. One hypothesis is that the immediate stabilisation after the bolus administration was primarily due to the bolus itself, with the continuous opioid dosage having minimal influence. Therefore, bypassing the feature values that are associated with a gradual change in membership values, as indicated by Figure 15. Considering existing research which suggests that the respiratory depressant effects of opioids are dose-dependent, one would anticipate the possibility of observing a more gradual change under different conditions [16].

Alternatively, it is possible that there might be a threshold beyond were the breathing pattern becomes irregular and no longer exhibits a linear or gradual degradation in features. As $EtCO_2$ indicates the time and severeness of OIRD, the observed difference in Figure 16 among the groups in $EtCO_2$ but the lack of pattern within, suggests that the irregularity in respiration does not increase gradually, but rather reaches a critical point where normal patterns are disrupted. It is also possible to consider that these hypotheses are not mutually exclusive but rather complementary, where low opioid doses might cause a gradual decline in respiratory function, while high doses lead to abrupt pattern disruptions.

Observed differences among subjects in the temporal plots were divided into four different classifications. However, the question why the subjects showed different behaviours still remains. For the collective-clustering the boxplots seemed to have a slight trend regarding the maximum reached EtCO2 difference relative to the baseline, suggestion that the severity of depression may have an influence on the observed differences in classification, however it was not significant and needs further investigation. Nevertheless, the significant differences found in the MV, especially between *complete switch* and *observed differences*, would also suggest the possible association of classification in relation to the severity of depression since a greater decline in MV is associated with more respiratory depression.

On the individual-clustering level, the differences in temporal behaviour of cluster membership were also observed. SHAP value analysis showed that subjects who were classified as *observed change* had periods of breaths similar to baseline breathing during the post-opioid period. Whether these subjects had these periods because the respiratory depression was less severe or the periods of good breaths within the post-opioid period were responsible for the lower $EtCO_2$ change could not be determined, however the latter would not be an explanation for the difference in observed MV change.

The observed instances of normal breathing during the post-opioid administration phase could also be influenced by certain external experimental conditions. For example, subjects were occasionally instructed to breathe more deeply or frequently if their oxygen saturation dropped too low or if they experienced extended periods of apnea, leading to transient periods of conscious breathing. Additionally, some subjects might have fallen asleep during the experiment, potentially altering their breathing patterns.

External variables might have also affected the breathing patterns observed at baseline and immediately after opioid administration. Notably, some subjects exhibited signs of nervousness, possibly leading to a less stable baseline breathing pattern. This could explain the unusual baseline respiratory patterns observed in some healthy participants. Awareness of opioid administration among subjects might have prompted them to consciously alter their breathing, potentially skewing the data towards an unnatural representation of their normal respiratory activity. Unfortunately, these external factors were not systematically documented during the study, limiting our ability to quantify their impact on the observed breathing patterns.

What could be determined was that the subjects who showed opposite behaviour in the SHAP values for *minute volume* were also experiencing nausea post-opioid administration. The increase in minute volume could be a physiological response to nausea, possibly indicating that these subjects were experiencing agitation as a side effect of the opioids.

Inter-individual differences were noted in both the preliminary and final studies. However, Sunshine and Fuller's research with Sprague-Dawley rats did not report similar inter-subject variability [11]. This lack of reported variability in the rat studies could be attributed to several factors that differentiate human and rat research contexts. Unlike humans, rats are not cognisant of receiv-

ing opioids, eliminating any potential anxiety or emotional responses that might affect the study's outcomes. They are also unaware of the timing of opioid administration, which means that any anticipatory psychological effects influencing breathing patterns, as likely observed in the subjects, do not apply to rats.

Moreover, Sprague-Dawley rats are known for their genetic consistency, contributing to reduced variability among subjects in experimental settings [25]. This genetic uniformity may be a factor in why lesser inter-subject differences were observed in their animal study compared to our study that used human data, where genetic and individual variability is broader.

The inter-subject variability resulted in discrepancy between the collective-clustering and individual-clustering results. As can be observed from the partial dependence plot (Figure 14), some subject fall outside the distribution resulted from the collective-clustering. However, removing the aberrant subjects did not alter the observed difference in assigned membership values between collective-clustering and individual-clustering.

The partial dependence plot showed significant correlation between the height of the subject and the SHAP value behaviour of *minute ventilation*, suggesting that the minute ventilation is higher if a subject is taller. The male subjects in this study were also significantly larger than the women. According to literature, men and women have different tidal volumes of 0.5 and 0.4 liter, respectively. As both breath approximately 12 breaths per minute, the minute volume in male subjects appear to be slightly larger [26].

5.2. Strengths and limitations of the study

The data obtained from the LUMC for this study is particularly distinctive. With information collected from healthy individuals using solely opioids under controlled conditions, it provides a valuable opportunity to examine the effects of opioids. While $EtCO_2$ and saturation are secondary parameters of the breathing quality, which may be influenced by other factors, the obtained flow data enables direct evaluation of respiratory performance. In combination with the high frequency time series and the negligible amount of leakage, in detail assessment of the breathing pattern has been made possible.

However, the use of this dataset also presents challenges. For example, such methods of data collection are not common in clinical settings and are not always feasible. Nevertheless, while the neglectful leakage deemed to be important as *minute volume* appeared to be the first or second most important feature in each subject, this study suggests that extensive data collection at high frequencies may not be essential, as the features used doesn't depend on the high frequency aspect of data.

Furthermore, due to the unique nature of the data, the dataset was relatively limited in size. Nonetheless bootstrapping showed stable clustering results when the clustering was obtained collectively, suggesting that the results were not heavily influenced by one subject. However, the sample size was likely insufficient to observe significant differences between the different classifications. While some test yielded significant p-values after considering the per-test family-wise error rate, corrections were not applied across all tests and should be interpreted with precaution.

Another disadvantage of the dataset was the discrepancy between the healthy subjects and the opioid users. While the healthy user were mainly young students, the opioid users were older adults. This can also be observed in the age difference between the two studies ORNAC (23.5 ± 1.8) and ROAR (39.7 ± 15.1), as only the ROAR included opioid users. However, it seemed that these differences did not influenced the study results.

A common problem in signal analysis and also recurring in this study is data quality. While the data contained less noise than clinic data, as it was ensured in a controlled setting and subjects were instructed to lay still, noise was not uncommon. Noise was mainly induced by talking and moving of the subject, as well as detachment of the device. The detecting algorithm used for artefact detection was not perfect. However, as for this pilot study optimising noise detection and filtering was not the priority due to time limitations, the choice was made to manually check and the remaining epochs and exclude the ones that showed to much noise for the data used for model development. While this ensured that the pilot study would not be unsuccessful due to insufficient data quality, it should be kept in mind that this manually checking was obtained by one researcher only and the absence of human error can therefore not be guaranteed.

In addition to noise, external factors also reduced the data authenticity. Additionally to the instructions given by the researcher and the phenomenon of the subject falling asleep mentioned before, another aspect was that some subject received

additional oxygen while others did not. While the saturation is an important feature of the respiratory depression in a majority of studies, it could not be used in this study to relate clustering results to the expected real-world situation. Therefore, it was decided to solely use $EtCO_2$ and MV.

5.3. Future recommendations

The first finding of this research that would require more research is the observation of the stagnation after the *complete switch*. It would be interesting to investigate whether the switch would be more transitional with lower doses of opioids. Therefore, a study should be obtained were no or lower boluses are given.

In addition, while the individual clustering method showed good results, it cannot be applied in practise and is solely for post-analysis as data of the subject is necessary to develop the model. Nonetheless, the collective clustering model of this study showed *no difference* in a large amount of the subjects, probably due to the inter-subject variability. However, the observation that the feature importance showed the same behaviour between subjects when clustering was obtained individually, however shifted in place and orientation, suggests there may be a possibility to allow correction by static features. Therefore, resulting a more generalisable model that is developed collectively and can by applied on new patients for live-analysis during clinical practise, potentially enabling the model to help prevent medical incidents.

Therefore, an expected follow-up to this study would be to test the method with more subjects. However, getting this kind of data in larger numbers would be difficult. Yet, as this study suggest that the high frequency data is not necessary, it would be interesting to test this approach with other datasets, such as capnography, which directly monitors the concentration or partial pressure of CO_2 inhaled and exhaled, serving as an indirect indicator of CO_2 partial pressure in arterial blood. The downside is that not all features used during this study can be retrieved from the capnography waveform. However, as obtaining capnography waveforms is more common practice there are already datasets available to test this approach. This way, a larger dataset would be available and the resulted model would be more applicable in current practise.

Although the FS approach showed promising results, the preliminary research showed slightly more preference towards the DTW approach. Nonetheless, while an effort has been made during this research to implement DTW, it was unsuccessfully and showed solely membership values of 0.5 regarding both clusters. However, it is known that using Fuzzy C-means with high dimensionality datasets, the majority of cluster centers may be pulled into the overall center of gravity. With an average breath having a length around 5 seconds and the dataset having a frequency of 40Hz, the dimension of the used time-series was quite high [27].

While the study did not delve into the predictions for the periods following naloxone administration, since it was not the primary focus, it would be interesting to investigate these results further. Given that the temporal plots exhibited different behaviours, it would be valuable to understand why these differences occur. As naloxone has a short half-life, the *temporary recovery* may indicate the return of respiratory depression and the need for repeating administration [21, 22]. Moreover, it is noteworthy that two subjects exhibited no recovery in the temporal plots yet demonstrated improvement in the $EtCO_2$ levels. This raises questions about what these observations signify.

6 | Conclusion

In this study, a comprehensive analysis of OIRD was performed using unsupervised machine-learning techniques. A fuzzy clustering model was implemented, incorporating SHAP value analysis to enhance the interpretability of the clustering results. Although the model successfully identified changes in respiratory flow patterns associated with OIRD and subsequent recovery after naloxone administration, it requires further refinement.

Moreover, while the questions regarding the cluster behaviour could not be directly addressed, the results introduced potential directions for future research, which included: exploring the possibility for dose-dependent effects, correcting for inter-subject variability using static features and trying to use an alternative data-set. These steps may contribute to a model which would be more generalisable among subjects and applicable in current practise.

7 | References

[1] Jocelynn L. Cook. "The opioid epidemic". In: *Best Practice amp; Research Clinical Obstetrics amp; Gynaecology* 85 (2022), pp. 53–58. DOI: 10.1016/j.bpobgyn.2022.07.003.

[2] *Understanding the Opioid Overdose Epidemic*. Jan. 2024. URL: https://www.cdc.gov/drugoverdose/.

[3] Jordan T. Bateman, Sandy E. Saunders, and Erica S. Levitt. "Understanding and countering opioid-induced respiratory depression". In: *British Journal of Pharmacology* 180.7 (2021), pp. 813–828. DOI: 10.1111/bph.15580.

[4] Jan-Marino Ramirez et al. "Neuronal mechanisms underlying opioid-induced respiratory depression: Our current understanding". In: *Journal of Neurophysiology* 125.5 (2021), pp. 1899–1919. DOI: 10.1152/jn.00017.2021.

[5] Lorri A. Lee et al. "Postoperative opioid-induced respiratory depression: A closed Claims Analysis". In: *Survey of Anesthesiology* 59.5 (2015), p. 250. DOI: 10.1097/sa.0000000000000172.

[6] Ashish K. Khanna et al. "Opioid-induced respiratory depression increases hospital costs and length of stay in patients recovering on the General Care Floor". In: *BMC Anesthesiology* 21.1 (2021). DOI: 10.1186/s12871-021-01307-8.

[7] Tong J. Gan et al. "Incidence, patient satisfaction, and perceptions of post-surgical pain: Results from a US national survey". In: *Current Medical Research and Opinion* 30.1 (2013), pp. 149–160. DOI: 10.1185/03007995.2013.860019.

[8] Shenhab Zaig, Carolina Scarpellini, and Gaspard Montandon. "Respiratory depression and analgesia by opioid drugs in freely-behaving larval zebrafish". In: *eLife* (2021). DOI: 10.1101/2020.09.30.320267.

[9] Sabry Ayad et al. "Characterisation and monitoring of postoperative respiratory depression: Current approaches and future considerations". In: *British Journal of Anaesthesia* 123.3 (2019), pp. 378–391. DOI: 10.1016/j.bja.2019.05.044.

[10] Bruce Friedman et al. "Identifying and monitoring respiratory compromise: Report from the Rules and Algorithms Working Group". In: *Biomedical Instrumentation amp;amp; Technology* 53.2 (2019), pp. 110–123. DOI: 10.2345/0899-8205-53.2.110.

[11] Sunshine M.D. and Fuller D.D. "Automated Classification of Whole Body Plethysmography Waveforms to Quantify Breathing Patterns". English. In: *Frontiers in Physiology* 12.(Sunshine) Rehabilitation Science Ph.D. Program, University of Florida, Gainesville, FL, United States (2021). Place: Switzerland Publisher: Frontiers Media S.A., p. 690265. ISSN: 1664-042X (electronic). DOI: 10.3389/fphys.2021.690265. URL: http://www.frontiersin.org/Physiology/archive/.

[12] Ermer S.C. et al. "An Automated Algorithm Incorporating Poincare Analysis Can Quantify the Severity of Opioid-Induced Ataxic Breathing". English. In: *Anesthesia and Analgesia* 130.5 (2020). Place: United States Publisher: Lippincott Williams and Wilkins, pp. 1147–1156. ISSN: 0003-2999. DOI: 10.1213/ANE.0000000000004498.

[13] Jungquist C.R. et al. "Identifying Patients Experiencing Opioid-Induced Respiratory Depression During Recovery From Anesthesia: The Application of Electronic Monitoring Devices". English. In: *Worldviews on evidence-based nursing* 16.3 (2019). Place: United States Publisher: NLM (Medline), pp. 186–194. ISSN: 1741-6787 (electronic). DOI: 10.1111/wvn.12362.

[14] Anh Pham. "Detecting Apnea Events Based on Pulse Oximetry's Red Photoplethysmography Waveform". In: *Utah Space Grant Consortium* (May 2018).

[15] Esmaeili N. et al. "Tracheal sound analysis for automatic detection of respiratory depression in adult patients during cataract surgery under sedation". English. In: *Journal of Medical Signals and Sensors* 8.3 (2018). Place: India Publisher: Isfahan University of Medical Sciences(IUMS) (Hezar Jerib Avenue, P.O. Box 81745-319, Isfahan, Iran, Islamic Republic of), pp. 140–146. ISSN: 2228-7477 (electronic). DOI: 10.4103/jmss.JMSS-67-16.

[16] Barbara Palkovic et al. "Multi-level regulation of opioid-induced respiratory depression". In: *Physiology* 35.6 (2020), pp. 391–404. DOI: 10.1152/physiol.00015.2020.

[17] Lynn R. Webster and Suzanne Karan. "The physiology and maintenance of respiration: A narrative review". In: *Pain and Therapy* 9.2 (Oct. 2020), pp. 467–486. DOI: 10.1007/s40122-020-00203-2.

[18] K.T.S. Pattinson. "Opioids and the control of respiration". In: *British Journal of Anaesthesia* 100.6 (June 2008), pp. 747–758. DOI: 10.1093/bja/aen094.

[19] Shivani Patel. *Respiratory acidosis*. June 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK482430/#:~:text=Respiratory%20acidosis%20typically%20occurs%20due,the%20pH%20of%20the%20blood..

[20] Eveline LA van Dorp, Ashraf Yassen, and Albert Dahan. "Naloxone treatment in opioid addiction: The risks and benefits". In: *Expert Opinion on Drug Safety* 6.2 (Mar. 2007), pp. 125–132. DOI: 10.1517/14740338.6.2.125.

[21] William R. Martin. "Drugs five years later: Naloxone". In: *Annals of Internal Medicine* 85.6 (Dec. 1976), p. 765. DOI: 10.7326/0003-4819-85-6-765.

[22] Jason M. White and Rodney J. Irvine. "Mechanisms of fatal opioid overdose". In: *Addiction* 94.7 (July 1999), pp. 961–972. DOI: 10.1046/j.1360-0443.1999.9479612.x.

[23] *Technology to the respective measuring principles*. URL: https://sensirion.com/products/technology/.

[24] Scott M. Lundberg et al. "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery". In: *Nature Biomedical Engineering* 2.10 (Oct. 2018), pp. 749–760. DOI: 10.1038/s41551-018-0304-0.

[25] Open Access Pub. *Sprague-Dawley rats: Journal of Current Scientific Research*. URL: https://openaccesspub.org/current-scientific-research/sprague-dawley-rats#:~:text=One%20of%20the%20primary%20advantages,prolonged%20animal%20contact%20is%20required..

[26] Joachim D Pleil et al. "The physics of human breathing: Flow, timing, volume, and pressure parameters for normal, on-demand, and ventilator respiration". In: *Journal of Breath Research* 15.4 (Sept. 2021), p. 042002. DOI: 10.1088/1752-7163/ac2589.

[27] Roland Winkler, Frank Klawonn, and Rudolf Kruse. "Fuzzy c-means in high dimensional spaces". In: *International Journal of Fuzzy System Applications* 1.1 (Jan. 2011), pp. 1–16. DOI: 10.4018/ijfsa.2011010101.

APPENDICES

This page intentionally left blank

A | Characteristics of the studies used.

A.1. ROAR

Reversal of opioid-induced respiratory depression with opioid antagonists

Opioids used	Fentanyl & sufentanil
Opioid administration	Bolus and continuous intravenous
Target decrease in respiration	40-60%
Intervention	Naloxone
Characteristics subjects	Opioid naïve individuals and chronic opioid users

Infusion of fentanyl and sufentanil whilst measuring minute ventilation and pupil diameter. Intranasal naloxone (IN, 4 mg) will be administered when ventilation has dropped by 40-60% (Saturation $>$ 85%) and repeated after 180 minutes. At the end of each experiment, 240min after first dose of naloxone, 0.4 mg naloxone will be administered intravenously to determine its effect on ventilation and to allow calculation of naloxone intranasal bioavailability.

A.2. ORNAC

L-NAC for reversal of opioid-induced respiratory depression

Opioids used	Fentanyl
Opioid administration	continuous intravenous
Target decrease in respiration	40%
Intervention	L-NAC (fluimicil) & placebo
Characteristics subjects	Healthy volunteers, aged 18-65 years of either sex and with a body mass index of 19-30 kg/m^2

Individualized intravenous fentanyl infusion will be initiated aimed at 40% respiratory depression compared to baseline. After 40% respiratory depression is attained, the L-NAC or placebo infusion will start over 1 h with dose 75 mg/kg. A second administration of L-NAC or placebo will be administered over hour the next hour with dose 150 mg/kg. The L-NAC dose may be adapted based on the results observed in previous subjects (max. increase with a factor of 2). Three hours after the first infusion of L-NAC or placebo, the experiment will end, and all infusions will be terminated. Measurements made are: minute ventilation on a breath-to-breath basis through a facemask for 3 hours, end-tidal carbon dioxide partial pressure, respiratory frequency, tidal volume, oxygen saturation (all obtained on a breath-to-breath basis), arterial blood gas analysis: pH, pO2, pCO2, oxygen saturation (obtained at 15 min intervals), blood pressure by cuff (at 30 min interval) and plasma concentrations of fentanyl and L-NAC (at regular intervals). Total volume of blood drawn is 125 mL.

B | Breath selection

B.1. Detected breaths

Table 3: Breath count detected by the algorithm per subject, divided in epochs detected during baseline, post-opioid and post-antagonist.Subjects reported twice have done both measurement days. The subjects highlighted in grey were excluded for further analysis, since they had too few epochs detected in baseline or post-opioid.

	Baseline		**Post-opioid**		**Post-antagonist**	
Subject	Duration	Breaths	Duration	Breaths	Duration	Breaths
R_S05	00:07:29	46	00:40:00	294	00:44:17	449
R_S06	01:25:42	556	00:07:00	0	00:00:00	0
R_S07	00:16:10	244	01:18:00	680	00:58:08	762
R_S08	00:04:43	76	00:42:00	400	02:09:44	1640
R_S09	00:50:23	614	00:50:00	356	00:00:00	0
R_S10	00:17:12	256	00:37:00	286	03:09:10	2108
R_S11	00:18:47	359	01:15:00	184	00:00:00	0
R_S12	00:18:32	298	00:41:00	322	04:34:48	2916
R_S13	00:31:19	399	00:20:00	200	04:51:44	3851
R_S14	00:32:26	416	00:22:00	190	04:46:18	3992
R_S15	00:25:47	309	00:22:00	164	04:38:35	3209
R_S15	00:14:11	185	00:25:00	152	01:06:43	648
R_S16	00:13:56	187	00:11:00	67	04:33:50	1895
R_S18	00:17:36	321	00:20:00	286	04:30:27	4549
R_S18	00:14:29	260	00:23:00	198	01:16:23	1200
R_S19	00:58:20	596	00:29:00	116	01:03:59	366
R_S20	00:38:16	265	00:33:00	270	02:55:36	1496
R_S20	00:19:43	112	00:38:00	274	04:35:51	3213
R_S21	00:06:44	45	00:30:00	223	02:10:18	1375
R_S21	00:17:55	234	00:30:00	203	00:37:06	525
R_S22	00:17:05	263	00:30:00	325	04:35:02	3385
R_S22	00:21:44	431	00:36:00	274	04:49:54	2827
R_S23	00:47:47	312	00:28:00	128	04:31:02	2715
R_S23	00:17:33	181	00:24:00	151	04:30:49	2780
R_S24	00:13:58	151	00:30:00	191	00:35:09	216
R_S25	00:22:50	166	00:39:00	277	01:13:57	716
R_S26	00:21:00	161	00:53:00	322	01:16:34	635
R_S27	00:36:30	283	00:54:00	366	03:01:20	1687
R_S27	00:14:58	101	00:31:00	188	04:32:31	2371
R_S28	01:01:26	334	00:05:00	3	00:54:40	342
R_S29	00:11:13	115	00:31:00	179	04:29:54	3034
R_S30	00:18:11	191	00:30:00	158	00:41:21	143
R_S31	00:15:40	164	00:33:00	221	04:29:43	3260
R_S32	00:15:55	165	00:30:00	317	03:39:31	2542
R_S32	00:23:52	292	00:30:00	221	00:38:04	179
R_S33	00:19:59	166	00:41:00	207	02:35:59	139
R_S34	00:11:10	130	00:35:00	318	00:37:48	23
R_S35	00:05:33	57	00:33:00	177	00:48:35	311
R_S36	00:08:50	39	00:30:00	110	04:45:49	2917
R_S36	00:13:19	107	00:31:00	196	05:02:40	3340
O_S01	00:18:30	327	00:33:00	408	02:27:29	1180
O_S01	00:07:18	118	00:30:00	306	02:30:52	1198
O_S02	00:27:12	414	00:30:00	198	02:30:57	705
O_S02	00:16:11	255	00:30:00	176	03:13:39	949
O_S03	00:15:32	205	00:30:00	277	02:45:51	1357
O_S04	00:06:36	70	00:30:00	377	03:20:37	1374

O_S04	00:19:58	299	00:30:00	329	03:44:55	1889
O_S05	00:14:42	247	00:30:00	421	03:00:59	2256
O_S05	01:14:56	1068	00:19:00	221	02:38:42	1875
O_S06	00:17:11	161	00:30:00	254	02:38:50	1338
O_S06	00:07:48	109	00:30:00	303	03:01:29	1213
O_S07	00:10:00	122	00:30:00	252	02:13:14	1288
O_S07	00:15:02	217	00:30:00	236	03:00:15	1584
O_S08	00:12:47	195	00:30:00	292	02:35:09	1275
O_S08	00:26:47	394	00:30:00	352	02:39:41	1474
O_S09	00:10:55	198	00:30:00	298	02:50:03	1892
O_S10	00:09:49	173	00:30:00	372	02:59:48	2640
O_S10	00:07:57	136	00:20:00	308	03:12:45	2874
O_S11	00:09:01	93	00:30:00	317	01:40:59	1006
O_S11	00:08:50	110	00:30:00	314	02:34:09	1746

B.2. Included breaths

Table 4: Overview of the included epochs after manual selection was obtained. The subjects highlighted in grey were excluded for further analysis, since they had too few epochs detected in baseline or post-opioid.

Subject	Baseline	Opioid
R_S07	100	300
R_S08	70	277
R_S09	100	300
R_S10	100	222
R_S11	100	172
R_S12	100	287
R_S13	100	144
R_S14	100	139
R_S15	100	116
R_S15	100	116
R_S18	100	251
R_S18	100	160
R_S19	100	93
R_S20	100	217
R_S20	76	186
R_S21	100	153
R_S22	100	283
R_S22	100	195
R_S23	100	87
R_S23	100	115
R_S24	100	153
R_S25	100	230
R_S26	100	179
R_S27	100	162
R_S27	63	87
R_S29	100	150
R_S30	100	134
R_S31	100	161
R_S32	100	219
R_S32	100	121
R_S33	100	131
R_S34	100	234
R_S35	30	88
R_S36	83	147

Subject	Baseline	Opioid
O_S01	100	300
O_S01	100	265
O_S02	100	118
O_S02	100	126
O_S03	100	211
O_S04	63	300
O_S04	100	300
O_S05	100	300
O_S05	100	172
O_S06	100	237
O_S06	84	279
O_S07	100	237
O_S07	100	219
O_S08	100	242
O_S08	100	295
O_S09	100	270
O_S10	100	300
O_S10	100	284
O_S11	79	264
O_S11	88	238

C | Data exploration

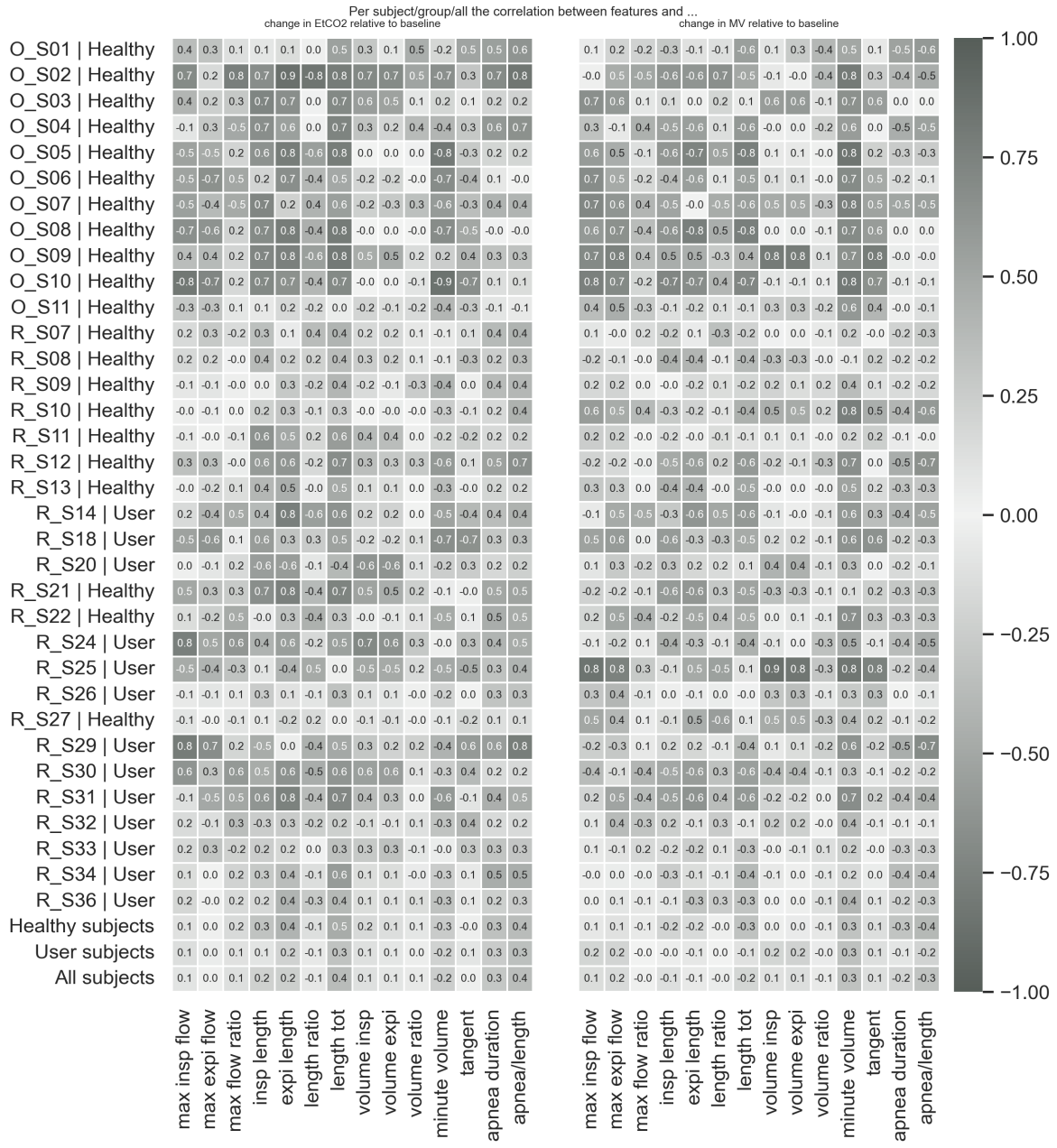


Figure 18: Correlation between features and EtCO2 (left) and Minute Volume (right) change relative to baseline for every individual subject, as well as collectively all subjects, all healthy subjects and all opioid users.

D | Preliminary data-analysis

D.1. Dendrograms

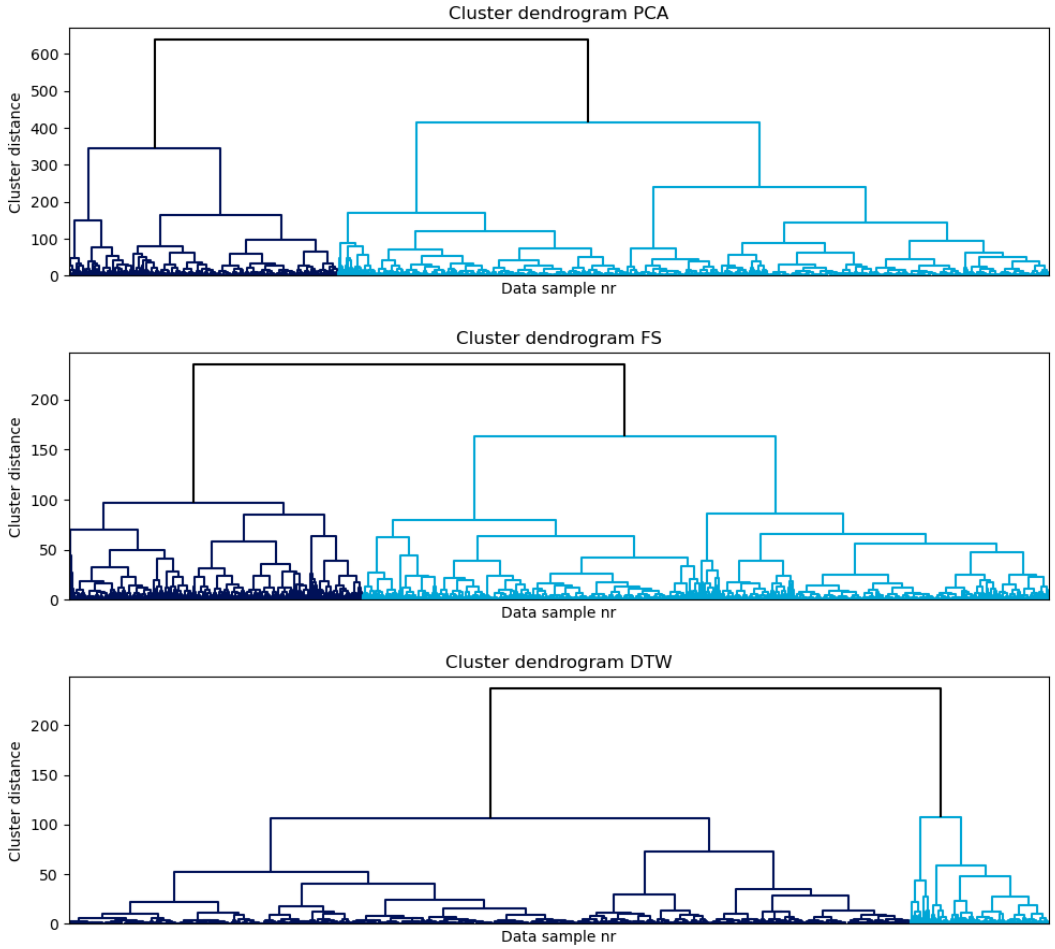


Figure 19: Dendrogram depicting the resulted linkage matrix of the hierarchical clustering for each approach.

D.2. Overview clustering

Table 5: Overview of the cluster count for each subject as well as each approach.

	PCA				FS				DTW			
	Cluster				Cluster				Cluster			
Subject	1	2	3	4	1	2	3	4	1	2	3	4
R_S07	103	29	16	47	37	77	23	58	98	62	32	3
R_S08	122	25	6	42	94	70	16	15	147	37	11	0
R_S09	81	72	4	38	79	38	46	32	106	69	20	0
R_S10	121	37	17	20	24	68	6	97	87	28	78	2
R_S11	111	69	1	14	177	8	8	2	192	2	1	0
R_S12	137	41	0	17	101	7	45	42	88	43	53	11
R_S13	48	51	24	72	34	147	4	10	97	97	1	0
R_S14	67	90	3	35	105	5	71	14	95	85	15	0
R_S18	131	57	0	7	101	63	9	22	176	18	1	0
R_S20	24	23	85	63	26	153	4	12	156	28	10	1
R_S21	16	90	33	56	3	140	10	42	135	41	17	2
R_S22	125	38	7	25	102	54	32	7	154	40	1	0
R_S24	40	100	6	49	50	52	49	44	24	114	49	8
R_S25	15	51	9	120	25	103	6	61	35	131	28	1
R_S26	42	56	14	83	43	77	19	56	56	98	37	4
R_S27	42	34	44	75	16	29	75	75	29	105	58	3
R_S29	43	62	19	71	14	71	28	82	77	64	46	8
R_S30	125	52	5	13	29	146	0	20	8	93	59	35
R_S31	21	41	54	79	32	79	29	55	117	62	15	1
R_S32	179	12	0	4	167	28	0	0	31	146	16	2
R_S33	40	106	1	48	106	18	25	46	41	92	33	29
R_S34	117	52	1	25	117	18	55	5	137	44	13	1
R_S36	70	82	4	39	55	57	38	45	53	113	27	2
O_S01	180	15	0	0	105	25	33	32	134	29	32	0
O_S02	61	97	2	35	90	36	9	60	113	23	49	10
O_S03	81	51	4	59	5	167	1	22	36	140	19	0
O_S04	115	69	0	11	103	16	44	32	138	32	25	0
O_S05	103	64	0	28	138	36	18	3	190	5	0	0
O_S06	15	27	70	83	49	63	65	18	129	63	3	0
O_S07	113	68	2	12	18	13	55	109	68	69	57	1
O_S08	95	42	1	57	118	75	2	0	34	161	0	0
O_S09	116	74	0	5	25	141	2	27	145	47	3	0
O_S10	125	47	0	23	99	94	1	1	119	76	0	0
O_S11	32	126	2	35	66	114	10	5	187	7	1	0

D.3. Clustering stability

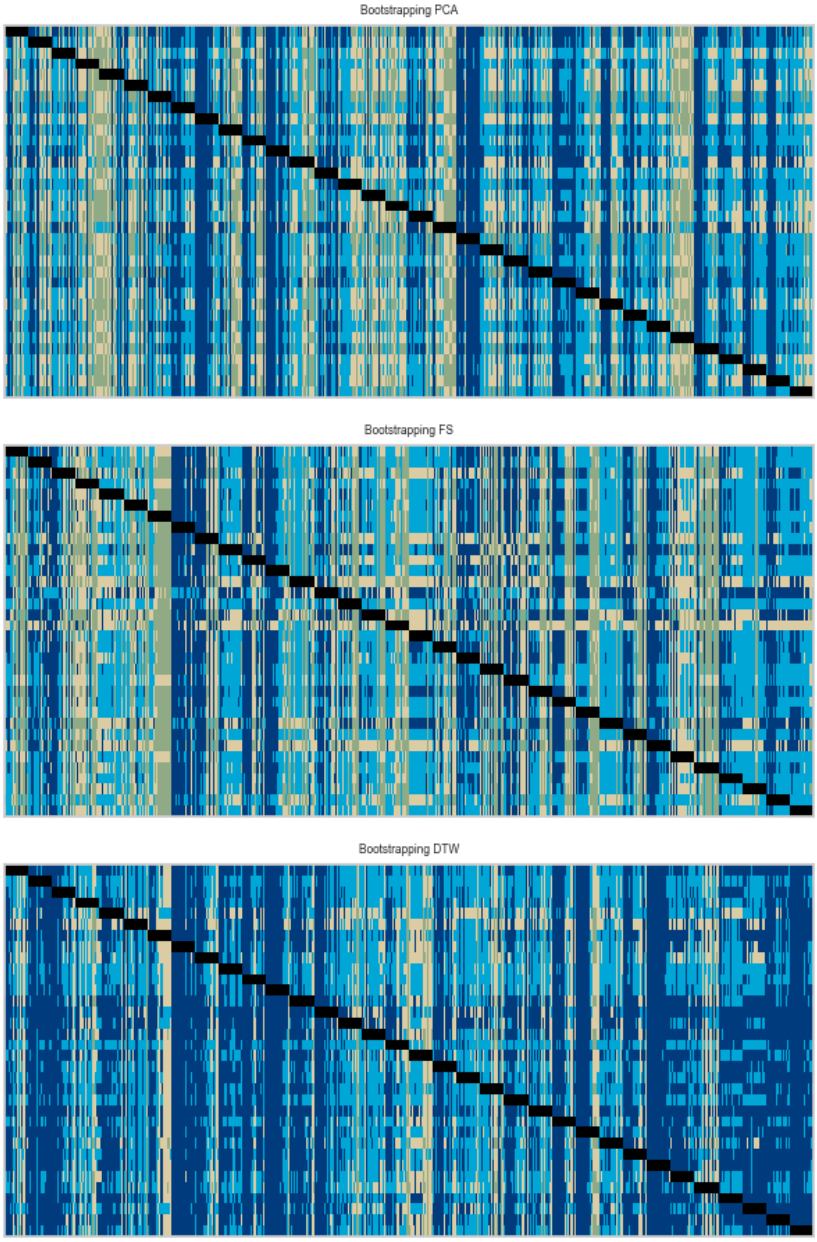


Figure 20: Bootstrapping results for each approach. Every row depicts an iteration and each row an epoch. Each color indicates a different color, which correspond to the colors used in Figure 10. The blacked-out eras are those of the excluded subject.

E | Main analysis

E.1. Matrix Kernel Density Estimate plot

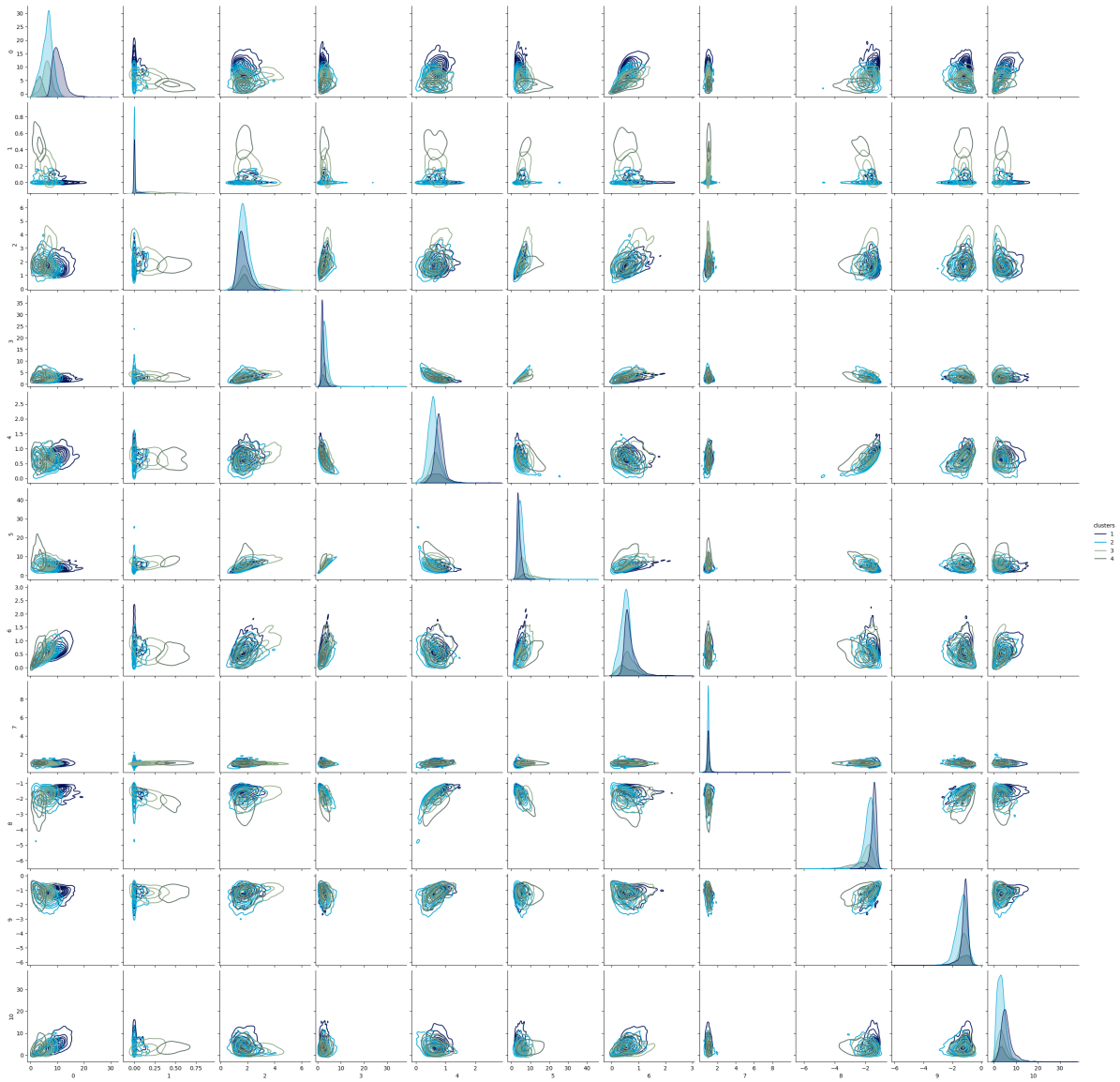
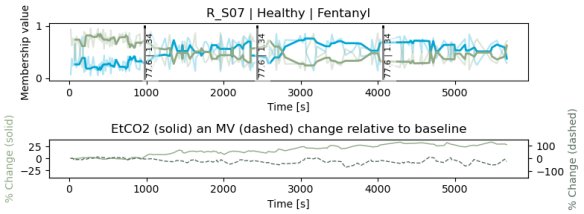
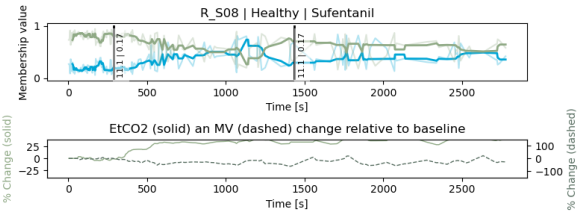


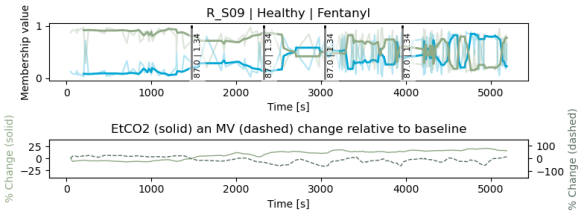
Figure 21: Example demonstrating a matrix Kernel Density Estimate plot among all utilised features. The epochs are coloured according to their assigned cluster, as determined by the FS approach.

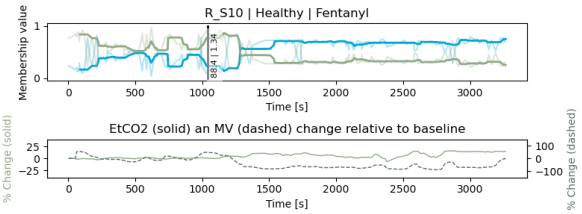
E.2. Clustering results - temporal plots

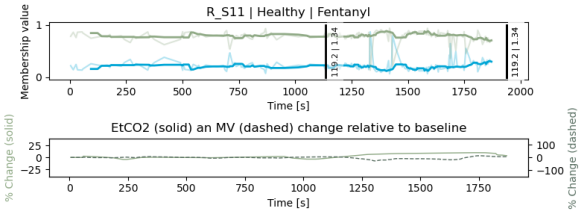
E.2.1. Collective-clustering

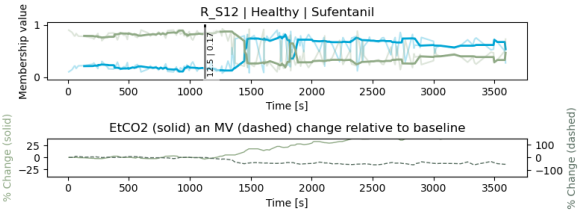


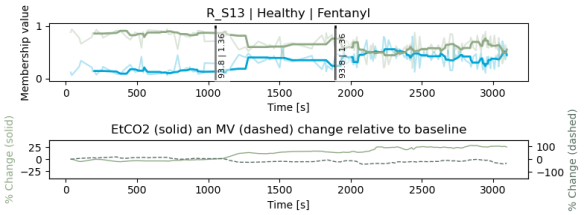


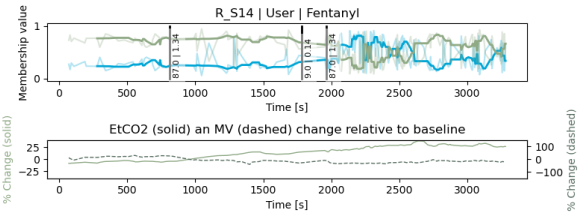


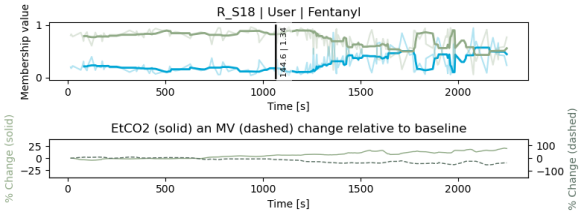


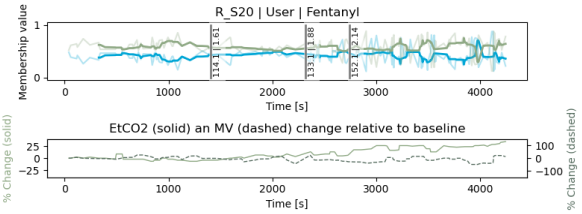


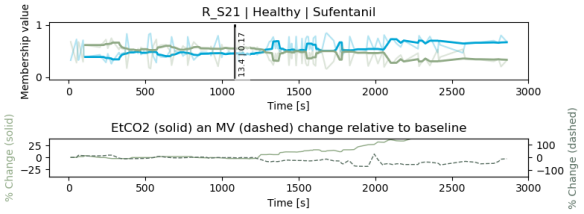


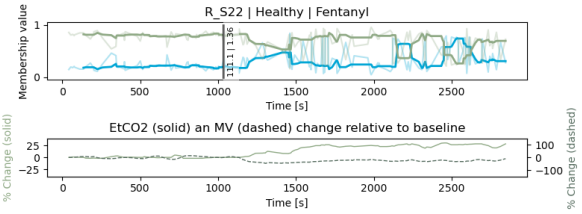


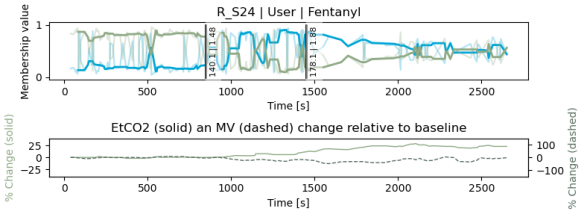


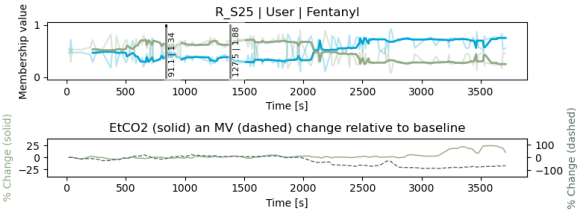


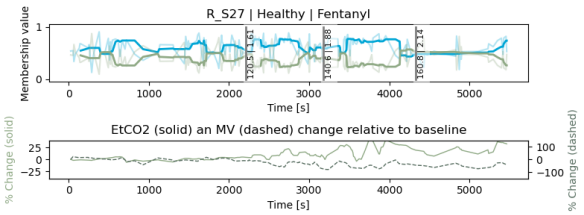


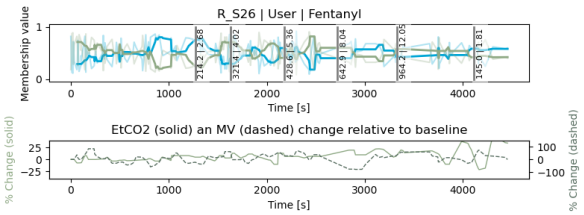


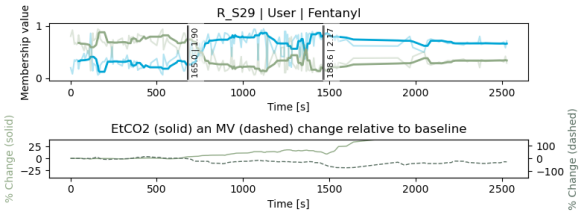


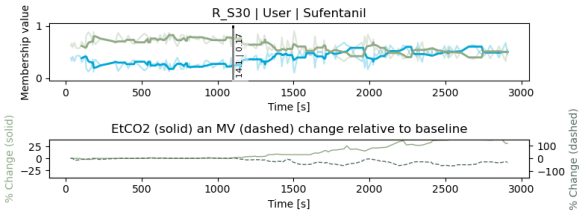


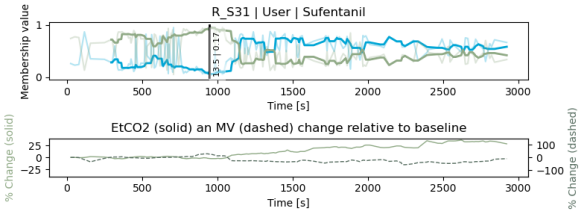


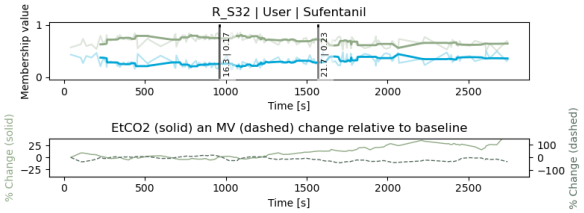

R_S27 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

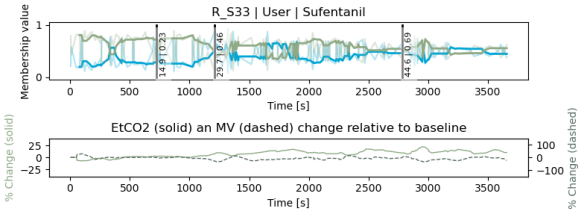

R_S26 | User | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

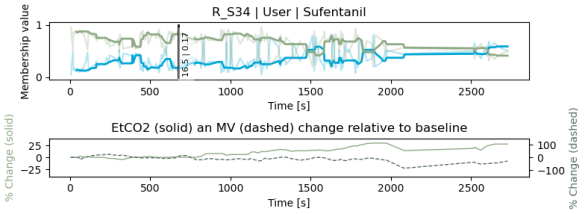

R_S29 | User | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

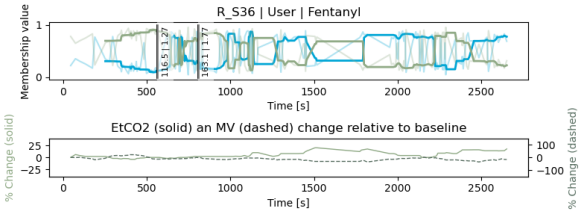

R_S30 | User | Sufentanil
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

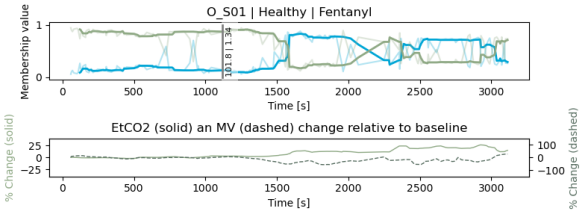

R_S31 | User | Sufentanil
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

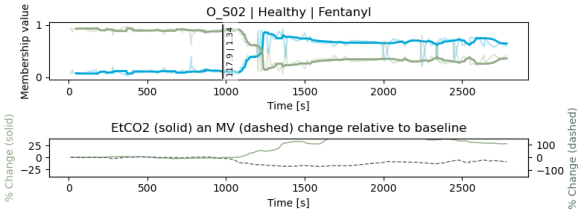

R_S32 | User | Sufentanil
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

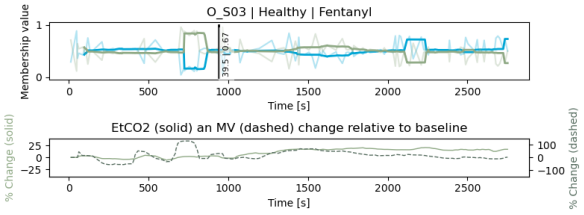

R_S33 | User | Sufentanil
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

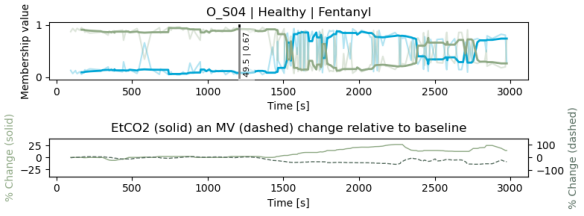

R_S34 | User | Sufentanil
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

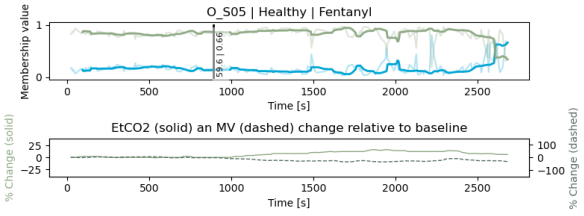

R_S36 | User | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)


O_S01 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)


O_S02 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)


O_S03 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)


O_S04 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)


O_S05 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)

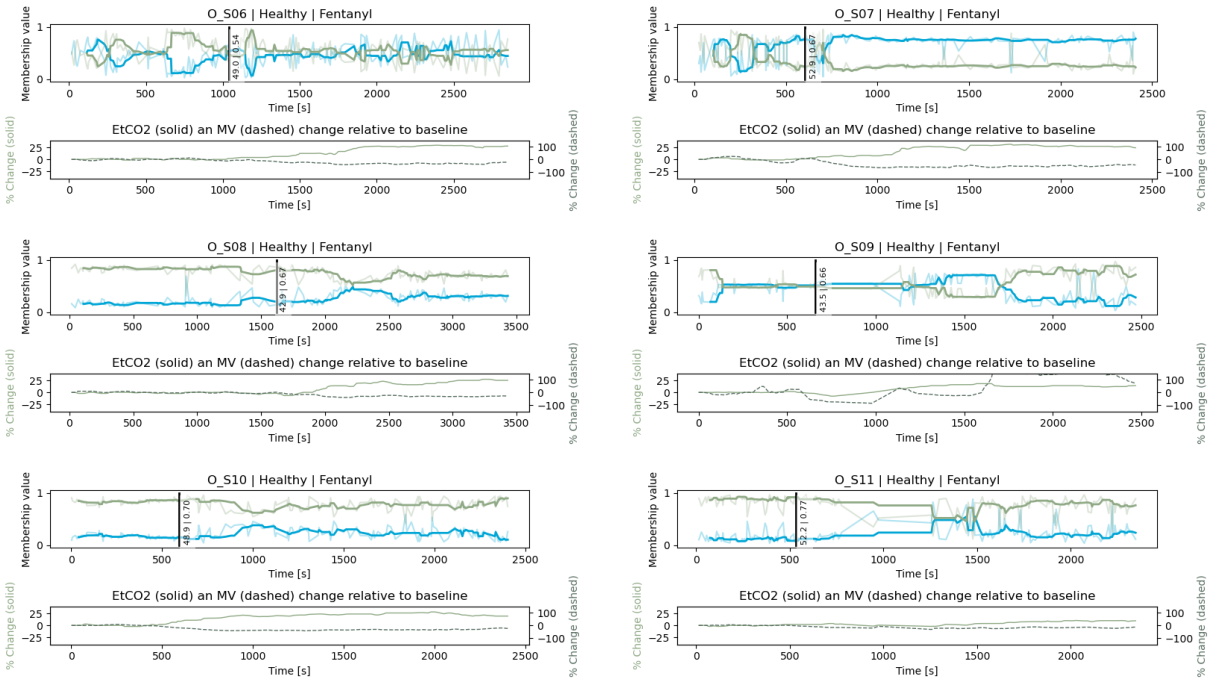
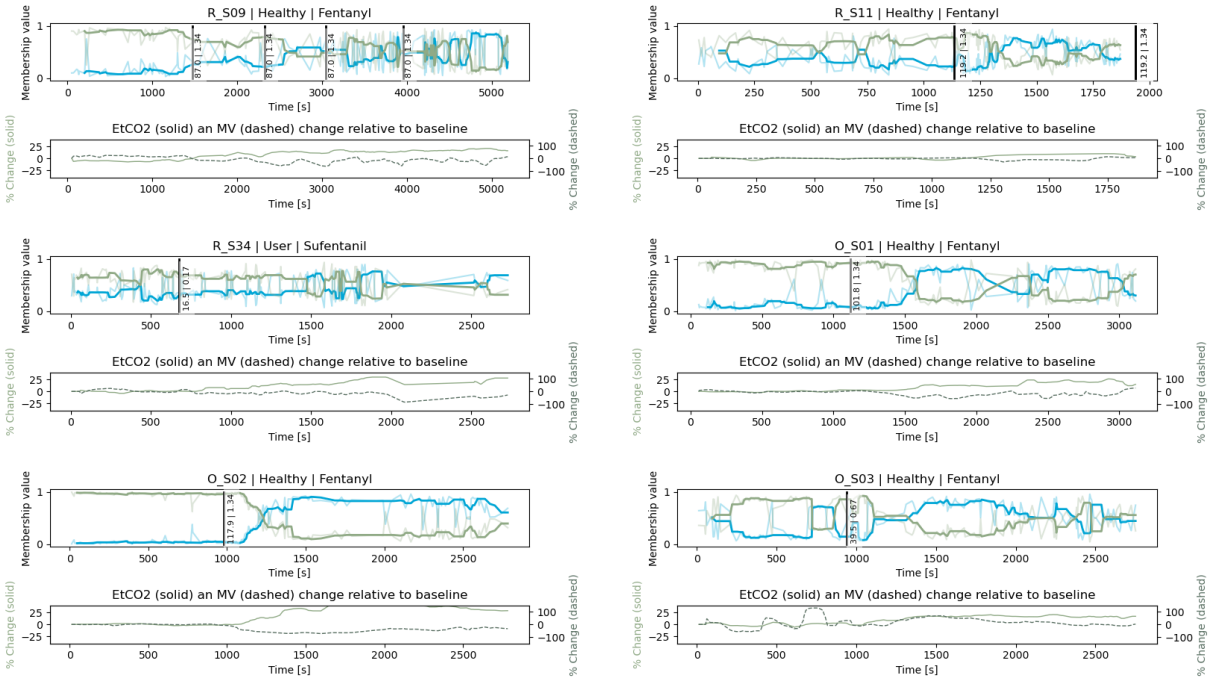


Figure 22: Temporal plots of the assigned membership-values for two clusters, where thicker lines represent the median over 7 epochs and the background lines depicts the original membership values. Moments of opioid administration are indicated by a solid line, including the given dose and dose per kilogram. Below the temporal plots, corresponding $EtCO_2$ and MV graphs are included.

E.2.2. Individual-clustering

Without antagonist administration



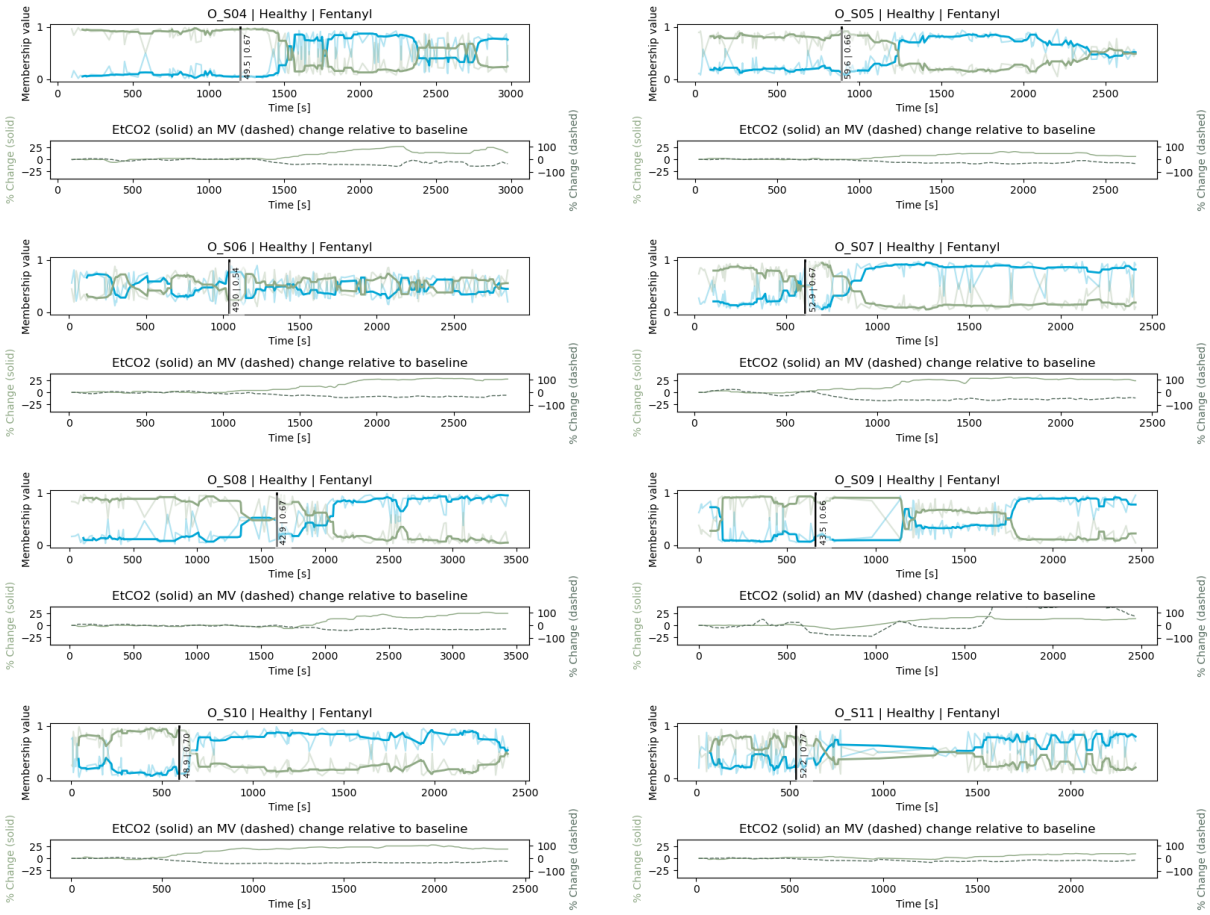
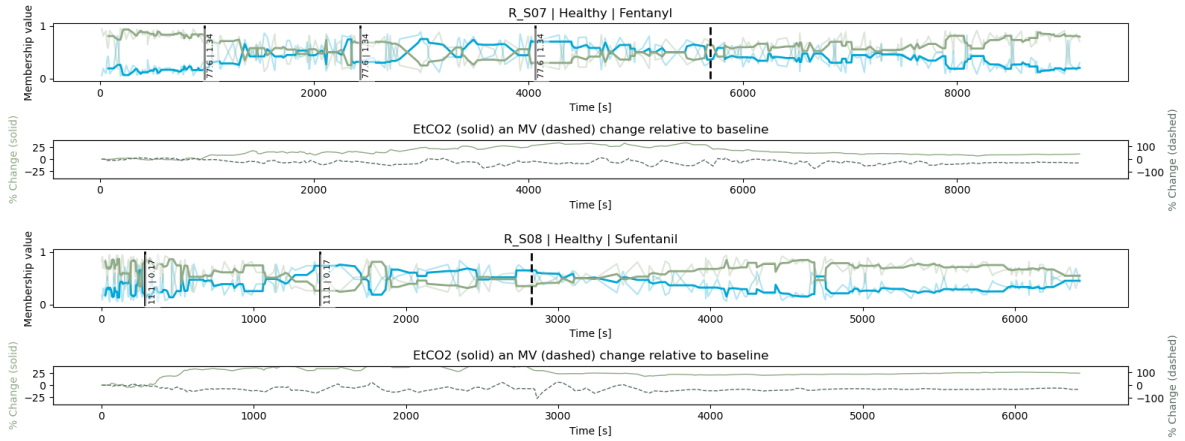
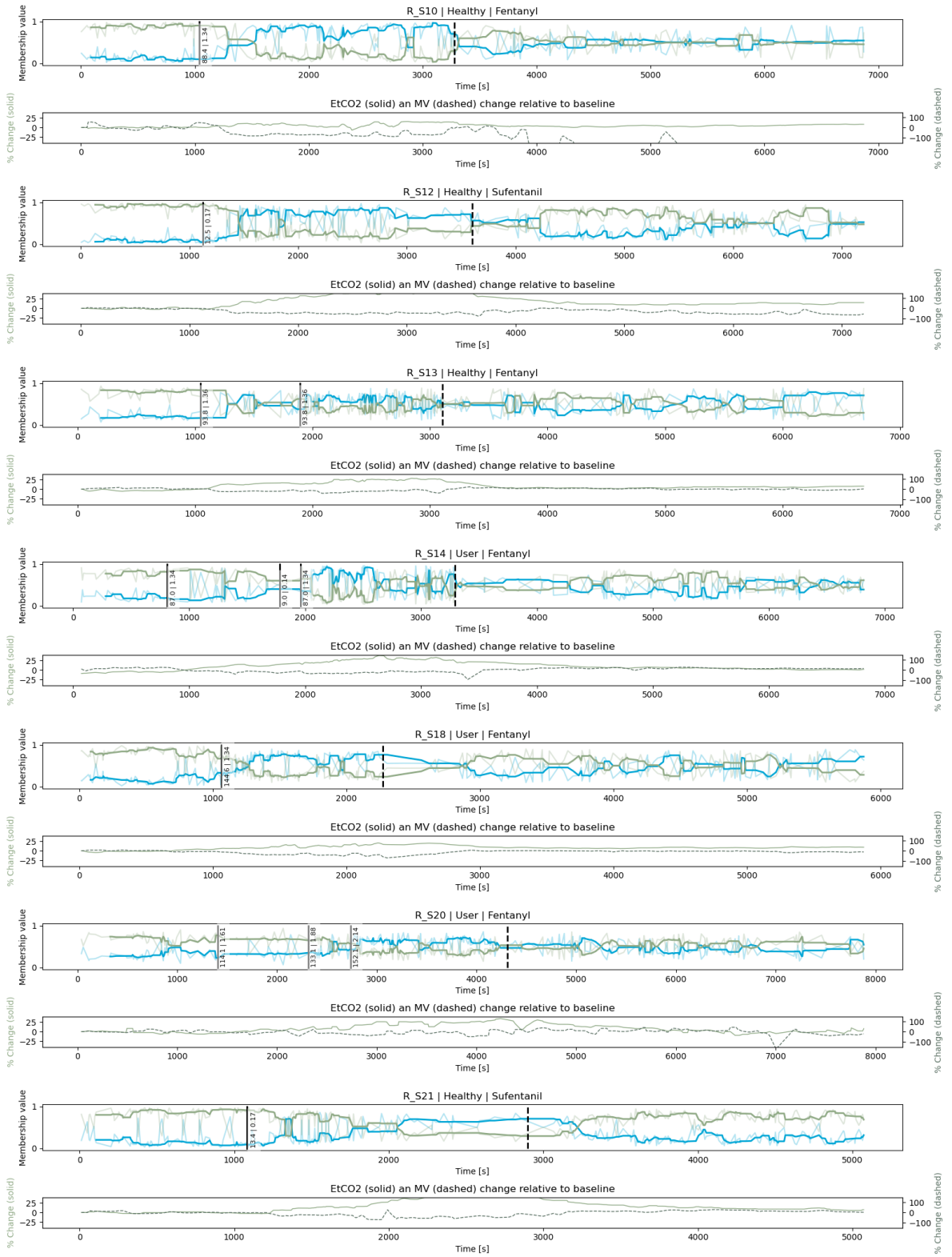
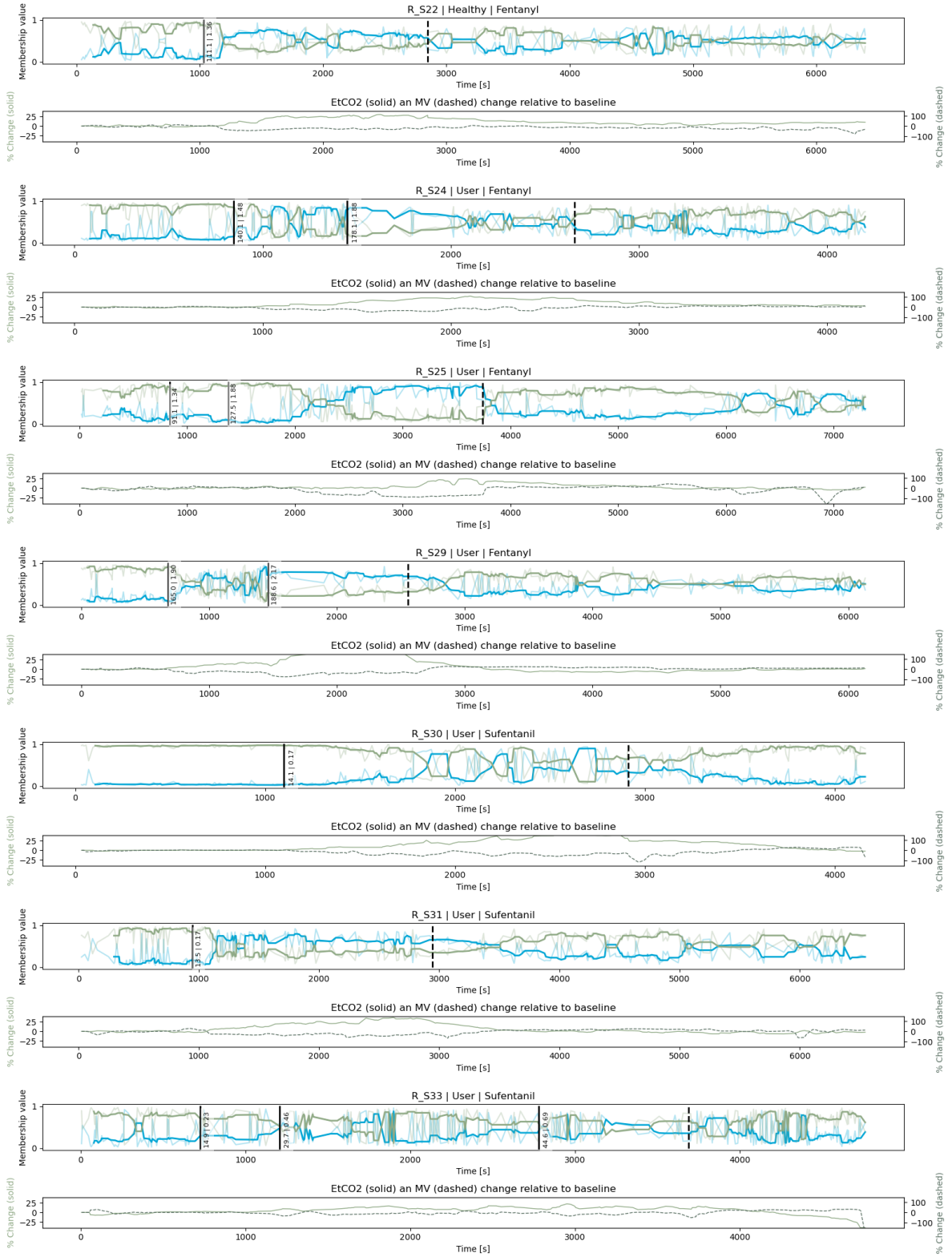


Figure 23: Temporal plots of the assigned membership-values for two clusters, where thicker lines represent the median over 7 epochs and the background lines depicts the original membership values. Moments of opioid administration are indicated by a solid line, including the given dose and dose per kilogram. Below the temporal plots, corresponding $EtCO_2$ and MV graphs are included.

With antagonist administration




R_S10 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)
R_S12 | Healthy | Sufentanil
R_S13 | Healthy | Fentanyl
R_S14 | User | Fentanyl
R_S18 | User | Fentanyl
R_S20 | User | Fentanyl
R_S21 | Healthy | Sufentanil


R_S22 | Healthy | Fentanyl
Membership value
Time [s]
EtCO2 (solid) an MV (dashed) change relative to baseline
% Change (solid)
% Change (dashed)
R_S24 | User | Fentanyl
R_S25 | User | Fentanyl
R_S29 | User | Fentanyl
R_S30 | User | Sufentanil
R_S31 | User | Sufentanil
R_S33 | User | Sufentanil

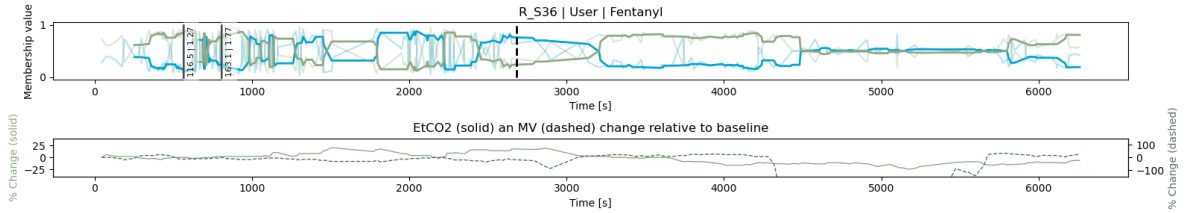
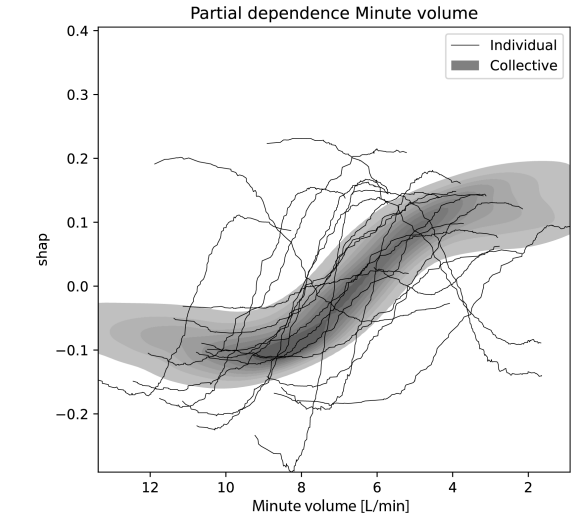


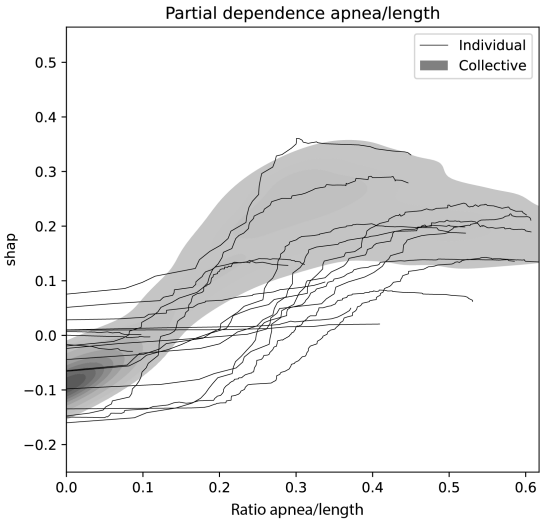
Figure 24: Temporal plots of the assigned membership-values for two clusters, where thicker lines represent the median over 7 epochs and the background lines depicts the original membership values. Moments of opioid administration are indicated by a solid line, including the given dose and dose per kilogram. The moment of antagonist administration is depicted by a dashed line. Below the temporal plots, corresponding $EtCO_2$ and MV graphs are included.

E.3. Partial dependence plots



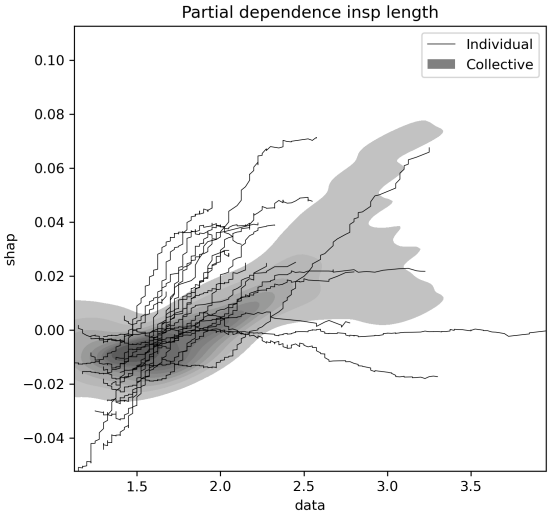
Corresponding correlation

	height	weight	BMI	age
middle data	0.53*	0.43	0.15	-0.16
middle shap	0.06	-0.03	-0.08	-0.02
slope	0.26	-0.01	-0.24	0.02
spread shap	-0.08	-0.25	-0.26	-0.26



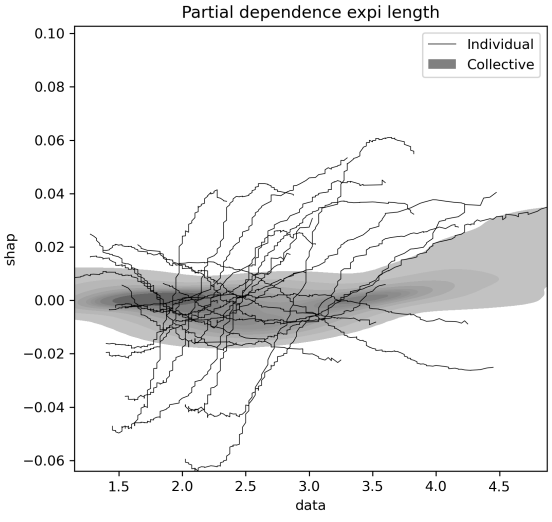
Corresponding correlation

	height	weight	BMI	age
middle data	-0.22	-0.10	0.05	0.31
middle shap	-0.11	0.11	0.26	-0.21
slope	0.35	0.28	0.10	0.20
spread shap	-0.15	0.09	0.25	0.29



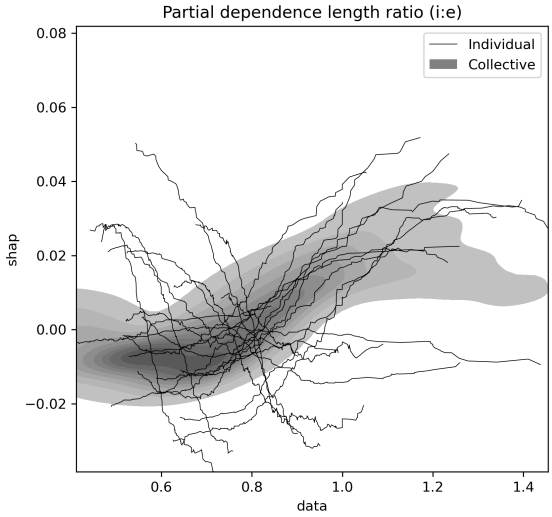
Corresponding correlation

	height	weight	BMI	age
middle data	-0.02	0.21	0.29	0.14
middle shap	-0.05	-0.02	0.03	-0.22
slope	0.04	-0.18	-0.26	-0.31
spread shap	-0.11	-0.10	-0.04	-0.25



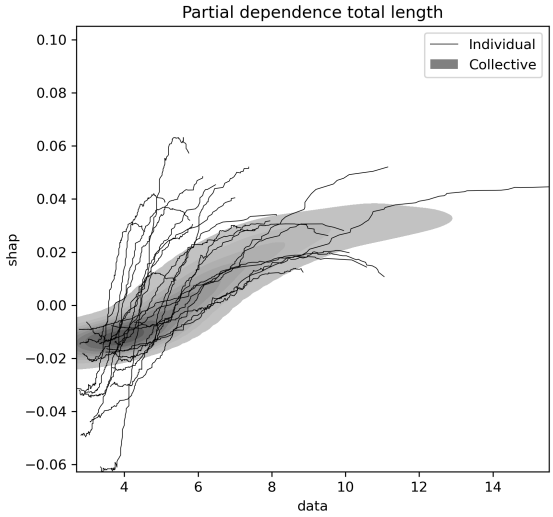
Corresponding correlation

	height	weight	BMI	age
middle data	-0.31	-0.13	0.06	0.11
middle shap	0.08	0.05	0.01	-0.11
slope	-0.01	-0.09	-0.09	-0.41
spread shap	0.25	0.06	-0.13	-0.45*



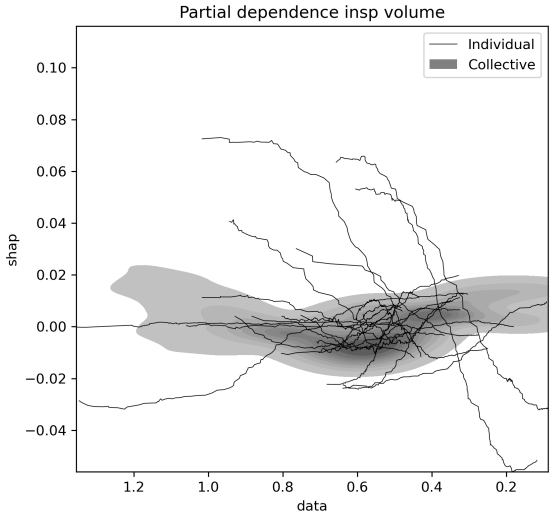
Corresponding correlation

	height	weight	BMI	age
middle data	0.15	0.17	0.11	0.17
middle shap	-0.06	0.04	0.11	-0.06
slope	-0.14	0.18	0.35	0.03
spread shap	0.25	-0.03	-0.24	-0.25



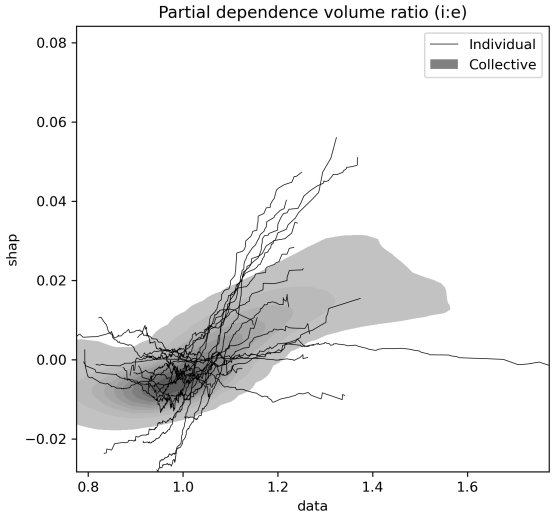
Corresponding correlation

	height	weight	BMI	age
middle data	-0.23	0.12	0.35	0.10
middle shap	0.17	0.21	0.17	-0.19
slope	0.12	-0.06	-0.19	0.05
spread shap	0.07	-0.17	-0.29	-0.06



Corresponding correlation

	height	weight	BMI	age
middle data	0.32	0.48*	0.40	0.17
middle shap	0.03	-0.02	-0.05	-0.42
slope	-0.27	-0.39	-0.29	0.07
spread shap	-0.15	-0.38	-0.39	-0.18



Corresponding correlation

	height	weight	BMI	age
middle data	-0.09	-0.01	0.08	0.36
middle shap	-0.19	0.08	0.30	-0.10
slope	-0.04	-0.08	-0.06	-0.14
spread shap	-0.14	0.24	0.46*	0.05

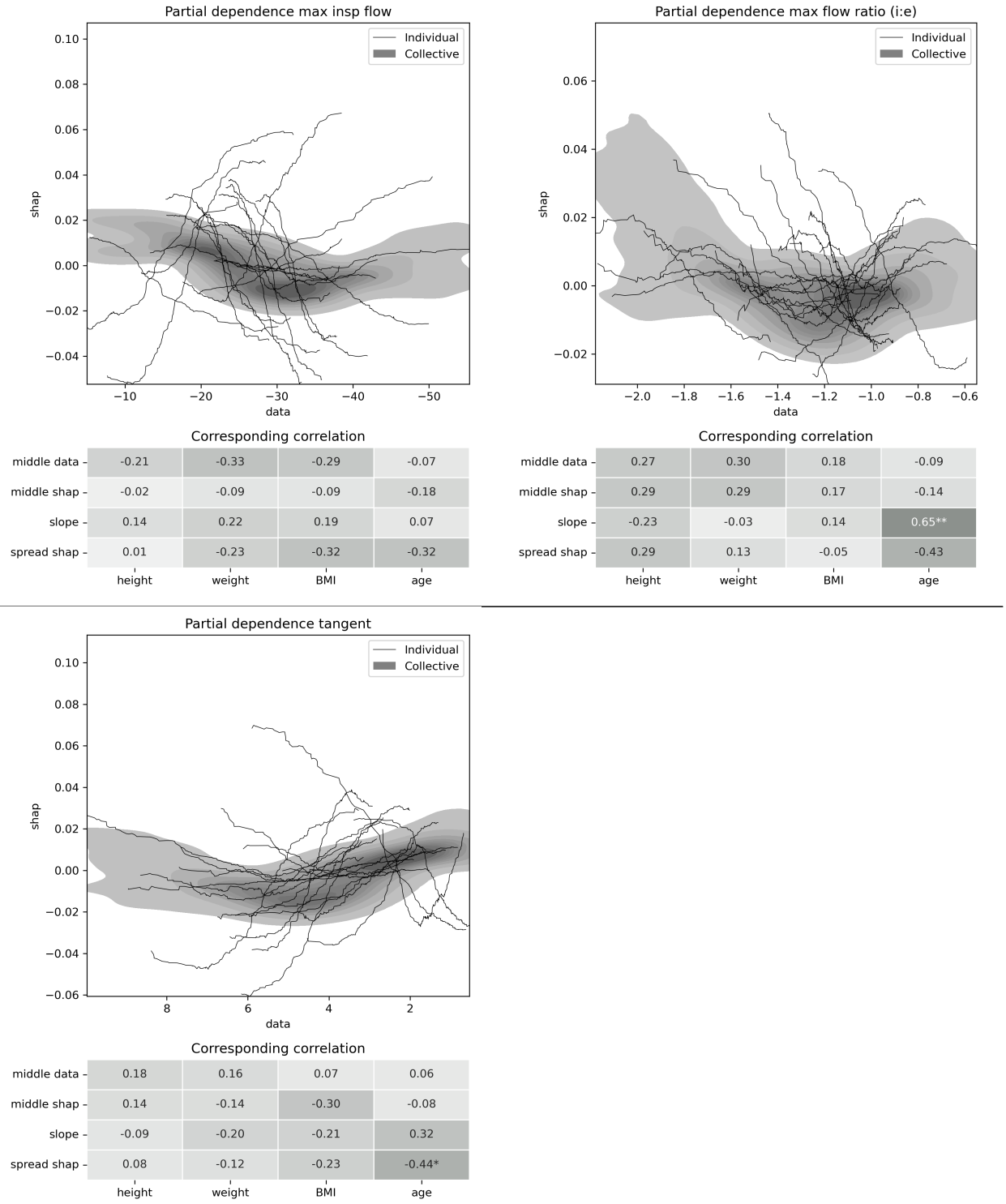


Figure 25: Partial dependence plot showing the collective-clustering results (grey area) as well as the individual-clustering results for every subject (black lines indicating the mean trend). Included are the corresponding correlations between various aspects of the individual-clustering result and static features of the subjects.
* : $0.01 < p <= 0.05$.
** : $0.001 < p <= 0.01$.
*** : $p <= 0.001$.